

OPEN

Nanopore sequencing of microbial communities reveals the potential role of sea lice as a reservoir for fish pathogens

Ana Teresa Gonçalves^{1,2}, Rayen Collipal-Matamal², Valentina Valenzuela-Muñoz^{1,2}, Gustavo Nuñez-Acuña^{1,2}, Diego Valenzuela-Miranda^{1,2} & Cristian Gallardo-Escárate^{1,2*}

Caligus rogercresseyi is a copepod ectoparasite with a high prevalence in salmon farms in Chile, causing severe welfare and economic concerns to the sector. Information on the parasite's underpinning mechanisms to support its life strategy is recently being investigated. Due to the critical role of microbiota, this study aimed to characterize the microbiota community associated with *C. rogercresseyi* from different regions with salmon aquaculture in Chile. Using third-generation sequencing with Nanopore technology (MinION) the full 16S rRNA gene from sea lice obtained from 8 areas distributed over the three main aquaculture regions were sequenced. Microbiota of the parasite is mainly comprised of members of phyla Proteobacteria and Bacteroidetes, and a core microbiota community with 147 taxonomical features was identified, and it was present in sea lice from the three regions. This community accounted for 19% of total identified taxa but more than 70% of the total taxonomical abundance, indicating a strong presence in the parasite. Several taxa with bioactive compound secretory capacity were identified, such as members of genus *Pseudoalteromonas* and *Dokdonia*, suggesting a possible role of the lice microbiota during the host infestation processes. Furthermore, the microbiota community was differentially associated with the salmon production, where several potential pathogens such as *Vibrio*, *Tenacibaculum*, and *Aeromonas* in Los Lagos, Aysén, and Magallanes region were identified. Notably, the Chilean salmon industry was initially established in the Los Lagos region but it's currently moving to the south, where different oceanographic conditions coexist with lice populations. The results originated by this study will serve as foundation to investigate putative role of sea lice as vectors for fish pathogens and also as reservoirs for antibiotic-resistant genes.

The copepod *Caligus rogercresseyi*, a caligid ectoparasite that is commonly found in southern Chilean salmon farms, is one of most impacting sanitary issues of the industry^{1,2}. *C. rogercresseyi* infestation has a profound effect on fish, producing irritation, loss of scales and skin wounds that may lead to secondary infections, producing strong physiological^{3,4} and molecular responses^{2,5} in the host. These contribute to a decrease in fish health causing an estimated production cost of 1.4USD per kilogram of produced salmon⁶. Despite the different types of available treatments, bath and oral application of delousing drugs are still preferred⁷. However, reduced sensitivity compromising treatment efficiency for organophosphates and pyrethroids have been recently reported in Chile⁸. The usage of advanced molecular biology tools such as high-throughput sequencing allowed a substantial increase of information regarding the parasite's invasive and defensive strategies within the parasite-host interface, and during exposure to delousing drugs over the past years (reviewed by Gallardo-Escárate, *et al.*²). Despite that, to date there is no available information regarding *C. rogercresseyi* associated microbiota, however, due to the key role played by these microorganisms in the host's well physiological functioning including homeostatic regulation, nutritional, metabolic and defensive performance^{9,10}, further research is necessary.

In parasites, the associated microbiota has also been reported as a potential reservoir for secondary pathogens¹¹. Known fish pathogens have been found in other caligid parasitic copepods such as *Vibrio* sp. found in

¹Interdisciplinary Center for Aquaculture Research, University of Concepción, Concepción, Chile. ²Laboratory of Biotechnology and Aquatic Genomics, Center of Biotechnology, University of Concepción, Concepción, Chile. *email: crisgallardo@udec.cl

*Caligus lalandei*¹², and there are reports referring to the Northern Hemisphere sea louse, *Lepeophtheirus salmonis*, as a potential vector for *Aeromonas salmonicida*^{13,14}. In copepods, microbial communities are most abundant close to the mouth and anus, as well as in digestive tract and eggs sacs¹⁵, and interactions with both transient and permanent epi or endobiotic associated microbiota seem to promote these organisms' health¹⁶. Marine copepods have host-specific microbial associations, symbiosis, or even a central natural core microbiome, that has been reported as more dependent of environment rather than diet¹⁶. In the case of *C. rogercresseyi*, information on these associations of core microbiome might bring light to the parasite life strategy and adaptation capacity and provide insights to develop adequate monitoring and control measures.

Evaluation of complex microbial communities has been performed using culture-dependent techniques, but the reduced proportion of bacteria that can be cultured, as well as the inability to reproduce complex microbial environments *in vitro* are limitations that were overcome by deep sequencing techniques. Presently, the most used method is 16S rRNA gene amplicon sequencing but also shotgun whole-genome metagenomics, the latter being more expensive due to requirements of higher sequencing depth¹⁷. Dalvin, *et al.*¹⁸ used shotgun metagenomics and reported that the majority of microorganisms associated with *L. salmonis* from the Atlantic and the Pacific Ocean belonged to the phyla Bacteroidetes and Proteobacteria, with a reduced incidence of known fish pathogens. On the other hand, Llewellyn, *et al.*¹⁹ applied 16S amplicon sequencing on the same species, and reported a high incidence of *Vibrio* sp. Although these are the only reports on sea lice microbiota to date, Moisaner, *et al.*¹⁶ reported in copepods a high abundance of organisms from Proteobacteria phylum, as well as Planctomycetes from PVC superphylum also using 16S amplicon sequencing.

The 16S rRNA gene has nine hypervariable regions with different characteristics, and the V3 and V4 regions have been the most sequenced for taxonomy assignation of the microorganisms associated with aquatic organisms²⁰, mostly performed on Illumina platforms. This sequencing strategy is based on sequencing by synthesis, and despite being frequently used, it is limited to a specific region of the 16S rRNA gene, reducing the estimation capacity of richness²¹. In this sense, the third-generation sequencing introduced a new technology based on single-molecule analysis, and sequencers have, therefore, the capacity to sequence full molecules or large DNA fragments²². In particular, the nanopore technology with the MinION™ system (Oxford Nanopore Technologies) presents an affordable and portable sequencing strategy that has great potential to evaluate microbiota communities. This technology allows the sequencing of the full-16S rRNA gene, increasing the capacity for taxonomical assignation, and it can be performed in real-time with reduced equipment, increasing its potential as a field tool for remote locations²³. However, despite these advantages, when comparing to short reads sequencing such as Illumina, the error rate is higher²², problem that has been progressively mitigated with technological advances and better bioinformatic tools^{24,25}. Regarding the study of microbiota, several studies have been dedicated to compare taxonomical analysis derived by this technology with Illumina sequencing, and have shown that error rate has been compensated with larger fragments, providing in some cases higher taxonomical resolution^{26,27}. Although the error rate remains the main limitation to the expansion of this sequencing methodology, it has been successfully applied in several species and tissues, identifying the microbiota present in complex samples from terrestrial^{28,29} and aquatic environment^{23,30}. Notably, nanopore sequencing has been applied to understand the interactions of human diseases³¹, environmental samples obtained from glacier regions²³, ocean water column³² and to identify the bacterial community associated with harmful microalgae³⁰, and these studies demonstrate the potential and applicability of nanopore sequencing for microorganism detection in all environments.

This study aimed to characterize the microbiota community associated with the salmon's ectoparasite *Caligus rogercresseyi* obtained from different geographical regions, using third-generation sequencing technology based on nanopore sequencing of full 16S rRNA gene. The results reported in this study are of high importance to support the development of sustainable salmon aquaculture worldwide, where the reduction of antibiotics and other drugs is pivotal in parallel with the characterization of potential reservoirs for fish pathogens.

Results

Nanopore sequencing output. A total of 4.945.864 reads passed the initial filter and were processed for taxonomical analysis. After the steps of demultiplexing, trimming, and length filter, an average of 94.6% were successfully classified in a taxonomical group (Supplementary Table S1) with at least 75% identity. All samples presented more than 300.000 reads except one sample from the Aysén region (ACS 30a) that presented fewer reads due to DNA quantity obtained from the sample. Since this study has a microbiome communities' exploratory character, this sample was evaluated for the relative abundance of the identified taxa but was excluded from statistical analysis. The average sequence length was 1438 bp, but more than 60% of the sequence's length was ranging from 1520 bp to 1900 bp (Fig. 1A), whereas the overall quality was 8.36 (Fig. 1B).

***C. rogercresseyi* microbial community structure.** Rarefaction analysis indicated that except for samples from the ACS 30a and ACS 15, samples reached a plateau denoting that the number of sequences was enough to cover all the identified taxonomical features (Fig. 2A). The microbial community associated with *C. rogercresseyi* from the Los Lagos region presented higher diversity than the communities from other regions, but this difference was significant only when measured by the Shannon diversity index (Fig. 2B) and not by Simpson's 1-D index (Fig. 2C). Microbial community from the ectoparasite original from Magallanes presented higher evenness than the community from Los Lagos (Fig. 2D), as well as lower richness (Fig. 2E), although the latter was not significant. Based on the Bray-Curtis distances between bacterial communities, Los Lagos (Pto Montt and Chiloe) separated from others (Fig. 2G). The most abundant phyla were Proteobacteria with more than 60% prevalence regardless of region (Fig. 3A), followed by Bacteroidetes and Cyanobacteria. When comparing the three areas of Los Lagos, the microbial community of *C. rogercresseyi* from Chiloe presents a higher abundance of Verrucomicrobia and fewer Bacteroidetes, whereas more Bacteroidetes were found in the ACS 23's community in Aysén (Fig. 3B). At class level, bacterial community from Los Lagos sea lice has high abundance

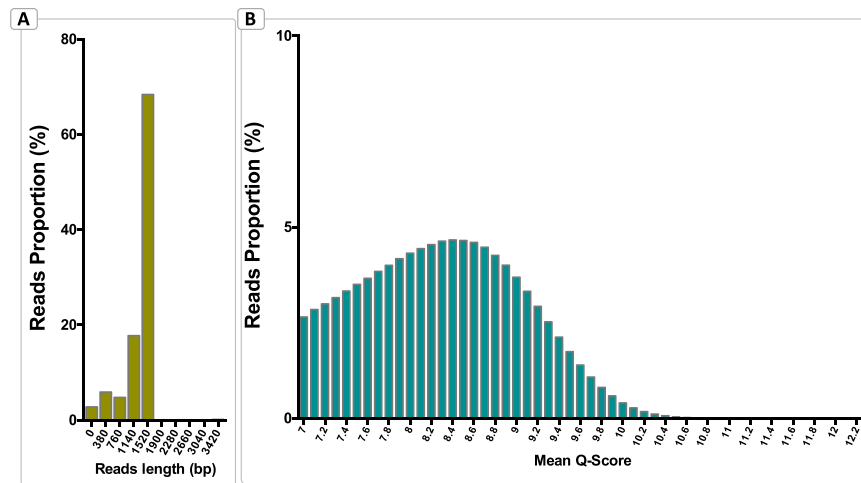


Figure 1. Distributions of reads according to sequence length (A), and distribution of reads according to their average Q-Score of quality (B).

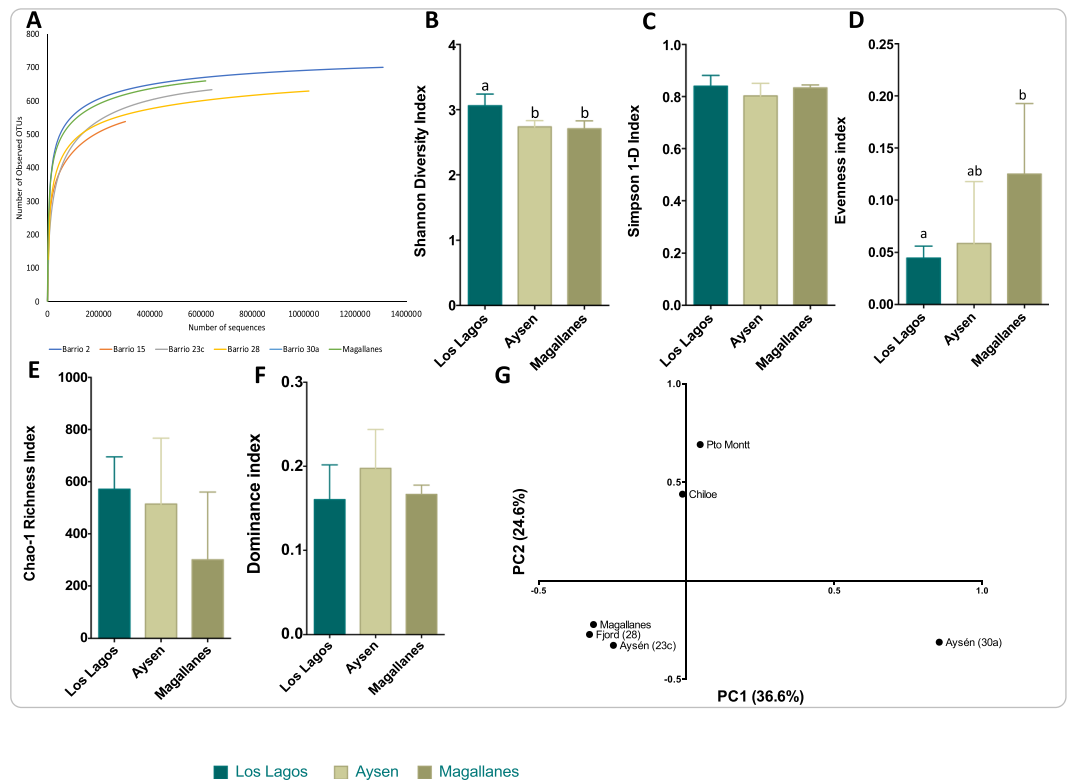


Figure 2. Diversity analysis of *Caligus rogercresseyi* microbiota community. Rarefaction curve (A), Shannon (B) and Simpson 1-D (C) diversity indices, evenness index (D), richness Chao1 estimator (E), dominance index (F) and principal coordinate analysis of the Bray Curtis distances between communities (G). Bars in B-F are mean \pm standard deviation and different letters indicate significant differences (One-Way ANOVA and post hoc Tukey HSD test, $P < 0.05$).

of Gammaproteobacteria followed by Flavobacteriia and Alphaproteobacteria, whereas in Aysén and Magallanes the communities have less Gammaproteobacteria and more Alphaproteobacteria and Saprospira (in the case of Aysén) (Fig. 4A). Within Aysén, communities from different areas have different class abundances, with a higher abundance of Saprospira in ACS 23c, as well as a higher abundance of Alphaproteobacteria in ACS 28 (Fig. 4B). At family level, microbial communities of sea lice from Pto Montt and Contao (ACS 2, Los Lagos region) are more similar, whereas communities from Chiloé clustered with communities from Aysén and Magallanes (Fig. 5A). Microbiota from Magallanes sea lice clustered with the community obtained from sea lice from fish produced in the fjords at ACS 28. When comparing the communities by regions at the genus level, 14 genera were significantly

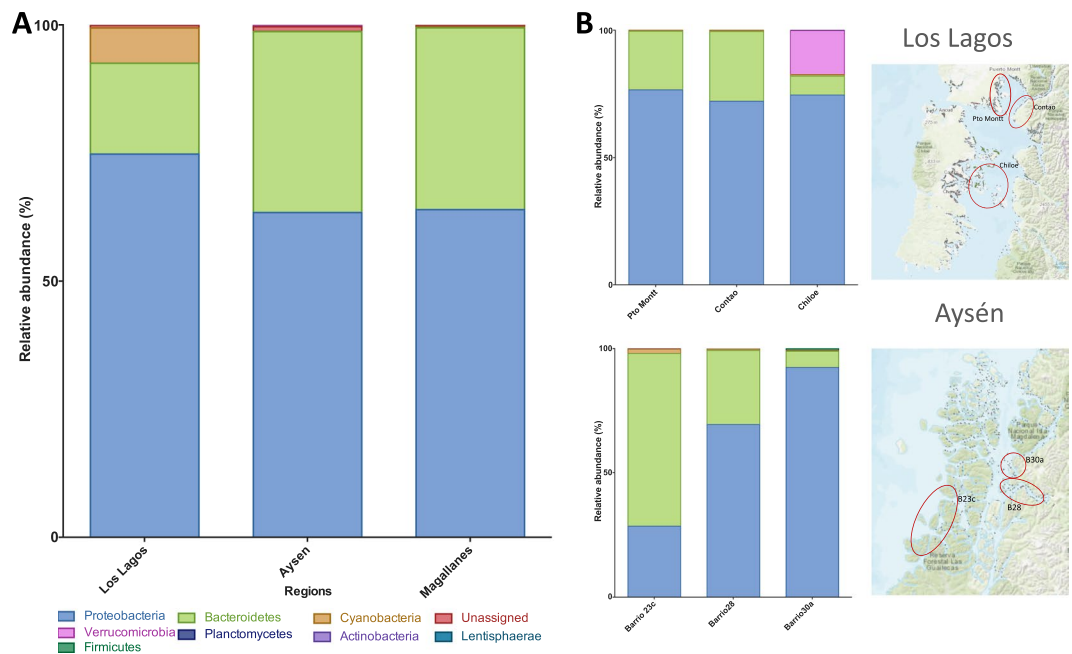


Figure 3. Relative abundance of *C. rogercresseyi* associated microbiota at phylum level by region (A), and among areas of Los Lagos (upper) and Aysén (lower) regions (B) (Map adapted from Daniel Jimenez).

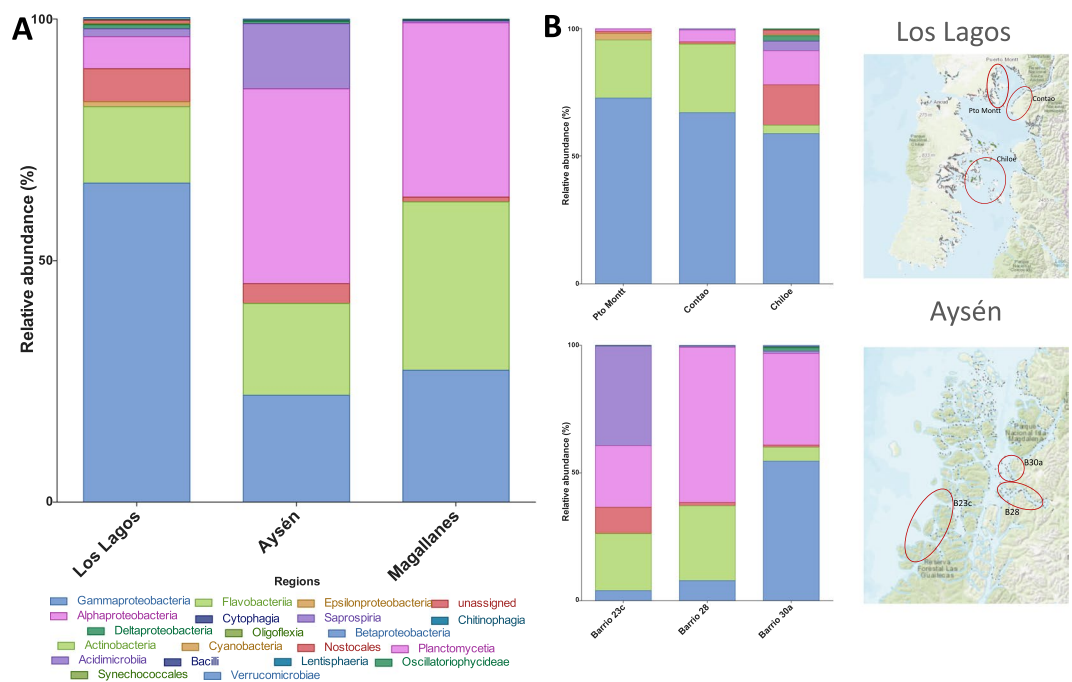


Figure 4. Relative abundance of *C. rogercresseyi* associated microbiota at class level by region (A), and among areas of Los Lagos (upper) and Aysén (lower) regions (B) (Map adapted from Daniel Jimenez).

different between Los Lagos and Aysén, and 9 between Los Lagos and Magallanes (Fig. 5B). However, when considering the different areas within each region other genus are significantly different (Fig. 5C) and these were *Pseudoalteromonas*, *Tenacibaculum*, and *Vibrio* with a significantly higher abundance in sea lice from Pto Montt and Contao, and also *Dokdonia* genus with higher abundance in Magallanes' sea lice.

Microbial taxa present in sea lice from all regions was considered part of the core microbial community and included 145 taxa (Fig. 6A) that constitute 19.7% of the total identified taxonomical features. This community includes only three phyla: Proteobacteria, Bacteroidetes and Cyanobacteria. The first two were found in similar abundances, being the family Flavobacteriaceae predominant within Bacteroidetes, whereas within Proteobacteria elements of Rhodobacteraceae also Vibrionaceae were the most abundant (Fig. 6B). The

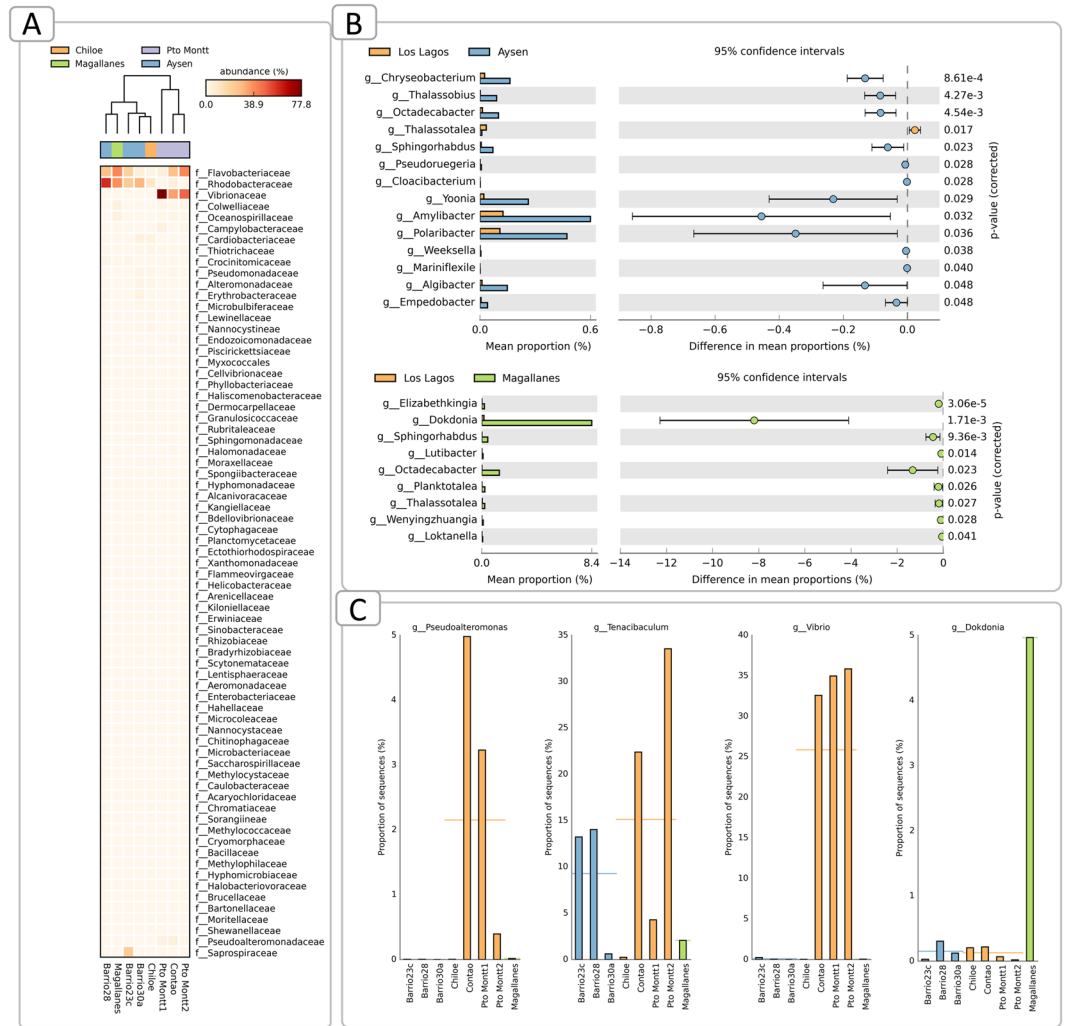


Figure 5. *Caligus rogercresseyi* associated microbial community characterization by hierarchical clustering at family level by UPGMA method (A); Comparison of microbiota relative abundance at genus level from Los Lagos, Aysén and Magallanes region and confidence intervals for the significant differences (B); Relative abundance of four known potential pathogens found with significant abundance in at least one region of the study (C).

abundance of the taxonomical features that comply with the *C. rogercresseyi* core community account for 70.3% of the obtained sequences from sea lice from the Los Lagos region, 86.8% of sequences from Aysén and 91.4% of sequences from Magallanes. Still, some features present different abundances between regions, and the most notable is the class Gammaproteobacteria which have higher abundance in sea lice from Los Lagos, Flavobacteriaceae and Rhodobacteraceae families with higher abundance in sea lice from Magallanes and the later also from Aysén, and the genus *Dokdonia* and *Tenacibaculum* with higher abundance in sea lice from Magallanes and from Los Lagos and Aysén, respectively (Fig. 6C). Moreover, the output of the mock community sequencing (Supplementary Table S2) used as internal technical control, indicated a good correspondence between the expected and the observed relative abundances.

Discussion

The sea louse *Caligus rogercresseyi* is an ectoparasite that affects salmon farming in southern Chile, not only by impacting fish health and performance with consequent increase in productions costs, but also by contributing to environmental jeopardy through the massive application of drugs to control the infestation^{6,33}. Knowledge of parasite's biological systems and strategies is crucial to consolidate adequate responses for sustainable management of the disease. In terms of microbiota evaluation, high-throughput sequencing is currently used to reveal information that is hardly available by culture-dependent methods. In this study, we sequenced the full-16S rRNA gene using third-generation sequencing technology to characterize for the first time the microbiota associated with *C. rogercresseyi* infecting Atlantic salmon from different regions in Chile. This information is especially relevant in parasites since it has been conjectured that these communities might play a fundamental role in the copepod fitness¹⁶ and, therefore, in its capacity to resist host defenses and successfully infect.

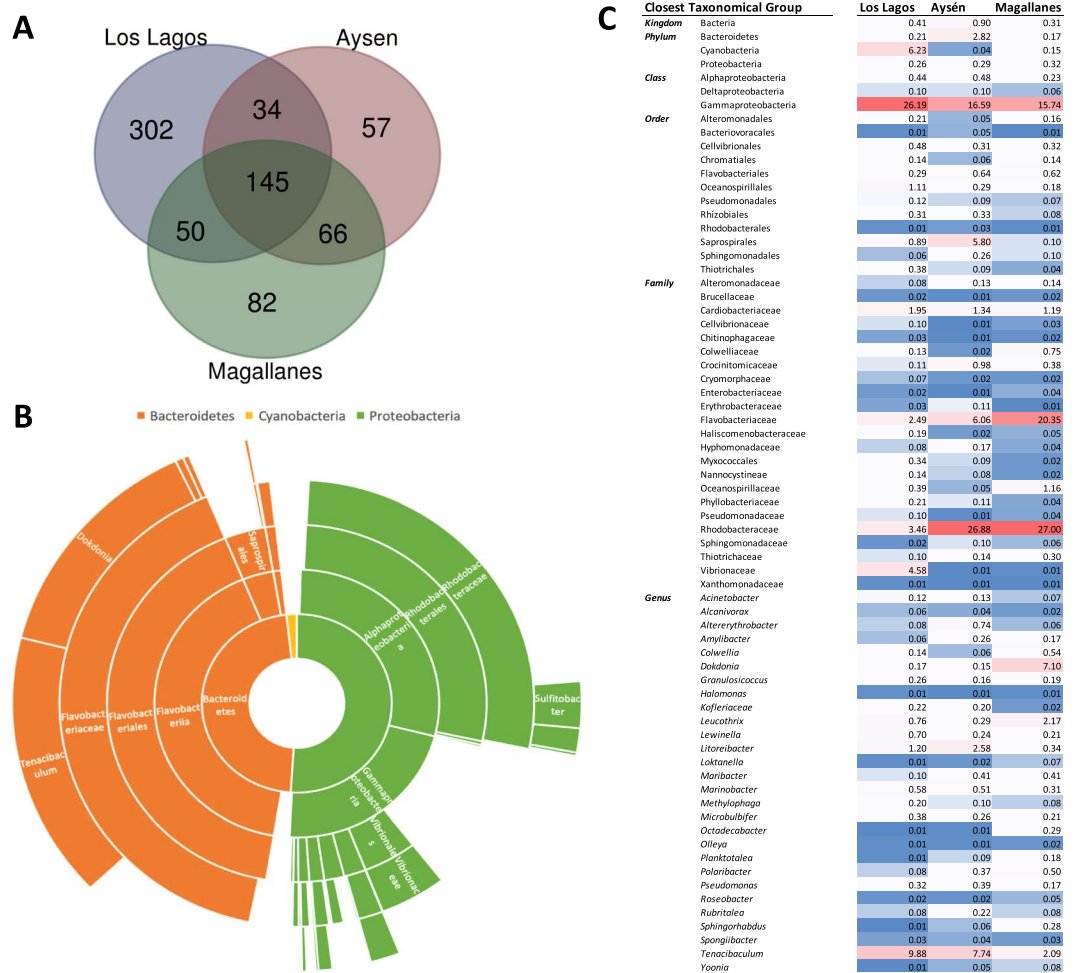


Figure 6. *Caligus rogercresseyi* core microbiota. Venn diagram representing the taxonomical features identified in the three regions of the study (A); Taxonomical distribution of taxa included in the 145 taxon shared by sea lice microbial community of the three regions based on the average relative abundance (B); Heat map of relative abundance of core microbiota taxa identified to the deepest level with a 95% confidence on identity, with colors ranging from blue to red representing lower to higher relative abundance respectively (C).

Regarding diversity, Shannon’s index was higher in the community of microorganisms from sea lice obtained in the Los Lagos Region, and this might be related to the higher taxonomical features found only in sea lice from this region (>300, Supplementary Fig. 3S). On the other hand, although not significantly, in Magallanes, sea lice microbiota community is the least rich. From the three regions in the study, Los Lagos is most influenced by anthropogenic activities since it has several cities in the surroundings, and it was one of the first places where the commercial salmon aquaculture started with higher presence nowadays³³. This fact contrasts with Magallanes, the most remote with less environmental pressure either anthropogenic or by aquaculture³³. Notably, the first official *C. rogercresseyi* infestation report was recently archived³⁴, suggesting some correlation between aquaculture pressure and *C. rogercresseyi* microbial community composition. However, more studies are required to explore the impacts of the sea lice microbiota in salmon farming and with particular emphasis to elucidate the capacity of lice’s microbiota to trigger infectious disease outbreaks.

More than 140 taxonomical features were identified in *C. rogercresseyi* core microbiota, and although these represent only 19% of the total features identified in the study, more than 70% of the sequences belonged to these core features, indicating the presence of robust resident microbiota in the parasite, which might be species-specific. Dalvin, *et al.*¹⁸ identified in *Lepeophtheirus salmonis* a dominance of members of the orders Flavobacteriales, Burkholderiales, Mycoplasmatales and Spirochaetales, and class Alphaproteobacteria in the associated microbiota. In *C. rogercresseyi* a high presence of Alphaproteobacteria and Flavobacteriales was also observed, but the other taxa, although present, were not highly abundant. On the other hand, contrary to what was described for *L. salmonis*¹⁸, in *C. rogercresseyi* microorganisms from orders Vibrionales and Alteromonadales were found in practically all samples. In another study conducted in *L. salmonis*, microorganisms from genus *Arcobacter*, *NS10* marine group, *Pseudomonas*, *Aeromonas*, *Vibrio*, *Rhizobiales*, and *Tenacibaculum* were identified¹⁹. Except for the first two above mentioned taxa, all the others were identified in *C. rogercresseyi*, which might indicate some level of similarities between microbiota associated to salmon ectoparasites.

The strong presence of Proteobacteria and Bacteroidetes in the core microbiota indicates dominance of Gram-negative bacteria, among which several known fish pathogens. It has been reported that the sea lice external surface area is covered by microorganisms³⁵, and this includes pathogens from *Tenacibaculum*, *Pseudomonas* and *Vibrio* genus³⁶, all of these found in *C. rogercresseyi* with noteworthy abundances. The origin of parasite's microbiota is expected to be the host mucus and skin (the parasite's dietary sources) and milieu water¹⁹. However, host mucus microbiota composition is altered by many factors such as salinity³⁷, diet³⁸, and even the simple presence of the parasite¹⁹. Therefore, it would be expected that these modulations would impact sea lice microbiota composition as well. However, reports indicate that sea lice and salmon mucus microbiota are not necessarily correlated¹⁹, and this supports the findings of this study where a stable, robust core microbiome was identified regardless region, center, producer practices or other factor. The factors influencing the stability of the core community are yet to be unraveled, but it is known that other parasites such as the nematode *Heligmosomoides polygyrus* can secrete antimicrobial compounds that shape the host-microbiota for the parasite's benefit, increasing its fitness³⁹. It is known that sea lice developed a complex secretome⁴⁰, that modulates host immune response⁴¹, and host mucus microbiome¹⁹.

The functional role of the identified core microbiota is of high interest for future studies since its potential to enhance sea lice fitness, and resistance is high. For instance, microbiota in insects actively participates in the peritrophic matrix – a chitin and proteins layer that surrounds gut epithelium and is present also in crustacean species – and this is thought to protect the host by reducing the influx of undesirable bacteria to the body⁴², possibly preventing an exacerbated immune response from host (in this case sea lice) to pathogenic bacteria present in the environment. On the other hand, adding to the high prevalence observed in pathogenic bacteria, *C. rogercresseyi* microbiota also showed a strong presence of taxa known as bioactive compounds producers. For example, organisms from genus *Cobetia* are nonpathogenic but produce biofilm and secrete exopolysaccharides, as well as microorganisms from the *Dokdonia* genus, a highly prevalent genus in Magallanes sea lice microbial community. Biofilm production is a characteristic that confers adhesion, protection, and robustness to the bacteria⁴³, and might be influencing the stability of the microbiota composition and fitness of the community. Members from the genus *Pseudoalteromonas* are microorganisms with the capacity to secrete bioactive compounds with antialgae and antimicrobial properties⁴⁴, and these were found in higher abundance in sea lice from Pto Montt and Contao in Los Lagos region. These are examples of the complex set of bioactive compounds producing microbiota present in *C. rogercresseyi* that might be increasing the parasite's resistance and persistence.

Using third-generation sequencing technology by Nanopore MinION in aquaculture is a novel approach that produced a meaningful set of information on a relevant pathogen for the sector. Despite the increasing amount of studies being reported using the technology released in 2014, this is the first study in an aquatic invertebrate. The higher error rate of nanopore sequencing compared to other platforms (i.e. Illumina) has been one of the main issues limiting the dissemination of this technology to broader fields. In microbial analysis however, several studies have revealed that the error rate has been somehow compensated by reads length, generating outputs comparable to Illumina^{26,27,45–48}, sharing more than 95% of total taxonomical abundance⁴⁹. These studies support the robustness of the results achieved with this technology and together with the low initial cost and high portability of the device, suggests that it will be a preferred to other technologies for microbial studies. Sequencing the full 16S rRNA gene increases identification power⁴⁷, and indeed in this study we achieved a high level of species identification indicating its potential for aquaculture pathogen detection or future monitoring and control of microbial dynamics.

Conclusions

Overall, the microbial composition of *C. rogercresseyi* is dominated by gram-negative microorganisms, including several pathogens and taxa associated with biofilm and bioactive compounds production, and these characteristics might be related to the high prevalence of the parasite and its resilience capacity. The presence of a stable and highly abundant core microbial community suggests a link to parasite fitness, and further research should be a pursuit to understand the role of these microorganisms in the maintenance of the core microbiota and its relationship with the parasite success and life strategies. Several taxa with bioactive compound secretory capacity were identified, suggesting a possible role during the host infestation processes. The microbiota community was differentially associated with the salmon production, where several potential pathogens in Los Lagos, Aysén and Magallanes region were identified. The putative role of sea lice as vectors for fish pathogens and also reservoirs for antibiotic-resistant genes should be further investigated.

Methods

Preliminary evaluation. The first step of this evaluation was to perform DNA extraction. Although different extraction kits and methods have been compared in other studies, the presence of exoskeleton in organisms such as copepods represents a challenge not usually addressed by the existent protocols⁵⁰. Most of the studies focused on the evaluation of symbiont microbiota in animals use fecal contents as biologic material, and in organisms such as *C. rogercresseyi*, such material is hardly accessible. Microbiota associated with gut tissue has been evaluated in insects by dissecting the abdomen segments that held the full digestive tract^{51,52}. This type of dissection requires equipment and technical precision that is not available when sampling in the field where conditions are not stable (as in the example in a salmon farm sea cages), and in these cases, its preferable using the whole animal to assure procedures homogeneity. According to Rubin, *et al.*⁵⁰, the DNA extraction method used in terrestrial arthropods did not affect the identification of the bacterial community structure of the organisms but affected the capacity of the sample to be successfully sequenced. This enhances the importance of optimizing extraction protocols before sequencing. Therefore, previously to sequencing, several preliminary tests were performed to select the DNA extraction method, and the minimum number of copepods to be pooled in a sample. To evaluate different DNA extraction methods, adult females of *C. rogercresseyi* from the wet lab of the National

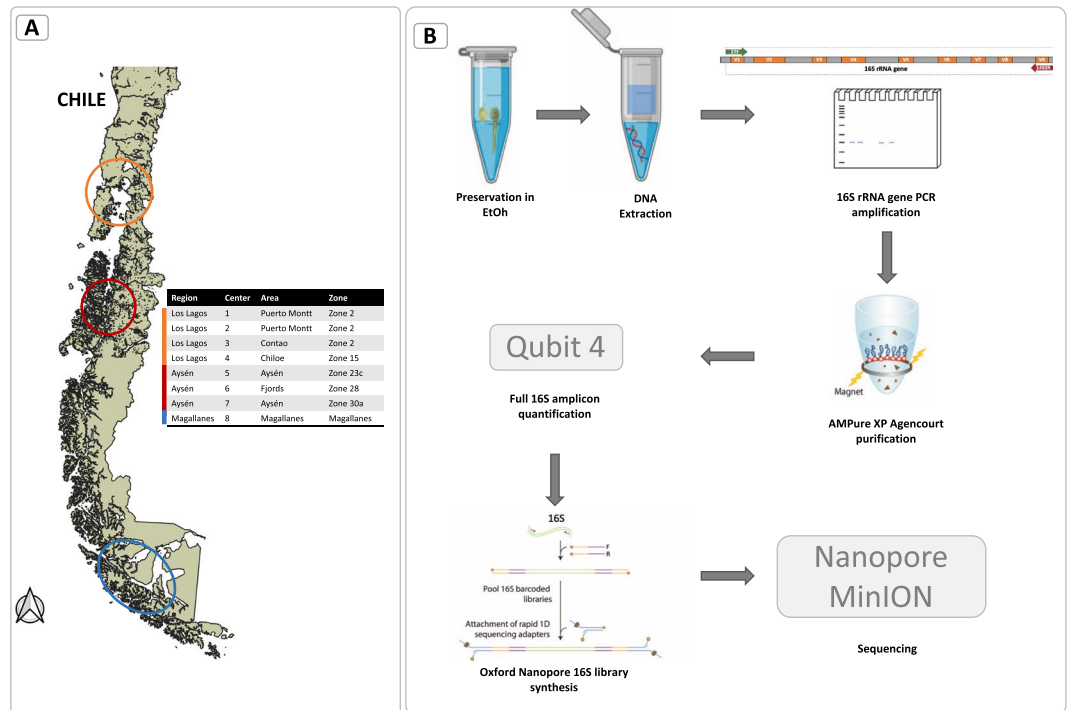


Figure 7. Chile's partial map with identification of sampling regions (A), and experiment workflow diagram (B) (Map adapted from Daniel Jimenez).

Reference Laboratory for this ectoparasite were obtained. Five protocols were tested; (a) commercial kit Qiagen DNeasy Fast Mini Stool (Qiagen, USA) with slight modification increasing incubation time and temperature, (b) commercial kit Qiagen DNeasy Blood and Tissue (Qiagen, USA) with modifications according to Michl, *et al.*⁵³ that includes a lysis buffer with lysozyme to increase Gram-positive bacteria DNA yield, (c) commercial kit Qiagen DNeasy Blood and Tissue with the manufacturer's modification for insects, (d) salt-extraction⁵⁴ and (e) salt-extraction protocol modified to include column purification and more extended incubation period. During a previous study, we identified difficulties in the complete dissociation of the copepod's exoskeleton using only mechanical disruption with ceramic beads, as previously mentioned by Rubin, *et al.*⁵⁰, and these were surpassed by increasing incubation periods. Regardless of the extraction method used, pools of 10 females preserved in 99–100% ethanol (molecular grade) were washed in sterilized by autoclave and UV PBS (phosphate-buffered saline solution) and were cut in small pieces in an ice cold sterilized petri dish. The lysis method was dependent on the protocol under analysis, and the five methods included a negative control with no sample and positive control with 10 mg of a commercial mixture of two *Bacillus subtilis* strains. Extracted DNA was measured by spectrophotometry in a Nanodrop (ThermoScientific, USA), and integrity was verified by a TAE (Tris-Acetic Acid-EDTA buffer) gel electrophoresis with 1% agarose. The presence of bacterial DNA after each extraction was evaluated by amplification of the V4 hypervariable region of the 16S rRNA gene using the universal primers 515F- GTGCCAGCMGCCGCGTAA, 806R-GGACTACHVGGGTWTCTAAT in a PCR with 50 ng of extracted genomic DNA as template in a Verity thermal cycler (Applied Biosystems, USA), producing a 390 bp amplicon as described in Gonçalves and Gallardo-Escárate⁵⁵. Extracted DNA purity and the qualitative evaluation of the intensity of PCR amplicon was used to decide the most appropriate method, and overall results are presented in Supplementary Fig. 1S. After choosing DNA extraction protocol, extraction was performed in triplicate to a single *C. rogercesseyi* adult and pools of two, three, and four specimens to obtain genomic DNA that was after used as a template in the above-described PCR for V4 amplification.

***Caligus rogercesseyi* sampling and sample processing.** Parasites were obtained from commercial salmon farms from the three regions with active industrial productions of salmon in Chile, Los Lagos (or X Region), Aysén (or XI Region) and Magallanes (Fig. 7A and Supplementary Fig. 2S). Samples were collected during Spring and Summer and centers were selected based on a) geographical distribution to include centers that would represent distinct aquaculture areas (denominated as "ACS" or Salmon farming concessions grouping by Chilean authorities) of the three major regions, b) production cycle in order to avoid centers with less than 8 months of seawater production that could still be under the influence of any antilousing prophylaxis administered during freshwater stage (e.g. Lufenuron), and c) drugs utilization, preferring centers with less or no drug (both delousing and antibacterial) usage in the cycle. Following this, sampling was performed in four centers of the Los Lagos region, three centers in the Aysén region, and one center in the Magallanes region. Samplings were performed during the morning, and procedures were previously standardized to ensure reproducibility between centers. When possible, samples were collected during the sampling performed by the centers as part of

the compulsory sea lice counting required by Chilean authorities (Sernapesca), and procedures were performed according to sampling protocols approved by the same entity. Atlantic salmon (*Salmo salar*) were confined in the sea cages using secondary nets and were collected with a soft net and sedated in a reservoir previously loaded with Benzocaine (20%) in seawater (15 ml per 100 L seawater). After approximately 2 minutes, fish were examined for parasites, and these were collected with plastic sterilized forceps and placed in vials with molecular grade ethanol. Vials were stored at -20°C and transported in cooled boxes by overnight express service to our laboratory where they were stored at -80°C until processing.

All operations involving fish handling were approved by the Ethical Committee of the University of Concepción and by the Chilean Authority SERNAPESCA, that regulates all operations in culture centers. *Caligus rogercresseyi* is a copepod, and since regulations for animal use in research do not cover this taxon, no ethical approval of these experiments was required. All methods and procedures were performed in accordance with the relevant guidelines and regulations.

DNA extraction. Pools of 10 adults were thawed, washed and sliced as above detailed and were processed as described in the study workflow (Fig. 7B), starting with DNA extraction following the protocol that presented the best results during the preliminary study. The minced copepods were mixed with 1 mL of lysis buffer (10 mM Tris-HCl, 400 mM NaCl, 100 mM EDTA, 0.4% SDS, and 24 $\mu\text{g}/\text{mL}$ Protease K, pH 8.0), after which were thoroughly vortexed and incubated overnight and 56°C . The contents of each tube were centrifuged at 5000 rpm for 1 min and supernatant mixed with molecular grade 99–100% ethanol and then added to a silica column that was centrifuged at 8,500 rpm for 1 min. The column was washed with the kit buffers (including a buffer with a low concentration of a guanidine based chaotropic salt), and after complete centrifugation, DNA was eluted in two steps of 15 μL of DEPC water finalizing with a total volume of 30 μL eluted DNA. DNA integrity was verified by electrophoresis as above mentioned, and concentration measured by fluorescence in a Qubit 4 (ThermoScientific, USA) using the Qubit dsDNA BR Assay Kit (ThermoScientific, USA) following the manufacturer's instructions. Genomic DNA samples were diluted to 100 ng/ μL and amplification of full-16S rRNA gene was performed by PCR with reaction volume of 25 μL , using the primers 27 F 5'-AGAGTTTGATCTGGCTCAG-3' γ 1492 R 5' GGTTACCTTGTTACGACTT-3', using Taq DNA polymerase LongAmp (NewEngland Biolabs, USA), with an initial denaturing step at 95°C for 1 min, followed by 25 cycles of 95°C for 20 sec, 56°C for 30 sec and 65°C for 2 min, and an extension at 65°C for 5 min. PCR products were evaluated by electrophoresis, as previously described.

Nanopore library synthesis and sequencing. After confirming the presence of full-16S rRNA amplicon, PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter, USA) in a 1:2 sample-beads ratio. The mixture was incubated for 5 min at room temperature, and the sample placed in the magnetic stand for ethanol washing (freshly prepared at 70%). The washed mixture was resuspended in ultrapure sterilized water, and the purified product was separated from beads by placing it in a magnetic stand. The purified amplicon obtained from pooled *C. rogercresseyi* DNA was quantified again in the fluorometer Qubit 4 (ThermoScientific, USA) and was used as a template for library synthesis using the 16S Barcoding Kit (SQK-RAB204, Oxford Nanopore Technologies), following the manufacturer's instructions for 1D sequencing strategy. Purified 16S amplicon was mixed with barcodes accordingly, and using the LongAmp Taq Polymerase (New England Biolabs), a PCR was performed as described in the previous step, with a final reaction volume of 50 μL . The PCR product was purified again with Agencourt AMPure XP beads and incubated in HulaMixer for 5 min at room temperature. After the magnetic beads washing step of the purified product was eluted in 10 μL of elution buffer (10 mM Tris-HCl pH8.0 with 50 mM NaCl). The final concentration of the library was quantified in fluorometer Qubit 4 (ThermoScientific, USA), and libraries were evaluated on a TapeStation Bioanalyzer 2100 (Agilent, USA), using DNA ScreenTape (Agilent, USA) following the manufacturer's instructions. A single library was synthesized from DNA of a pool of ectoparasites per center, and libraries were pooled in multiplex mode following the protocol of Oxford Nanopore Technologies, and the flowcell MK1 Spot-ON FLO-MIN107-R9 was loaded accordingly. Sequencing efficiency was monitored through the software MinKNOW 2.0 (Oxford Nanopore Technologies). Finally, as internal control, the DNA of a microbial mock community (ZymoBiomix Microbial Community Standard) was extracted and sequenced following the same describe procedures and observed abundances of taxa were compared with expected abundance.

Bioinformatic analysis. After the sequencing run, generated fast5 files were base-called using Guppy (version 3.2.2, Oxford Nanopore Technologies, UK), and a filter step was applied to retain only sequences with Q-score ≥ 7 (quality filter). Sequences were then evaluated using the Fastq 16S v3.2.1 (Oxford Nanopore Technologies, UK) pipeline that included demultiplexing and primers and barcode trimming. Sequences with ambiguous identification were excluded from subsequent analysis. BLASTN aligned cleaned sequences against NCBI RefSeq for 16S ribosomal RNA database for bacteria and archaea within the EPI2ME software package (Oxford Nanopore Technologies), and only the sequences with identity $\geq 75\%$ were retained (confidence level filter) as mentioned by Edwards, *et al.*²³. The occurrence of each NCBI taxonomy identifier was evaluated for each sample using R⁵⁶, and the relative abundance of each NCBI taxa id was calculated per sample to normalize. Taxonomical identifiers were ordered, and to minimize singletons, doubletons product of sequencing errors taxon with $<0.01\%$ of relative abundance in at least one sample were eliminated. For each sample, an equivalent to OTU (Operative Taxonomic Unit) frequency table was built using R, and rarefaction analysis were performed to assess sequencing depth performance. The alpha diversity of microbiota communities associated with *C. rogercresseyi* was compared between region by calculating the Shannon and Simpson 1-D diversity indices, the Chao-1 richness index, community Evenness, and dominance index using PAST3 v3.25⁵⁷. Differences between regions were evaluated with One-Way ANOVA followed by the posthoc multiple comparison Tukey-HSD test when applicable, considering

a P-value < 0.05. Microbial community structure was analyzed using Bray–Curtis distances based on taxon relative abundance data, and principal coordinate analysis (PCoA) were performed as a multivariate unsupervised data exploration. Variations of the microbial community along geographical distribution were explored by plotting relative abundance at Phylum and Class level, and multivariate exploratory analysis was performed through Principal Component Analysis of the OTU's relative abundance at the order level.

Distance between communities of different areas was evaluated at family level by hierarchical clustering using UPGMA method, and genus differential abundances between regions were evaluated by Welch's t-test considering an interval of confidence based on the Newcombe-Wilson method⁵⁸ with P-value correction of Storey's FDR⁵⁹. When applicable, comparisons of bacterial communities' taxon abundance were performed between two samples (i.e., between two areas) by assessing the Fisher's exact test⁶⁰ with P-value correction of Storey's FDR. Besides, to complement the first assessment of *C. rogercresseyi* associated microbiota communities, the "core" microbiome was defined as the set of OTU's present in samples from all the geographical regions sampled in the study, and these OTU's were evaluated following the deepest taxonomical assignment with an identity $\geq 95\%$.

Data availability

Data are available at NCBI-SRA BioProject PRJNA598707.

Received: 31 October 2019; Accepted: 3 February 2020;

Published online: 19 February 2020

References

- Burka, J. F., Fast, M. D. & Revie, C. W. In *Fish Parasites: Pathobiology and Protection* (eds Woo, P. T. K. & Buchanan, K.) 360–380 (CAB International, 2012).
- Gallardo-Escárate, C. *et al.* The race between host and sea lice in the Chilean salmon farming: a genomic approach. *Rev. Aquaculture* **11**, 325–339, <https://doi.org/10.1111/raq.12334> (2019).
- González, M. P., Marín, S. L. & Vargas-Chacoff, L. Effects of *Caligus rogercresseyi* (Boxshall and Bravo, 2000) infestation on physiological response of host *Salmo salar* (Linnaeus 1758): Establishing physiological thresholds. *Aquaculture* **438**, 47–54, <https://doi.org/10.1016/j.aquaculture.2014.12.039> (2015).
- Vargas-Chacoff, L. *et al.* Neuroendocrine stress response in Atlantic salmon (*Salmo salar*) and Coho salmon (*Oncorhynchus kisutch*) during sea lice infestation. *Aquaculture*, <https://doi.org/10.1016/j.aquaculture.2019.04.046> (2019).
- Gallardo-Escárate, C. *et al.* The *Caligus rogercresseyi* miRNome: Discovery and transcriptome profiling during the sea lice ontogeny. *Agri. Gene* **4**, 8–22, <https://doi.org/10.1016/j.aggene.2017.03.002> (2017).
- Dresdner, J. *et al.* Impact of *Caligus* treatments on unit costs of heterogeneous salmon farms in Chile. *Aquaculture Econ. Manag.* **23**, 1–27, <https://doi.org/10.1080/13657305.2018.1449271> (2019).
- Aaen, S. M., Helgesen, K. O., Bakke, M. J., Kaur, K. & Horsberg, T. E. Drug resistance in sea lice: a threat to salmonid aquaculture. *Trends Parasitol.* **31**, 72–81, <https://doi.org/10.1016/j.pt.2014.12.006> (2015).
- Arriagada, G. *et al.* First report of the reduction in treatment efficacy of the organophosphate azamethiphos against the sea lice *Caligus rogercresseyi* (Boxshall & Bravo, 2000). *Aquaculture Research*, <https://doi.org/10.1111/are.14334> (2019).
- Leulier, F. *et al.* Integrative Physiology: At the Crossroads of Nutrition, Microbiota, Animal Physiology, and Human Health. *Cell Metab.* **25**, 522–534, <https://doi.org/10.1016/j.cmet.2017.02.001> (2017).
- Nieuwdorp, M., Gilijamse, P. W., Pai, N. & Kaplan, L. M. Role of the microbiome in energy regulation and metabolism. *Gastroenterology* **146**, 1525–1533, <https://doi.org/10.1053/j.gastro.2014.02.008> (2014).
- Abraham, N. M. *et al.* Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proc. Natl Acad. Sci. USA* **114**, E781–E790, <https://doi.org/10.1073/pnas.1613422114> (2017).
- 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. *Journal of Fish Diseases*, n/a–n/a, <https://doi.org/10.1111/jfd.12582> (2016).
- Nese, L. & Enger, O. Isolation of *Aeromonas salmonicida* from salmon lice *Lepeophtheirus salmonis* and marine plankton. *Dis. Aquat. Org.* **16**, 79–81 (1993).
- 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).
- Carman, K. R. & Dobbs, F. C. Epibiotic microorganisms on copepods and other marine crustaceans. *Microsc Res Tech* **37**, 116–135, [10.1002/\(SICI\)1097-0029\(19970415\)37:2<116::AID-JEMT2>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1097-0029(19970415)37:2<116::AID-JEMT2>3.0.CO;2-M) (1997).
- Moisander, P. H., Sexton, A. D. & Daley, M. C. Stable Associations Masked by Temporal Variability in the Marine Copepod Microbiome. *PLoS one* **10**, e0138967, <https://doi.org/10.1371/journal.pone.0138967> (2015).
- Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical Biophysical Res. Commun.* **469**, 967–977, <https://doi.org/10.1016/j.bbrc.2015.12.083> (2016).
- Dalvin, S. *et al.* Microbial communities associated with the parasitic copepod *Lepeophtheirus salmonis*. *Marine Genomics*, <https://doi.org/10.1016/j.margen.2019.05.003> (2019).
- Llewellyn, M. S. *et al.* Parasitism perturbs the mucosal microbiome of Atlantic Salmon. *Sci. Rep.* **7**, 43465, <https://doi.org/10.1038/srep43465> (2017).
- Ortiz-Estrada, Á. M., Gollas-Galván, T., Martínez-Córdova, L. R. & Martínez-Porchas, M. Predictive functional profiles using metagenomic 16S rRNA data: a novel approach to understanding the microbial ecology of aquaculture systems. *Rev. Aquaculture* **11**, 234–245, <https://doi.org/10.1111/raq.12237> (2019).
- Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**, 635, <https://doi.org/10.1038/nrmicro3330>, <https://www.nature.com/articles/nrmicro3330#supplementary-information> (2014).
- Tyler, A. D. *et al.* Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* **8**, 10931, <https://doi.org/10.1038/s41598-018-29334-5> (2018).
- Edwards, A. *et al.* In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota. *bioRxiv*, 073965, <https://doi.org/10.1101/073965> (2018).
- Li, C. *et al.* INC-Seq: accurate single molecule reads using nanopore sequencing. *Gigascience* **5**, 34, <https://doi.org/10.1186/s13742-016-0140-7> (2016).
- Calus, S. T., Ijaz, U. Z. & Pinto, A. J. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *Gigascience* **7**, <https://doi.org/10.1093/gigascience/giy140> (2018).

26. Cusco, A., Catozzi, C., Vines, J., Sanchez, A. & Francino, O. Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and whole *rrn* operon. *Fl1000Res* **7**, 1755, <https://doi.org/10.12688/fl1000research.16817.1> (2018).
27. Leggett, R. M. *et al.* Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. *Nat Microbiol*, <https://doi.org/10.1038/s41564-019-0626-z> (2019).
28. Shin, J. *et al.* Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Scientific Reports* **6**, 29681, <https://doi.org/10.1038/srep29681>, <https://www.nature.com/articles/srep29681#supplementary-information> (2016).
29. Bertrand, D. *et al.* Nanopore sequencing enables high-resolution analysis of resistance determinants and mobile elements in the human gut microbiome. *bioRxiv*, 456905, <https://doi.org/10.1101/456905> (2018).
30. Shin, H. *et al.* Elucidation of the bacterial communities associated with the harmful microalgae *Alexandrium tamarense* and *Cochlodinium polykrikoides* using nanopore sequencing. *Sci. Rep.* **8**, 5323, <https://doi.org/10.1038/s41598-018-23634-6> (2018).
31. Kafetzopoulou, L. E. *et al.* Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Sci.* **363**, 74, <https://doi.org/10.1126/science.aau9343> (2019).
32. Hamner, S. *et al.* Metagenomic Profiling of Microbial Pathogens in the Little Bighorn River, Montana. *Int J Environ Res Public Health* **16**, <https://doi.org/10.3390/ijerph16071097> (2019).
33. Quiñones, R. A., Fuentes, M., Montes, R. M., Soto, D. & León-Muñoz, J. Environmental issues in Chilean salmon farming: a review. *Reviews in Aquaculture* **11**, 375–402, <https://doi.org/10.1111/raq.12337> (2019).
34. Arriagada, G. *et al.* First report of the sea louse *Caligus rogercresseyi* found in farmed Atlantic salmon in the Magallanes region, Chile. *Aquaculture* **512**, 734386, <https://doi.org/10.1016/j.aquaculture.2019.734386> (2019).
35. Harasimczuk, E., Overgard, A. C., Grotmol, S., Nilsen, F. & Dalvin, S. Characterization of three salmon louse (*Lepeophtheirus salmonis*) genes with fibronectin II domains expressed by tegumental type 1 glands. *Mol. Biochem. Parasitol.* **219**, 1–9, <https://doi.org/10.1016/j.molbiopara.2017.12.002> (2018).
36. 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. *Parasitology Reserach* **105**, 1173–1177, <https://doi.org/10.1007/s00436-009-1523-9> (2009).
37. Lokesh, J. & Kiron, V. Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. *Sci. Rep.* **6**, 19707, <https://doi.org/10.1038/srep19707> (2016).
38. Minniti, G. *et al.* The Skin-Mucus Microbial Community of Farmed Atlantic Salmon (*Salmo salar*). *Frontiers in Microbiology* **8**, <https://doi.org/10.3389/fmicb.2017.02043> (2017).
39. Rausch, S. *et al.* Parasitic Nematodes Exert Antimicrobial Activity and Benefit From Microbiota-Driven Support for Host Immune Regulation. *Front. Immunol.* **9**, 2282, <https://doi.org/10.3389/fimmu.2018.02282> (2018).
40. Hamilton, S. *et al.* Characterisation of proteins in excretory/secretory products collected from salmon lice, *Lepeophtheirus salmonis*. *Parasit. Vectors* **11**, 294, <https://doi.org/10.1186/s13071-018-2885-6> (2018).
41. Gallardo-Escarate, C. *et al.* Catching the complexity of salmon-lice interactions. *Fish. Shellfish. Immunol.* **90**, 199–209, <https://doi.org/10.1016/j.fsi.2019.04.065> (2019).
42. Rodgers, F. H., Gendrin, M., Wyer, C. A. S. & Christophides, G. K. Microbiota-induced peritrophic matrix regulates midgut homeostasis and prevents systemic infection of malaria vector mosquitoes. *PLoS Pathog.* **13**, e1006391, <https://doi.org/10.1371/journal.ppat.1006391> (2017).
43. Dang, H. & Lovell, C. R. Microbial Surface Colonization and Biofilm Development in Marine Environments. *Microbiol. Mol. Biol. Rev.* **80**, 91–138, <https://doi.org/10.1128/MMBR.00037-15> (2016).
44. Holmström, C. & Kjelleberg, S. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiology Ecol.* **30**, 285–293, <https://doi.org/10.1111/j.1574-6941.1999.tb00656.x> (1999).
45. Leggett, R. M. & Clark, M. D. A world of opportunities with nanopore sequencing. *J. Exp. Bot.* **68**, 5419–5429, <https://doi.org/10.1093/jxb/erx289> (2017).
46. Shin, J. *et al.* Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci. Rep.* **6**, 29681, <https://doi.org/10.1038/srep29681> (2016).
47. Benítez-Paez, A., Portune, K. J. & Sanz, Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer. *Gigascience* **5**, 4, <https://doi.org/10.1186/s13742-016-0111-z> (2016).
48. Kai, S. *et al.* Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION nanopore sequencer. *FEBS Open. Bio* **9**, 548–557, <https://doi.org/10.1002/2211-5463.12590> (2019).
49. Ma, X., Stachler, E. & Bibby, K. Evaluation of Oxford Nanopore MinION Sequencing for 16S rRNA Microbiome Characterization. *bioRxiv*, 099960, <https://doi.org/10.1101/099960> (2017).
50. Rubin, B. E. *et al.* DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. *Microbiologyopen* **3**, 910–921, <https://doi.org/10.1002/mbo3.216> (2014).
51. Jones, R. T., Sanchez, L. G. & Fierer, N. A cross-taxon analysis of insect-associated bacterial diversity. *PLoS One* **8**, e61218, <https://doi.org/10.1371/journal.pone.0061218> (2013).
52. Kautz, S., Rubin, B. E. R. & Moreau, C. S. Bacterial Infections across the Ants: Frequency and Prevalence of *Wolbachia*, *Spiroplasma*, and *Asaia*. *Psyche* **2013**, 11, <https://doi.org/10.1155/2013/936341> (2013).
53. Michl, S. C. *et al.* The malleable gut microbiome of juvenile rainbow trout (*Oncorhynchus mykiss*): Diet-dependent shifts of bacterial community structures. *PLoS One* **12**, e0177735, <https://doi.org/10.1371/journal.pone.0177735> (2017).
54. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**, 4692–4693 (1997).
55. Gonçalves, A. T. & Gallardo-Escarate, C. Microbiome dynamic modulation through functional diets based on pre- and probiotics (mannan-oligosaccharides and *Saccharomyces cerevisiae*) in juvenile rainbow trout (*Oncorhynchus mykiss*). *J. Appl. Microbiology* **122**, 1333–1347 (2017).
56. RCoreTeam. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, 2018).
57. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**, 9 (2001).
58. Newcombe, R. G. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat. Med.* **17**, 873–890 (1998).
59. Storey, J. D., Taylor, J. E. & Siegmund, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Society: Ser. B* **66**, 187–205, <https://doi.org/10.1111/j.1467-9868.2004.00439.x> (2004).
60. Rivals, I., Personnaz, L., Taing, L. & Potier, M.-C. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinforma.* **23**, 401–407, <https://doi.org/10.1093/bioinformatics/btl633> (2006).

Acknowledgements

This study was funded by the Program for Health Management of Aquaculture, FIE-2015-V014, and by the FONDAP-INCAR #15110027 and FONDECYT #1180867 from CONICYT-Chile. We thank Dr. Daniel Jimenez for providing the maps for geographical distribution of sampling areas.

Author contributions

A.T.G., C.G.E., G.N.A., V.V.M. and D.V.M. designed the study; A.T.G. and C.G.E. supervised the project and analyzed data; A.T.G., R.C.M., V.V.M. and G.N.A. sampled, processed samples, sequenced samples; A.T.G. wrote the manuscript; C.G.E., R.C.M., V.V.M., G.N.A. and D.V.M. revised and edited the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-59747-0>.

Correspondence and requests for materials should be addressed to C.G.-E.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020