

RESEARCH ARTICLE

From the Andes to the desert: 16S rRNA metabarcoding characterization of aquatic bacterial communities in the Rimac river, the main source of water for Lima, Peru

Pedro E. Romero^{1*}, Erika Calla-Quispe², Camila Castillo-Vilcahuaman¹, Mateo Yokoo³, Hammerly Lino Fuentes-Rivera², Jorge L. Ramirez⁴, André Ampuero⁵, Alfredo J. Ibáñez², Paolo Wong^{3*}

1 Departamento de Ciencias Biológicas y Fisiológicas, Facultad de Ciencias y Filosofía, Universidad Peruana Cayetano Heredia, Lima, Peru, **2** Instituto de Ciencias Ómicas y Biotecnología Aplicada (ICOBA), Pontificia Universidad Católica del Perú, Lima, Peru, **3** Departamento de Ciencias de la Medicina, Facultad de Medicina Humana, Universidad de Piura, Lima, Peru, **4** Departamento de Biología Celular y Genética, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Peru, **5** Departamento de Malacología y Carcinología, Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima, Peru

* pedro.romero@upch.pe (PER); paolo.wong.@udep.pe (PW)



OPEN ACCESS

Citation: Romero PE, Calla-Quispe E, Castillo-Vilcahuaman C, Yokoo M, Fuentes-Rivera HL, Ramirez JL, et al. (2021) From the Andes to the desert: 16S rRNA metabarcoding characterization of aquatic bacterial communities in the Rimac river, the main source of water for Lima, Peru. *PLoS ONE* 16(4): e0250401. <https://doi.org/10.1371/journal.pone.0250401>

Editor: João Carlos Nabout, Universidade Estadual de Goiás - Campus de Ciências Exatas e Tecnológicas Henrique Santillo, BRAZIL

Received: August 26, 2020

Accepted: April 6, 2021

Published: April 22, 2021

Copyright: © 2021 Romero et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data are available from the NCBI BioProject database (accession number PRJNA646070).

Funding: This work was funded by the Facultad de Medicina Humana (Universidad de Piura) [grant number PI2002, "Metabarcoding del Río Rímac, el principal afluente de la ciudad de Lima"]; GenLab del Perú SAC [Programa de Incentivos 2019,

Abstract

The Rimac river is the main source of water for Lima, Peru's capital megacity. The river is constantly affected by different types of contamination including mine tailings in the Andes and urban sewage in the metropolitan area. In this work, we aim to produce the first characterization of aquatic bacterial communities in the Rimac river using a 16S rRNA metabarcoding approach which would be useful to identify bacterial diversity and potential understudied pathogens. We report a lower diversity in bacterial communities from the Lower Rimac (Metropolitan zone) in comparison to other sub-basins. Samples were generally grouped according to their geographical location. Bacterial classes Alphaproteobacteria, Bacteroidia, Campylobacteria, Fusobacteriia, and Gammaproteobacteria were the most frequent along the river. *Arcobacter cryaerophilus* (Campylobacteria) was the most frequent species in the Lower Rimac while *Flavobacterium succinicans* (Bacteroidia) and *Hypnocyclus* (Fusobacteriia) were the most predominant in the Upper Rimac. Predicted metabolic functions in the microbiota include bacterial motility and quorum sensing. Additional metabolomic analyses showed the presence of some insecticides and herbicides in the Parac-Upper Rimac and Santa Eulalia-Parac sub-basins. The dominance in the Metropolitan area of *Arcobacter cryaerophilus*, an emergent pathogen associated with fecal contamination and antibiotic multiresistance, that is not usually reported in traditional microbiological quality assessments, highlights the necessity to apply next-generation sequencing tools to improve pathogen surveillance. We believe that our study will encourage the integration of omics sciences in Peru and its application on current environmental and public health issues.

“Aplicaciones de la secuenciación masiva (NGS) en metagenómica y secuenciación de genes”]; and the Max Planck Society [Max Planck Partner Groups: Chemical-Ecology + Pontificia Universidad Católica del Perú]. Additionally, PER was funded by the Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (Fondecyt - Perú) [grant number 34-2019, “Proyecto de Mejoramiento y Ampliación de los Servicios del Sistema Nacional de Ciencia, Tecnología e Innovación Tecnológica”]. JLR received a grant from Concytec - Banco Mundial, through the Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (Fondecyt) [grant number 022-2019]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. PW works for Universidad de Piura.

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

Introduction

Worldwide, water quality problems are associated with poverty conditions and lack of efficient sanitation, especially in developing countries [1]. These problems are recurrent in highly populated cities where most of the waste is directly washed to nearby rivers [2]. Lima, the capital city of Peru, is the second largest desert city in the world [3]. Its Metropolitan area is inhabited by more than 10 million people creating enormous challenges for environmental and public health [4]. For instance, it has been shown that water quality in Lima is a significant risk factor for pathogenic infections in children [5].

The Rimac river is the main source of drinking water for the Lima Metropolitan area. Recently, more than 700 pollution sources were identified by the Peruvian National Authority of Water (ANA) [6]. The river is constantly polluted by mine tailings in the Upper Rimac closer to the central Peruvian Andes, by agricultural wastewater in its middle region, and by industrial wastewater and urban sewage in the Lower Rimac within the Metropolitan area nearby the Pacific Ocean [7]. The lack of an efficient wastewater treatment in the Metropolitan region promotes the presence of potentially pathogenic bacteria such as *Escherichia coli* or *Salmonella typhi*, associated with fever and diarrhea symptoms [2, 8].

Traditionally, assessments of water quality and bacterial contamination in Peru are focused on evaluating the presence of common coliforms (e. g. *Citrobacter*, *Enterobacter*, *Escherichia*, *Klebsiella*) [9]. However, there are a plethora of likely pathogens that are not currently studied using classic methods because of taxonomic assignment problems or the existence of non-culturable phenotypes [10]. In recent years, advances in the study of bacterial communities and diversity have occurred because of the improvement of next-generation sequencing (NGS) methodologies. One of these techniques, 16S rRNA metabarcoding, uses a fragment of the 16S ribosomal gene to obtain a diversity profile of the bacterial community in a specific environment. Therefore, it has been used to study different types of samples from the internal microbiota of several species, including humans to environmental community surveys [11].

16S rRNA metabarcoding has been also used to study bacterial communities from several rivers. For instance, a study in the Danube river, which crosses many countries from central Europe, found a higher microbial community richness in the Upper basin, while in the Lower basin, there was a predominance of only a few free-living and particle associated bacteria [12]. Studies in rivers have also looked for the occurrence of potential pathogens. A report from the river Tama (Tokyo, Japan) showed that the predominant bacteria genus was *Flavobacterium* (Bacteroidia), a freshwater fish pathogen [13]. Recently, a study in the Pinheiros river (Sao Paulo, Brazil), one of the most polluted Brazilian rivers, found the predominance of *Arcobacter cryaerophylus* (Campylobacteria). This species is considered an emerging pathogen and an indicator of fecal contamination [14, 15].

Information from metabolomics provides additional insight into chemical contaminants in water that may influence bacterial diversity and would be useful to indicate the health status of an aquatic ecosystem and understand interactions between microbial communities and their environment [16, 17]. For instance, a study in the Brisbane river (Australia) found that human interference was associated with a population increase of Actinomycetes and Pseudomonadales. This increment was linked to higher levels of sugar alcohols, short-chain fatty acids and aromatic amino acids which contribute towards biofilm production [16].

Aquatic bacteria play a key role in water ecosystems being involved in biogeochemical processes and attenuation of biological and chemical pollutants [18]. Although some bacteria can have an adverse impact causing water quality deterioration and public health problems. We agree with Paruch et al. [18] that the understanding of aquatic microbial diversity,

composition, and dynamics is important for water quality assessments and management strategies for sustainable ecosystem functions.

Thus, we aim to provide for the first time an overview of the microbiota from the Rimac river. We aim to compare diversity patterns from the Rimac river sub-basins and show differences in the bacteria composition from Andean to Metropolitan areas. Our study is exploratory and could be used as evidence for future rapid biodiversity surveys or water quality assessments focused on microbial diversity.

Materials & methods

Area of study and sampling

In January 2020, we sampled river water in 13 points along the extension of the Rimac river, within provinces of Huarochiri, Lima and Callao (Fig 1). According to the Peruvian National Authority of Water, the Rimac basin is divided in nine sub-basins [19]. We sampled from six sub-basins, namely, Upper Rimac (Chicla), Parac-Upper Rimac (San Mateo, Tamboraque), Santa Eulalia-Parac (Huanchor, Chacahuaro), Santa Eulalia (Santa Eulalia), Jicamarca-Santa Eulalia (Huachipa, Chaclacayo), Lower Rimac, near main city bridges (Libertadores, Nuevo,

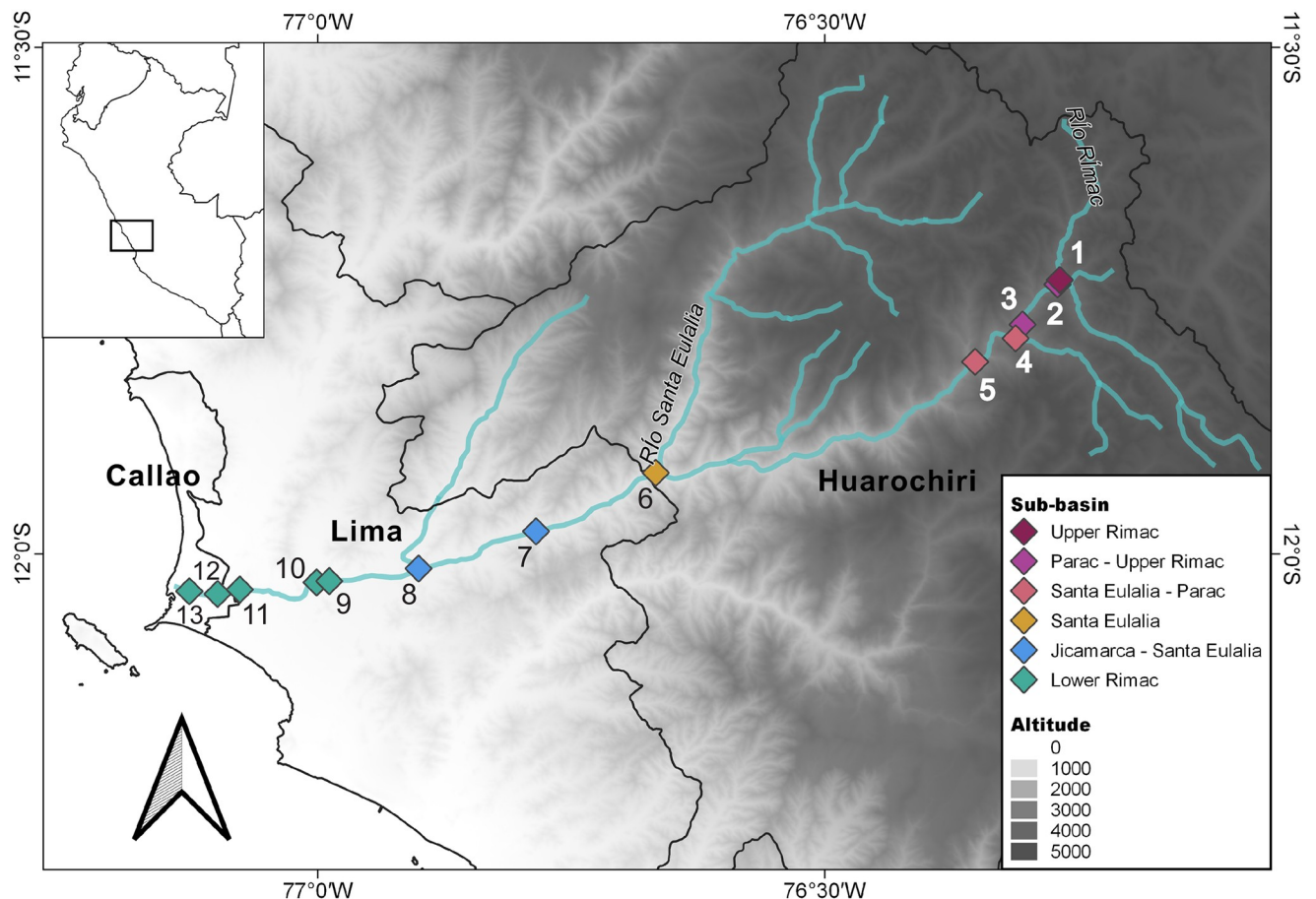


Fig 1. Area of study and sampling localities. The Rimac basin is situated between the central Peruvian Andes and the Pacific Ocean. The river runs across Callao, Huarochiri and Lima provinces in the Lima region. We sampled in different localities from the following sub-basins: Upper Rimac, 1. Chicla; Parac-Upper Rimac, 2. San Mateo, 3. Tamboraque; Santa Eulalia-Parac, 4. Huanchor, 5. Chacahuaro; Santa Eulalia, 6. Santa Eulalia; Jicamarca-Santa Eulalia, 7. Chaclacayo, 8. Huachipa; Lower Rimac (Metropolitan area), 9. Libertadores, 10. Nuevo, 11. Universitaria, 12. Faucett, 13. Gambetta.

<https://doi.org/10.1371/journal.pone.0250401.g001>

Universitaria, Faucett and Gambetta). Sterile plastic bottles were used to sample approximately 1.5 L of superficial water. When we sampled from a bridge, we attached the plastic bottle to another flask linked to a 30 m rope. In this case, we sampled superficial water (ca. 50 cm below the surface) from the middle of the bridge. Water samples were transported to the lab in coolers and were processed in the same day. Ten milliliters of water were removed for the metabolomics study. The rest of the sample was filtered river using 0.2 μM Sterivex filters for DNA extraction.

DNA isolation, amplification, and preparation of genomic libraries. Total DNA was isolated from the Sterivex filter using the PowerWater DNA kit (Qiagen) following manufacturer's instructions. Additionally, DNA isolation was performed from two empty Sterivex filters, both samples were used as blanks. Then, we quantified DNA quality using a Qubit 3.0 (Thermo Fisher Scientific, USA) fluorometer. We followed the standard Illumina 16S rRNA genomic library preparation protocol [20] which consisted in the amplification of the 16S rRNA V3-V4 region (ca. 420 bp). A second amplification was performed to attach oligo adapters (indexes) to each amplicon sample. Indexes were informative to differentiate samples after the sequencing step. Each step was followed by amplicon cleaning using AMPure XP beads (Beckman Coulter, USA). In the final step, all samples were pooled and sequenced using the MiSeq v3 Reagent Kit (Illumina, USA) on the MiSeq instrument (Illumina). Raw sequences from the genomic libraries were deposited in the NCBI BioProject database (accession number: PRJNA646070) (S1 Table).

Bioinformatics and data analyses

Sequencing reads per each sample were analyzed using DADA2 [21] in R v3.6.3 [22]. Reads from replicate samples, with the exception of Chacahuaro and Universitaria localities were also integrated in the analysis. We filtered reads using the parameter $\text{maxEE} = 2$ (maximum number of expected errors) for quality threshold as suggested before [23, 24]. DADA2 was used to infer Amplicon Sequence Variants (ASV) which are groups of identical sequences. We also used DADA2 for taxonomic assignment, comparing ASV sequences against the ribosomal database SILVA v.138 [25] with the function "*assignTaxonomy*". Identification up to the species (or multispecies) was done using the function "*assignSpecies*". Only two ASVs were identified in the blanks corresponding to ASV_2999 (*Wolbachia*) and ASV_3029 (*Rickettsia*) and were removed from all samples in further analyses. The code for bioinformatic analyses can be accessed from GitHub [26]. We used the Chao1 index, a non-parametric richness index that calculates the minimal number of taxa present in a sample, and the Shannon index, which estimates the taxa diversity in a sample taking into account both abundance and evenness [27]. Both indexes were estimated using the package phyloseq [28].

To compare Chao1 and Shannon indices from each sub-basin we used the one-way ANOVA and the Tukey's post hoc tests. Assumptions of normality and heteroscedasticity were checked with Shapiro-Wilks and Breusch-Pagan tests, implemented in the packages agricolae [29] and car [30], respectively [31–33]. To compare beta diversity among localities, we normalized the ASV matrix using the variance stabilizing transformation (VST) method available in the package DESeq2 [34], as suggested before for microbiome analyses [35]. Then, we used a multidimensional scaling (MDS) approach based on using a Bray-Curtis similarity matrix to compare beta diversity among localities.

In addition, we used the taxonomic and occurrence information to produce a stacked bar plot of the most predominant classes. Then, we explored the most predominant genera using the package ampvis2 [36]. Next, we selected ASV with more than 1 000 counts and a taxonomic assignment until species level. We used this information to look for potential human

and other animal pathogens in the Risk Group database [37], specifically if predicted species are pathogens for humans or other animals. Finally, we used Piphillin [38, 39] to predict functional content based on the frequency of the 16S rRNA sequences comparing them to annotated genomes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Predicted KEGG orthologues (KO) occurrence was retrieved from KEGG (May 2020) using a 97% cutoff threshold to create a gene feature table. We used the 10 most frequent unique KO per locality, then we grouped localities by sub-basin and created a Venn diagram to look for shared KOs for each region using DeepVenn [40].

Finally, we did a redundancy analysis (RDA) using information of environmental parameters (pH, temperature, electric conductivity [EC], nitrate concentration) obtained from a report of the Peruvian National Authority of Water (ANA) [41] (S2 Table). Eight sample sites were located to the near points evaluated in the ANA report so environmental information could be comparable. Significant identified compounds from the metabolomic analysis were also added to the RDA. The RDA was performed using a Bray-Curtis distance matrix estimated from logarithmic transformed count data that prevents placing too much weight on extremely high values of ASVs. Forward selection was carried out to calculate the relative significances of abiotic variables, using the package *adespatial* [42] in R. We also performed a Spearman correlation in R among environmental parameters, significant identified compounds in the metabolomic analyses, and bacterial classes frequency in the sampling locations. P-values were corrected for multiple comparisons using the FDR correction.

Metabolomic analyses

Ten milliliters of each water sample and Milli Q water (Blank) were filtered through a Nal-gene™ 0.22 µm syringe filter (Thermo Scientific) to remove suspended particulates. Each water sample was immediately extracted with dichloromethane (ACS grade, J.T. Baker) (3 times x 10 mL). After that, dichloromethane extracts were concentrated in vacuo and placed in a vial with 150 µL of dichloromethane and 2 µL of 26.7 ppb of toluene (GC grade, Sigma-Aldrich) (internal standard). The standard solution consisted in 2 ppm of toluene in dichloromethane, 2 µL of this solution was diluted with 150 µL dichloromethane. All water samples were analyzed except Santa Eulalia (6), Chaclacayo (7), Huachipa (8) and Gambetta (13) because these water samples were used completely in the previous filtering step. Water samples, blanks and internal standards were analyzed by gas chromatography (GC) coupled to an APPI-Q-Exactive HF mass spectrometer (Thermo Fisher Scientific, USA). GC was equipped with a DB-5 column (30 m x 0.25 mm i.d., 0.25 µm thickness film). The oven temperature was programmed as follows: aliquots of 2 µL sample were injected at 45 °C for 2 min, increase at 10 °C/min until 270 °C, and hold for 7.5 min. Injections were made in splitless-mode with helium as the carrier gas (1.5 mL/min), injector temperature at 270 °C, and detector temperature at 270 °C. High-accuracy MS data were acquired in positive data-dependent acquisition (DDA) with scan range m/z 50–750. Raw data from this GC-MS experiment was first converted into ABF format. Then, peak spotting was performed by exploring retention time and accurate mass. MS-DIAL v4.16 provided peak alignments of all samples and normalizes data based on TIC (total ion current). Final filtering was performed with a principal component analysis (PCA, p -anova < 0.05) using MATLAB vR2019b to track data quality, reduce the data dimensionality, identify potential outliers in the dataset, as well as to identify sample clusters. Additionally, the software Compound Discovery was used for tentative identification using MS/MS data and comparing our results against several databases such as ChemBioFinder, Chemspider, Kegg, LipidBank, LipidMaps, Metlin and NIST. Besides, a PCA was performed to observe differences in mass profiles among samples group.

Ethics statement

The study did not involve experiments in humans nor animal subjects. We did not sample in protected areas nor protected species were sampled.

Results

Characterization of bacterial communities in the Rimac sub-basins

16S rRNA metabarcoding produced an average of 244 393 [105 680–287437] reads per sample. After discarding low quality sequences and merging forward and reverse reads, we obtained an average of 92 064 [39 496–124 58] final reads per sample (S1 Table). From this subset, we identified a total of 23 682 ASV. The first 268 ASV were represented by more than 1 000 reads and the first 19, by more than 10 000 reads (S3 Table). Richness (Chao1 index) and diversity (Shannon index) was significantly lower for the Lower Rimac sub-basin in comparison to the other sub-basins (Fig 2). In the MDS analysis, samples from the Lower Rimac clustered tightly together and are less similar to other sub-basins. Samples from Parac-Upper Rimac and Santa Eulalia-Parac sub-basins did not clustered together.

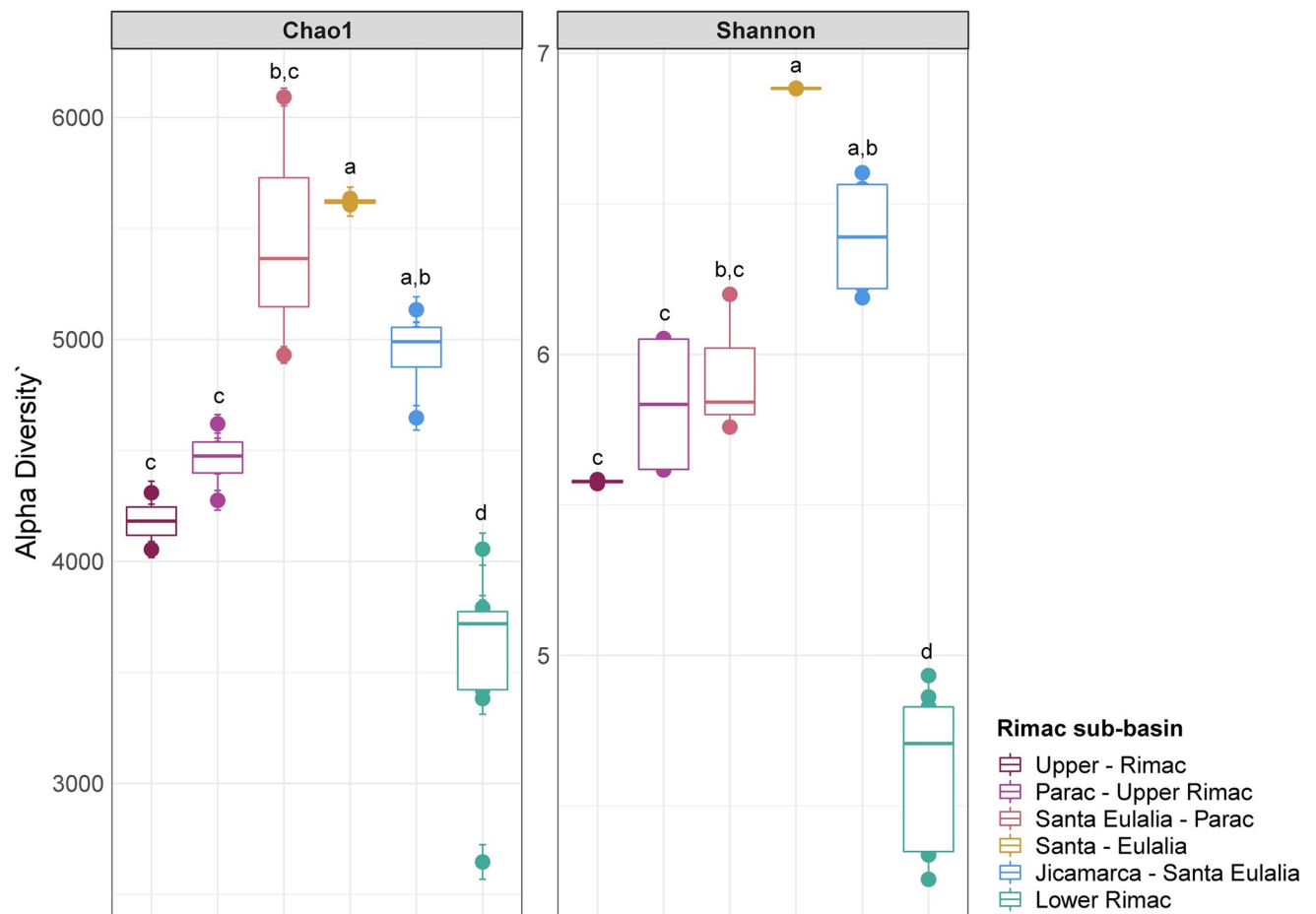


Fig 2. Alpha diversity indexes, Chao1 (richness) and Shannon (diversity) in the samples. The Lower Rimac localities consistently showed lower richness and diversity values in comparison to other sub-basins. Letters above boxplots represent significant differences ($p < 0.05$) between average values of each sub-basin.

<https://doi.org/10.1371/journal.pone.0250401.g002>

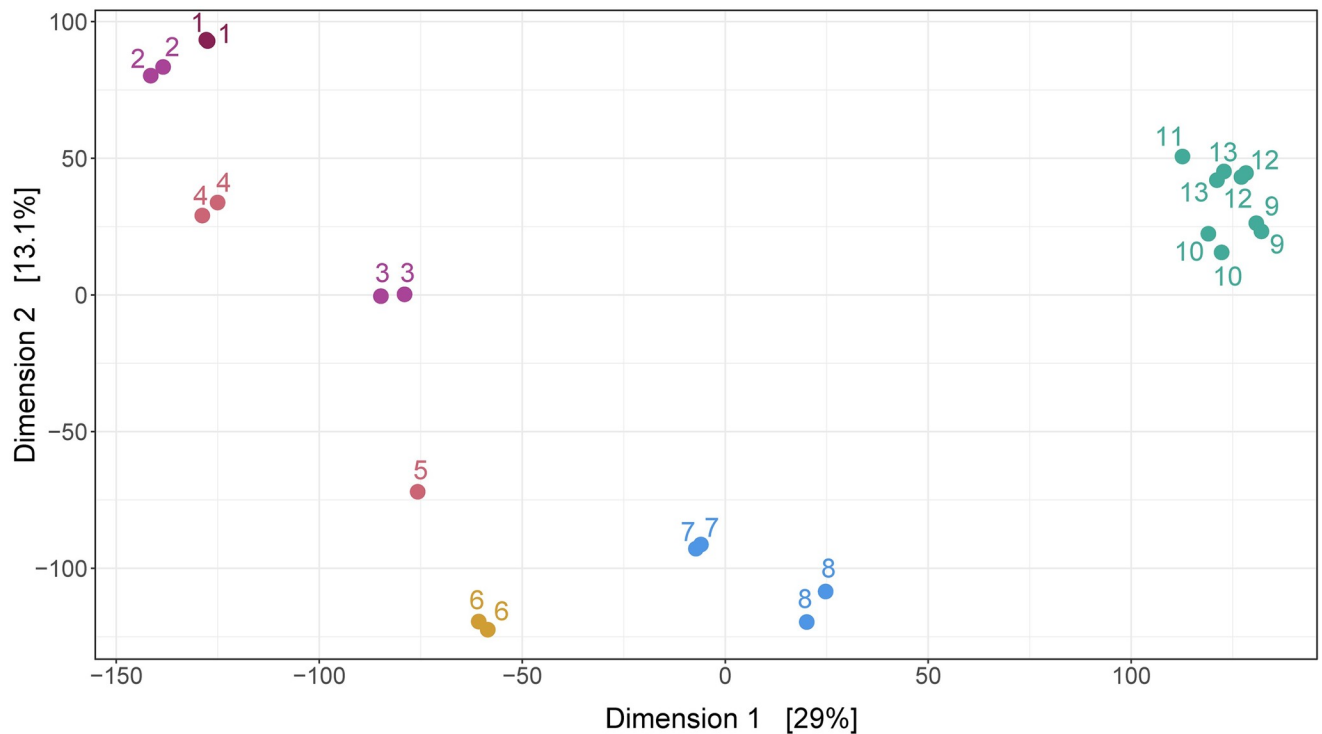


Fig 3. Multi-Dimensional Scaling (MDS). Principal coordinate analysis (PCoA) of ASVs shows similarities among groups of samples. Lower Rimac localities (green) clustered together and are less similar to other sub-basins. Samples are numbered according to the information in Fig 1.

<https://doi.org/10.1371/journal.pone.0250401.g003>

Aquatic bacteria community diversity seems higher in the Upper and middle Rimac than in the Lower Rimac (Fig 3).

Phyla Bacteroidota, Campylobacterota, Firmicutes, Fusobacteriota and Proteobacteria represented 75.75% (18 078 from 23 864) of the total ASVs (S4 Table). Taxonomic identification of ASV was higher for phylum, class, order, and family ranks (95–99%), and moderate for genus and species ranks (84 and 66%, respectively) (S4 Table). Classes Alphaproteobacteria, Bacteroidia, Campylobacteria, Clostridia, Fusobacteriia, and Gammaproteobacteria were the consistently common along all zones (Fig 4). Fusobacteriia declines from Chacahuaro (Santa Eulalia-Parac) to Huachipa (Jicamarca- Santa Eulalia). Campylobacteria frequency rises in the Lower Rimac zone (Metropolitan area). The most frequent genera (Fig 5) in the Upper Rimac, Parac-Upper Rimac, and Santa Eulalia-Parac sub-basins were *Hypnocyclicus* (ASV2) (Fusobacteriia) and *Flavobacterium* (ASV3) (Bacteroidia), the latter was also predominant in Santa Eulalia and Jicamarca-Santa Eulalia sub-basins. In the Lower Rimac, we found a dominance of *Arcobacter* (Campylobacteria). Human pathogenic bacteria identified to species level were *Arcobacter cryaerophilus* and *Prevotella copri* (S5 Table), and to the genus level were *Aeromonas*, *Escherichia*, *Pseudomonas* and *Shigella*. Other animal pathogens identified were *Flavobacterium succinicans* and *Faecalibacterium prausnitzii*.

Piphillin functional prediction of the top unique KEGG orthologies (KO) for each locality revealed 7 KO in the Upper Rimac sub-basin, 11 in Parac-Upper Rimac, 18 in Santa Eulalia-Parac, 10 in Santa Eulalia, 12 in Jicamarca-Santa Eulalia, and 9 in the Lower Rimac (S6 Table). The Venn diagram shows two KO (K02030, K03406) shared by the six sub-basin zones, and two KO (K05874, K01999) were shared among all sub-basins except the Lower Rimac (S1 Fig). The most abundant predicted pathways shared by the six sub-basins (S6 Table) were related to

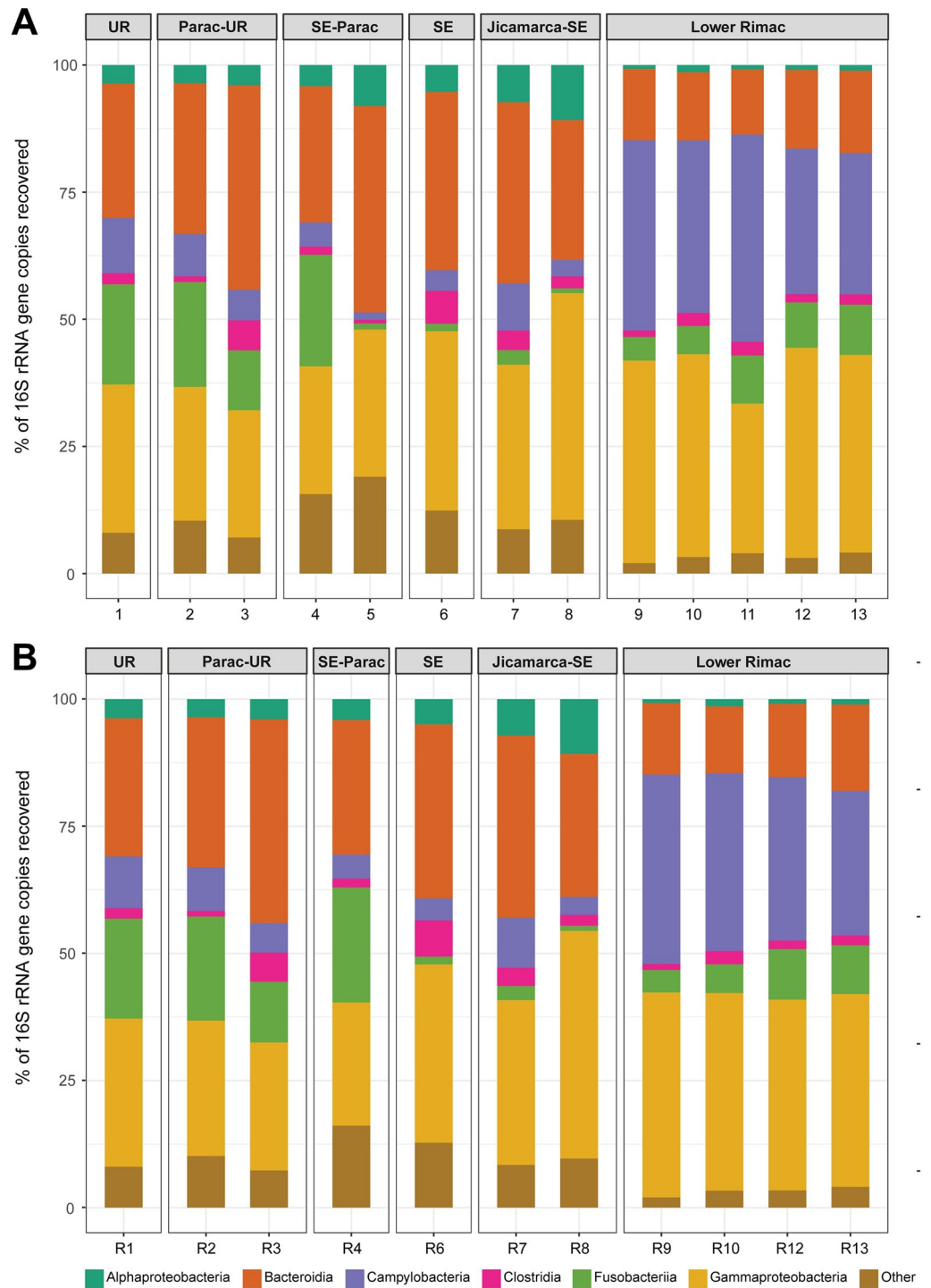


Fig 4. Stack barplot of the frequencies from the most common bacterial classes in each locality. Samples are numbered according to the information in Fig 1. UR: Upper Rimac, SE: Santa Eulalia. A. Samples, B. Replicates.

<https://doi.org/10.1371/journal.pone.0250401.g004>

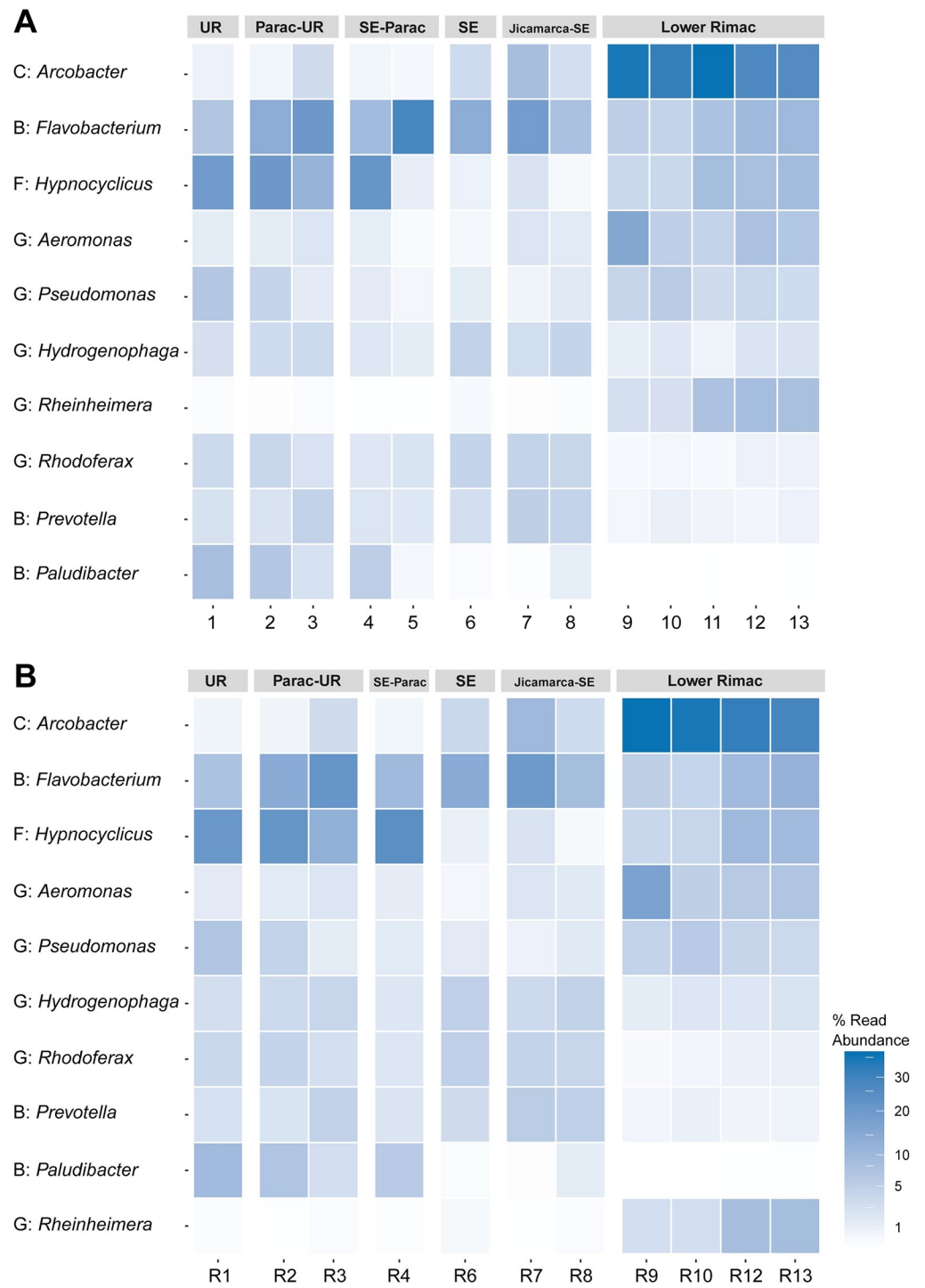


Fig 5. Heatmap of the abundance from the ten most frequent bacterial genera. Classes codes: B, Bacteroidota; C, Campylobacteria; F, Fusobacteriia; G, Gammaproteobacteria. Samples are numbered according to the information in Fig 1. UR: Upper Rimac, SE: Santa Eulalia. A. Samples, B. Replicates.

<https://doi.org/10.1371/journal.pone.0250401.g005>

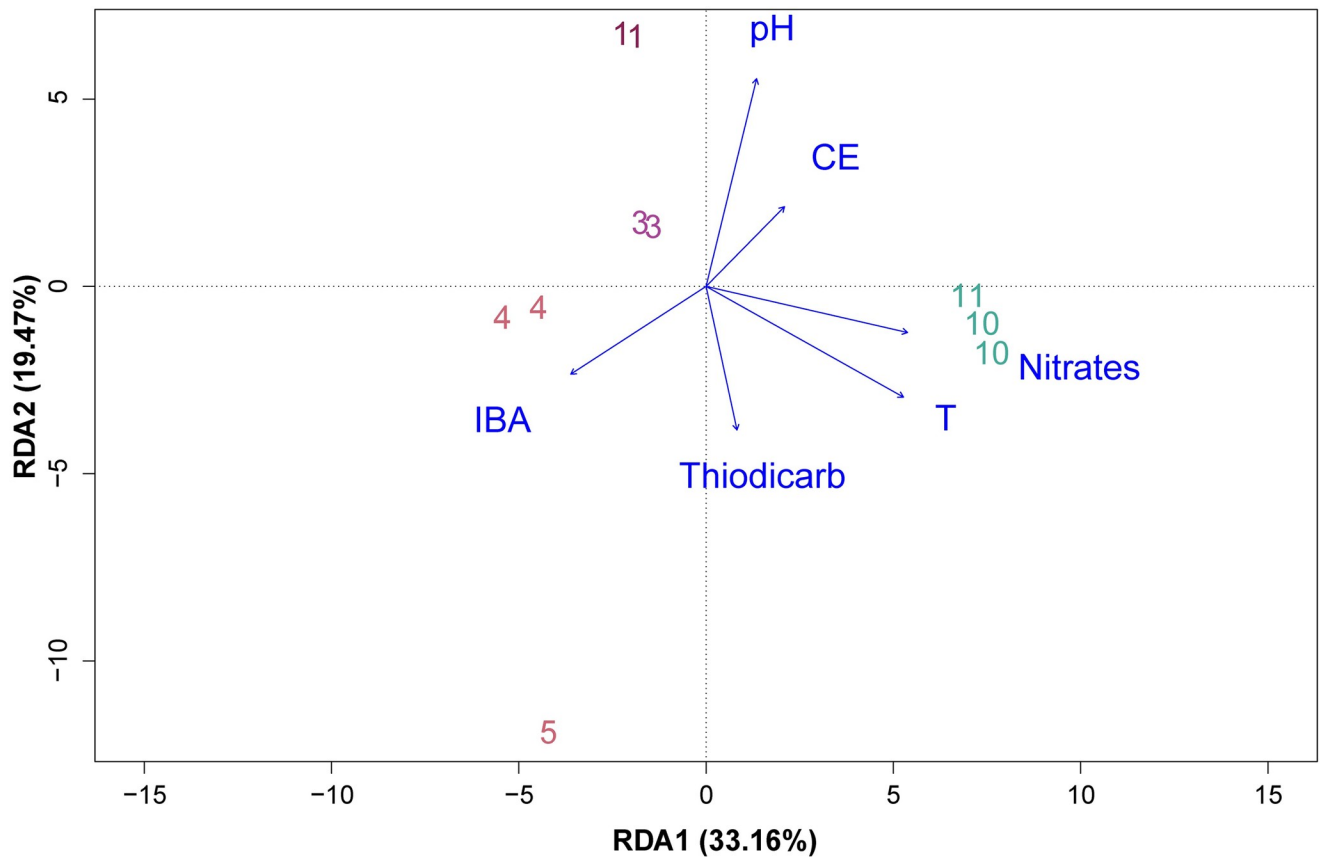


Fig 6. Redundance analysis (RDA) between species with environmental factors. Samples are numbered according to the information in Fig 1. Environmental data was taken from Ref. 34. T: Temperature, EC: Electrical conductivity. IBA: 3-indolebutyric acid.

<https://doi.org/10.1371/journal.pone.0250401.g006>

the ATP-binding cassette transporters (polar amino acid transport system substrate-binding protein, K02030), bacterial chemotaxis (methyl-accepting chemotaxis protein/MCP, K03406), and quorum sensing (K01999).

The redundant analysis (RDA) showed that the first and second axes accounted for 33.16% and 19.47% respectively, of the total variation in the community, the first axis showed that microbial communities were associated with temperature and nitrates (longest arrows), while the second axis was associated to pH (Fig 6). Temperature was found to be the most significant among explanatory variables ($p = 0.001$). Finally, the Spearman correlation coefficient (S2 Fig) showed that Gammaproteobacteria and Campylobacteria frequencies were positively correlated with higher nitrate concentration. Gammaproteobacteria frequency seems to be positively correlated with temperature.

Metabolomic analyses

The untargeted GC-MS (APPI) analysis in positive mode resulted in a curated data matrix, comprised of 166 features (i.e. m/z values) (S7 Table). Moreover, PCA performed from these results (S3 Fig), showed distinguishable separation between the Upper Rimac, Parac-Upper Rimac, Santa Eulalia-Parac, Lower Rimac, blanks and internal standard samples. Indeed, some localities from the Parac-Upper Rimac sub-basin (San Mateo, Tamboraque) and Santa Eulalia-Parac (Huanchor, Chacahuaro) were well differentiated from the other areas. However, the

Upper Rimac sample (Chicla) was clustered within the Lower Rimac because of similar chemical composition. Additionally, we tentatively identified five compounds from thirteen features statistically significant (p -anova < 0.05, total score > 95%) using Compound Discovery and MS-DIAL softwares (S8 Table), such as thiodicarb (peak 93), 3-indolebutyric acid (IBA) (peak 104) and benzenemethanamine derivatives (peaks 71, 112 and 119). The loading plot of PCA coefficients from metabolomics analysis (S4 Fig) showed that peak 104 was identified in Santa Eulalia-Parac localities and peaks 71, 92, 93, 112 and 119 were identified in Parac-Upper Rimac and in Santa Eulalia-Parac localities. Whereas the Lower Rimac samples were influenced by peak 43. Additionally, peaks 22, 23, 29–31 and 87 were identified in all zones (S8 Table).

Discussion

Our study provides the first overview of the bacterial community diversity in the Rimac river (Lima, Peru) using a 16S rRNA metabarcoding approach. We described bacterial community composition shifts along an altitude gradient (0–4 500 msl), and compared these results with environmental and metabolomics data available. We found a clear separation among bacterial communities from the Upper Rimac (> 3 000 msl) and Lower Rimac sub-basins (< 300 msl), while this was more subtle for the intermediate sub-basins. We mainly found the occurrence of phyla Bacteroidota, Campylobacterota, Fusobacteriota and Proteobacteria along the Rimac river basin. This result is similar to a previous report in the polluted Pinheiros river in Sao Paulo, Brazil [14]. The authors associated the high presence of these phyla with freshwater environments and domestic sewage sludges.

Bacterial communities in high-altitude aquatic environments are still not thoroughly studied. We found only few characterizations using 16S rRNA sequencing, especially in lakes. For instance, community profiling of Tibetan and Pyrenean lakes showed dominance of classes Actinobacteria, Alphaproteobacteria, Betaproteobacteria, and phyla Bacteroidota [43–45]. Our samples did have a major presence of the Bacteroidia class, part of the latter phylum, and class Alphaproteobacteria. However, we found no Betaproteobacteria. In the Upper Rimac and Parac-Upper Rimac sub-basins, we found a predominance of *Hypnocyclicus* and *Flavobacterium*, the latter has been related to several issues in fish such as “cold water” and *columnaris* disease [46]. In particular, *Flavobacterium succinicans*, the most frequent *Flavobacterium* (S4 Table), has been associated with bacterial gill disease in trout [47]. Additional surveys in high-altitude rivers are necessary to have a broader view of bacterial communities in these kinds of habitats.

The Lower Rimac is the essential water source for the Lima Metropolitan area, a mega-city situated on the hyper-arid Peruvian desert [4, 48]. Community composition in the Lower Rimac is influenced by similar pollution factors such as human feces and domestic sewage that are produced in the Metropolitan area [49]. This could explain why samples from the Lower Rimac looked more similar among each other than samples from other sub-basins (MDS, Fig 3). The rise of Campylobacteria frequency in the Lower Rimac is correlated with a higher nitrate concentration (S2 Table). Source of the latter are fertilizers, sewage sludge and discharges from municipal wastewater treatment [50, 51]. Thus, urban pollution in the last part of the Rimac river is influencing the bacterial diversity and the occurrence of potential pathogens.

Arcobacter cryaerophilus, the predominant species in the Lower Rimac, has been reported as very frequent in other rivers such as the Yangtze (China) [49], many rivers in Nepal [52], the Llobregat (Spain) [53], and the Pinheiros river [14]. As mentioned before, *Arcobacter* is an emergent enteropathogenic bacterium and an indicator of fecal contamination. It can also

occur in water treatment and sewage systems [54] and survive adverse conditions imposed by food processing and storage [55]. *Arcobacter* is also abundant in effluents from wastewater treatment plants [56]. Moreover, *Arcobacter* has shown resistance to several antibiotics [55] and it has been proposed to be involved in the exchange of resistance genes between gram-negative and gram-positive phyla [57] and to carry several virulence genes [58]. Notwithstanding, reports in Peru on the pathogenicity of this genus are scarce, e.g. we found only one published study that mentioned *Arcobacter* in stool from children with diarrhea [59].

The official Peruvian environmental quality standards for surface water evaluate only two microorganisms, *Escherichia coli* and *Vibrio cholerae* [60]. Our results are aligned with previous work [61, 62] that corroborate that next-generation sequencing is capable to identify potential bacterial pathogens in the water samples. Thus, we believe that a future national water surveillance system shall include evidence from amplicon sequencing and metagenomics.

The high occurrence of bacterial chemotaxis pathway, a necessary function to move towards nutrients or away toxins, would be connected to bacterial growth and survival in aquatic environments [63]. A study in the Pinheiros river found similar high occurrence of bacterial chemotaxis and flagellar assembly functions and hypothesized that this could be influenced by the elevated concentration of nutrients, e.g. ammonia and phosphates [14]. In our study, one of the most predicted KO was the MCP protein (S7 Table). This protein is important for bacterial motility because it triggers the activation of flagella and has been also encountered in other highly polluted rivers such as the Yamuna, a major tributary of the Ganges river in India [64]. In addition, we found a high occurrence of predicted ABC transporters, ubiquitous proteins involved in several processes including nutrient uptake and chemotaxis [65]. According to the KEGG annotation these proteins also take part in quorum sensing functions. Quorum sensing has been profusely studied in free-living bacteria in laboratory conditions. However, it is now known that, in nature, bacteria can use quorum sensing mechanisms in fluid environments such as rivers, streams, intertidal and marine areas by forming biofilms [66]. We found the glutathione metabolism pathway was common in Santa Eulalia and Jicamarca-Santa Eulalia. Glutathione S-transferase is involved in biodegradation of xenobiotics, defense against chemical and oxidative stress, and antibiotic resistance [67]. Rivers have been shown as reservoirs of genes related to antibiotic resistance influenced by anthropogenic causes [64, 68]. Further research is necessary to obtain environmental metagenomes from the Rimac and to look for possible genes linked to resistance to antibiotics in the Lower Rimac or resistance to heavy metals in the Upper Rimac.

An untargeted GC-MS metabolomics coupled to multivariate statistical analysis showed differences in chemical composition between some localities from the Parac-Upper Rimac and Santa Eulalia-Parac with respect to the Lower Rimac. This result may be due to more population density in the Lower Rimac and to the distance from the Metropolitan area. Samples from the Lower Rimac clustered together in both, metabarcoding and metabolomics analyses (Fig 3, S3 Fig), suggesting similar factors such as pollution affecting bacterial and chemical compound diversity, especially, in the high populated Metropolitan area.

Additionally, compound identification was attempted for the 5 features of the discriminant panel. The GC-MS (APPI) analysis allowed generating elemental formulae for unknown compounds, which together with tandem MS capability contributed to their identification. Indeed, we found thiodicarb, 3-Indolebutyric acid and benzenemethanamine derivatives in the Parac-Upper Rimac and Santa Eulalia-Parac localities. The first compound is an agricultural insecticide against some species of Lepidoptera, Coleoptera and Hemiptera [69], the second one is a plant hormone, an ingredient in many commercial horticultural plant rooting products and a food-specific compound that was detected in human urine [70], and the third one is commonly used for the manufacture of plastic. Further sampling efforts should also consider the

interaction between microbiota and chemical pollutants, product of anthropogenic activities [16].

Our study could be complemented with DNA metabarcoding analyses of other pathogenic protozoa [7] or invertebrates [71] which are also used as indicators of water quality. Additionally, it would be advisable to sample more localities between sites 5 to 8 to provide a better assessment of local diversity in the middle Rimac. Besides, we sampled in January 2020, rainy season in the Andes. Thus, it would be desirable collecting also during the dry season to compare temporal diversity patterns. Likewise, this research would be supported by other MS-based techniques like hyphenated and nonhyphenated MS to ensure the identification of family compounds in order to understand the complex interaction between microbial communities and environmental changes, natural or anthropogenic [16, 72, 73].

We showed that the integration of omics approaches with environmental sciences will be very beneficial to improve water quality assessments. We hope that the information from this work could be useful for further public health studies that could influence public policies on rural and sub-urban areas in Lima which depend on the Rimac river water. Finally, we believe that this initial work will promote more similar studies in our country and in the Latinamerican region.

Supporting information

S1 Fig. Venn diagram of most common KEGG orthologues (KO) shared among the Rimac sub-basins.

(PDF)

S2 Fig. Spearman correlation coefficient values. Bacterial ASVs frequencies were correlated with environmental parameters and significant identified chemical compounds. GC-MS: Gas chromatography—Mass spectrometry. BD: Benzenemethanamine derivative. IBA: 3-indolebutyric acid.

(PDF)

S3 Fig. Principal Component Analysis (PCA) diagram of metabolomics analysis of water samples from the Rimac sub-basins by GC-MS (APPI).

(TIF)

S4 Fig. Loading plot of PCA coefficients. The peaks represent metabolites described in S11 Table. Peak 22, Peak 23 (C₇H₁₉N₃O₂), peak 29 (C₇H₁₄F₂O₂), peak 30 (C₉H₁₉F₂N₅), peak 31, peak 43 (C₁₂H₂₁N₃S), peak 71 Benzenemethanamine derivative, C₁₆H₁₄BrN₅, peak 87 (C₁₁H₁₄N₃P), peak 92 (C₁₃H₇N₆S), peak 93 (Thiodicarb, C₁₆H₁₄BrN₅), peak 104 (3-Indolebutyric acid, C₁₂H₁₃N₂O₂), peak 112 (Benzenemethanamine derivative, C₁₆H₁₄BrN₅), peak 119 (Benzenemethanamine derivative, C₈H₂₀F₂N₄O₄). These peaks were selected using a p-anova < 0.05 and total score > 95%. These peaks were selected using a p-anova < 0.05 and total score > 95%.

(TIF)

S1 Table. Metadata associated with each sample. Summary of the original and filtered reads. Geographical location. Sample names. BioProject (SRA) accession codes.

(XLSX)

S2 Table. Environmental and metabolomic parameters for each sample. We evaluated four parameters: pH, temperature, electric conductivity (EC) and nitrate concentration; and five peaks corresponding to the metabolites benzenemethanamine derivatives, thiodicarb, and

3-indolebutyric acid.
(XLSX)

S3 Table. Frequency of ASVs in each sample.
(XLSX)

S4 Table. Taxonomic assignment for each ASV.
(XLSX)

S5 Table. Most common potential pathogens present in our samples. The biosafety information was obtained from the Risk Group database (American Biological Safety Association).
(XLSX)

S6 Table. Description of KEGG orthologues (KO) present in the Rimac sub-basins.
(XLSX)

S7 Table. Metabolomic analysis by GC-MS (APPI). Data was normalized using the total ion current (TIC) and Log10-transformation.
(XLSX)

S8 Table. Tentative identification of chemical compounds obtained by GC-MS for each in water samples from the Rimac sub-basins by GC-MS (APPI). Data was filtered using a score threshold > 95% and p-anova < 0.05.
(XLSX)

Acknowledgments

PER thanks to Yulissa Estrada and Grecia Valdivia (Universidad de Ingeniería y Tecnología, Lima, Peru) for their support during sampling in the Upper Rimac. In addition, PER thanks to Raul Condori and Enrique Santisteban for their kind support during sampling in the Metropolitan area. AJI thanks Thermo Fisher Scientific for its support in the metabolomic analyses. PER and PW thank to Arturo Gonzáles and Stefany Infante (Universidad de Piura) for their help during lab work. PER and CCV thank to Guillermo Trujillo (GenLab) for his guidance in the preparation of genomic libraries.

Author Contributions

Conceptualization: Pedro E. Romero, Camila Castillo-Vilcahuaman, Jorge L. Ramirez.

Data curation: Pedro E. Romero, Camila Castillo-Vilcahuaman.

Formal analysis: Pedro E. Romero, Erika Calla-Quispe, Camila Castillo-Vilcahuaman, André Ampuero, Alfredo J. Ibáñez.

Funding acquisition: Pedro E. Romero, Camila Castillo-Vilcahuaman, Hammerly Lino Fuentes-Rivera, Alfredo J. Ibáñez, Paolo Wong.

Investigation: Pedro E. Romero, Camila Castillo-Vilcahuaman, Hammerly Lino Fuentes-Rivera, Jorge L. Ramirez.

Methodology: Pedro E. Romero, Erika Calla-Quispe, Camila Castillo-Vilcahuaman, Mateo Yokoo, Hammerly Lino Fuentes-Rivera, André Ampuero.

Writing – original draft: Pedro E. Romero, Erika Calla-Quispe, Camila Castillo-Vilcahuaman, Mateo Yokoo, André Ampuero, Paolo Wong.

Writing – review & editing: Pedro E. Romero, Erika Calla-Quispe, Camila Castillo-Vilcahuan, Jorge L. Ramirez, Alfredo J. Ibáñez, Paolo Wong.

References

1. Chalchisa D, Megersa M, Beyene A. Assessment of the quality of drinking water in storage tanks and its implication on the safety of urban water supply in developing countries. *Environmental Systems Research*. 2018; 6:12.
2. Abraham WR. Megacities as sources for pathogenic bacteria in rivers and their fate downstream. *Int J Microbiol*. 2011; 2011:1–13. <https://doi.org/10.1155/2011/798292> PMID: 20885968
3. Edelman DJ. Managing the Urban Environment of Lima, Peru. *Advances in Applied Sociology*. 2018; 08:233–84.
4. Silva J, Rojas J, Norabuena M, Molina C, Toro RA, Leiva-Guzmán MA. Particulate matter levels in a South American megacity: the metropolitan area of Lima-Callao, Peru. *Environmental Monitoring and Assessment*. 2017; 189:635. <https://doi.org/10.1007/s10661-017-6327-2> PMID: 29134287
5. Wesley IV. Helicobacter and Arcobacter: Potential human foodborne pathogens? *Trends in Food Science & Technology*. 1997; 8:293–9.
6. Autoridad Nacional del Agua. Diagnóstico inicial para el Plan de gestión de recursos hídricos en el ámbito de las cuencas Chillón, Rímac, Lurín y Chilca. Lima, Peru: Ministerio de Agricultura; 2019. p. 151.
7. Bautista M, Bonatti TR, Fiuza V, Terashima A, Canales-Ramos M, Jose J, et al. Occurrence and molecular characterization of *Giardia duodenalis* cysts and *Cryptosporidium* oocysts in raw water samples from the Rimac River, Peru. *Environ Sci Pollut Res Int*. 2018; 25(12):11454–67. <https://doi.org/10.1007/s11356-018-1423-6> PMID: 29423699
8. Grothen DC, Zach SJ, Davis PH. Detection of Intestinal Pathogens in River, Shore, and Drinking Water in Lima, Peru. *Journal of Genomics*. 2017; 5:4–11. <https://doi.org/10.7150/jgen.18378> PMID: 28138344
9. Buccheri MA, Salvo E, Coci M, Quero GM, Zoccarato L, Privitera V, et al. Investigating microbial indicators of anthropogenic marine pollution by 16S and 18S High-Throughput Sequencing (HTS) library analysis. *FEMS Microbiol Lett*. 2019; 366(14). <https://doi.org/10.1093/femsle/fnz179> PMID: 31418783
10. Cannon MV, Craine J, Hester J, Shalkhauser A, Chan ER, Logue K, et al. Dynamic microbial populations along the Cuyahoga River. *PLoS One*. 2017; 12(10):e0186290. <https://doi.org/10.1371/journal.pone.0186290> PMID: 29049324
11. Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data*. 2019; 6:190007. <https://doi.org/10.1038/sdata.2019.7> PMID: 30720800
12. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, et al. Bacterial diversity along a 2600 km river continuum. *Environ Microbiol*. 2015; 17(12):4994–5007. <https://doi.org/10.1111/1462-2920.12886> PMID: 25922985
13. Reza MS, Mizusawa N, Kumano A, Oikawa C, Ouchi D, Kobiyama A, et al. Metagenomic analysis using 16S ribosomal RNA genes of a bacterial community in an urban stream, the Tama River, Tokyo. *Fisheries Science*. 2018; 84(3):563–77.
14. Godoy RG, Marcondes MA, Pessoa R, Nascimento A, Victor JR, Duarte A, et al. Bacterial community composition and potential pathogens along the Pinheiros River in the southeast of Brazil. *Sci Rep*. 2020; 10(1):9331. <https://doi.org/10.1038/s41598-020-66386-y> PMID: 32518363
15. McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ Microbiol*. 2010; 12(2):378–92. <https://doi.org/10.1111/j.1462-2920.2009.02075.x> PMID: 19840106
16. Beale DJ, Karpe AV, Ahmed W, Cook S, Morrison PD, Staley C, et al. A Community Multi-Omics Approach towards the Assessment of Surface Water Quality in an Urban River System. *Int J Environ Res Public Health*. 2017; 14(3).
17. Yang L, Li Y, Su F, Li H. A study of the microbial metabolomics analysis of subsurface wastewater infiltration system. *RSC Advances*. 2019; 9:39674–83.
18. Paruch L, Paruch AM, Eiken HG, Skogen M, Sorheim R. Seasonal dynamics of lotic bacterial communities assessed by 16S rRNA gene amplicon deep sequencing. *Sci Rep*. 2020; 10(1):16399. <https://doi.org/10.1038/s41598-020-73293-9> PMID: 33009479
19. Autoridad Nacional del Agua. Estudio Hidrológico y Ubicación de la Red de Estaciones Hidrométricas en la Cuenca del Río Rímac. Lima, Peru: Ministerio de Agricultura; 2010. p. 225.
20. Illumina Inc. 16S Metagenomic Sequencing Library Preparation. USA: Illumina; 2020.

21. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016; 13(7):581–3. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047
22. R: A language and environment for statistical computing. Austria: R Foundation for Statistical Computing; 2020 [cited 2020 07 December]. <https://www.R-project.org/>.
23. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*. 2015; 31:3476–82. <https://doi.org/10.1093/bioinformatics/btv401> PMID: 26139637
24. Lee MD. A full example workflow for amplicon data. Canadá: Happy Belly Bioinformatics; [cited 2020 26 August]. https://astrobiomike.github.io/amplicon/dada2_workflow_ex.
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013; 41(Database issue):D590–6. <https://doi.org/10.1093/nar/gks1219> PMID: 23193283
26. Romero P. Microbial diversity in the Rimac river [cited 2020 26 August]. <https://github.com/quipupe/metabarcoding/wiki>.
27. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LF. Rethinking microbial diversity analysis in the high throughput sequencing era. *J Microbiol Methods*. 2011; 86(1):42–51. <https://doi.org/10.1016/j.mimet.2011.03.014> PMID: 21457733
28. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013; 8(4):e61217. <https://doi.org/10.1371/journal.pone.0061217> PMID: 23630581
29. Mendiburu F. Statistical Procedures for Agricultural Research using R Peru2017 [2021 13 March]. <https://tarwi.lamolina.edu.pe/~fmendiburu/>.
30. Fox J, Weisberg S, Adler D, Bates D, Baud-Bovy G, Ellison S, et al. Package 'car' Austria: R Foundation for Statistical Computing; 2012 [
31. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*. 2012; 10(2):486–9. <https://doi.org/10.5812/ijem.3505> PMID: 23843808
32. Giacomini Sari B, Dal'Col Lúcio A, Souza Santana C, Olivoto T, Diel MI, Ketzer Kryszczun D. Nonlinear growth models: An alternative to ANOVA in tomato trials evaluation. *European Journal of Agronomy*. 2019; 104:21–36.
33. Schützenmeister A, Jensen U, Piepho HP. Checking Normality and Homoscedasticity in the General Linear Model Using Diagnostic Plots. *Communications in Statistics—Simulation and Computation*. 2012; 41(2):141–54.
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
35. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol*. 2014; 10(4):e1003531. <https://doi.org/10.1371/journal.pcbi.1003531> PMID: 24699258
36. Albertsen M, Andersen KS, Kirkegaard RH. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data [cited 2020 26 August]. <https://madsalbertsen.github.io/ampvis2/>.
37. American Biological Safety Association. Risk Group Database 2020 [cited 2020 07 December]. <https://my.absa.org/Riskgroups>.
38. Narayan NR, Weinmaier T, Laserna-Mendieta EJ, Claesson MJ, Shanahan F, Dabbagh K, et al. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences. *BMC Genomics*. 2020; 21:56. <https://doi.org/10.1186/s12864-019-6427-1> PMID: 31952477
39. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, et al. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLOS ONE*. 2016; 11: e0166104. <https://doi.org/10.1371/journal.pone.0166104> PMID: 27820856
40. Hulsen T. DeepVenn 2020 [cited 2020 07 December]. <https://www.deepvenn.com/>.
41. Autoridad Nacional del Agua. Tercer monitoreo participativo 2013 de la calidad de agua superficial de la cuenca del río Rímac: Informe técnico. Lima, Peru.2014.
42. Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, et al. adespatial: Multivariate Multi-scale Spatial Analysis. 2020.
43. Sommaruga R, Casamayor EO. Bacterial 'cosmopolitanism' and importance of local environmental factors for community composition in remote high-altitude lakes. *Freshwater Biology*. 2009; 54:994–1005.
44. Casamayor EO. Towards a Microbial Conservation Perspective in High Mountain Lakes. In: J C, J N, M A, editors. *High Mountain Conservation in a Changing World Advances in Global Change Research* 2017. p. 157–80.

45. Xing P, Hahn MW, Wu QL. Low Taxon Richness of Bacterioplankton in High-Altitude Lakes of the Eastern Tibetan Plateau, with a Predominance of Bacteroidetes and *Synechococcus* spp. *Applied and Environmental Microbiology*. 2009; 75:7017–25. <https://doi.org/10.1128/AEM.01544-09> PMID: 19767472
46. Loch TP, Faisal M. Emerging flavobacterial infections in fish: A review. *J Adv Res*. 2015; 6(3):283–300. <https://doi.org/10.1016/j.jare.2014.10.009> PMID: 26257926
47. Good C, Davidson J, Wiens GD, Welch TJ, Summerfelt S. *Flavobacterium branchiophilum* and *F. succinicans* associated with bacterial gill disease in rainbow trout *Oncorhynchus mykiss* (Walbaum) in water recirculation aquaculture systems. *J Fish Dis*. 2015; 38(4):409–13. <https://doi.org/10.1111/jfd.12249> PMID: 24720801
48. Arana C, Carlo TA, Salinas L. Biological soil crust in Peru: first record and description. *Zonas Áridas*. 2016; 16(1):112.
49. Cui Q, Huang Y, Wang H, Fang T. Diversity and abundance of bacterial pathogens in urban rivers impacted by domestic sewage. *Environmental Pollution*. 2019; 249:24–35. <https://doi.org/10.1016/j.envpol.2019.02.094> PMID: 30877966
50. Nemčić-Jurec J, Jazbec A. Point source pollution and variability of nitrate concentrations in water from shallow aquifers. *Applied Water Science*. 2016; 7(3):1337–48.
51. Gan Y, Zhao Q, Ye Z. Denitrification performance and microbial diversity of immobilized bacterial consortium treating nitrate micro-polluted water. *Bioresour Technol*. 2019; 281:351–8. <https://doi.org/10.1016/j.biortech.2019.02.111> PMID: 30831514
52. Shrestha RG, Tandukar S, Bhandari D, Sherchan SP, Tanaka Y, Sherchand JB, et al. Prevalence of *Arcobacter* and Other Pathogenic Bacteria in River Water in Nepal. *Water*. 2019; 11:1416.
53. Collado L, Kasimir G, Perez U, Bosch A, Pinto R, Saucedo G, et al. Occurrence and diversity of *Arcobacter* spp. along the Llobregat River catchment, at sewage effluents and in a drinking water treatment plant. *Water Research*. 2010; 44:3696–702. <https://doi.org/10.1016/j.watres.2010.04.002> PMID: 20427071
54. Fisher JC, Levican A, Figueras MJ, McLellan SL. Population dynamics and ecology of *Arcobacter* in sewage. *Front Microbiol*. 2014; 5:525. <https://doi.org/10.3389/fmicb.2014.00525> PMID: 25426103
55. Ferreira S, Queiroz JA, Oleastro M, Domingues FC. Insights in the pathogenesis and resistance of *Arcobacter*: A review. *Critical Reviews in Microbiology*. 2015:1–20.
56. Kristensen JM, Nierychlo M, Albertsen M, Nielsen PH. Bacteria from the Genus *Arcobacter* Are Abundant in Effluent from Wastewater Treatment Plants. *Applied and Environmental Microbiology*. 2020; 86. <https://doi.org/10.1128/AEM.03044-19> PMID: 32111585
57. Jacquiod S, Brejnrod A, Morberg SM, Abu Al-Soud W, Sorensen SJ, Riber L. Deciphering conjugative plasmid permissiveness in wastewater microbiomes. *Mol Ecol*. 2017; 26(13):3556–71. <https://doi.org/10.1111/mec.14138> PMID: 28390108
58. Barboza K, Cubillo Z, Castro E, Redondo-Solano M, Fernández-Jaramillo H, Echandi MLA. First isolation report of *Arcobacter cryaerophilus* from a human diarrhea sample in Costa Rica. *Revista do Instituto de Medicina Tropical de São Paulo*. 2017; 59. <https://doi.org/10.1590/S1678-9946201759072> PMID: 29116292
59. Zerpa Larrauri R, Alarcón Villaverde JO, Lezama Vigo PE, Patiño Gabriel L, Reyes Dioses A, Valencia Ramírez AM, et al. Identificación de *Arcobacter* en heces de niños y adultos con/sin diarrea y en reservorios animales. *Anales de la Facultad de Medicina*. 2014; 75(2).
60. Ministerio del Ambiente. Decreto Supremo N° 004-2017-MINAM. Lima2017.
61. Vadde KK, Feng Q, Wang J, McCarthy AJ, Sekar R. Next-generation sequencing reveals fecal contamination and potentially pathogenic bacteria in a major inflow river of Taihu Lake. *Environ Pollut*. 2019; 254(Pt B):113108. <https://doi.org/10.1016/j.envpol.2019.113108> PMID: 31491696
62. Garner E, Davis BC, Milligan E, Blair MF, Keenum I, Maile-Moskowitz A, et al. Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res*. 2021; 194:116907. <https://doi.org/10.1016/j.watres.2021.116907> PMID: 33610927
63. Pandey G, Jain RK. Bacterial Chemotaxis toward Environmental Pollutants: Role in Bioremediation. *Applied and Environmental Microbiology*. 2002; 68:5789–95. <https://doi.org/10.1128/aem.68.12.5789-5795.2002> PMID: 12450797
64. Mittal P, Prasoodanan PK V, Dhakan DB, Kumar S, Sharma VK. Metagenome of a polluted river reveals a reservoir of metabolic and antibiotic resistance genes. *Environmental Microbiome*. 2019; 14:5.
65. Detmers FJM, Lanfermeijer FC, Poolman B. Peptides and ATP binding cassette peptide transporters. *Research in Microbiology*. 2001; 152:245–58. [https://doi.org/10.1016/s0923-2508\(01\)01196-2](https://doi.org/10.1016/s0923-2508(01)01196-2) PMID: 11421272
66. Emge P, Moeller J, Jang H, Rusconi R, Yawata Y, Stocker R, et al. Resilience of bacterial quorum sensing against fluid flow. *Scientific Reports*. 2016; 6:33115. <https://doi.org/10.1038/srep33115> PMID: 27650454

67. Allocati N, Federici L, Masulli M, Di Ilio C. Glutathione transferases in bacteria. *FEBS Journal*. 2009; 276:58–75.
68. Amos GCA, Zhang L, Hawkey PM, Gaze WH, Wellington EM. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. *Veterinary Microbiology*. 2014; 171:441–7. <https://doi.org/10.1016/j.vetmic.2014.02.017> PMID: 24636906
69. Richardson SD, Ternes TA. Water Analysis: Emerging Contaminants and Current Issues. *Analytical Chemistry*. 2017; 90(1):398–428. <https://doi.org/10.1021/acs.analchem.7b04577> PMID: 29112806
70. Reisdorph NA, Hendricks AE, Tang M, Doenges KA, Reisdorph RM, Tooker BC, et al. Nutrimetabolomics reveals food-specific compounds in urine of adults consuming a DASH-style diet. *Scientific Reports*. 2020; 10(1). <https://doi.org/10.1038/s41598-020-57979-8> PMID: 31980691
71. Kuntke F, de Jonge N, Hesselsøe M, Lund Nielsen J. Stream water quality assessment by metabarcoding of invertebrates. *Ecological Indicators*. 2020; 111:105982.
72. Beale DJ, Barratt R, Marlow DR, Dunn MS, Palombo EA, Morrison PD, et al. Application of metabolomics to understanding biofilms in water distribution systems: a pilot study. *Biofouling*. 2013; 29(3):283–94. <https://doi.org/10.1080/08927014.2013.772140> PMID: 23458161
73. Calla-Quispe E, Fuentes-Rivera HL, Ramirez P, Martel C, Ibanez AJ. Mass Spectrometry: A Rosetta Stone to Learn How Fungi Interact and Talk. *Life (Basel)*. 2020; 10(6). <https://doi.org/10.3390/life10060089> PMID: 32575729