Metagenomic Analyses of Surface Waters and Wastewater in the Colombian Andean Highlands: implications in health and disease

Vanessa del Pilar Urrea Messa

Universidad del Rosario
Facultad de Ciencias Naturales
Bogotá, Colombia
2024

Metagenomic Analyses of Surface Waters and Wastewater in the Colombian Andean Highlands: implications in health and disease

Estudiante

Vanessa del Pilar Urrea Messa

Tesis presentada como requisito para obtener el título de:

Magíster en Ciencias Naturales

Director

Marina Muñoz Diaz, Ph.D.
Profesora Asociada
Instituto de Biotecnología-UN (IBUN)
Universidad Nacional de Colombia

Co-director

Juan David Ramírez, Ph.D.
Profesor Titular
Facultad de Ciencias Naturales
Universidad del Rosario

Facultad de Ciencias Naturales
Maestría en Ciencias Naturales
Universidad del Rosario
Bogotá, Colombia
2024

Agradecimientos

En este enriquecedor y desafiante proceso, quisiera expresar mi profundo agradecimiento a todas las personas que me han acompañado a lo largo de un camino lleno de descubrimientos, crecimiento y experiencias inolvidables. En primer lugar, quiero dedicar unas palabras de gratitud a mi madre, *Pilar Messa*, quien ha sido mi motor y mi guía constante. Su fuerza, determinación y sacrificio han sido cruciales en cada momento, motivándome a superar cada obstáculo y alcanzar mis metas. Mamá, gracias por siempre estar a mi lado, impulsándome a ser mejor, gracias por motivarme con tu ¡sí se puede!, por tu amor y dedicación, hoy puedo decir que lo logramos porque esta meta fue también nuestro sueño y esto es por nosotras. Gracias por todo, gracias por ser ese pilar esencial en mi vida. También quiero expresar mi gratitud a mis hermanos y a mi sobrina, cuyo ánimo y compañía fue esencial en este camino. No puedo dejar de agradecer a mis tres peluditos, quienes han estado a mi lado durante cada etapa de este proceso.

Por otro lado, quiero expresar mi profundo agradecimiento a mi directora de tesis, *Marina Muñoz*, quien me ha brindado su apoyo incondicional a lo largo de este camino y ha sido una excelente guía. Siempre me ha impulsado y retado a alcanzar la excelencia profesional. Gracias por brindarme todas aquellas herramientas para hacer de esta meta y sueño realidad. Mary, gracias porque desde el pregrado me permitiste confiar en mí, me ensañaste cosas muy valiosas y definitivamente tuve la oportunidad de conocer a una excelente persona y profesional. A mi codirector de tesis, *Juan David Ramírez*, te agradezco por la confianza depositada en mi al permitirme continuar con esta línea de investigación que tanto me apasiona. Gracias por brindarme los espacios y recursos para hacer de este proyecto una realidad. Juan, siempre tendré palabras de total y absoluto cariño y agradamiento hacia ti. A ambos, quiero expresarles mi infinita gratitud, ya que han sido un ejemplo a seguir en mi vida. Trabajar con profesionales tan inspiradores como ustedes me ha permitido alcanzar metas que nunca imaginé posibles. Gracias por mostrarme mi potencial y por ayudarme a desplegar mis alas, demostrándome siempre que tengo la capacidad de conquistar todos mis sueños.

También, quiero agradecer a mis colegas y compañeros del Grupo de Investigación en Microbiología de la UR – GIMUR, ya que, durante toda mi estancia en el grupo, cada uno logró aportar un grano de arena en la construcción de la profesional que soy hoy en día. Agradecer su apoyo, participación, animo, cariño y compromiso aportando siempre para mejorar cada uno de nuestros proyectos. Además, deseo agradecer especialmente a *Luisa Páez* y *Natalia Velázquez*, quienes no solo fueron unas amigas, sino también un apoyo incondicional durante todo este proyecto. Gracias, Luisa y Natuchis, por su sabiduría y por cada consejo certero que me brindaron. Su contribución fue invaluable y significó mucho para mí.

Quiero agradecer al comité de posgrados de la Universidad del Rosario, al Grupo de Investigación en Materiales Funcionales y Catálisis – GIMFC y a los jurados, Dra. *Victoria Villegas* y Dra. *Patricia Acosta*, por dedicar valioso tiempo e invaluables aportes durante la revisión de esta tesis. Finalmente, mencionar que este proceso fue muy enriquecedor, y que el apoyo de cada persona que contribuyó en este proyecto me impulsa a seguir adelante con mayor motivación y determinación en mi camino de investigación, especialmente a nivel microbiológico y de salud ambiental.

Resumen

El agua es un recurso esencial en nuestro planeta, especialmente en los entornos de agua dulce, fuente vital tanto para humanos, animales y ecosistemas; además es importante destacar que es un recurso limitado. Sin embargo, dichos entornos se han visto impactados significativamente por la contaminación y degradación generada por las actividades humanas. El objetivo de este estudio fue caracterizar el microbioma de aguas residuales y superficiales del Río Pasto, representando el primer panorama del estado actual de este recurso en las tierras altas andinas del suroeste de Colombia. Para lograr esto, realizamos análisis fisicoquímicos y microbiológicos tradicionales, incluida la detección de material genético de los parásitos protozoarios *Giardia* spp. y *Cryptosporidium* spp. Además, utilizamos tecnología Illumina para secuenciar las muestras para análisis metagenómicos. Estos análisis nos permitieron explorar el microbioma de las aguas residuales y superficiales a niveles taxonómicos y funcionales utilizando dos enfoques. El primer enfoque fue a partir de lecturas que revelaron las familias más abundantes en cada punto de muestreo, junto con especies que exhiben características patógenas potenciales, como *Aeromonas media*, y aquellas con roles potencialmente benéficos, como *Polaromonas naphthalenivorans*. También se identificaron marcadores moleculares de importancia en salud en cada punto de muestreo, incluidos marcadores de resistencia antimicrobiana como tetraciclinas y aminoglucósidos, así como factores de virulencia. El segundo enfoque involucró el ensamblaje de genomas a partir de metagenomas (MAGs), que condujo a obtener 270 ensamblajes de alta calidad y la identificación de 16 especies bacterianas, que incluyen especies reportadas en heces y nativas del cuerpo, para las cuales se obtuvieron sus perfiles funcionales. Este estudio proporciona información del estado actual del Río Pasto, que abarca tanto las aguas superficiales como las aguas residuales no tratadas, y ofrece una herramienta valiosa para evaluar los riesgos potenciales asociados con el reúso del agua y las implicaciones de la descarga directa de contaminantes en los cuerpos de agua.

Artículo

Vanessa Urrea, Luisa Páez-Triana, Natalia Velásquez-Ortiz, Milena Camargo, Luz H Patiño, Laura Vega, Nathalia Ballesteros, Arsenio Hidalgo-Troya, Luis-Alejandro Galeano, Juan David Ramírez and Marina Muñoz. Metagenomic Analyses of Surface Waters and Wastewater in the Colombian Andean Highlands: implications in health and disease. *Current Microbiology* **(SOMETIDO)**

Metagenomic Analyses of Surface Waters and Wastewater in the Colombian Andean Highlands: implications in health and disease

Vanessa Urrea[1], Luisa Páez-Triana[1], Natalia Velásquez-Ortiz[1], Milena Camargo[1,2], Luz H Patiño[1], Laura Vega[1], Nathalia Ballesteros[1], Arsenio Hidalgo-Troya[3], Luis-Alejandro Galeano[4], Juan David Ramírez[1,5] and Marina Muñoz[1]*

[1] Centro de Investigaciones en Microbiología y Biotecnología -UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, 110221, Colombia
[2] Centro de Tecnología en Salud (CETESA), Innovaseq SAS, Funza, Cundinamarca, 250027, Colombia
[3] Grupo de Investigación Salud Pública, Departamento de Matemáticas y Estadística, Universidad de Nariño, 520002 Pasto, Colombia.
[4] Grupo de Investigación en Materiales Funcionales y Catálisis (GIMFC), Departamento de Química, Universidad de Nariño, 520002 Pasto, Colombia.
[5] Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

* **Corresponding autor**: claudia.munoz@urosario.edu.co
Carrera 26 # 63B - 48, Bogotá – Colombia
(+57-1) 297 0200 Ext. 3359

ABSTRACT

Freshwater is an essential resource on our planet; however, human activities have significantly impacted it. This study aimed to assess the microbiome of Pasto River's wastewater and surface waters in the southwestern Colombian Andean highlands. Traditional physicochemical and microbiological analyses, alongside *Giardia* spp. and *Cryptosporidium* spp parasites were detected by real time PCR. Illumina technology was used for metagenomic analyses. allowing to explore both taxonomic and functional profiling using two approaches. The first approach was from reads reveling the most abundant families, along with species exhibiting potential pathogenic characteristics, like *Aeromonas media*, and those with potential beneficial roles, such as *Polaromonas naphthalenivorans*. It also uncovered molecular markers of health significance, including common antimicrobial resistances markers linked to tetracyclines and aminoglycosides, as well as virulence factors. The second approach involved genome assembly using metagenomic data, leading to 270 high-quality assemblies and the identification of 16 bacterial species, encompassing both fecal-derived and water-body-native species, including insights into their functional relationships. This study provides a current state of the Pasto River, encompassing both surface and untreated wastewater, and offers a valuable tool for assessing potential risks associated with water reuse and the implications of the direct wastewater discharge into surface waters bodies.

Key words

Microbiome, metagenomics, rivers, molecular markers, surface water, wastewater.

INTRODUCTION

Water is a vital resource on Earth, covering approximately 71% of the Earth's surface. However, only 2.5% of this water is freshwater (1), which is essential for human beings, animals, and ecosystems (2). Surface waters refer to freshwater bodies found on the Earth's surface (3), and they have played a crucial role in human development, providing social and economic benefits that contribute to urban growth. Despite their significance, the demand for water has continuously risen, leading to negative impacts on the state of these water bodies due to human activities (4). These impacts are not limited to urbanization near rivers but also encompass agricultural, industrial, and domestic activities, particularly pronounced in developing countries (5), where resources for implementing water source conservation and treatment schemes are limited. This contamination has contributed to a surge in waterborne diseases affecting both humans and animals, while also adversely affecting the ecology of the aquatic ecosystem. Ultimately, these challenges limit the sustainable use of this precious resource (6).

In Latin America, there is a growing concern about the state of water resources because it has been affected by the increasing population density. Moreover, the current strategies for treatment of wastewater seems inadequate to deal with emerging contaminants. Pollutants detected in water bodies often result from the discharge of domestic, hospital, and wastewater, which frequently contain personal care products, pharmaceuticals, and various other substances (7). Assessing the condition of surface water quality enables us to understand the impact of human-induced activities on this vital resource. Furthermore, it aids in comprehending the consequences of this contamination on human health, animal welfare, and ecosystems (8). This is due to water bodies hosting a variety of microorganisms, including a wide diversity of protozoan, bacteria, and viruses, some of which are associated with fecal contamination (9) and antibiotic resistance (10). Additionally, protozoan parasites like Giardia spp. and Cryptosporidium spp. are a fecal contamination indicative and have been linked to waterborne intestinal diseases, which can negatively impact human health (11). Moreover, it is essential to monitor physicochemical parameters that can influence these microbial communities (9–15). Assessing the condition of surface water quality enables us to understand the impact of human-induced activities on this vital resource. Molecular markers of interest in health, such as those conferring antibiotic resistance and encoding virulence factors have been detected, posing a risk to human and animal well-being, as aquatic environments can facilitate the spread of these contaminants to other areas (16). As a result, considerable efforts have been directed towards characterizing the microbiome (17). This not only provides information about the diverse microbial communities present in these water bodies but also offers insights into their interactions and genetic material, shedding light on ecosystem processes, functionality, adaptability, and the services they can provide (18).

Currently, advances in sequencing have allowed us to understand the composition of the microbiome in environmental samples. Metagenomics has become a popular tool for characterizing the microbiome of multiple samples, because it provides insights into taxonomic profiling of all microorganisms present in the sample, including those that cannot be cultured, without the biases associated with the amplification process. For instance, in the United States, Chopyk and colleagues conducted a study where they characterized the microbial composition in both freshwater and saline surface waters over time. Metagenomics, in addition to taxonomic classification of the

microorganisms, facilitated the identification of antibiotic resistance genes, enhancing the understanding of the risks associated with the use of these water bodies in activities like agriculture (19). Conversely, in Southern China, a study was carried out to profile pathogenic microorganisms and antibiotic resistance genes in wastewater and surface waters. Through metagenomic analysis, they discovered that the composition in these water bodies was similar, suggesting that the presence of these pathogens in surface waters might be linked to the direct discharge of untreated wastewater (20). In Korea, metagenomic assembled genomes (MAGs) from surface water samples provided detailed information about the microbial members, the impact of resistomes, and the understanding species without known cultivated representatives (21). A study in Brazil also highlighted the importance of conducting tests to identify microorganisms with pathogenic potential in surface waters. This stands out the need to enhance surveillance of these water bodies and monitoring of health-relevant protozoa such as *Cryptosporidium* spp. and *Giardia* spp. (22). In Mexico, next-generation sequencing was employed for the taxonomic characterization of surface waters. This enabled the identification of microorganisms with pathogenic potential and their associations with water quality (23). Finally, in Colombia, in addition to characterizing bacterial communities in wastewater, the researchers elucidated molecular markers of health interest, including those associated with antibiotic resistance. This emphasizes that wastewater and its treatment plants are critical points for discovering antibiotic resistance genes (24).

In Colombia, the drinking water quality risk is evaluated using the IRCA - Water Quality Risk Index, which considers physicochemical and microbiological characteristics. Based on the results, states (departments in Colombia) are classified into categories of high, low, medium, or no risk regarding water quality. Specifically, the department of Nariño in Southwest Colombia reports medium level of risk in water quality according to this index, and there are complementary reports of health issues related to water quality in the region, including waterborne diseases (25). This concern is reinforced by the fact that Pasto River, one of the main rivers in the capital city of Nariño, supplies about 85% of the city's drinking water (27). However, only routine microbiological methods have been applied to understand the health risk imposed in the Pasto River.

Therefore, this study aims to create a comprehensive monitoring and surveillance tool for the Pasto River, covering both surface and wastewater. This was achieved by using advanced techniques, such as shotgun metagenomics to describe microbiome characterization, combined with traditional physicochemical analysis and the detection of *Giardia* spp. and *Cryptosporidium* spp. Additionally, we include analyses to identify molecular markers of health significance, particularly those related to antibiotic resistance and virulence factors. The research was proposed as a longitudinal study for the Pasto River, which is one of the main tributaries in the department of Nariño and has been identified as facing a high level of water quality risk. Understand the composition of the circulating microorganisms will allow the identification of potential health hazards for both animals and humans. This is crucial, as the river's water is utilized in various sectors, including domestic and industrial applications.

METHODS

Sample Collection

Pasto River has played a crucial role in the city's development, but it is currently facing challenges due to the impact of various pollutants resulting from the growing demand for this resource. The river is divided into three main basins. The upper basin serves as a productive area for the agricultural sector. The middle basin, mostly located in the city of San Juan de Pasto, is a residential and urban area with discharges primarily from domestic and industrial sources. The lower basin is predominantly used for agricultural crops (27–29). It has been reported that approximately 90% of the domestic wastewater from the municipality is discharged without any prior treatment as it flows out of the city of San Juan de Pasto. Moreover, in the northern part of the city, there is a significant pollution load from the Juan XXIII collector, which gathers wastewater from the urban area (28).

To determine the necessary number of samples, we used EpiDat software (v.3.1) (30) for sample size calculation. This calculation was based on the research conducted by Köchling *et al* (31), who collected samples from a river in northeastern Brazil exposed continuously to industrial and domestic contaminants. Building on their findings, we assessed bacterial diversity, considering an expected diversity of 79%, a precision of 5%, and a 95% confidence interval. According to these parameters, it was determined that a minimum of five water samples is required for each type of sample (surface water and wastewater). With this information in hand, we proceeded to collect samples.

We obtained a total of 50 surface water samples, which covered the upper basin in the city's south (S1), the middle basin in the urban area (S2), and the lower basin in the city's north (S3). Additionally, we collected 50 wastewater samples from the Juan XXIII wastewater collector located in the northern side of Pasto city (R4). This extensive study spanned seven months, running from August 2022 to March 2023, with an exception for January due to logistical constraints (Figure 1). This was done to better comprehend the shifts in microbiome composition over time and to facilitate comparisons across the collection points.

Sample collection was accomplished using a collection model that eased access to water bodies via bridges or the riverbank (Figure S1-A information in the following link: GitHub - gimur/Metagenomics-of-Surface-Water-and-Wastewater-for-Environmental-Monitoring). This model featured a plastic container connected to a 6-meter-long rope. Additionally, different collection models were employed for gathering samples from both surface water and wastewater sources. Before sampling, the container was cleaned with potable water and submerged multiple times in the surface water or wastewater to obtain the final sample. Two distinct sampling methods were utilized. First, a single-sample collection involved gathering the entire sample (400 mL) at a specific location during a designated time. The second method employed was composite collection. In this case, the sample volume was reach from sub-samples obtaining a total of 400 mL.. For surface water samples, collection took place at the same location with varying collection times. Two sub-samples were obtained every 15 minutes to ensure representation of the basin. Regarding wastewater samples, composite collection was conducted at the Juan XXIII wastewater collector, comprising 12 sub-samples taken at hourly intervals to constitute the complete sample. Finally, a total of 102 samples were gathered, comprising 16 from point S1, 20 from point S2, 19 from point S3, and 47 from point R4, making a total of 47 wastewater samples and 55 surface water samples (Figure S1-B information in the following link: GitHub - gimur/Metagenomics-of-Surface-Water-and-Wastewater-for-Environmental-Monitoring). To preserve sample integrity during transport to the laboratory, they

were stored in properly labeled sterile glass containers and placed in a portable refrigerator, maintaining a temperature of approximately 4°C.

Physicochemical analysis of surface and wastewater samples

For the physicochemical analyses, each month an aliquot of the samples composed of points S1, S2, S3 and R4 was implemented to analyze all the variables. Additionally, we conducted *in-situ* measurements using a multiparameter instrument (HANNA instruments). This instrument recorded pH values, water and ambient temperatures, conductivity, and dissolved oxygen levels. Furthermore, a sample aliquot was analyzed in a specialized laboratory to obtain additional physicochemical measurements, such as turbidity, suspended solids, alkalinity, acidity, electrical conductivity, nitrites, chlorides, phosphates, among others. All of this was accomplished through specific tests to measure each variable and by utilizing specialized equipment. To highlight the specific months where these significant differences occurred, we conducted a Kruskal-Wallis test and a Dunn's test with Benjamini-Hochberg adjustment. For this, we include significant differences * (0.01 - 0.05), and very significant differences **, *** (0.001 - 0.01, < 0.001).

Sample preprocessing and DNA extraction

In the laboratory, we concentrated the biomass from the samples through a series of filtration steps. Initially, we used 3.0 µm membranes (Millipore) to filter each sample, performing filtrations for every 100 mL until the entire sample was processed. Following this first filtration, we repeated the process with 0.45 µm membranes and subsequently with 0.22 µm membranes to ensure the recovery of microorganisms of various sizes was maximized. Following this, the membranes were placed in sterile 50 mL Falcon tubes, and the biomass was collected by gently scraping them with 15 mL of molecular-grade water, repeating this process twice for each membrane, with each round of scraping lasting 15 minutes. After centrifugation, we obtained the pellet. Subsequently, we carried out DNA extraction from this pellet using the Qiagen DNeasy Power Soil kit, following the manufacturer's instructions. We prewarmed the elution solution to 65°C for optimal results. Once we had obtained the nucleic acids from each sample, we evaluated their quality and quantity using the Nanodrop equipment.

qPCR and Analysis

To detect protozoan parasites *Giardia spp*. and *Cryptosporidium spp.*, real-time PCR was used. For *Giardia spp*., the analysis involved the use of GdF (5'-CATGCATGCCCGCTCA-3'), GdR (5'-AGCGGTGTCCGGCTAGC-3'), and GdP (6FAM/AGGACAACGGTTGCAC/MGB) primers and probes. The thermal profile included an initial 15-minutes phase at 50°C, followed by 10 minutes at 95°C. Subsequently, 40 cycles were carried out: 15 seconds at 95°C, 1 minute at 58°C, and 1 minute at 60°C. For the detection of *Cryptosporidium spp*., CcF18S (5'-GTTTTCATTAATCAAGAACGAAAGTTAGG-3'), CcR18S (5'-GAGTAAGGAACAACCTCCAATCTCTAG-3'), and CsP (6HEX/TCAGATACCGTCGTAGTCTTAACCATAAACTATGCC/TAMRA) primers and probes were employed (32), with a thermal profile of 95°C for 10 minutes. This was followed by 45 cycles: 10 seconds at 95°C and 30 seconds at 60°C. Some modifications were made to the protocol during its execution. These adjustments specifically referred to the thermal profile for the amplification of

Cryptosporidium spp. and were evaluated through laboratory tests, in which 45 cycles were implemented.

Metagenomic Sequencing and Data Analysis

After meeting the quality standards, integrity was assessed through horizontal agarose gel electrophoresis, and concentration and purity were determined using the Nanodrop equipment (Thermo Fisher Scientific). The samples were sent to Novogene for metagenomic sequencing. The objective was to attain a raw sequencing of at least four Gb per sample, and the sequencing was executed using the Illumina Novaseq platform in a Paired-End format, with each read having a minimum length of 150 base pairs.

After obtaining the results, the quality of the reads was assessed using FastQC v.0.11.9 (33) along with MultiQC v.1.6 (34) for a comprehensive evaluation. Various factors like adapter content, N content per base, and quality scores were considered. In case any of these characteristics were detected, read trimming was performed using Trimmomatic v.0.38 (35). Once the data had undergone preprocessing, clean and high-quality reads were obtained. Subsequently, taxonomic assignment was carried out using the Centrifuge software v.1.0.3-beta (36), Once the data was collected, it was visualized through Pavian (37) and RStudio. Furthermore, with these clean and high-quality reads, markers related to health were identified. For markers associated with antibiotic resistance, the Resistance Gene Identifier (RGI) software was utilized, incorporating databases like CARD v.3.1.3 (38). On the other hand, markers linked to virulence factors were identified using the Virulence Factor Database (VFDB) (39).

Assembly of Genomes from Metagenomes

For genome assembly-based analyses from metagenomes (MAGs), a series of steps were performed. Initially, clean, and high-quality reads were assembled using the Metaspades software v3.15.3 (40). Subsequently, sequences were binned using the MetaBAT, Maxbin, and Concoct software (41–43). To ensure the high quality of the bins, a refinement was conducted through DASTool (44). To assess the assembly quality and proceed with those of high quality, data processing was done with CheckM v1.1.3 (45). Following this process, the GTDB-Tk software v.1.7.0 (46) was used for the taxonomic assignment of the assemblies. The annotation of the assemblies was carried out using Prokka (47), and the results were visualized through Proksee. Additionally, for identifying both commonalities and distinctions among the various genomes found in the samples, a comparative genomic analysis was performed using the Roary software (48). Furthermore, markers of health interest were also identified using the ABRICATE software, both for those related to antibiotic resistance and those linked to virulence factors. Finally, to explore functional aspects, analyses were conducted using eggNOG mapper v.2 (49). Metabolic pathways were elucidated using KofamKOALA (50).

RESULTS

Monitoring physicochemical parameters in surface and wastewater samples

For the physicochemical analyses, we subjected the data to a normality test – Shapiro-Wilks, and it turned out that only certain variables like conductivity, dissolved oxygen percentage, pH, and water

temperature exhibited a normal distribution. However, other factors, such as fluorides, sulfates, suspended solids, among others, did not conform to this normal distribution pattern. Subsequently, we assessed whether there were any significant differences (P-value < 0.05) in the measured variables across the sampled months. When dealing with non-parametric data, the Kruskal-Wallis test was conducted, and no significant differences were detected. On the other hand, for parametric data, an ANOVA test was performed, and it showed that significant differences existed only for the pH variable. After this, a Dunn´s test with Benjamini-Hochberg adjustment was carried out, revealing significant variations in pH during the months of March and November.

Despite the absence of statistically significant differences, a descriptive comparison was made between wastewater and surface water. Wastewater shows higher average water and ambient temperatures, with a difference of approximately 3.5°C compared to surface water. Concerning phosphates, both wastewater (16.4 mg/L) and surface water (5.8 mg/L) samples exhibit elevated concentrations. In contrast, nitrites have a higher average value in surface water (0.2 mg/L) compared to wastewater (0.1 mg/L). Suspended solids in wastewater (113.6 mg/L) exceed those in surface water (72 mg/L) by roughly 41.6 mg/L. Lastly, both BOD - Biochemical oxygen demand and COD - Chemical Oxygen Demand values are higher in wastewater (311.2 mg/L; 426.7 mg/L) in comparison to surface water (163.9 mg/L; 269.4 mg/L) (Table_S4 information in the following link: GitHub - gimur/Metagenomics-of-Surface-Water-and-Wastewater-for-Environmental-Monitoring).

Detection of* Giardia *spp and* Cryptosporidium *spp. in surface and wastewater samples

Regarding the microbiological component of the study, we identified the genetic material of the protozoan parasites *Giardia* spp. and *Cryptosporidium* spp. using real-time PCR. These parasites have been responsible for widespread gastrointestinal diseases, sometimes leading to fatalities, especially among children and individuals with weakened immune systems (51). The DNA of *Giardia* spp. was detected in 76 out of 102 samples (75%), with 34 positive samples from wastewater and 42 from surface water. Samples without detectable *Giardia* spp. DNA were concentrated in three specific months. In November, six negative samples were obtained, two from the S3 surface water sampling point and four from the R4 wastewater point. Another instance where *Giardia* spp. was not detected in certain samples was in February. During this month, 12 negative samples were distributed across all sampling points, with two at point S1, one at point S2, three at S3, and six at point R4. Finally, in March, seven negative samples were identified, with two at S1, two at S3, and three at R4. Regarding the detection of *Cryptosporidium* spp., it was found in 96 out of 102 samples (94%). Notably, the detection of *Cryptosporidium* spp was higher throughout the study months, except for October (where one sample from the R4 wastewater point tested negative), November (with four negative samples at the R4 point), and March (with no detection at the S1 surface water sampling point).

Description of the aquatic microbiome: a taxonomic profiling

Afterwards, we sequenced and performed the metagenomic analysis which was conducted on the 74/102 (72,5%) samples that successfully passed the sequencing quality control checks. Sequences were not trimmed, as they exhibited high quality with scores exceeding 34, and the analysis employed the entire dataset (Table S1, S2, S3). An average of 52.8% of the reads were taxonomically classified using the Centrifuge software. The sample composition was predominantly microbial, with bacteria accounting for over 98% of the reads. The remaining fraction was divided among eukaryotes, archaea,

and viruses. Notably, no reads were associated with protozoa or fungi. In pursuit of these, additional analyses were conducted, entailing the extraction of the 18S-rRNA region and subsequent comparison with the Kraken2 database (52).

- **Temporal Analysis of Bacterial Family Dynamics in Wastewater and Surface Waters**

For this section, we examined the prevalence of the most common bacterial families at each of the sample points across the map of San Juan de Pasto. Figure 1 illustrates the locations: Point S1 is positioned in the city's southern area, Point S2 in the central district, and Point S3 in the northern part. The wastewater sampling location, R4, is also in the northern region of the city but precedes Point S3, indicating potential contamination from this point to S3 (Figure 1). No comparisons were performed for Site S1 due to the limited number of available samples (N=1). To investigate potential variations among microbial families across different sampling months, we performed a Kruskall-Wallis analysis. Interestingly, no significant differences (P-value > 0.05) were observed for Site S2. However, for Sites S3 and R4, differences did emerge. To highlight the specific months where these significant differences occurred, we conducted a Dunne test with Benjamini-Hochberg adjustment. Significant differences were noted at site S3, as illustrated in Figure S2, and marked with an asterisk (*), with a P-value of 0.04 for all families. Variations within the Bacteroidaceae family were evident between February and September. Likewise, notable differences were identified in the Rhodocyclaceae and Pseudomonadaceae families between February and November (Figure S2).

Figure 1. Top 10 most abundant families at the sampling points in the Pasto River. (S1 = wastewater sampling point 1, S2 = surface water sampling point 2, S3 = surface water sampling point 3, R4 = wastewater collector point).

Moreover, variations in family composition were observed at wastewater point R4 across the months, including significant differences (*), and very significant differences (**, ***), which were present in all months (Figure S3). It is worth highlighting the significance of very significant differences, indicated by three asterisks (***) and characterized by P-value values below 0.001, as illustrated in Figure S3. Of note, the Bacteroidaceae family exhibited significant differences, with a P-value of 0.002 when comparing September to February, and 0.006 when comparing September to March. Additionally, variations were evident between October and February (P-value 0.009). The Burkholderiaceae family displayed variations in October compared to February (P-value 0.0006) and November (P-value 0.002). In contrast, the Campylobacteraceae family presented significant differences in February compared to August (P-value 0.00002) and December (P-value 0.00009). Similarly, the Comamonadaceae family showed variations between February and October (P-value 0.0002) and during the months of August and February (P-value 0.001). The Enterobacteriaceae family revealed significant differences in March compared to October (P-value 0.0007) and between March and September (P-value 0.002). Concerning the Moraxellaceae family, changes in abundance were evident in October compared to December (P-value 0.008) and in relation to March (P-value 0.008). When examining changes in the Pseudomonadaceae family, very significant differences were identified in November compared to February (P-value 0.0003) and in relation to October (P-value 0.00004). Lastly, the Shewanellaceae family, in addition to its distinctive profile relative to surface water points, exhibited variations in February and December (P-value 0.00008) and between February and November (P-value 0.00001).

- **Assessing Species for Pathogenic and Beneficial Traits**

Then, broadly we selected the 20 most abundant bacterial species from each site, resulting in a total of 44 highly abundant species shared among them. Figure 2A presents all the bacteria for which there are reports of being potentially pathogenic. Particularly abundant species include *Bacteroides cellulosilyticus, Raoultella ornithinolytica*, and *Pseudomonas aeruginosa*. Particularly, *Aeromonas media* stands out as the most abundant species in this classification. On the other hand, although the *Burkholderia cepacia* complex is not be the most abundant, but its presence is significant among the top 44. Conversely, in the second category encompassing bacteria with potential benefits, *Thauera* sp. MZ1T and *Alicycliphilus denitrificans* emerge as notable species. Lastly, *Polaromonas naphthalenivorans* was the most abundant bacterium in this classification (Figure 2B).

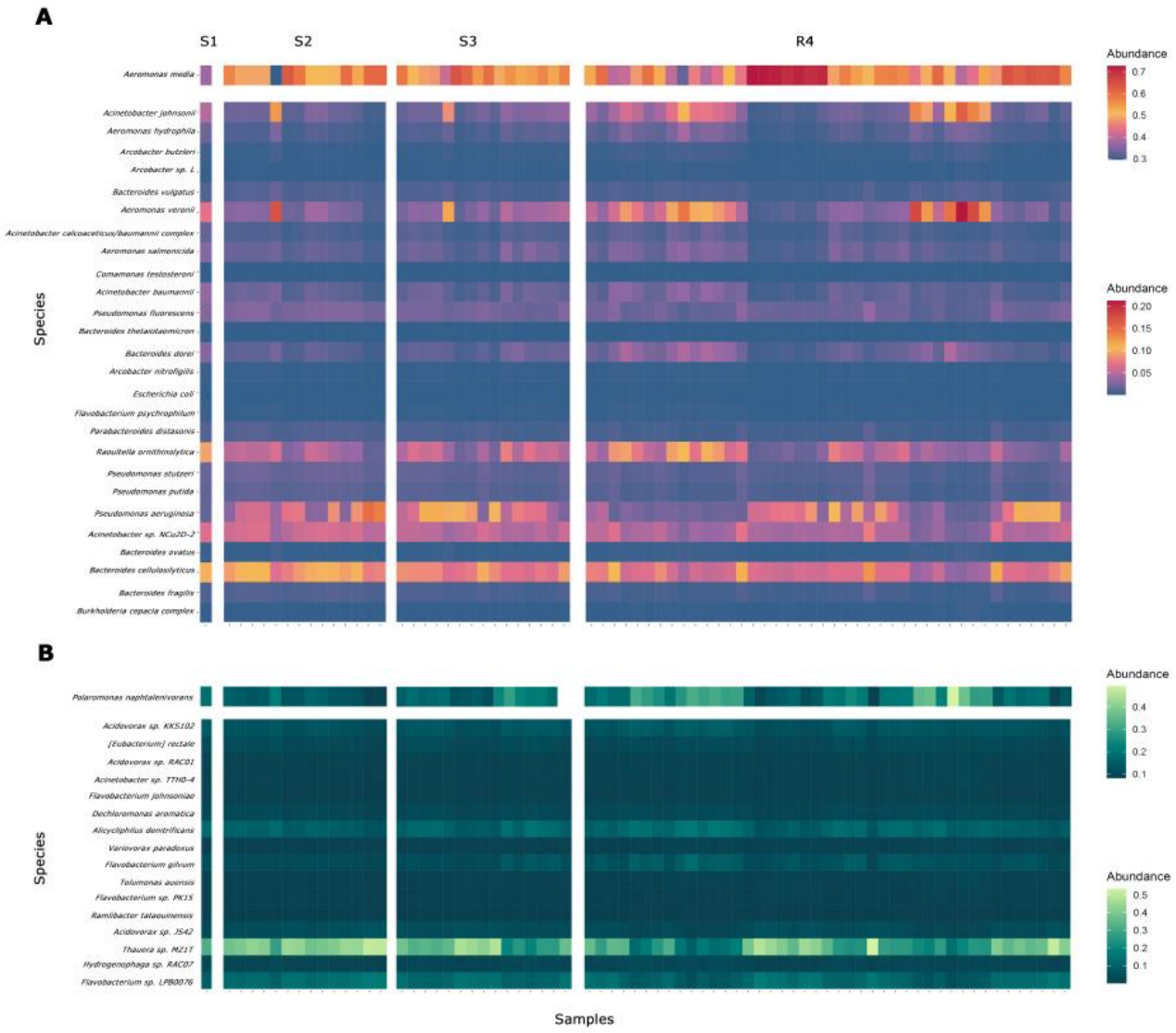

Figure 2. Global overview of the most abundant bacteria at all sampling points: A) Bacteria with reported infections, B) Bacteria with beneficial potential.

- **Microbial abundance: exploring diversity un aquatic systems**

Lastly, the reads that did not classified as Bacteria corresponded to *Homo sapiens*, and their prevalence was higher in wastewater (0.60) compared to surface water (0.39). As for the most abundant archaea species found in surface water samples, *Methanobrevibacter smithii* and *Methanococcus maripaludis* were prominent. In wastewater, *Methanosarcina barkeri* was the dominant species. Lastly, the most abundant assigned viruses globally in the samples were *Aeromonas virus phiO18P* and *Acanthamoeba polyphaga mimivirus*. Regarding the extraction of 18S-rRNA and taxonomic assignment, reads from protozoa and fungi were obtained. However, these accounted for less than 0.08% of the total, prompting the decision not to pursue these analyses due to potential inaccuracies in taxonomic assignment.

Antibiotics and virulence factors: Molecular markers of health interest associated with reads

On the other hand, for the analysis of molecular markers with reads we analyzed our data using internal databases like CARD and Resistomes & Variants to identify molecular markers with health-related implications, particularly those associated with antibiotic resistance. Our research revealed the ten most frequently detected drugs in both wastewater and surface water samples. In the wastewater samples, indicated by dark green in the outer ring (R4), we found that the most common drugs, shown by their wide bands in the circus (Figure 3A), were tetracyclines, aminoglycosides, and macrolides. Conversely, in the surface water samples (S1 - S3), marked in light blue on the outer ring, aminoglycosides and tetracycline consistently dominated. Additionally, sulfonamides were frequently observed at points S1 and S3. Reflecting the trends in wastewater, macrolides were most prominent at point R4, following the other two drugs (Figure 3A). Moreover, we selected the top 15 molecular markers frequently detected in both wastewater and surface water samples. In the upper panel marked in pink in Figure 3B, the health-related markers associated with antibiotic resistance can be observed. Each marker's frequency is indicated with a maximum frequency of 29 for surface waters and 42 for wastewater. This emphasizes the prevalence of markers *emrK* and *emrY* in surface waters compared to wastewater. In contrast, sul1 and sul2 were more common in wastewater than in surface waters. Regarding the molecular markers associated with virulence factors, located in the lower panel marked in green, we found that markers *hlyA*, *mshE*, and *flrC* were more abundant in wastewater than in surface waters. Conversely, *gacS* and *sefC* were predominant in surface waters but absent in wastewater.

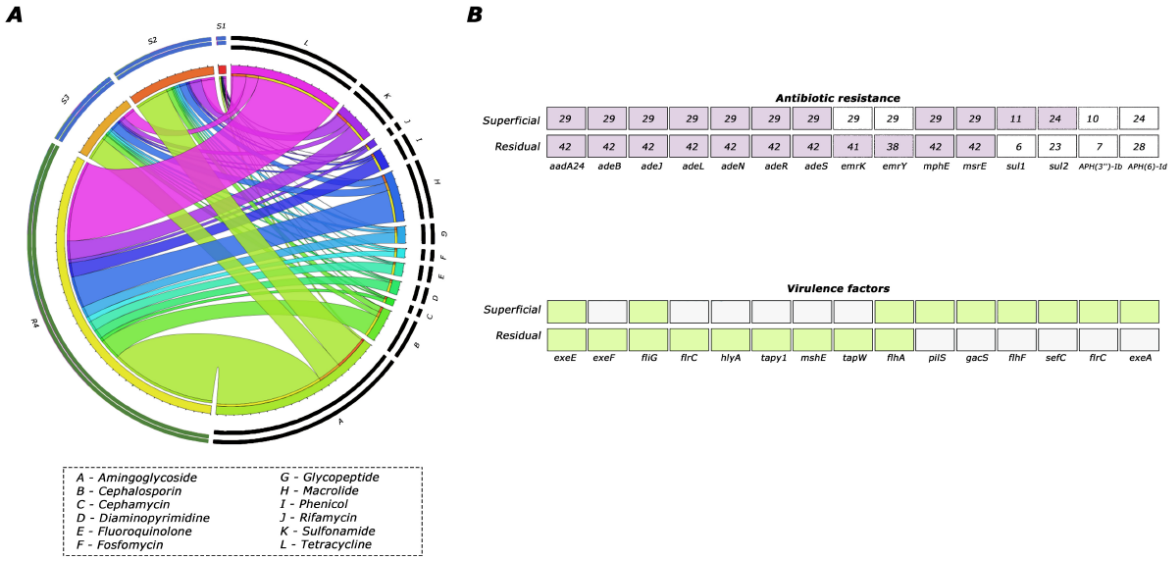

Figure 3. Health-related molecular markers associated with readings. A) Relationship between the frequency of the genes that code for the most common drugs in each of the sampled points (S1-S3 surface water samples, R4 = wastewater samples). B) Molecular markers associated with resistance (upper panel); molecular markers associated with virulence factors (lower panel).

Assembly and taxonomic assignment of aquatic metagenomes

Then, we used the clean reads for the assemblages, where we obtained a total of 270 high-quality Assembled Genomes from Metagenomes (MAGs), following the guidelines established by Bowers *et al*. in 2017. These guidelines ensured that the genomes had a completeness level exceeding 90% and contamination below 5%. In cases where the taxonomic assignment remained unresolved after applying the GTDB-TK software, we conducted cross-references with PubMLST (52) to identify potential matches with known microorganism species. However, no matches were found. Then, our subsequent analysis focused exclusively on MAGs that had received taxonomic assignments down to the species level, resulting in a final count of 175 genomes. To visualize the genomes of these species, we conducted searches for reference genomes corresponding to each assigned species in BV-BRC: Bacterial and Viral Bioinformatics Resource Center (53). Our selection criteria prioritized genomes that were both complete and of high quality. In instances where certain species lacked complete high-quality genomes, we turned to high-quality Whole Genome Sequencing (WGS) data as an alternative. Finally, species that lacked available reference genomes were excluded from further analyses.

Among the genomes we assigned, 50 were from wastewater samples, and 19 came from surface water samples. These genomes were linked to 16 different bacterial species, and we visualized the genomes with the best quality for each species using Proksee. In Figure 4A, it can be observed the genome assemblies from metagenomes of some species. While moving from the outermost to the innermost part of the genomes, it can be noticed the antibiotic resistance markers in red which align with CARD. In purple, it can be found the reference genomes for each species, and in gray it can be seen the assembled genome from the study, along with its respective contigs. Our study successfully identified various microbial species. For instance, we detected *Acinetobacter johnsonii* in six different genome bins. When we examined its genome, we found 939 contigs and three molecular markers in CARD. Similarly, we identified *Megamonas funiformis* in two genome bins, revealing a genome consisting of 505 contigs and one molecular marker in CARD. Additionally, the genome of *Phocaeicola vulgatus*, identified in three bins, comprised 689 contigs and four molecular markers in CARD. Our study also revealed the presence of other species, including *Aliarcobacter cryaerophilus_A*, *Sphaerotilus montanus, Prevotella copri, Aeromonas media, Aeromonas rivipollensis, Bacteroides uniformis*, and *Phascolarctobacterium_A_succinatutens*. Furthermore, through analyses using high-quality WGS data, we assigned the following species: *Kaistella chaponensis, Trichococcus flocculiformis, Zoogloea ramigera, Alistipes putredinis, Lactococcus A raffinolactis, and Faecalibacterium prausnitzii*.

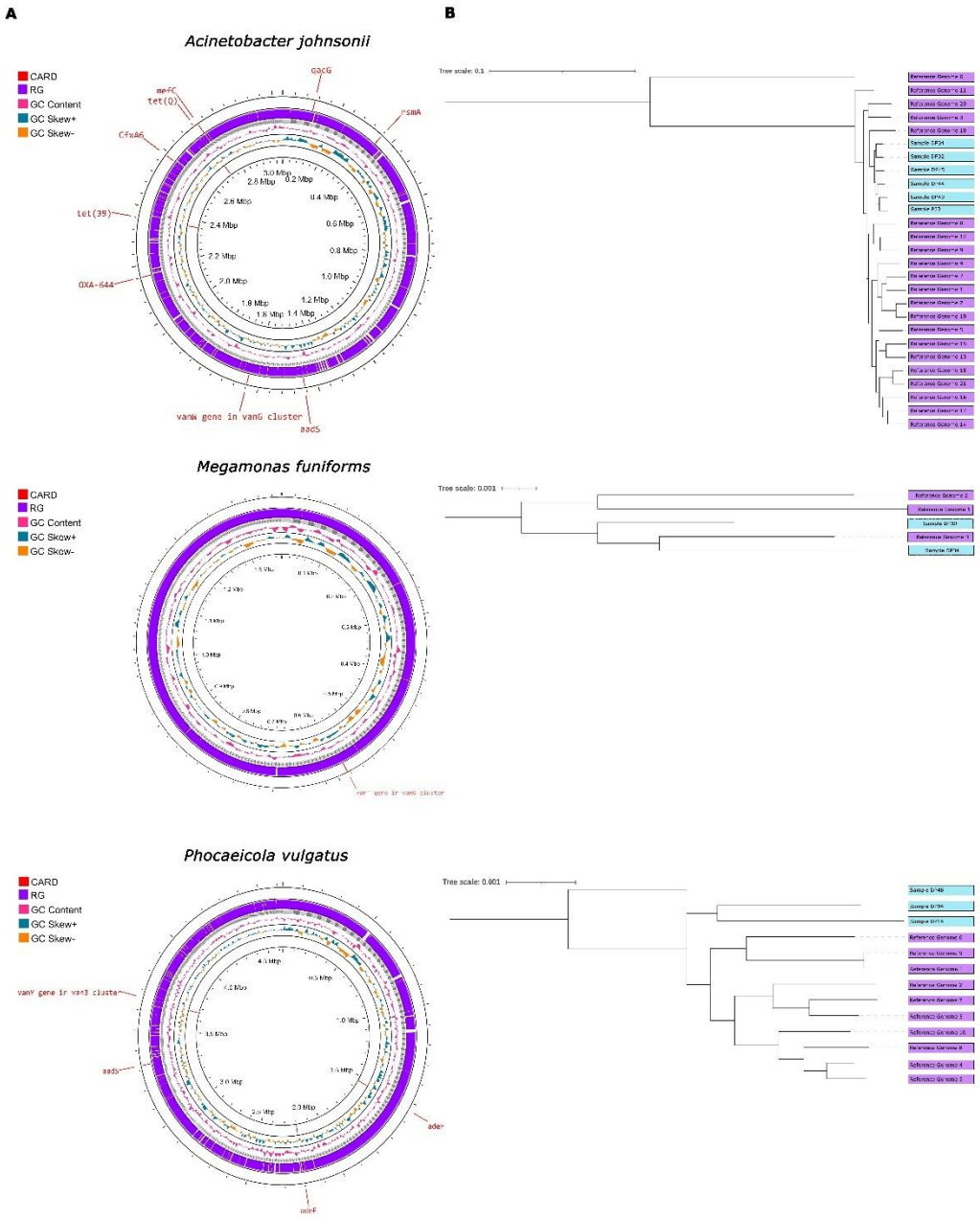

Figure 4. Genome assembly from metagenomes. A) MAGs, from the outside in antibiotic resistance marker assignment by CARD, reference genome, study genome. B) Reconstruction of phylogenetic trees, midpoint-rooted, study genomes in blue, reference genomes in purple.

During our analysis of the MAGs, we observed notable patterns in the COGs (Clusters of Orthologous Groups) of the species under examination. Figure 5 presents the following categories: "Unclassified" in gray, "Function unknown" in green, "Metabolism" in purple, "Information storage and processing" in pink, and "Cellular process and signaling" in blue. This overview highlights the notably high frequency within the "Function unknown" category. Moreover, across all samples, there was a more

extensive range of subcategories within the overarching "Metabolism" category. For *A. johnsonii* displayed in the first panel of the figure, the COGs L and J from the information storage and processing category were the most prevalent. Additionally, the specie *P. vulgatus* consistently presented COG M from the cellular process and signaling category. Moreover, *M. funiformis* shown in the third panel of the figure showed that within the COGs the category "Metabolism" predominated over the others. Also, we observed significant gene activity related to the Drug resistance: antimicrobial metabolic pathway. *Acinetobacter johnsonii,* for instance, exhibited a total of 11 genes associated with the beta-Lactam resistance pathway, six genes linked to the Vancomycin resistance pathway, and five genes involved in the Cationic antimicrobial peptide (CAMP) pathway. Similarly, in *Megamonas funiformis,* we identified five genes within the beta-Lactam pathway, five in the Vancomycin pathway, and 5 in the CAMP pathway.

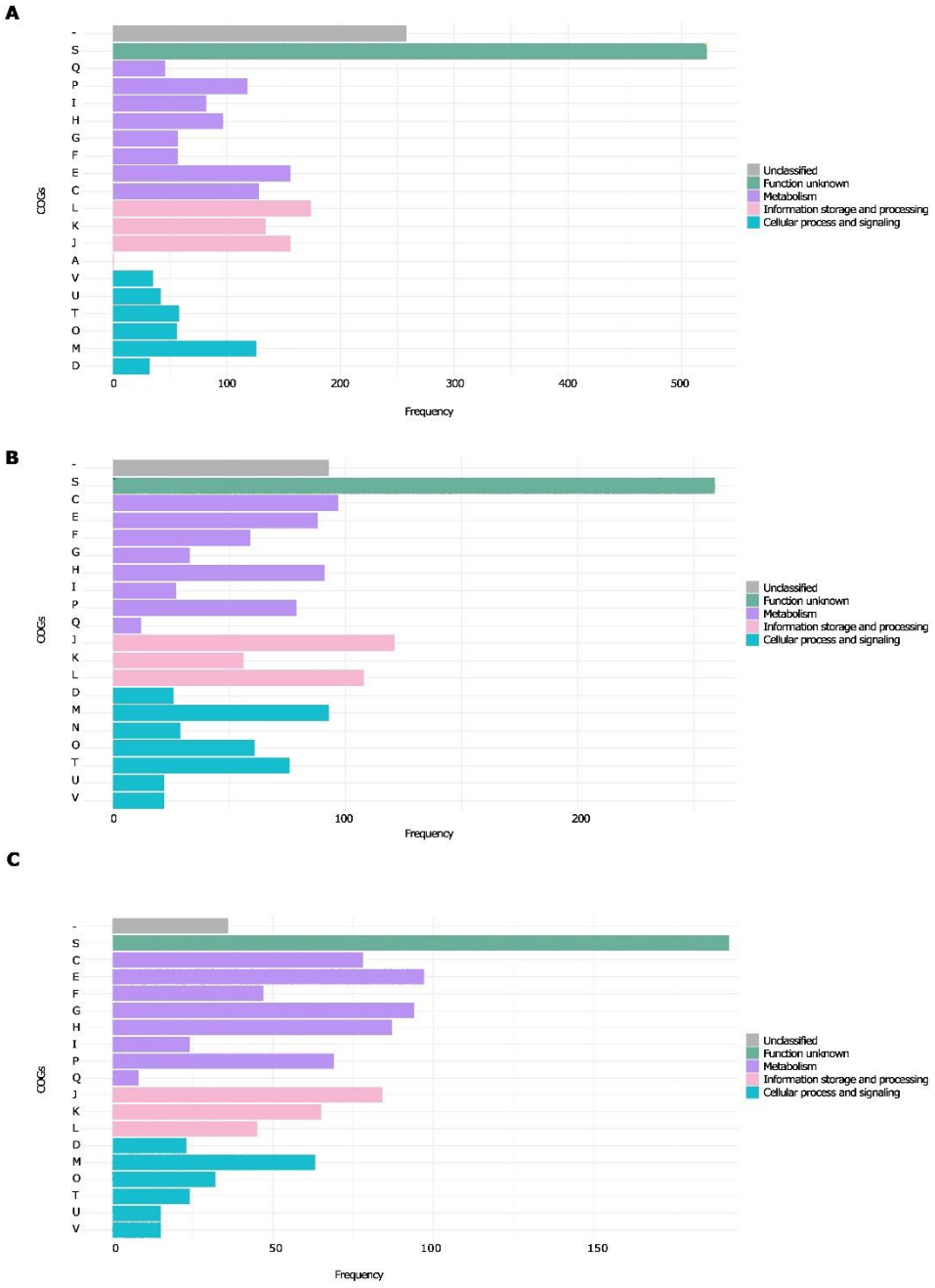

Figure 5. Cluster of Orthologous Groups (COGs) for *Acinetobacter johnsonii*, *Phocaeicola vulgatus*, and *Megamonas funiformis* (from top to bottom).

Antibiotics and virulence factors: Molecular markers of interest in health associated with assemblages

Furthermore, using the previously obtained assemblages, we identified molecular markers related to health, specifically those associated with antibiotic resistance and virulence factors in the MAG species. For resistance markers, we categorized them by species and their respective sampling points. This approach helped us point out the most found drugs within each species. In Figure 6, it can be observed genome assemblies for each species, depicted as circles. The color reflects the source of each assembly, indicating whether the sample originated from wastewater or surface water, and in the latter scenario, the specific sampling point. In wastewater (R4), the frequently detected drugs included aminoglycoside, tetracycline, cephamycin, and cephalosporin, each found in three species. In contrast, in surface waters (S3, S2), cephalosporin was present in five species, while macrolide appeared in two species. Furthermore, we identified virulence factors, which were exclusively found in the species *Zoogloea ramigera*. In this case, we detected the genes *hsiC1/vipB, hsiC1/vipA*, and *pilT*.

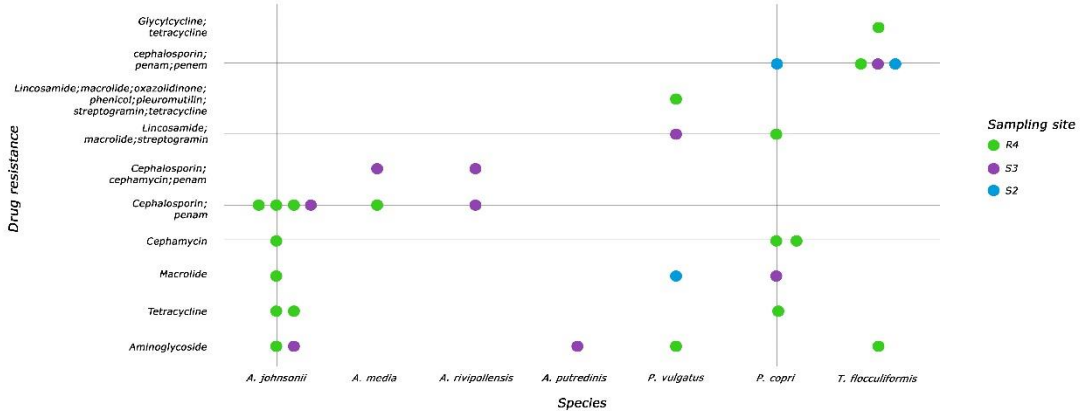

Figure 6. Molecular health-related markers associated with the MAGs. Each point represents an assembled genome of each of the species.

Pangenome

Finally, with the genomes of assembled species, we analyzed them alongside previously described reference genomes. For this analysis, we included MAGs with more than one bin and more than one assigned reference genome, excluding MAGs of the species *Aeromonas rivipollensis*, *Faecalibacterium prausnitzii*, and *Phascolarctobacterium_A_succinatutens.* We used iTOL (53) to generate trees for each species, rooted at the midpoint for intraspecies comparisons. Additionally, the Roary software was employed to identify both the core genome and accessory genome for each species. In Figure 4B, you can see midpoint-rooted trees, each with its respective scale. In purple, it can be found the reference genomes (to access these, please refer to the following link in the Table S5 GitHub - gimur/Metagenomics-of-Surface-Water-and-Wastewater-for-Environmental-

Monitoring), and in blue, the genomes corresponding to each assembled genome. Regarding *Acinetobacter johnsonii*, all genomes from our study clustered together within the same clade. Interestingly, the reference genome 6 did not clustered with none of the other reference genomes or ours. For *Megamonas funiformis*, most study genomes formed a distinct cluster. Lastly, in the case of *Phocaeicola vulgatus*, two of our genomes were closely grouped, while the remaining genome (Sample DP46) stood out as the most distant and distinct compared to the other genomes, including the reference genomes.

DISCUSSION

Studying the microbiome in aquatic environments is essential to understand the microorganisms and their interactions, especially those that could impact public health (54). From the findings of the physicochemical analyses, we can draw several conclusions. First, the elevated nitrite values in surface waters indicate contamination related to agricultural activities (55). In contrast, the high values of temperature, suspended solids, and BOD in wastewater imply a greater degree of contamination, primarily stemming from the presence of domestic, hospital, and stormwater contaminants, among others (56–58). The identification of parasites like *Giardia* spp. and *Cryptosporidium* spp. aligns with findings from other studies. These parasites have been detected in various regions, including Spain in wastewater (59), Poland in surface waters (60), and within the Nariño (Colombia), as reported by Sánchez (63). This underscores the concern associated with reusing surface water, as water serves as a means for the transmission of these parasites (61). These findings emphasize the potential public health risks and draw attention to the potential impact of the Juan XXIII wastewater collector, which discharges untreated water into surface waters.

In our study, we aimed to identify the most abundant families at each collection point to understand how they change over time. When examining the families present at the S3 surface water point (Figure 1, Figure S2), we noticed the widespread distribution of the Bacteroidaceae family in the environment. This is attributed to their large genomes, which enable them to adapt to diverse conditions (62). Nevertheless, it is worth noting that some members of this family can act as opportunistic pathogens, potentially causing infections in humans. This leads us to infer that in September, there was a higher load of contaminants associated with mammal excretions, where these family members may thrive (63). Similarly, we observed significant changes in the Rhodocyclaceae family. Members of this family have been identified in aquatic environments, as demonstrated by Chernitsyna *et al*. They also found that these family members play a role in bioremediation, particularly in denitrification processes. This aligns with our research, as we detected elevated levels of nitrites in surface waters, suggesting the possible involvement of these species in denitrification processes (64). We also observed a similar pattern in the Pseudomonadaceae family. While some species are commonly found in the environment, it is essential to highlight the presence of pathogenic species that could potentially pose public health concerns. The findings in surface waters provide valuable insights into the identification of bacteria present in the environment, particularly emphasizing bacterial families that include members with pathogenic potential. This underscores the importance of monitoring this environment to gain a comprehensive understanding of the associated risks when in contact with this valuable resource.

In terms of the families identified at wastewater sampling point R4 (Figure 1, Figure S3), we observed significant variations in the abundance of all families across different months. The Burkholderiaceae

family exhibited higher abundance in September, suggesting increased contamination from hospital, domestic, and industrial wastewater sources during that month (65). We also noticed fluctuations in the Rhodocyclaceae family, which has been previously reported in wastewater treatment systems by Wang and colleagues. They have identified some members of this family participating in denitrification processes (66). Another noteworthy finding was the higher abundance of the Campylobacteraceae family in February. This finding is significant as certain members of this family are associated with diarrheal diseases, indicating the presence of relevant pathogens (67). Furthermore, the identification of families such as Comamonadaceae and Moraxellaceae highlights the importance of monitoring the microbiome. Some members of these families have been reported as opportunistic pathogens. However, it is essential to note that they also include species with potential applications in biotechnology due to their advantages (68). The Enterobacteriaceae family comprises members commonly found in the intestines of animals and humans. Fluctuations in the abundance of this family across different months may be linked to variations in the volume of wastewater discharges from homes and hospitals (69). Finally, the Shewanellaceae family predominantly appeared exclusively at point R4. Considering the characteristics of this family, its presence at the wastewater point is expected, as some of its members are typically found in aquatic environments and play a fundamental role in the decomposition of organic matter (70). This information emphasizes the diverse array of bacteria present in wastewater, including both pathogenic and beneficial organisms. Furthermore, it emphasizes the concern that discharging wastewater into the Pasto River might be causing a detrimental impact on the environment.

Additionally, it is crucial to highlight species of particular interest, particularly those that may represent a risk to public health due to their pathogenic (Figure 2). In our research, we identified *Bacteroides cellulosilyticus*, which is widely distributed in the human intestine (71), suggesting fecal contamination in the water. Furthermore, *Raoultella ornithinolytica*, known to be found in bodies of water and even in drinking water systems, as demonstrated by Zou *et al* (72), raises concerns as this species has been associated with hospital-acquired infections and in immunosuppressed individuals (73). Equally important to note is *Pseudomonas aeruginosa*, a species that is not only found in a wide range of environments, including hospitals and sewage systems, but also represents a significant public health concern as a nosocomial pathogen and a multi-resistant to antibiotics (74). The most abundant species identified was *Aeromonas media*, a bacterium capable of causing various diseases in both humans and animals, and it has also been reported to exhibit resistance to specific antibiotics (75). Lastly, we identified the *Burkholderia cepacia* complex, a group of bacteria widely distributed and highly adaptable, with a tendency for rapid genetic and phenotypic mutations. Despite their environmental origin, these species are of concern as opportunistic pathogens (76). The identification of these species provides us with valuable tools for regulating the use of water resources and estimating potential risks associated with their presence. It also emphasizes the critical importance of continuous monitoring of this resource, as this surveillance is essential for ensuring the control and prevention of public health emergencies. Water contaminated with specific microorganisms can serve as a means of propagating these pathogens, potentially leading to severe illnesses.

Additionally, we would like to highlight species with notable beneficial potential (Figure 2). For example, *Thauera sp. MZ1T*, which is distinguished by its unique ability to produce a specific exopolysaccharide, consequently improving the viscosity of its environment. This distinctive feature significantly improves the efficiency of disinfection processes in industrial wastewater treatment

plants (77). In contrast, we have also identified *Alicycliphilus denitrificans*, capable of biodegrading a wide range of organic contaminants, suggesting its potential to remediate xenobiotic substances, particularly in contaminated water bodies (78). Furthermore, among the most abundant species, we have identified *Polaromonas naphthalenivorans*, a bacteria reported in various water bodies with the ability to decompose naphthalene, a common component in fuels, pesticides, and other substances (79,80). Our findings go beyond simply identifying species with promising attributes for applications in biotechnology, including bioremediation and the decomposition of harmful compounds. They also shed light on the possible applications of these species in biotechnology and provide valuable information on contaminants that may be present in both wastewater and surface water. This underlines the critical importance of recognizing and characterizing microbial species in these environments with such substantial potential. Additionally, we identified archaea in surface waters, including *Methanobrevibacter smithii*, which is prevalent in the human intestine (81), and *Methanococcus maripaludis*, a marine methanogenic organism (82). In wastewater we found *Methanosarcina barkeri*, which is widely studied as a model of methanogenesis (83). Regarding viruses, we identified the *Aeromonas virus phiO18P viruses*, known for their ability to transmit resistance and virulence (84), and the *Acanthamoeba polyphaga mimivirus*, responsible for infecting amoebas (85). Our research highlights a diverse spectrum of microorganisms in wastewater and surface water, some of which are associated with human and animal fecal matter, indicating contamination. We recommend continuing such analyzes and improving public databases to collect more complete information on all microorganisms, including parasites and fungi. It would also be highly beneficial to increase the sequencing capacity by including more gigabytes (Gb) of data. This would provide a deeper understanding of the eukaryotic communities, as most of the data is currently allocated to bacterial communities due to their significant diversity.

Rivers have encountered significant challenges due to human activities (85,86), and one of these challenges is pollution caused by the improper disposal of antibiotics. Antibiotics are considered emerging contaminants because they cannot be broken down, leading to their accumulation in the aquatic environment and the development of selection pressure (87,88). We examined genes responsible for conferring resistance to antibiotics (Figure 3), and we observed a high frequency of genes related to tetracyclines and aminoglycosides in wastewater. This pattern is consistent with the global context, as both types of drugs are widely used in human, animal, and agricultural healthcare (87). Furthermore, we noticed a substantial frequency of genes conferring resistance to macrolides, likely linked to their increasing use in livestock and agriculture. As a result, the aquatic environment has become a conduit for the spread of antibiotic resistance genes (89). Conversely, in the surface waters of the Pasto River, we identified a high prevalence of genes associated with aminoglycoside and tetracycline resistance, which underscores the potential impact of wastewater discharge directly into surface waters. Additionally, at point S1, we frequently detected sulfonamide resistance genes, suggesting a notable contribution to the aquatic environment from contaminants such as urine and feces from both humans and animals due to their use in medical treatment (90).

In the other analysis, we identified specific molecular markers in both wastewater and surface water (Figure 3). One of these markers is *EmrK*, responsible for conferring resistance to tetracyclines (86) and *EmrY*, which confers resistance to macrolides, streptogramins and lincosamides. *Emr* markers are typically found on plasmids, which facilitate horizontal gene transfer, enhancing the transmission of these genes to other pathogens (38,87). This finding aligns with our previous results, where we

observed a high prevalence of genes that confer resistance to tetracyclines. In the case of surface waters, we observed a higher frequency of the *sul1* and *sul2* markers, responsible for conferring resistance to sulfonamides. This corresponds with our previous findings, where we identified the gene associated with sulfonamide resistance as commonly present in surface waters (38). Notably, both markers have been reported in bacterial genera such as *Acinetobacter*, *Aeromonas*, and *Pseudomonas*, which were also abundant in our study. Furthermore, identifying molecular markers associated with virulence factors allows us to understand the enzymes, toxins, and other elements that enable pathogenic microorganisms to cause more harm (88). For instance, we detected markers like *hlyA* (89) in wastewater, which through an exotoxin can damage the cell membrane. We also found markers like *mshE* (90), which modify the pilus to enhance adhesion for colonization in aquatic environments and biofilm formation. Additionally, we identified the flagellar transcriptional activator gene *flrc* (91). In surface waters, we found significant markers involved in regulating virulence factors related to stress tolerance, pathogen motility, and more, such as *gacS* (92). We also identified markers like *sefC* (93), which facilitate adhesion and colonization of various substrates and cells. While identifying microorganisms in water bodies is crucial, it is equally important to detect the molecular markers that enhance their virulence and potential harm.

The generation of metagenome-assembled genomes - MAGs, played a crucial role in identifying previously unknown microorganisms. In environmental samples, MAGs provide essential information about the environment, specific pressures, and the potential to understand non-cultivable or challenging-to-isolate species (94). In this study, a diverse range of bacterial species was identified, including some originating from the intestine, gastrointestinal system, and feces. These species include *Phocaeicola vulgatus* (95), *Megamonas funiformis* (96), *Prevotella copri* (97), *Bacteroides uniformis* (93), *Phascolarctobacterium_A_succinatutens* (98), *Alistipes putredinis* (99), *Faecalibacterium prausnitzii* (100) y *Aliarcobacter cryaerophilus_A* (101). Additionally, species commonly found in water bodies, biofilms, and related environments were detected, such as *Acinetobacter johnsonii (102)*, *Sphaerotilus montanus* (103), *Aromonas media* (104), *Aeromonas rivipollensis* (105), *Kaistella chaponensis* (106), *Trichococcus flocculiformis* (107) y *Zoogloea ramigera* (108). Finally, *Lactococcus A_raffinolactis* distributed in fermented foods (109). These findings illustrate the diverse range of genomes within the assembled bacterial species and their respective origins. They contribute to a global understanding that the dominant species are those originating from the digestive system and those commonly found in the environment.

A key focus of our research centers on the genome of *A. johnsonii*, a microorganism widely distributed in aquatic environments, yet also reported as an opportunistic pathogen. Additionally, when examining its functional characteristics, we note that the category "Information storage and processing" predominates. This indicates that this species primarily allocates its coding potential to J (translation, ribosomal structure, and biogenesis) and L (replication, recombination, and repair) (50). Furthermore, the assembly of the genomes of *Megamonas funiformis* and *Phocaeicola vulgatus* provides a valuable tool to enhance our comprehension of their functions in the human intestine (96,110). In terms of functional characteristics, the bacterium *P. vulgatus* displays a higher frequency of COGs related to "cell processing and signaling," particularly in M (cell wall/membrane/envelope biogenesis). This underscores the crucial role of this process in maintaining and protecting these species within their environment. *M. funiformis* directs its coding potential toward “metabolism.” All of this contributes to a better understanding of the sets of genes that could potentially impact the

adaptation and survival of these species in aquatic environments (111,112). At a broader level, the most frequent categories were "Unknown function" and "Unclassified," underscoring the importance of exploring and contributing information to public databases to obtain deeper knowledge about these functional attributes. Finally, it is essential to note that genes associated with antibiotic resistance pathways have been identified in all three species, including beta-lactam resistance, vancomycin resistance, and cationic antimicrobial peptide (CAMP) resistance (50). Therefore, we encourage further comprehensive studies on these functional characteristics, with a specific focus on identifying the precise functions of each gene within these pathways. This effort will allow us to provide comprehensive information on potential risks and shed light on the acquisition of resistance by certain pathogenic species that currently lack detailed characterization (Figure 4, Figure 5).

The identification of antibiotic resistance genes in the genomes of the assembled species (Figure 6), particularly those conferring resistance to aminoglycosides and tetracyclines, raises concerns about the potential public health risks associated with some of these species, especially since they are classified as opportunistic pathogens. Moreover, the detection of genes related to resistance to cephamycins, a drug used in the treatment of infections (113), in addition to cephalosporins (which constitute approximately 50% of the antibiotics used in human health) in wastewater, is a cause for concern. This not only implies the creation of selective pressure in this environment but also suggests fecal contamination in surface waters (114). These findings highlight the inadequate elimination and accumulation of these drugs in water (115). Regarding markers associated with virulence factors, they were only identified in the genome of *Zoogloea ramigera*. Within this genome, we identified genes such as *hsiC1/vipB* and *hsiC1/vipA*, responsible for tubule-forming proteins of the type VI secretion system *VipB* and type VI *VipA*. *PilT* is associated with the sporadic motility protein PilT (39). In conclusion, while these analyses provide an initial perspective on health-related markers in the aquatic environment, it is advisable to conduct more comprehensive studies in future research to obtain robust information for resource monitoring and surveillance. In conclusion, although these analyzes provide an initial view of health-related markers in the aquatic environment, it is advisable to conduct more comprehensive studies in future research to obtain solid information for resource monitoring and surveillance.

Pangenome analysis provides valuable insights into the relationships among the samples. When we conducted the pangenome analysis for the species *A. johnsonii*, we observed that reference genome 6 is not clustered with the other reference genomes and the assembled genomes from the study. This could be attributed to the percentage of the core genome identified in this study, which amounts to 218 genes, while the accessory genome accounts for 12.958 genes out of a total of 13.176 genes. This hypothesis is proposed because reference genome 6 was isolated from water in 2019 (116), and its extensive array of accessory genes might have led to its distinct separation from the others. Furthermore, this species is characterized by having high plasticity in the genome (117). On the contrary, *M. funiformis* displayed a cluster of diverse samples with no significant grouping differences. However, it is worth noting that this species has a limited number of reference genomes. Lastly, in the case of *P. vulgatus*, sample DP46 from the study appeared distinctly separated from the other samples. This sample was collected from point R4 in wastewater. Nevertheless, is important to acknowledge that this genome might be highly fragmented, and its assignment may not be entirely accurate. Besides, it may also be influenced by selection pressures such as antibiotic resistance.

Likewise, it is important to highlight the importance of understanding this species since although it is part of the human intestine, it has stood out for being an opportunistic pathogen (118).

CONCLUSION

This study provides vital information about the current state of the Pasto River. We carry out a comprehensive analysis of the river and its wastewater, with a specific focus on the evaluation of physicochemical and microbiological parameters. In essence, this pioneering study has great importance for the region, as it emphasizes the importance of continuous monitoring and surveillance of this critical resource, especially in the context of possible water reuse. Furthermore, this study represents the first microbiological characterization at a regional level, covering both taxonomic and functional aspects, made possible thanks to next-generation sequencing techniques, which provides valuable information on its relevance to public health. This contributes to informed decision-making by relevant authorities responsible for managing these water resources. However, we recommend that future studies incorporate complementary techniques to evaluate the viability of these microorganisms circulating in these bodies of water using next-generation sequencing techniques. Furthermore, it is essential not to ignore traditional physicochemical and microbiological analyzes to correlate these results with sequencing data. Finally, we encourage the scientific community to continue with the characterization studies of water bodies and thus nourish public databases, to be able to identify and characterize even more microorganisms in future studies.

DATA AVAILABILITY

The data obtained and analyzed in this study are available in the European Nucleotide Archive – ENA repository under project number PRJEB64340.

FUNDING

This work was financed by the Colombian Fund for Science, Technology and Innovation of *Sistema General de Regalías* [project BPIN: 2020000100770]. Neither of thefunding entities participated in the development of the study.

DECLARATION OF COMPETING INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Rajasulochana P, Preethy V. Comparison on efficiency of various techniques in treatment of waste and sewage water – A comprehensive review. Resource-Efficient Technologies. 2016 Dec;2(4):175–84.

2. Kanakoudis V, Tsitsifli S. Potable water security assessment-a review on monitoring, modelling. 2017.

3. IDEAM. 2015. AGUAS SUPERFICIALES.

4. Gad M, Saleh AH, Hussein H, Farouk M, Elsayed S. Appraisal of Surface Water Quality of Nile River Using Water Quality Indices, Spectral Signature and Multivariate Modeling. Water (Switzerland). 2022 Apr 1;14(7).

5. Páll E, Niculae M, Kiss T, Şandru CD, Spînu M. Human impact on the microbiological water quality of the rivers. Vol. 62, Journal of Medical Microbiology. 2013. p. 1635–40.

6. Xu H, Gao Q, Yuan B. Analysis and identification of pollution sources of comprehensive river water quality: Evidence from two river basins in China. Ecol Indic. 2022 Feb 1;135.

7. Morin-Crini N, Lichtfouse E, Liu G, Balaram V, Ribeiro ARL, Lu Z, et al. Worldwide cases of water pollution by emerging contaminants: a review. Vol. 20, Environmental Chemistry Letters. Springer Science and Business Media Deutschland GmbH; 2022. p. 2311–38.

8. Lugli GA, Longhi G, Mancabelli L, Alessandri G, Tarracchini C, Fontana F, et al. Tap water as a natural vehicle for microorganisms shaping the human gut microbiome. Environ Microbiol. 2022 Sep 1;

9. Sobsey MD. Absence of virological and epidemiological evidence that SARS-CoV-2 poses COVID-19 risks from environmental fecal waste, wastewater and water exposures. J Water Health. 2022 Jan 1;20(1):126–38.

10. Kolundžija S, Cheng DQ, Lauro FM. RNA Viruses in Aquatic Ecosystems through the Lens of Ecological Genomics and Transcriptomics. Vol. 14, Viruses. MDPI; 2022.

11. Kodera SM, Sharma A, Martino C, Dsouza M, Grippo M, Lutz L, et al. Microbiome response in an urban river system is dominated by seasonality over wastewater treatment upgrades. Available from: https://doi.org/10.1101/2022.06.30.498375

12. Amarasiri M, Sano D, Suzuki S. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. Crit Rev Environ Sci Technol. 2020 Oct 1;50(19):2016–59.

13. Feng C, Xu Z, Li Y, Zhu N, Wang Z. Research progress on the contamination status and control policy of Giardia and Cryptosporidium in drinking water. Vol. 11, Journal of Water Sanitation and Hygiene for Development. IWA Publishing; 2021. p. 867–86.

14. Zhu K, Ren H, Lu Y. Potential Biorisks of Cryptosporidium spp. and Giardia spp. from Reclaimed Water and Countermeasures. Current Pollution Reports. Springer Science and Business Media Deutschland GmbH; 2022.

15. Rothman JA, Loveless TB, Kapcia J, Adams ED, Steele JA, Zimmer-Faust AG, et al. RNA viromics of Southern California wastewater and detection of SARS-CoV-2 single-nucleotide variants. Appl Environ Microbiol. 2021 Nov 1;87(23).

16. Song W, Chen H, Xue N, Wang S, Yang Y. Metagenomic binning and assembled genome analysis revealed the distinct composition of resistome and mobilome in the Ili River. Ecotoxicol Environ Saf. 2022 Sep 1;242.

17. Payne JT, Millar JJ, Jackson CR, Ochs CA. Patterns of variation in diversity of the Mississippi river microbiome over 1,300 kilometers. PLoS One. 2017 Mar 1;12(3).

18. G Sutherland BJ, Finke JF, Saunders R, Warne S, Schulze AD, Strohm T, et al. Metatranscriptomics reveals a shift in microbial community composition and function during summer months in a coastal marine environment 2 3. Available from: https://doi.org/10.1101/2022.01.21.477286

19. Chopyk J, Nasko DJ, Allard S, Bui A, Treangen T, Pop M, et al. Comparative metagenomic analysis of microbial taxonomic and functional variations in untreated surface and reclaimed waters used in irrigation applications. Water Res. 2020 Feb 1;169.

20. Tang Y, Liang Z, Li G, Zhao H, An T. Metagenomic profiles and health risks of pathogens and antibiotic resistance genes in various industrial wastewaters and the associated receiving surface water. Chemosphere. 2021 Nov 1;283.

21. Kim JJ, Seong HJ, Johnson TA, Cha CJ, Sul WJ, Chae JC. Persistence of antibiotic resistance from animal agricultural effluents to surface water revealed by genome-centric metagenomics. J Hazard Mater. 2023 Sep 5;457.

22. Scherer GS, Guiguet Leal DA, Goulart JAG, Araújo RS, Beux MR, Moreira NM. Parasitological, microbiological, and antimicrobial resistance profiles of raw and drinking water in a tourist city in the tri-border region of South America. J Water Health. 2022 Feb 1;20(2):385–95.

23. Breton-Deval L, Juárez K, Sánchez-Flores A, Vera-Estrella R. Water quality assessments and metagenomic analysis of the polluted river Apatlaco, Mexico. Available from: http://dx.doi.org/10.1101/536128

24. Rodríguez EA, Ramirez D, Balcázar JL, Jiménez JN. Metagenomic analysis of urban wastewater resistome and mobilome: A support for antimicrobial resistance surveillance in an endemic country. Environmental Pollution. 2021 May 1;276.

25. Lucía M, Martínez O, Carolina A, Sánchez F, González Hormiga M, Del A, et al. INSTITUTO NACIONAL DE SALUD. 2022.

26. Sánchez C, López MC, Galeano LA, Qvarnstrom Y, Houghton K, Ramírez JD. Molecular detection and genotyping of pathogenic protozoan parasites in raw and treated water samples from southwest Colombia. Parasit Vectors. 2018 Oct 26;11(1).

27. Empopasto. https://sig.empopasto.com.co/arcgisportal/apps/webappviewer/index.htmL?id=799f69ae0bb3482e94c498ebbd843523. 2021. Cuencas de abastecimiento.

28. Corponariño. “PROGRAMA MANEJO INTEGRAL DEL RECURSO HÍDRICO EN EL DEPARTAMENTO DE NARIÑO.” 2011. Plan de Ordenación del Río Pasto.

29. EPIDAT 3.1 EPIDEMIOLOGICAL ANALYSIS FROM TABULATED DATA.

30. Köchling T, Sanz JL, Galdino L, Florencio L, Kato MT. Impact of pollution on the microbial diversity of a tropical river in an urbanized region of northeastern Brazil. International Microbiology. 2017;20(1):11–24.

31. Sánchez C, López MC, Galeano LA, Qvarnstrom Y, Houghton K, Ramírez JD. Molecular detection and genotyping of pathogenic protozoan parasites in raw and treated water samples from southwest Colombia. Parasit Vectors. 2018 Oct 26;11(1).

32. 36. Andrews S. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data .

33. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016 Oct 1;32(19):3047–8.

34. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

35. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res. 2016 Dec 1;26(12):1721–9.

36. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. 2020 Feb 15;36(4):1303–4.

37. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020 Jan 1;48(D1):D517–25.

38. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: A general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022 Jan 7;50(D1):D912–7.

39. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 2017 May 1;27(5):824–34.

40. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;2015(8).

41. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016 Feb 15;32(4):605–7.

42. 48. Johannes Alneberg BSBI de BMSJQUZILLNJLAFA& CQuince. Binning metagenomic contigs by coverage and composition. . Nat Methods. 2014;

43. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018 Jul 1;3(7):836–43.

44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul 1;25(7):1043–55.

45. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36(6):1925–7.

46. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

47. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 May 25;31(22):3691–3.

48. Cantalapiedra CP, Herņandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9.

49. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020 Apr 1;36(7):2251–2.

50. Sammarro Silva KJ, Sabogal-Paz LP. Cryptosporidium spp. and Giardia spp. (oo)cysts as target-organisms in sanitation and environmental monitoring: A review in microscopy-based viability assays. Vol. 189, Water Research. Elsevier Ltd; 2021.

51. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019 Nov 28;20(1).

52. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.

53. Zhu L, Li R, Yan Y, Cui L. Urbanization drives the succession of antibiotic resistome and microbiome in a river watershed. Chemosphere. 2022 Aug 1;301.

54. Ustaoğlu F, Tepe Y, Taş B. Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. Ecol Indic. 2020 Jun 1;113.

55. Verma A. Assessment of Water quality and Pollution Status of Kalpi (Morar) River, Gwalior, Madhya Pradesh: with special reference to Conservation and Management plan Water Treatment and Water Technology View project [Internet]. 2014. Available from: https://www.researchgate.net/publication/266065376

56. Pattanayak S, Priyadarsini S, Paul A, Kumar PR, Sahoo PK. Diversity of virulence-associated genes in pathogenic Aeromonas hydrophila isolates and their in vivo modulation at varied water temperatures. Microb Pathog. 2020 Oct 1;147.

57. Yu Q, Han Q, Shi S, Sun X, Wang X, Wang S, et al. Metagenomics reveals the response of antibiotic resistance genes to elevated temperature in the Yellow River. Science of the Total Environment. 2023 Feb 10;859.

58. Ramo A, Del Cacho E, Sánchez-Acedo C, Quílez J. Occurrence and genetic diversity of Cryptosporidium and Giardia in urban wastewater treatment plants in north-eastern Spain. Science of the Total Environment. 2017 Nov 15;598:628–38.

59. Lass A, Szostakowska B, Korzeniewski K, Karanis P. Detection of Giardia intestinalis in water samples collected from natural water reservoirs and wells in northern and north-eastern Poland using LAMP, real-time PCR and nested PCR. J Water Health. 2017 Oct 1;15(5):775–87.

60. Hemati S, MMF, MHA, NM, & FM. Occurrence of Giardia and Cryptosporidium in effluents of urban wastewater treatment plants: A global systematic review and meta-analysis. Journal of Cleaner Production, 134555. 2022;

61. Zhang L, Ben Said L, Hervé N, Zirah S, Diarra MS, Fliss I. Effects of drinking water supplementation with Lactobacillus reuteri, and a mixture of reuterin and microcin J25 on the growth performance, caecal microbiota and selected metabolites of broiler chickens. J Anim Sci Biotechnol. 2022 Dec 1;13(1).

62. Jean S, WMJ, DG, & BCAD. Time for some group therapy: update on identification, antimicrobial resistance, taxonomy, and clinical significance of the Bacteroides fragilis group. Journal of clinical microbiology, 60(9), e02361-20. 2022;

63. Chernitsyna S, Elovskaya I, Pogodaeva T, Bukin S, Zakharenko A, Zemskaya T. Bacterial Communities in a Gradient of Abiotic Factors Near a Sulfide Thermal Spring in Northern Baikal. Diversity (Basel). 2023 Feb 1;15(2).

64. Kerber-Diaz JC, Leos-Ramírez MA, Flores-Ceron AA, Ponce-Mendoza A, Estrada-de los Santos P, Ibarra JA. Distribution of CRISPR-Cas systems in the Burkholderiaceae family and its biological implications. Arch Microbiol. 2022 Dec 1;204(12).

65. Wang Z, Li W, Li H, Zheng W, Guo F. Phylogenomics of Rhodocyclales and its distribution in wastewater treatment systems. Sci Rep. 2020 Dec 1;10(1).

66. Sazan Qadr Amin1 HJM 2 and HKZ. Campylobacteriosis. International Journal of Agriculture and Biosciences [Internet]. 2023;2:87–93. Available from: https://uniquescientificpublishers.com/wp-content/uploads/2023/oht-v2/87-93.pdf

67. Knoot CJ, Wantuch PL, Robinson LS, Rosen DA, Scott NE, Harding CM. Discovery and characterization of a new class of O-linking oligosaccharyltransferases from the Moraxellaceae family. Glycobiology. 2023 Jan 8;33(1):57–74.

68. Singh AK, Das S, Kumar S, Gajamer VR, Najar IN, Lepcha YD, et al. Distribution of Antibiotic-Resistant Enterobacteriaceae Pathogens in Potable Spring Water of Eastern Indian Himalayas: Emphasis on Virulence Gene and Antibiotic Resistance Genes in Escherichia coli. Front Microbiol. 2020 Nov 5;11.

69. Xu S, Yu K, Su H, Chen B, Huang W, Liang J, et al. Proposal of Parashewanella gen. nov. to accommodate Parashewanella curva sp. nov. and Parashewanella spongiae comb. nov. in the Shewanellaceae. Int J Syst Evol Microbiol. 2019 May 1;69(5):1259–64.

70. Christensen L, Sørensen C V., Wøhlk FU, Kjølbæk L, Astrup A, Sanz Y, et al. Microbial enterotypes beyond genus level: Bacteroides species as a predictive biomarker for weight change upon controlled intervention with arabinoxylan oligosaccharides in overweight subjects. Gut Microbes. 2020;12(1).

71. Zou H, Berglund B, Xu H, Chi X, Zhao Q, Zhou Z, et al. Genetic characterization and virulence of a carbapenem-resistant Raoultella ornithinolytica isolated from well water carrying a novel megaplasmid containing blaNDM-1. Environmental Pollution. 2020 May 1;260.

72. Zou H, Berglund B, Xu H, Chi X, Zhao Q, Zhou Z, et al. Genetic characterization and virulence of a carbapenem-resistant Raoultella ornithinolytica isolated from well water carrying a novel megaplasmid containing blaNDM-1. Environmental Pollution. 2020 May 1;260.

73. Tuon FF, Dantas LR, Suss PH, Tasca Ribeiro VS. Pathogenesis of the Pseudomonas aeruginosa Biofilm: A Review. Vol. 11, Pathogens. MDPI; 2022.

74. Talagrand-Reboul E, Roger F, Kimper JL, Colston SM, Graf J, Latif-Eugenín F, et al. Delineation of taxonomic species within complex of species: Aeromonas media and related species as a test case. Front Microbiol. 2017 Apr 18;8(APR).

75. Tavares M, Kozak M, Balola A, Sá-Correia I. Burkholderia cepacia Complex Bacteria: a Feared Contamination Risk in Water-Based Pharmaceutical Products. 2020; Available from: https://doi.org/10.1128/CMR

76. Jiang K, Sanseverino J, Chauhan A, Lucas S, Copeland A, Lapidus A, et al. Complete genome sequence of Thauera aminoaromatica strain MZ1T. Stand Genomic Sci. 2012;6(3):325–35.

77. Fuentes-Jaime J, Vargas-Suárez M, Cruz-Gómez MJ, Loza-Tavera H. Concerted action of extracellular and cytoplasmic esterase and urethane-cleaving activities during Impranil biodegradation by Alicycliphilus denitrificans BQ1. Biodegradation. 2022 Aug 1;33(4):389–406.

78. Padilla-Del Valle R, Morales-Vale LR, Ríos-Velázquez C. Unraveling the microbial and functional diversity of Coamo thermal spring in Puerto Rico using metagenomic library generation and shotgun sequencing. Genom Data. 2017 Mar;11:98–101.

79. Hwang K, Choe H, Nasir A, Kim KM. Complete genome of Polaromonas vacuolata KCTC 22033T isolated from beneath Antarctic Sea ice. Mar Genomics. 2021 Feb 1;55.

80. Bang C, Weidenbach K, Gutsmann T, Heine H, Schmitz RA. The intestinal archaea Methanosphaera stadtmanae and Methanobrevibacter smithii activate human dendritic cells. PLoS One. 2014 Jun 10;9(6).

81. Palabikyan H, Ruddyard A, Pomper L, Novak D, Reischl B, Rittmann SKMR. Scale-up of biomass production by Methanococcus maripaludis. Front Microbiol. 2022 Nov 23;13.

82. Ye J, Yu J, Zhang Y, Chen M, Liu X, Zhou S, et al. Light-driven carbon dioxide reduction to methane by Methanosarcina barkeri-CdS biohybrid. Appl Catal B. 2019 Nov 15;257.

83. Subirats J, Sànchez-Melsió A, Borrego CM, Balcázar JL, Simonet P. Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes. Int J Antimicrob Agents. 2016 Aug 1;48(2):163–7.

84. Ferek JD, Thoden JB, Holden HM. Biochemical analysis of a sugar 4,6-dehydratase from Acanthamoeba polyphaga Mimivirus. Protein Science. 2020 May 1;29(5):1148–59.

85. Christaki E, Marcou M, Tofarides A. Antimicrobial Resistance in Bacteria: Mechanisms, Evolution, and Persistence. Vol. 88, Journal of Molecular Evolution. Springer; 2020. p. 26–40.

86. Abushaheen MA, Muzaheed, Fatani AJ, Alosaimi M, Mansy W, George M, et al. Antimicrobial resistance, mechanisms and its clinical significance. Disease-a-Month. 2020 Jun 1;66(6).

87. Leitão JH. Microbial virulence factors. Vol. 21, International Journal of Molecular Sciences. MDPI AG; 2020. p. 1–6.

88. Noori AA. Detection of the, hlyA, fimH PgaD gene in Escherichia coli isolated from different sources in Kirkuk city Sundus jassim Mohamad. Vol. 10, Journal of Survey in Fisheries Sciences. 2023.

89. Zhang YL, Wu F, Huang YC, Jian JC, Cai SH. Mechanisms Underlying the Virulence and Intestinal Colonization of the Vibrio alginolyticus HY9901 DctP Protein with Proteomic Analysis. Appl Biochem Microbiol [Internet]. 2023 Oct 2;59(5):646–58. Available from: https://link.springer.com/10.1134/S0003683823050198

90. Syed KA, Beyhan S, Correa N, Queen J, Liu J, Peng F, et al. The Vibrio cholerae flagellar regulatory hierarchy controls expression of virulence factors. J Bacteriol. 2009 Nov;191(21):6555–70.

91. Strateva T, Mitov I. Contribution of an arsenal of virulence factors to pathogenesis of Pseudomonas aeruginosa infections. Vol. 61, Annals of Microbiology. 2011. p. 717–32.

92. El-Tayeb MA, Ibrahim ASS, Maany DA, Elbadawi YB, Al-Salamah AA. Detection of virulence genes and investigation of the immunogenicity of outer membrane proteins of salmonella enterica. J Appl Biol Biotechnol. 2019 Sep 1;7(5):62–73.

93. Pinhassi J, Farnelid H, García SM, Teira E, Galand PE, Obernosterer I, et al. Functional responses of key marine bacteria to environmental change – toward genetic counselling for coastal waters. Front Microbiol. 2022 Dec 1;13.

94. Neff A, Lück R, Hövels M, Deppenmeier U. Expanding the repertoire of counterselection markers for markerless gene deletion in the human gut bacterium Phocaeicola vulgatus. Anaerobe. 2023 Jun 1;81.

95. Chen G, Zeng Z, Xie M, Peng Y, Zhou W, Xu W, et al. Fermentation characteristics and probiotic activity of a purified fraction of polysaccharides from Fuzhuan brick tea. Food Science and Human Wellness. 2022 May 1;11(3):727–37.

96. Abdelsalam NA, Hegazy SM, Aziz RK. The curious case of Prevotella copri. Vol. 15, Gut Microbes. Taylor and Francis Ltd.; 2023.

97. Shigeno Y, Kitahara M, Shime M, Benno Y. Phascolarctobacterium wakonense sp. nov., isolated from common marmoset (Callithrix jacchus) faeces. Int J Syst Evol Microbiol. 2019 Jul 1;69(7):1941–6.

98. Tyrrell KL, Warren YA, Citron DM, Goldstein EJC. Re-assessment of phenotypic identifications of Bacteroides putredinis to Alistipes species using molecular methods. Anaerobe. 2011 Jun;17(3):130–4.

99. Leylabadlo HE, Ghotaslou R, Feizabadi MM, Farajnia S, Moaddab SY, Ganbarov K, et al. The critical role of Faecalibacterium prausnitzii in human health: An overview. Microb Pathog. 2020 Dec 1;149.

100. Gungor C, Hizlisoy H, Ertas Onmaz N, Gundog DA, Barel M, Disli HB, et al. Profile of Aliarcobacter spp. from edible giblets: Genetic diversity, antibiotic resistance, biofilm formation. Int J Food Microbiol. 2023 Feb 2;386.

101. Jia J, Liu M, Feng L, Wang Z. Comparative genomic analysis reveals the evolution and environmental adaptation of Acinetobacter johnsonii. Gene. 2022 Jan 15;808.

102. Nott MA, Driscoll HE, Takeda M, Vangala M, Corsi SR, Tighe SW. Advanced biofilm analysis in streams receiving organic deicer runoff. PLoS One. 2020 Jan 1;15(1).

103. Dubey S, Ager-Wick E, Peng B, Evensen Ø, Sørum H, Munang'andu HM. Characterization of virulence and antimicrobial resistance genes of Aeromonas media strain SD/21–15 from marine sediments in comparison with other Aeromonas spp. Front Microbiol. 2022 Dec 1;13.

104. Fono-Tamo EUK, Kamika I, Dewar JB, Lekota KE. Comparative Genomics Revealed a Potential Threat of Aeromonas rivipollensis G87 Strain and Its Antibiotic Resistance. Antibiotics. 2023 Jan 1;12(1).

105. Ai J, He X, Ren M, Yu T, Liu X, Jiang Y, et al. Kaistella yananensis sp. nov., a novel indoleacetic acid producing bacterium isolated from the root nodules of Sophora davidii (Franch.) Skeels. 2023; Available from: https://doi.org/10.21203/rs.3.rs-3289031/v1

106. Shim J gon, Cho SG, Kim SH, Chuon K, Meas S, Choi A, et al. Heliorhodopsin Helps Photolyase to Enhance the DNA Repair Capacity. Microbiol Spectr. 2022 Dec 21;10(6).

107. Bharti M, Nagar S, Khurana H, Negi RK. Metagenomic insights to understand the role of polluted river Yamuna in shaping the gut microbial communities of two invasive fish species. Arch Microbiol. 2022 Aug 1;204(8).

108. Jung MY, Lee C, Seo MJ, Roh SW, Lee SH. Characterization of a potential probiotic bacterium Lactococcus raffinolactis WiKim0068 isolated from fermented vegetable using genomic and in vitro analyses. BMC Microbiol. 2020 May 27;20(1).

109. Neff A, Lück R, Hövels M, Deppenmeier U. Expanding the repertoire of counterselection markers for markerless gene deletion in the human gut bacterium Phocaeicola vulgatus. Anaerobe. 2023 Jun 1;81.

110. Zhang W, Xu X, Cai L, Cai X. Dysbiosis of the gut microbiome in elderly patients with hepatocellular carcinoma. Sci Rep. 2023 Dec 1;13(1).

111. Michelle L, Dalmacio M. Bacterial composition and antibiotic resistome profile of water in the Manila Bay South Harbor [Internet]. Available from: https://www.researchgate.net/publication/373238250

112. Yamashiro H, Kasamatsu Y, Anan N, Takemura M, Yamano Y. In vitro efficacy of humanized regimen of flomoxef against extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae . Antimicrob Agents Chemother. 2023 Sep 19;67(9).

113. Das N, Madhavan J, Selvi A, Das D. An overview of cephalosporin antibiotics as emerging contaminants: a serious environmental concern. Vol. 9, 3 Biotech. Springer Verlag; 2019.

114. Li J, Li W, Liu K, Guo Y, Ding C, Han J, et al. Global review of macrolide antibiotics in the aquatic environment: Sources, occurrence, fate, ecotoxicity, and risk assessment. Vol. 439, Journal of Hazardous Materials. Elsevier B.V.; 2022.

115. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res. 2023 Jan 6;51(1 D):D678–89.

116. Jia J, Liu M, Feng L, Wang Z. Comparative genomic analysis reveals the evolution and environmental adaptation of Acinetobacter johnsonii. Gene. 2022 Jan 15;808.

117. Vu H, MY, HM, NH, TK, & YY. Complete Genome Sequences of Three Phocaeicola vulgatus Strains Isolated from a Healthy Japanese Individual. Microbiology Resource Announcements, 11(2), e01124-21. 2022;

SUPPLEMENTARY INFORMATION

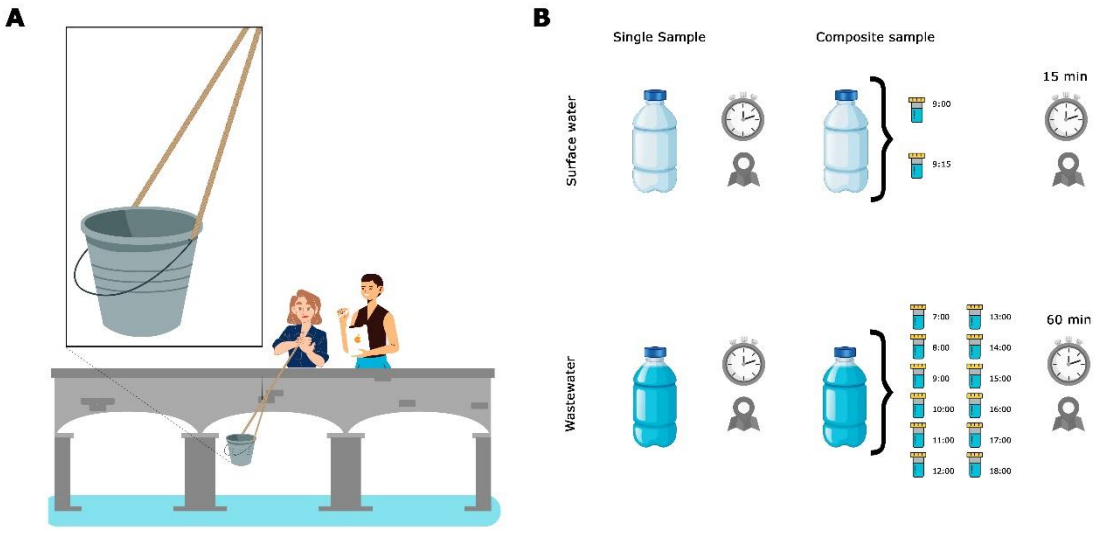

Figure S1. Schematic representation of A) Model implemented for the collection of wastewaters and surface water samples from pedestrian and vehicular bridges B) System used for sample collection from simple collection and composite collection for wastewater and surface water.

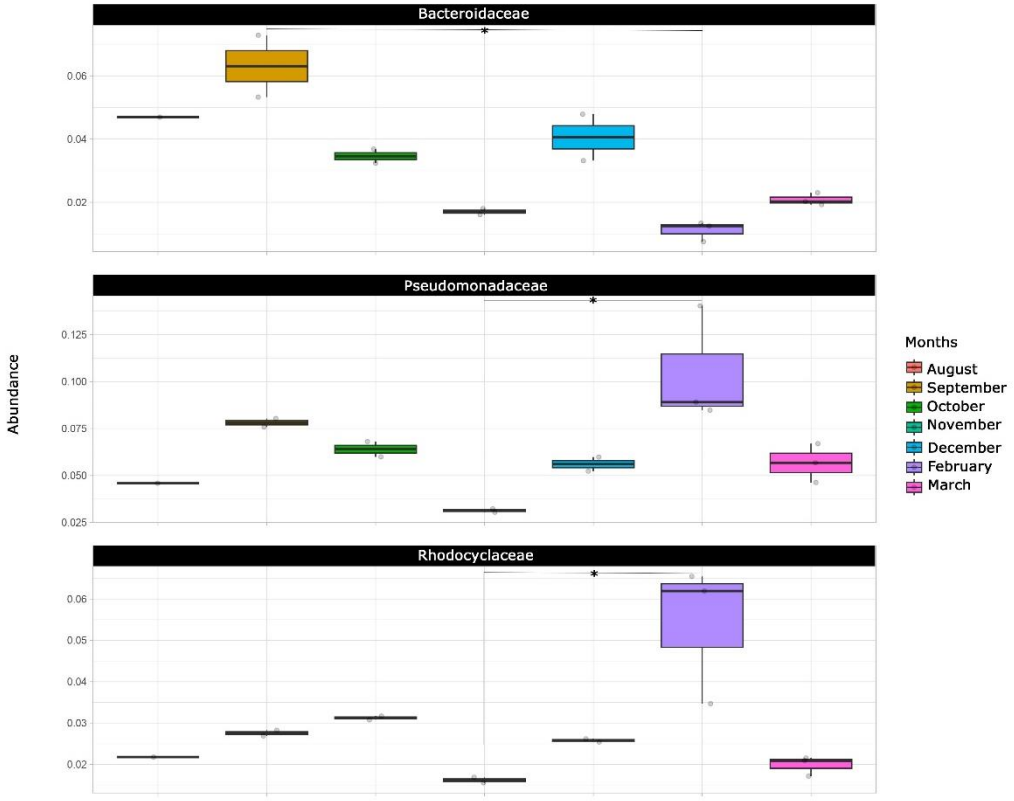

Figure S2. Boxplot diagram illustrating families at point S3 with statistically significant differences*.

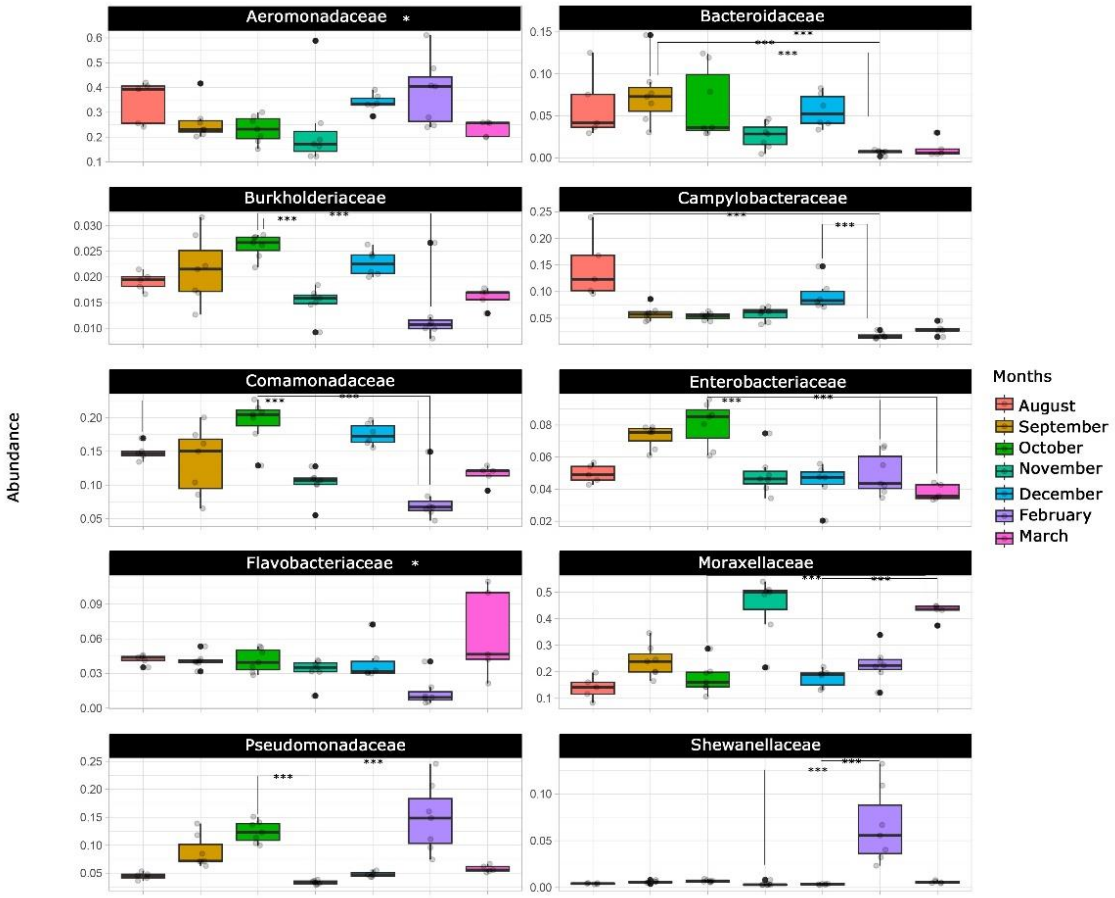

Figure S3. Boxplot diagram illustrating families at point R4 with statistically significant differences***.

Link para consultar figuras en alta resolución:

https://uredu-my.sharepoint.com/personal/vanessa_urrea_urosario_edu_co/_layouts/15/onedrive.aspx?sw=bypass&bypassReason=abandoned&ga=1&id=%2Fpersonal%2Fvanessa%5Furrea%5Furosario%5Fedu%5Fco%2FDocuments%2FBIOLOG%C3%8DA%20UR%2F02%20%2D%20Maestr%C3%ADa%2F1%2E%20Tesis%20Maestr%C3%ADa%2F00%20%2D%20Tesis%2F01%20%2D%20Metagen%C3%B3mica%2F04%20%2D%20Tesis%20maestr%C3%ADa%20UR%2F1%2E%20Figuras&view=0

Contribuciones del estudio:

Enfatiza en la importancia del monitoreo y vigilancia del recurso hídrico, tanto en aguas residuales como superficiales
Se muestran los potenciales riesgos de la reutilización del recurso hídrico
Primera caracterización microbiológica a nivel regional abarcando aspectos taxonómicos y funcionales
Brinda un panorama actualizado del estado actual de los cuerpos hídricos a nivel microbiológico

Contribuciones específicas del estudiante dentro de la investigación llevada a cabo. Señalar las actividades en las cuáles estuvo involucrado activamente el estudiante:

La concepción del proyecto de investigación	(X)
El diseño del estudio	(X)
La adquisición de los datos a través de la experimentación	(X)
Análisis e interpretación de los datos	(X)
Elaboración del borrador del artículo	(X)
Revisión y aprobación definitiva de la versión que se presenta	(X)