



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Ecología, Genética y Evolución

**Estudios genéticos y ecológicos de la estructura espacial del
bacteriplancton asociada a la contaminación de origen fecal
en una red hidrológica del AMBA.**

Tesis para optar por el título de doctor de la Universidad de Buenos Aires,
en el área de Ciencias Biológicas

Lic. Martín Saraceno

Director: Dr. Nicolás Frankel

Director Adjunto: Dr. Martín Graziano

Consejera de estudios: Dra. Irina Izaguirre

Lugar de trabajo: Laboratorio de Limnología.

Instituto de Ecología, Genética y Evolución de Buenos Aires (UBA-
CONICET) – Departamento de Ecología, Genética y Evolución

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

2023

Índice

Glosario de abreviaturas	5
Resumen/ <i>Abstract</i>	8
Agradecimientos.....	14
Introducción General	16
Objetivos.....	30
Objetivo general	31
Objetivos específicos por capítulo.....	31
Hipótesis de trabajo	33
Capítulo 1	37
1.1. Resumen	38
1.2. Materiales y Métodos	40
1.2.1 Selección de sitios y muestreo.....	40
1.2.2 Determinaciones fisicoquímicas.....	42
1.2.3 Variables demográficas, hidráulicas e infraestructuras sanitarias	43
1.2.4 Parámetros biológicos.....	44
1.2.5 Índice de heterogeneidad espacial	45
1.2.6 Análisis estadístico	45
1.3. Resultados.....	47
1.3.1 Distribución de las características urbanas y del hábitat ecológico local en la cuenca	47
1.4. Anexo Capítulo 1.....	57
Capítulo 2	58
2.1. Resumen	59
2.2. Materiales y métodos.....	62
2.2.1 Secuenciación de ADN ambiental y procesamiento bioinformático.....	62
2.2.2 Identificación del core de ASVs a partir de curvas de rango y abundancia ..	63
2.2.3 Determinación de secuencias potencialmente patógenas	63
2.2.4 Determinación de secuencias vinculadas a microbiomas fecales humanos del AMBA	64
2.2.5 Procesamiento estadístico.....	65
2.3. Resultados.....	70
2.3.1 Estructura taxonómica global de la comunidad bacteriana en la cuenca	70
2.3.2 Análisis de la distribución espacial de la comunidad bacteriana.....	74
2.3.3 Redes de co-ocurrencia de la comunidad bacteriana.....	78

2.3.4 Identificación de bacterias de origen fecal en la cuenca	81
2.3.5 Rastreo de taxones potencialmente patógenos	85
2.3.6 Asociaciones entre las distintas comunidades bacterianas analizadas	89
2.4. Anexo	91
Capítulo 3	99
3.1. Resumen	100
3.2. Materiales y métodos.....	102
3.2.1Anotación filogenética de aislamientos ambientales de <i>E. coli</i>	102
3.2.2 Análisis de datos.....	103
3.3. Resultados.....	106
3.3.1 Abundancia de <i>E. coli</i> y asociación con las características del hábitat ecológico local y de la infraestructura urbana	106
3.3.2 Distribución espacial de los grupos filogenéticos de <i>E. coli</i>	108
3.3.3 Asociación de factores ambientales, urbanos y espaciales en los patrones de abundancia y composición filogenética de <i>E. coli</i>	110
3.4. Anexo 3	114
Capítulo 4	118
4.1. Resumen	119
4.2. Materiales y métodos.....	120
4.2.1 Ensayo de supervivencia	120
4.2.2 Análisis de los ensayos de supervivencia.....	121
4.2.3 Ensayo de formación de biofilm.....	122
4.2.4 Análisis estadístico general	123
4.3. Resultados.....	125
4.3.1 Ensayos de supervivencia bacteriana	125
4.3.2 Ensayo de formación de biofilm.....	126
4.3.3 Análisis integral del comportamiento experimental de los filogrupos	127
Capítulo 5	130
5.1. Resumen	131
5.2. Materiales y métodos.....	132
2.1 Extracción de ADN genómico y secuenciación	132
5.2.2 Procesamiento bioinformático	133
5.3. Resultados.....	135
5.3.1 Análisis genómico del clado críptico IV hallado en la cuenca de estudio ..	135
5.3.2 Relaciones filogenéticas entre el genoma 339_SF y otros miembros del clado críptico IV	136

5.3.3 Análisis genómico comparativo de aislamientos del clado críptico IV.....	139
5.4. Anexo	142
Discusión general	146
Conclusiones generales.....	161
Bibliografía.....	165
Financiamiento	183

Glosario de abreviaturas

Glosario de abreviaturas

ACV: absorbancia de cristal violeta (del inglés, *absorbance of crystal violet*)

AMR: Resistencia antimicrobrial (del inglés, *antimicrobial resistance*)

ANOVA: Análisis de la varianza (del inglés, *analysis of variance*)

APL: absorbancia de células planctónica (del inglés, *absorbance of planctonic cells*)

ASV: Variante de secuenciación de amplicón (del inglés, *amplicon variant sequence*)

CART: Árbol de clasificación y regresión (del inglés *classification and regression tree*)

CC: clado críptico

CRISPR: Repeticiones palindrómicas cortas agrupadas y regularmente inter-espaciadas (del inglés, *clustered regularly interspaced short palindromic repeats*)

DIN: Nitrógeno inorgánico disuelto (del inglés, *dissolved inorganic nitrogen*)

DO₅₉₅: Densidad óptica a 595 nm

DOC: Carbono orgánico disuelto (del inglés, *dissolved organic carbon*)

EAEC: *E. coli* enteroaggregativa

EDTA: Ácido etilendiaminotetraacético

EHEC: *E. coli* enterohemorrágica

EIEC: *E. coli* enteroinvasiva

EPA: Agencia de Protección Ambiental de Estados Unidos (del inglés, *U.S. environmental protection agency*)

EPEC: *E. coli* enteropatogénica

ETEC: *E. coli* enterotoxigénica

FV: Factor de virulencia

ICP-MS: espectrometría de masas con plasma acoplado inductivamente (del inglés, *inductively coupled plasma-mass spectrometry*)

LB: Luria-Bertani

LC: Límite de cuantificación

NCBI: Centro Nacional para la Información Biotecnológica - EE. UU. (del inglés, *national center of biotechnology information*)

ND: No detectado

NMEC: *E. coli* asociada a la meningitis neonatal

NTU: Unidad nefelométrica de turbidez (del inglés, *nephelometric turbidity unit*)

OD: Oxígeno disuelto

PCA: Análisis de componentes principales (del inglés, *principal component analysis*)

PLSs: Secuencias similares a patógenos a nivel de especie (del inglés, *pathogen-like sequences at species level*)

pRDA: Análisis de redundancia parcial (del inglés, *partial redundancy analysis*)

ROI: Regiones de interés (del inglés, *regions of interest*)

rpm: revolución por minuto

SBF: Índice de Formación de Biofilm Específico (SBF, del inglés *specific biofilm formation index*)

SNP: Polimorfismo de nucleótido único (del inglés, *single nucleotide polymorphism*)

SRP: Fósforo reactive soluble (del inglés, *soluble reactive phosphorus*)

STEC: *E. coli* productora de la toxina Shiga

TE: Tris-EDTA

TP: Fósforo total (del inglés, *total phosphorus*)

UFC: Unidad formadora de colonias

UPEC: *E. coli* uropatogénica

VFDB: Bases de datos de factores de virulencia (del inglés, *virulence factor database*)

WGS: Secuenciación de genoma completo (del inglés, *whole genome sequencing*)

Resumen/*Abstract*

Resumen

Las ciudades son sistemas socio-ecológico-tecnológicos complejos en los que el grado de avance y calidad del desarrollo urbano ejerce un rol clave en la definición de las condiciones ambientales predominantes. Los arroyos urbanos ofrecen beneficios culturales, económicos, paisajísticos y ambientales. Alojan una importante diversidad biológica y son centrales en los ciclos biogeoquímicos de los nutrientes en el ámbito urbano. Fuentes antropogénicas como la escorrentía superficial y la exfiltración de aguas residuales introducen importantes cargas de bacterias alóctonas en los sistemas hidrológicos superficiales. En estos casos las bacterias de origen fecal y/o patogénicas suelen ser ubicuas en la cuenca, lo que genera riesgos sanitarios y efectos negativos sobre el bacteriplancton nativo. En este trabajo de Tesis intentamos comprender los procesos ecológicos, genéticos y espaciales que condicionan la estructura del bacteriplancton (con énfasis en el estudio de los grupos bacterianos de origen fecal y/o patogénicos) en una red hidrológica predominantemente urbana de la zona sur del Área Metropolitana de Buenos Aires (cuenca de los arroyos San Francisco, Las Piedras y Santo Domingo). En primer lugar, se analizaron las condiciones demográficas e infraestructurales de la cuenca, y las condiciones del hábitat local de los cauces. Se detectó una importante variabilidad en relación con la cobertura de infraestructuras sanitarias. Las características del hábitat local también fueron variables en la extensión de la cuenca. Sin embargo, la estructuración de las condiciones del hábitat reveló un significativo grado de homogeneización a escala local (tramo). Se estudió el bacteriplancton a lo largo de la cuenca mediante secuenciación del gen ARNr 16S. La partición aditiva de la riqueza de taxones bacterianos demostró que su recambio fue mayor al esperado entre las locaciones de la cuenca (β_2) y menor al esperado entre muestras provenientes de una misma locación (β_1), lo que sugiere una importante estructuración de nichos a escala de la cuenca y una mayor homogeneidad en las condiciones locales. La partición de la varianza de la riqueza y la equitatividad de grupos bacterianos se halló principalmente estructurada por las condiciones del hábitat local. A través de un análisis de redes de co-ocurrencia se observó que los sitios con menor grado de disturbio presentaron una red más cohesiva entre los taxones del bacteriplancton, sugiriendo un mayor grado de resiliencia ecológica, en contraste con los sitios con mayor intensidad de infraestructuras urbanas. Además, se identificaron taxones de origen fecal y potencialmente patogénicos en el bacteriplancton de todas las locaciones. Los determinantes de la riqueza de estos taxones fueron en mayor

medida las condiciones del hábitat local y en menor medida los factores urbanos. Además, se observaron correlaciones fuertes, positivas y significativas entre las abundancias y riquezas de taxones fecales y de taxones potencialmente patógenos, y la presencia de *E. coli*. Esto confirma el rol indicador de riesgo sanitario que cualquiera de estas determinaciones posee sobre las otras. Por otro lado, se analizaron las asociaciones entre las condiciones de la cuenca y la estructura poblacional de *E. coli*. Se demostró la influencia del ambiente local y las infraestructuras en explicar la abundancia y composición filogenética de *E. coli*. Los grupos filogenéticos A y B1 se detectaron en forma ubicua y abundante, predominando procesos de exclusión competitiva, mientras que otros taxones de baja frecuencia y distribución se asociaron a sitios más disturbados. Además, se identificó una correlación negativa entre la abundancia de *E. coli* y su riqueza filogenética y una correlación positiva entre esta riqueza y la equitatividad entre los grupos. Teniendo en cuenta la evidencia previa sobre la capacidad de *E. coli* de persistir en ambientes secundarios, se evaluó esta característica en el laboratorio. Se hicieron ensayos de supervivencia y de formación de biofilm con aislamientos de la cuenca de distintos filogrupos. Los grupos A y B1 fueron los únicos que presentaron tasas de crecimiento positivas. El índice de formación de biofilm fue significativamente mayor en los aislamientos del filogrupo B1. Los resultados demuestran que ciertos grupos de *E. coli* poseen una aptitud diferencial para persistir en ambientes secundarios, sugiriendo un mecanismo por el cual se genera el proceso de exclusión competitiva. Por último, uno de los aislamientos de la cuenca se identificó como perteneciente al *clado criptico IV* del género *Escherichia*, mayormente asociado a condiciones de vida libre. Este aislamiento es el primero de su filiación en ser detectado en un ambiente natural de Sudamérica. Se secuenció el genoma completo del aislamiento y se analizó su posición filogenética dentro del género *Escherichia*. El genoma del aislamiento presentó una baja cantidad de factores de virulencia y de resistencia antibiótica, rasgo que se asocia al estilo de vida libre y que ya se había observado en otros genomas del *clado criptico IV*. Además, analizando genomas disponibles de origen comensal y ambiental pertenecientes al *clado criptico IV*, se identificaron genes potencialmente asociados al fenómeno de persistencia ambiental dentro del género *Escherichia*. En su totalidad, este trabajo contribuye al entendimiento del fenómeno de persistencia bacteriana dentro del género *Escherichia* y aporta nuevas hipótesis que pueden servir de base para indagaciones futuras.

Abstract

Cities are complex socio-ecological-technological systems where the degree and quality of urban development plays a crucial role in shaping the prevailing environmental conditions. Urban streams provide cultural, economic and environmental benefits. Streams harbor significant biological diversity and play a central role in nutrient biogeochemical cycles in the urban environment. Anthropogenic sources such as surface runoff and wastewater exfiltration introduce significant loads of allochthonous bacteria into surface hydrological systems. In these cases, bacteria of fecal-origin and/or pathogenic bacteria are often ubiquitous in the watershed, posing sanitary risks and generating negative effects on native bacterioplankton. In this thesis, we aimed at understanding the ecological, genetic and spatial processes that influence the structure of bacterioplankton (with an emphasis on the study of bacteria of fecal-origin and/or pathogenic bacteria) in a predominantly urban hydrological network in the southern region of the Buenos Aires Metropolitan Area (San Francisco, Las Piedras and Santo Domingo streams watershed). We analyzed the demographic and infrastructural conditions of the watershed, as well as the local habitat conditions of the streams. We observed heterogenous sanitation infrastructure throughout the watershed. In addition, the characteristics of the local habitat also exhibited variation across the watershed. However, the structuring of habitat conditions revealed a notable degree of homogenization on the local scale (reach). Bacterioplankton was studied with the 16S rRNA gene. The additive partitioning of taxon richness demonstrated that taxon turnover was higher than expected between locations in the watershed (β_2), while lower than expected within samples from the same location (β_1). This suggests niche structuring at the watershed scale and greater homogeneity in local conditions. The variance partitioning of taxon richness and evenness was primarily structured by local habitat conditions. Through a co-occurrence network analysis, we observed that sites with lower disturbance exhibited a more cohesive network among bacterioplankton taxa, suggesting a higher degree of ecological resilience compared to sites with higher intensity of urban infrastructure. Furthermore, taxa of fecal origin and potentially pathogenic taxa were identified in the bacterioplankton across all locations. The determinants of richness for these taxa were primarily associated with local habitat conditions and, to a lesser extent, urban factors. Strong, positive, and significant correlations were also observed between the abundances and richness of fecal taxa, potentially pathogenic taxa, and the presence of *E. coli*. This confirms the role as

indicators of sanitary risk that each of these determinations has on the others. On the other hand, we analyzed possible associations between watershed conditions and *E. coli* population structure. We detected an influence of the local environment and infrastructure in explaining the abundance and phylogenetic composition of *E. coli*. Phylogenetic groups A and B1 were ubiquitous and abundant, predominating competitive exclusion processes, while other taxa with low-frequency and limited distribution were associated with more disturbed sites. Furthermore, a negative correlation was identified between *E. coli* abundance and its phylogenetic richness, and a positive correlation was observed between richness and evenness among phylogenetic groups. Considering available evidence on the ability of *E. coli* to persist in secondary environments, we decided to evaluate this characteristic in the laboratory. Survival and biofilm formation assays were conducted using isolates from different phylogroups of the watershed. Groups A and B1 were the only ones that exhibited positive growth rates. The biofilm formation index was significantly higher in isolates from phylogroup B1. These results demonstrate that certain groups of *E. coli* have a differential aptitude for persistence in secondary environments, suggesting a mechanism by which the process of competitive exclusion is generated. Lastly, one of the isolates from the watershed was identified as belonging to *cryptic clade IV* of the genus *Escherichia*, which is mostly associated with a free-living lifestyle. This isolate is the first of its kind to be detected in a natural environment in South America. The complete genome of the isolate was sequenced, and its phylogenetic position within the genus *Escherichia* was obtained. The genome of the isolate exhibited a low number of virulence factors and antibiotic resistance genes, a trait associated with a free-living lifestyle, which had been observed in other genomes of *cryptic clade IV*. Additionally, by analyzing available *cryptic clade IV* genomes from commensal and environmental sources, we identified genes potentially associated with the phenomenon of environmental persistence within the genus *Escherichia*. Overall, this study contributes to understanding the bacterial persistence phenomenon within the genus *Escherichia* and provides new hypotheses that can serve as the basis for future investigations.

Agradecimientos

Agradezco enormemente a mis directores, los Dres. Nicolás Frankel y Martín Graziano por haber depositado su confianza, inspirarme con su compromiso y pasión por la tarea que realizamos, y ser referentes en cómo realizar ciencia crítica y de calidad. Además, quiero agradecerles en forma conjunta por haber armado un gran equipo de trabajo, apostando a llevar adelante nuestros objetivos, cada uno desde un rol constructivo y sinérgico.

Particularmente, agradezco a Nicolás por su continua disposición a colaborar en cada etapa de este proyecto de Tesis Doctoral, desde el transmitir su conocimiento en la biología molecular y genómica, discutir sobre cómo llevar adelante cada uno de nuestros objetivos, y ser una figura presente y dedicada en instancias de escritura y publicación. Tu creatividad a la hora de ofrecer consejos metodológicos y resolver escollos de mesada, la inmensa variedad de recursos facilitados como espacios y materiales de trabajo al cabo de estos años, y tu receptividad para hacer oído cuando aconteció la necesidad, vuelven enormemente grata mi experiencia.

Agradezco a Martín por ser un invaluable norte y compañero cotidiano en la sintonía fina de este trabajo. Me has ayudado a crecer profesional y personalmente desde tu mirada humana, crítica y paciente. Has cultivado motivación y disciplina de trabajo, sin descuidar los objetivos detrás del por qué hacemos lo que hacemos desde el lugar que ocupamos. Has puesto todo tu talento al servicio de lo que te apasiona y priorizas desde tu práctica laboral: una Ciencia con perspectiva regional y ciudadana, al servicio de construir respuestas innovadoras para nuestro territorio y gente; y esto te vuelve una persona admirable. Inmensa gratitud y orgullo por ser parte del *grupo arroyos* que encabezás dentro del Laboratorio de Limnología, cuyas líneas de investigación aplicada gestadas al cabo de los últimos años, con seguridad, aguardan un futuro más que promisorio.

Es enorme mi agradecimiento con cada uno de los miembros del Laboratorio de Limnología, quienes más que compañerxs de trabajo, se han vuelto mis maestrxs y amistades. A las Directoras del Laboratorio, Irina, Inés y Haydée, a quienes admiro por su garra y dedicación en hacer una ecología situada, y ser mentoras de un grupo humano diverso, innovador y colaborativo. A Sebas, por ser un apoyo sostenido desde casi los comienzos de este trabajo y un gran amigo de la vida. A Carmen, por su espíritu docente a la hora de ofrecer sus aprendizajes y guía, a la vez que amadrinando con calidez y atención cada etapa del proceso. Y a cada una/o de las compañeras y compañeros que,

entre mates, almuerzos y campañas, hemos forjado lazos de cariño, compartido anécdotas, nervios, *scripts* y alegrías.

Quiero agradecer también al personal técnico y administrativo de nuestro Instituto y Departamento, por sus colaboraciones en el día a día de nuestro trabajo, atentos siempre a accionar en equipo e, incluso, a muchas veces hacer magias con lo que hay. Particularmente, quiero mencionar a nuestro técnico bioinformático David Kachanovsky, cuyas colaboraciones indispensables se plasman principalmente en los capítulos 2 y 5 de esta Tesis. Agradezco tu avidez puesta al servicio de colaborar en nuestro proyecto, a la vez que tu paciente escucha y asertividad a la hora de ayudarnos a resolver nuestras preguntas.

A mi familia, que es mi vieja y mis hermanos. No hacen falta muchas palabras para describir cuánto los amo y agradezco su apoyo infinito, acompañando y cuidando de cerca. Frente a toda adversidad, estuvimos, estamos y estaremos siempre el/la unx para el/la otrx.

A mis amigas y amigos, con quienes no compartimos sangre, pero estamos fortuitamente hermanadxs. Muchxs ya peinamos canas, pero hay algo del alma joven y fresca que no se quita. Gracias por tanto amor, alegría y contención.

A mi compañero de vida, por su maestría en amor incondicional y dulzura superlativa. Germinar, crecer, cambiar cortezas y florecer junto a vos es de lo más sagrado y expansivo que pueda concebir.

Y, por último, más que agradecer, deseo valorar a la educación y al sistema de ciencia públicos de nuestro país. Sin estas oportunidades y apoyos, es probable que no hubiera podido desarrollarme en este campo. Aspiro estar a la altura de poder devolver a nuestra sociedad la gran inversión realizada en mí.

Introducción General

Introducción General

Las ciudades son sistemas socio-ecológico-tecnológicos complejos en los que las poblaciones humanas juegan un rol clave en la definición de las condiciones ambientales predominantes¹. Esto implica una densa trama de componentes demográficos, infraestructurales y de otros sistemas tecnológicos en intersección con el sistema natural². La intensidad y calidad del avance urbano moderno impone alteraciones profundas sobre los procesos biogeoquímicos y comunidades bióticas asociadas al territorio. En particular, las cuencas hidrográficas y sus cauces son receptores de graves deterioros en su integridad ecológica. Diversos autores han propuesto la metáfora del *síndrome del arroyo urbano* para identificar a la serie de alteraciones que comparten estos sistemas³. Entre los síntomas comunes se encuentra (i) un mayor escurrimiento superficial (caudal) por la impermeabilización del suelo, vinculada a una aceleración de la velocidad con la que el exceso de agua drena sobre el cauce, (ii) elevados niveles de nutrientes y materia orgánica en la columna de agua y sedimentos, (iii) presencia de sustancias tóxicas tales como metales pesados y agroquímicos, (iv) y mayor simplicidad en la estructura del canal. A nivel biológico, (v) una mayor eutrofización de las aguas, (vi) mayor abundancia de especies tolerantes y pérdida de las más sensibles, (vii) y modificaciones en el metabolismo ecosistémico tales como una menor captación de nutrientes y materia orgánica⁴. Estos aspectos contribuyen de distinta manera a la alteración de los procesos fundamentales para la integridad ecológica de los cuerpos de agua urbanos, a la vez que vulneran los servicios ecosistémicos que proveen⁵. A pesar del grave riesgo que suponen estas transformaciones, la diversidad de efectos ecológicos que tiene la urbanización sobre los arroyos no ha sido estudiada en profundidad.

Un rasgo característico de los sistemas acuáticos urbanos es su **hiperconectividad** con el paisaje aledaño. Diversos autores han señalado que las distintas infraestructuras características del ámbito urbano tales como las redes de drenaje pluvial y la trama de caminos actúan como vías efectivas de traslocación de contaminantes por escorrentía hacia las aguas superficiales¹⁰. Es decir, estas conexiones actúan como nexos entre la actividad humana en las áreas secas de la cuenca y los cursos de agua, volviendo a los últimos en sumideros. En efecto, la **presencia de infraestructura de drenaje urbano, así como el aumento de las superficies impermeabilizadas, han sido positivamente asociadas a concentraciones elevadas de DOC, SRP, TP, amonio, nitratos y nitritos, y a elevada conductividad eléctrica en arroyos urbanos**¹¹⁻¹³. A su vez, las concentraciones elevadas

de contaminantes en estos sistemas no son resultado únicamente del ingreso incrementado de fuentes puntuales y difusas, sino también como producto de la disminución en la eficiencia de remoción de nutrientes en zonas ribereñas y cauces hidrológicamente desconectados, dada la simplificación sistemática de los mismos¹⁴⁻¹⁶. A la vez, se sabe que la presencia y cobertura de vegetación acuática es fundamental para la remoción de contaminantes, tales como los nutrientes, incorporándolos de manera directa de la columna de agua a su biomasa, o indirectamente, aumentando la superficie disponible para el crecimiento de algas epífitas o bacterias que los metabolicen¹⁷. Además, su ocurrencia afecta a otros factores como la velocidad de corriente y la retención de sedimentos finos, e incrementa la heterogeneidad física proporcionando de hábitat a otras comunidades¹⁸.

Las presiones antropogénicas actúan sobre distintas escalas espaciales -globales, regionales y locales-, determinando patrones biogeoquímicos y ecológicos particulares en cada red hidrológica, en tanto que modulan su heterogeneidad espacial e intensidad en los disturbios^{3,5-7}. A pesar de la tendencia homogeneizadora que impone el desarrollo urbano sobre las características de los ecosistemas fluviales, es esperable que la respuesta ecológica de cada red varíe entre regiones climatológicas y fisiográficas del planeta⁸. Para el estudio de los factores que modulan las condiciones en cada cuenca, así como para la generación de propuestas factibles de manejo, se requiere de un enfoque que aborde la multiplicidad de dimensiones que confluyen sobre las mismas. La visión del paisaje urbano como un sistema socio-ecológico-tecnológico posibilita considerar explícitamente las conexiones existentes entre los procesos antropogénicos en el territorio con los procesos ecológicos que co-ocurren en él⁹.

Factores determinantes en la distribución y persistencia de contaminantes en cuencas urbanas

La distribución y concentración de los contaminantes en una cuenca depende de un gran número de factores. Un factor principal es el tiempo de vida media de los mismos, íntimamente relacionado con su naturaleza química y las interacciones que acontezcan entre los contaminantes y los demás elementos del medio¹⁹⁻²¹. Por ejemplo, contaminantes comúnmente hallados en cuencas urbanas como los metales pesados tienen mayores tiempos de residencia en forma activa en el medio dado que sus concentraciones son moduladas principalmente por procesos fisicoquímicos (por ejemplo, redox y

adsorción) y procesos hidrológicos. Por el contrario, las sustancias xenobióticas biodegradables o las bacterias alóctonas a estos sistemas, son susceptibles a los procesos naturales y, en consecuencia, persisten durante menos tiempo en el ambiente. Además de esta distinción, la distribución y persistencia de los contaminantes en una red hidrológica dependerá en cada caso del tipo de uso del territorio, de factores climáticos, topográficos, hidrológicos, y de factores fisicoquímicos como la temperatura, pH, concentración de oxígeno y la presencia de sólidos en suspensión, entre otros^{7,22}. Adicionalmente, estos patrones responden a procesos vinculados a las especificidades de cada red hidrológica, en los que la conectividad y direccionalidad en los flujos influyen sobre la distribución de los constituyentes solubles o en suspensión, o bien responden a procesos de carácter más regional en los que pesa mayormente la distancia euclíadiana de las locaciones entre sí, sin estar vinculados a las dinámicas hidrológicas²³. Como resultado, cada compuesto sigue un tipo particular de patrón espacial, el cual es función tanto de su naturaleza como de las presiones que modelan su ocurrencia. Comprender estos patrones espaciales y los factores incidentes es de fundamental importancia para un manejo efectivo de la calidad de aguas, ya que este conocimiento provee un enorme potencial diagnóstico y predictivo⁷.

El rol de la comunidad bacteriana y su asociación con los disturbios urbanos

La compresión de los procesos ecológicos que caracterizan a los ecosistemas está fundamentalmente sujeta al estudio de sus flujos metabólicos, es decir, a la cuantificación de la materia y energía que se fija dentro de los mismos o que se importa o exporta a través de sus límites. Tal comprensión está bien desarrollada para los ecosistemas terrestres y lóticos, mientras que la fenología ecosistémica de los ecosistemas lóticos se encuentra escasamente descripta en general, y en menor grado para aquellos en contextos de elevada complejidad como lo son los arroyos urbanos²⁴. En esta línea, en los últimos años se ha avanzado en la comprensión del rol clave de la comunidad bacteriana como mediadora de los ciclos biogeoquímicos y de los flujos energéticos, pero la relación entre la estructura de la **comunidad microbiana y las funciones del ecosistema** sigue siendo **poco conocida**²⁵. Estudios recientes han demostrado que los patrones globales de la comunidad microbiana correlacionan con factores ambientales como la salinidad, el pH y el tipo de hábitat^{25,26}. Adicionalmente, las relaciones entre estructura y función bacteriana y los procesos ecosistémicos en los que participan son variables entre tipos de ambiente, incluso evidenciando rasgos específicos a escala local²⁷. Ahondar en la

compresión de la estructura de la comunidad bacteriana en cuerpos de agua urbanos puede generar predicciones sobre procesos biogeoquímicos de enorme interés en contextos de fuerte presión antrópica, como por ejemplo esclareciendo su vínculo con las tasas de ciclado de carbono y nitrógeno²⁵.

Con respecto a la biodiversidad en los sistemas acuáticos urbanos, se ha descripto que las presiones antropogénicas generan la pérdida significativa de insectos, peces y anfibios. Sin embargo se conoce poco sobre cómo los factores antropogénicos afectan la composición y funcionamiento de la comunidad microbiana²⁸. Disturbios frecuentes para estos sistemas, como el ingreso de aguas pluviales, grises y negras conllevan una inmensa carga de bacterias alóctonas, lo que representa uno de los mayores factores de cambio en la composición de la comunidad microbiana nativa. Por un lado, la escorrentía de las superficies impermeabilizadas aporta microorganismos asociados a las actividades antropogénicas en las áreas drenadas, a la vez que los nichos artificiales como los sistemas pluviales subterráneos contribuyen con la impronta de su diversidad y estructura microbiana²⁹. La presencia de pozos sépticos en zonas aledañas a los cauces también ha sido identificada como una fuente importante de contaminación no-puntual debido a la carencia del aislamiento adecuado o por su mantenimiento deficitario³⁰. La fuga de materiales desde pozos sépticos actúa como una fuente de bacterias fecales y patógenas, nutrientes disueltos y otros contaminantes hacia los acuíferos subterráneos y, potencialmente, hacia las aguas superficiales lindantes³¹. De igual manera, los sistemas cloacales urbanos pueden presentar fugas en sus trayectos o verse superada su capacidad hidráulica, y exfiltrar aguas residuales hacia los cuerpos de agua aledaños³². Por ende, las condiciones locales del nicho ecológico y los procesos de dispersión facilitada mediados por infraestructuras humanas jugarían un potencial rol central en modular a las comunidades microbianas en el ambiente, en particular controlando la distribución de bacterias alóctonas en las aguas superficiales asociadas a aglomerados urbanos. La liberación masiva de microorganismos alóctonos en aguas naturales podría alterar los flujos de nutrientes en las redes alimentarias y favorecer la existencia de puntos calientes para el intercambio de factores genéticos de virulencia y de resistencia antibiótica entre taxones bacterianos, lo que contribuiría a la proliferación de patógenos emergentes para las poblaciones humanas^{29,33}.

El estudio del bacteriplancton de aguas superficiales urbanas puede contribuir a comprender los efectos de la contaminación, informar sobre la existencia de riesgos para

la salud humana, o formular predicciones sobre la estabilidad de esos ecosistemas en el largo plazo³⁴. La secuenciación del gen 16S ARNr es una herramienta clave en el estudio estructural y funcional de la comunidad bacteriana³⁵. Este gen contiene regiones que presentan un elevado grado de variabilidad entre taxones bacterianos, la cual es aprovechada como base para la discriminación y asignación taxonómica de variantes de secuenciación obtenidas a partir de muestras de ADN ambiental³⁶. Si bien este enfoque metodológico permite una buena resolución taxonómica (generalmente alcanza hasta el nivel de género bacteriano), la información que aporta se restringe a lo estrictamente composicional, es decir, no refleja las abundancias muestrales absolutas de cada taxón³⁷. Sin embargo, el enorme caudal de información que provee y la progresiva disminución en sus costos, han vuelto clave el empleo de esta tecnología en el campo de la ecología microbiana acuática.

Escherichia coli en ambientes acuáticos urbanos y el fenómeno de persistencia ambiental

Escherichia coli es una especie bacteriana gram negativa perteneciente a la familia *Enterobacteriaceae*, que está asociada principalmente con el comensalismo entérico de animales homeotermos³⁸. Existen cepas con capacidad patogénica que son clasificadas en perfiles de patogenicidad según el tejido que colonizan y el tipo de síntoma principal asociado a su infección³⁹. La detección de *E. coli* en el ambiente ha sido históricamente empleada como indicadora de contaminación fecal reciente⁴⁰. Sin embargo, en los últimos años se reportó una voluminosa cantidad de evidencia sobre su capacidad de persistir en nichos secundarios (hábitats dulceacuícolas y marinos, sedimentos y otras superficies colonizables)⁴¹⁻⁴³. Los procesos ecológicos que influencian la estructura poblacional de *E. coli* en hábitats acuáticos urbanos son poco conocidos, pero se hipotetiza que su ubicuidad sería el producto de un complejo equilibrio entre el aporte externo, condiciones ecológicas permisivas y procesos de nicho actuando sobre estos sitios^{38,44}.

En cuanto a las condiciones abióticas que contribuirían al fenómeno de persistencia ambiental de *E. coli* en estos ambientes, estudios previos han sugerido la influencia positiva de factores tales como la disponibilidad de nutrientes y las temperaturas medias favorables; se ha demostrado que los niveles elevados de nutrientes solubles tales como el DOC o el SRP tienen una influencia positiva sobre la sobrevida y el crecimiento de *E. coli* en la columna de agua de arroyos urbanos⁴⁴. Esta evidencia sugiere que procesos de nicho tales como la exclusión competitiva y la partición de nicho, así como fenómenos

de dispersión facilitada, serían claves en el modelado de la distribución espacial de la diversidad filogenética de *E. coli* en ambientes secundarios como los arroyos urbanos. Sin embargo, no existe aún evidencia *in-situ* que valide el rol de estos procesos.

Relación entre la diversidad filogenética de Escherichia coli y su persistencia en ambientes acuáticos urbanos: aspectos genómicos, ecológicos y sanitarios

Dentro de *E. coli* se han descripto 8 grupos filogenéticos o filogrupos, identificados en la literatura con letras: A, B1, B2, C, D, E, F y G⁴⁵⁻⁴⁷. A pesar de la detallada descripción genotípica y fenotípica de las cepas de laboratorio y patogénicas, los vínculos entre los distintos filogrupos y sus nichos ecológicos permanece aún en discusión. En este sentido, los patrones de abundancia informados en estudios descriptivos realizados sobre cuerpos de agua dulceacuícolas superficiales evidencian en forma consistente la predominancia de los filogrupos A y B1. Por contrario, la distribución del resto de los filogrupos es de menor frecuencia relativa y se encontraron asociaciones con zonas de elevada presión urbana (por ejemplo, para los filogrupos D y F)⁴⁸⁻⁵⁰. La sobrerepresentación de ciertos grupos filogenéticos en el ambiente puede atribuirse a un *fitness* diferencial⁵¹.

La evidencia basada en estudios observacionales asocia generalmente al filogrupo abundante A, y también a grupos de menor representación como B2, al comensalismo entérico de animales⁵². En tanto, la presencia del filogrupo A en ambientes secundarios podría estar mediada tanto por mecanismos de dispersión facilitada a través de las heces de hospedadores, así como por rasgos diferenciales que favorezcan su persistencia ambiental. Por su parte, estudios realizados en microcosmos con aguas provenientes de un estuario han revelado que el filogrupo B1 tiene un mayor tiempo de sobrevida y tolera temperaturas más bajas en comparación con el resto de los filogrupos, lo que representa un indicio de una mayor capacidad de permanencia en ambientes secundarios⁵³.

Van Elsas *et al*⁵⁴ argumentan que si bien no hay evidencia explícita sobre una relación intrínseca entre la filiación intraespecífica de *E. coli* y su capacidad de supervivencia en el ambiente, las diferencias genómicas entre filogrupos permitirían hipotetizar una aptitud diferencial. Se sabe que el genoma común a todas las cepas de *E. coli* cuenta con alrededor de 2.000 genes que determinan características estructurales y metabólicas transversales a la especie. Sin embargo, su pangenoma sería de aproximadamente 75.000 genes, entre los cuales estarían los genes que determinan los fenotipos divergentes entre grupos

filogenéticos^{55,56}. Además, Touchon *et al.*⁵⁶ describió que el tamaño promedio de los genomas de filogrupos A y B1, que son más generalistas, son más pequeños que los del resto de los filogrupos, que se presumen especialistas, dado que se detectan en menor abundancia relativa y confinados a ciertos nichos específicos. Por otro lado, al comparar los pangenomas de todos los filogrupos, se observa que A y B1 tienen los pangenomas más grandes, y en esta característica estaría la base funcional para su generalismo ecológico, indicio de un ciclo de vida bifásico de comensalismo-vida libre mediado por mecanismos de dispersión y reinfección.

Además, se ha reportado que algunas cepas de *E. coli*, principalmente pertenecientes al filogrupo B1, exhiben rasgos que contribuirían a favorecer su persistencia ambiental, como una mayor capacidad en la formación de biofilms⁵². Los biofilms se definen como agregados bacterianos producto de la secreción de sustancias protectoras (sustancias poliméricas extracelulares) que les confieren a los microorganismos la posibilidad de alojarse sobre superficies sólidas y asegurar su persistencia y resguardarse de condiciones adversas comunes a los ambientes no entéricos (deshidratación, radiación UV y depredación por protozoos)^{52,57}. En referencia al desarrollo de biofilms por parte de *E. coli*, este aspecto guarda una relevancia sanitaria adicional, dado que estas estructuras pueden conferir carácter infectivo a las cepas, como fue detectado en cepas uropatogénicas de *E. coli* (UPEC, del inglés *UroPathogenic Escherichia coli*)⁵⁸. Adicionalmente, existe abundante bibliografía en relación con la variedad de reservorios de *E. coli* en ambientes naturales y artificiales bajo la forma de biofilms. Por ejemplo, se han observado biofilms sobre hojas de plantas macrófitas y plantas terrestres (vector vinculado a brotes epidémicos de *E. coli* patogénicas por consumo de vegetales frescos) y también en tuberías de sistemas de distribución de agua potable. Además, se ha descripto que la formación de biofilms de *E. coli* puede comprometer los esfuerzos de esterilización de los materiales quirúrgicos, siendo una importante fuente de infecciones ligadas a procedimientos médicos^{52,58-60}.

Clados crípticos de Escherichia: filogenia, factores genómicos y la hipótesis ambiental

Dentro del género *Escherichia* se han descripto otros grupos monofiléticos conocidos como “clados crípticos” (identificados con números romanos del I al V) en base a la imposibilidad de distinguirlos de aislamientos representativos de *E. coli* mediante reacciones bioquímicas de diagnóstico típicas⁶¹. Los estudios filogenéticos ubican al

clado críptico I en proximidad a *E. coli* y la evidencia genómica indica que los aislamientos de este grupo portan factores de virulencia comunes a *E. coli*, por lo que también podrían ser patógenos^{61,62}. Los factores de virulencia bacterianos codifican para componentes determinantes en el establecimiento de perfiles de infección, contribuyendo, por ejemplo, con toxinas o elementos para la adherencia bacteriana a los tejidos diana^{60,63,64}. La distribución de estos factores en taxones bacterianos, así como de los factores de resistencia antibiótica, suele vincularse a presiones de selección específicas de la vida asociada a huéspedes, mientras que la vida libre no favorecería la posesión de este tipo de genes^{65,66}. En contraparte, la evidencia sostiene la hipótesis de que los clados crípticos restantes, del II al V, son grupos oportunistas como patógenos. Esto se basa en que, en general, no han sido vinculados a casos de infección humana y/o animal y en que, mayormente, carecen de los factores de virulencia específicos de *E. coli* para los perfiles de infección intestinal y extra-intestinal^{67,68}. Además, en un estudio genómico comparativo se sugirió que los genomas de los clados crípticos ligados a la vida libre estarían enriquecidos en genes funcionales que confieren una mayor sobrevida en el ambiente⁶⁹. Esta información, en conjunto con una sobrerepresentación por parte de estos grupos en muestras ambientales, sirvieron de base para asociar a los mismos al estilo de vida libre, lo que se conoce como la “hipótesis ambiental”, aunque la evidencia es aún acotada^{53,61,69,70}. El aumento de los estudios genómicos sobre estos taxones permitiría, por un lado, alcanzar un mayor discernimiento en cuanto a los caracteres distintivos a nivel filogenético, fisiológico y ecológico de estos grupos filogenéticos, y a la vez indagar sobre los factores genéticos asociados al fenómeno de persistencia ambiental dentro del género *Escherichia*.

Área de estudio

El Área Metropolitana de Buenos Aires (AMBA) es una de las regiones más pobladas de Latinoamérica, con aproximadamente 14 millones de habitantes en una superficie de 13285 km²⁷¹. Los cursos principales que drenan esta región se encuentran afectados por múltiples estresores antropogénicos. Por un lado, la mayoría de sus redes hidrológicas sufrieron modificaciones hidráulicas tales como canalizaciones, rectificaciones, desviaciones, y entubamientos parciales o totales. Además, es particular de estas aglomeraciones la heterogénea e insuficiente cobertura en el desarrollo de infraestructuras sanitarias y tratamientos sobre los residuos urbanos, lo cual conlleva a un elevado grado

de contaminación industrial, cloacal y doméstica⁷². En consecuencia, los cuerpos de agua superficiales del AMBA reúnen un notable estado de deterioro en la calidad de sus aguas y de sus comunidades biológicas, reuniendo percepciones socio-ambientales fuertemente negativas y ejerciendo un riesgo latente al desarrollo de la vida, incluida la humana⁷³⁻⁷⁵.

La cuenca de los arroyos San Francisco, Las Piedras y Santo Domingo se ubica en la llanura pampeana argentina, en el sur del AMBA, y recorre parte de los partidos de Avellaneda, Quilmes, Florencio Varela y Almirante Brown (Figura 1). Esta región conurbana, que circunvala a la Ciudad de Buenos Aires, presenta un alto grado de urbanización. El AMBA es irrigada por numerosas subcuenca exorreicas de bajo caudal y escasa pendiente que escurren sus aguas hacia el estuario del Río de la Plata, que son agrupadas bajo el dominio de un sistema hidrográfico mayor a nivel regional, la Cuenca del Plata.

El clima de esta región es templado, con temperaturas medias anuales que varían de 14 a 20 °C. La precipitación se concentra durante los meses de primavera y verano, y disminuye de NE a SO (de 1.000 a 400 mm). La región se corresponde con un régimen pluviométrico moderadamente húmedo⁷⁶. En cuanto a su clasificación fitogeográfica, el área pertenece al dominio chaqueño de la región neotropical, en particular, a la Provincia Pampeana⁷⁷.

La cuenca en estudio comprende un área superficial total de aproximadamente 160 km², con una extensión longitudinal cercana a los 23 km. El área se encuentra drenada por una red de arroyos de bajo orden, cuyos cursos principales son los arroyos San Francisco, Las Piedras y el resultante de la confluencia de estos dos junto al arroyo Las Perdices (entubado), el arroyo Santo Domingo (Figura 1). Adicionalmente, en las cabeceras de la cuenca se observan numerosos cursos intermitentes, tributarios principalmente al arroyo Las Piedras. La red de arroyos se halla afectada por numerosos disturbios antropogénicos vinculados principalmente a la actividad urbana, dado que el uso del suelo predominante de la cuenca es de tipo urbano y periurbano (Figura 2). La sucesiva ocupación del territorio por la urbanización alteró el drenaje natural de las cuencas de los arroyos en estudio, además de provocar la impermeabilización del suelo, con el consiguiente incremento del escurrimiento superficial. El área posee una red de desagües pluviales más o menos desarrollada, complementada con un conjunto de conducciones menores que drenan barrios más alejados, llegando incluso a conducciones en zanja en aquellas zonas de asentamientos más precarios o calles sin pavimentar⁷⁴.

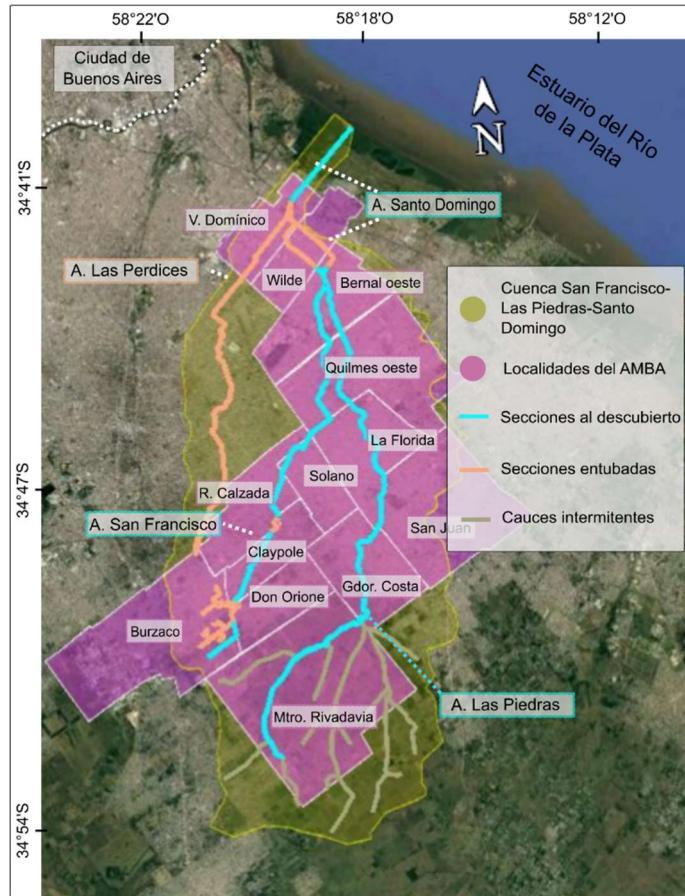


Figura 1: Ubicación de la cuenca San Francisco-Las Piedras-Santo Domingo y sitios de muestreo. La cuenca de estudio se ubica en el lado sur del Área Metropolitana de Buenos Aires.

El desagüe natural de la red pluvial son los mencionados arroyos, y finalmente por éstos, al Río de la Plata. Este sistema de drenaje urbano se encuentra ubicado en una zona muy poblada y de importante desarrollo industrial, recibiendo además de los aportes pluviales, efluentes domiciliarios, cloacales e industriales, por lo cual el agua conducida posee niveles altos de contaminación. La falta de conservación y mantenimiento, sobre todo de las canalizaciones, contribuye a esta contaminación y a la degradación del entorno. Por su parte, los cursos contenidos han sido altamente modificados desde una perspectiva hidráulica: se han realizado rectificaciones, excavaciones/incisiones de los cauces, a la vez que se han soterrado numerosos tramos de estos. Tal es el grado de modificación de los cauces originales que incluso uno de los arroyos comprendidos en esta red, el Arroyo Las Perdices, fue traslocado hacia esta cuenca (Figura 1). En cuanto a la calidad de la urbanización extendida en la zona, se observa que el crecimiento demográfico no ha sido debidamente acompañado por inversiones en infraestructuras sanitarias y, como resultado de ello, los cuerpos de aguas aledaños reciben constantes aportes residuales sin el tratamiento de mitigación adecuado⁷⁴. Además, se advierte que gran parte del área

involucrada tiene graves problemas de inundación y anegamiento tras eventos de precipitaciones intensas, principalmente en los tramos medios e inferiores de la cuenca. Este problema hídrico sería consecuencia principalmente de una geomorfología local de escasa pendiente que dificulta el drenaje, sumada a la simplificación de las canalizaciones naturales. A su vez, el crecimiento demográfico no planificado sobre la planicie de inundación de los cuerpos de agua ha llevado a la formación de asentamientos informales que sufren las inundaciones frecuentes^{78,79}.

En este sentido, estudios previos dan cuenta de la heterogeneidad en el estado de integridad ecológica de los arroyos San Francisco, Las Piedras y Santo Domingo en sus distintos tramos. Siendo relativamente elevada la contaminación orgánica en la totalidad del sistema, la cuenca alta, principalmente las nacientes del arroyo Las Piedras y, en menor grado, la cuenca alta del arroyo San Francisco, mantienen las mejores condiciones ecológicas (mínimos de modificaciones geomorfológicas y máximos de cobertura de plantas macrófitas y conectividad con el hábitat ripario). Conforme los cauces atraviesan el entorno urbano, abruptamente los parámetros típicos de calidad de agua se ven desmejorados, alcanzando el máximo grado de deterioro en los tramos de la cuenca baja, próximos a la confluencia de los arroyos San Francisco y Las Piedras, y en el arroyo Santo Domingo⁸⁰⁻⁸². El elevado grado de impacto por efluentes tanto domésticos como industriales se resume en una gran carga orgánica, lo que contribuye a un bajo nivel de oxígeno disuelto, a la presencia de patógenos de origen fecal (bacterias, protozoos y nematodos) y a la detección de concentraciones de compuestos químicos (Cd, Cu, Pb, Cr e hidrocarburos, entre otros) por encima de los niveles guía para las convenciones de calidad ambiental internacionales⁷⁴.

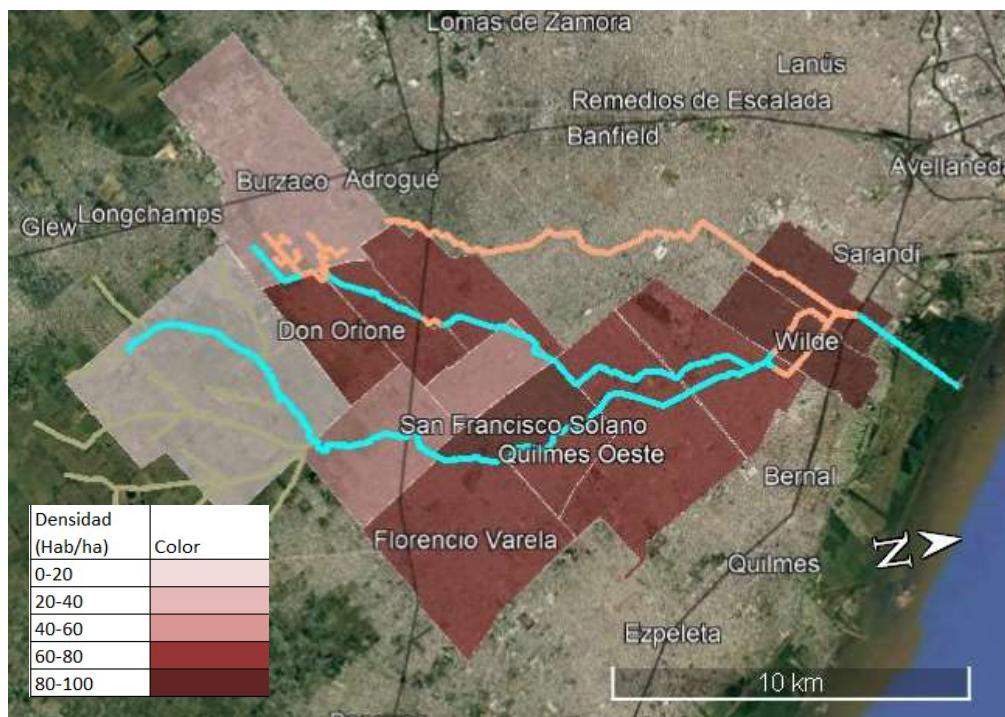


Figura 2: Distribución de la densidad poblacional por municipio en la cuenca.

Objetivos

Objetivos

Objetivo general

Comprender la estructura espacial del bacteriplancton en una red hidrológica de arroyos con un elevado grado de urbanización y, en particular, de las poblaciones asociadas a la contaminación de origen fecal, como *E. coli*, y el efecto de factores ecológicos y genéticos que podrían contribuir favorablemente a fenómenos de persistencia ambiental en estos ambientes.

Objetivos específicos por capítulo

Capítulo 1

1.1 Analizar la distribución de las condiciones socio-ecológico-tecnológicas que confluyen en la red hidrológica de los arroyos San francisco, Las Piedras y Santo Domingo.

Capítulo 2

2.1 Analizar la diversidad y distribución espacial del bacteriplancton en la red hidrológica.

2.2 Detectar el núcleo (*core*) de taxones bacterianos de la cuenca basado en curvas de rango y abundancia.

2.3 Analizar la influencia de los factores ambientales, urbanos y espaciales sobre la composición del bacteriplancton.

2.4 Estudiar cambios en la co-ocurrencia entre taxones bacterianos según el grado de disturbio del sistema hidrológico.

2.5 Identificar los factores que podrían afectar a la distribución espacial de taxones bacterianos potencialmente patógenos y de taxones asociados a la contaminación fecal humana en la cuenca.

Capítulo 3

3.1 Analizar el impacto de las condiciones de infraestructura, hábitat y fisicoquímicas concurrentes sobre la abundancia y estructura filogenética de *E. coli* en la cuenca de estudio.

3.2 Identificar procesos ecológicos que puedan influenciar la dinámica poblacional de *E. coli* en la cuenca.

Capítulo 4

4.1 Analizar experimentalmente la supervivencia y la formación de biofilm de diferentes filogrupos de *E. coli* en aguas provenientes de la cuenca.

4.2 Comparar la supervivencia y la formación de biofilm en el laboratorio con relación a la posible persistencia ambiental de los diferentes filogrupos de *E. coli*.

Capítulo 5

5.1 Analizar el genoma de un aislamiento correspondiente al clado críptico IV, proveniente de la cuenca en estudio.

5.2 Realizar un análisis comparativo entre el genoma del aislamiento correspondiente al clado críptico IV con otros genomas del género *Escherichia* de origen ambiental y comensal.

Hipótesis de trabajo

Hipótesis de trabajo

Capítulo 1

H1.1 Las características y la intensidad de la urbanización en la cuenca son heterogéneas, tendiendo a alcanzar su máxima intensidad en los tramos de la cuenca baja.

H1.2 La contaminación orgánica es ubicua en la cuenca de estudio.

H1.3 La heterogeneidad ambiental dentro de los tramos analizados es un componente de menor relevancia respecto a la variabilidad global de las condiciones ambientales de la cuenca.

Capítulo 2

H2.1 La diversidad específica de la comunidad bacteriana es modelada tanto por la presencia de infraestructuras urbanas como por las condiciones de los ambientes locales.

H2.2 La topología de la red de co-ocurrencia entre taxones bacterianos se verá alterada conforme varíe el grado de disturbio en los sitios de muestreo.

H2.3 La presencia de bacterias de origen fecal y/o patógenas es ubicua en la cuenca.

H2.4 La contaminación fecal es la principal fuente de patógenos en la cuenca.

Capítulo 3

H3.1 La abundancia de *E. coli* es influenciada de manera positiva por ambientes ricos en nutrientes (DOC y DIN) y con baja velocidad de flujo como los característicos de arroyos urbanos de nuestra región.

H3.2 La prevalencia de *E. coli* y su estructura genética poblacional en ecosistemas lóticos se ve afectada tanto por factores de infraestructura urbana como de control ecológico.

H3.3 La ocurrencia del filogrupo B1 aumenta con mayores concentraciones de nutrientes disueltos y a una mayor cobertura de vegetación acuática.

H3.4 La distribución de filogrupos de baja frecuencia como D y F se asocia a condiciones de elevada urbanización.

Capítulo 4

H4.1 En condiciones experimentales con agua de la cuenca, los aislamientos ambientales de *E. coli* presentan capacidades diferenciales de supervivencia en suspensión y de formación de *biofilm* en función de su identidad filogenética.

H4.2 Los aislamientos de *E. coli* pertenecientes a los filogrupos A y B1 expresarán mayores capacidades de supervivencia en suspensión y de formación de *biofilm* en agua de la cuenca que los pertenecientes al resto de los grupos filogenéticos evaluados.

Capítulo 5

H5.1 La detección de factores de virulencia y de resistencia antibiótica en los genomas de aislamientos del clado críptico IV es baja en comparación a los genomas de *E. coli*.

H5.2 La frecuencia de genes de interés ecológico y sanitario en los genomas del clado críptico IV está relacionada con la procedencia del aislamiento (ambiental o ligada al comensalismo entérico).

El Capítulo 1 fue retirado a pedido del autor

Capítulo 2

Análisis de la distribución espacial del bacteriplancton, con énfasis en la contaminación de origen fecal y en grupos potencialmente patógenos.

Capítulo 2

Análisis de la distribución espacial del bacteriplancton, con énfasis en la contaminación de origen fecal y en grupos potencialmente patógenos.

2.1. Resumen

Con el objetivo de analizar la estructura y distribución espacial del bacteriplancton en la red hidrológica de estudio se secuenció el gen del ARN ribosomal 16S de muestras obtenidas de la columna de agua de los 14 sitios y sus respectivas 3 ubicaciones intra-sitio (42 muestras en total). *Proteobacteria* (47%), *Bacteroidota* (20%), *Campylobacterota* (11%), *Firmicutes* (10%), *Patescibacteria* (7%) y *Actinobacteriota* (3%) fueron los principales filos detectados. A partir del análisis de curvas de rango y abundancia, de las 5867 secuencias únicas y taxonómicamente asignables (ASV), 364 taxones fueron identificados como ubicuos (*core*) a nivel de la cuenca. El análisis de los factores demográficos, infraestructurales, ecológicos y espaciales a partir de un análisis de partición de la varianza evidenció que la riqueza de ASVs es explicada mayormente por las condiciones fisicoquímicas y estructurales del hábitat (R^2 ajustado= 42%). Además, por análisis de la partición aditiva de la riqueza de ASVs, se observó que el recambio de especies intra-sitio (β_1) es menor al esperado por la teoría neutral, evidenciando un grado de homogeneización mayor al esperado a escala de sitio. Por el contrario, el recambio de taxones inter-sitio (β_2) fue mayor a lo esperado, lo cual es indicio de factores estructurantes de la distribución de los taxones bacterianos a lo largo de la red hidrológica. Por otro lado, se analizaron las redes de co-ocurrencia entre taxones bacterianos resultantes de agrupar a los sitios de la cuenca según su grado de disturbio antropogénico. La topografía de la red de co-ocurrencia asociada a los sitios menos impactados (cuenca alta) fue globalmente más cohesiva, mientras que a mayor presencia de disturbios la comunidad bacteriana se observó más fragmentada en pequeños clústeres con elevada conectividad interna. Este rasgo antagónico plantea escenarios hipotéticos de resiliencia ecológica disímiles ante cambios ambientales, hipotetizándose como más resilientes los sitios de la cuenca alta. Por último, se buscaron asociaciones entre la distribución espacial de taxones bacterianos asociados al origen fecal y a taxones potencialmente patógenos, y las condiciones socio-ecológico-tecnológicas en el área de estudio. Para ello se emplearon secuencias disponibles del gen del ARN ribosomal 16S de 30 microbiomas intestinales humanos del Área Metropolitana de Buenos Aires pertenecientes a personas sanas y sin sobrepeso, con igual representación de géneros y edad media 35.23 ± 8.26 años⁹¹. De este

set se obtuvo un núcleo de ASVs identitario de los microbiomas humanos de la región basado también en un análisis de curvas de rango y abundancia. Este núcleo fue de 316 ASVs, de los cuales 139 estaban presentes en las muestras de la columna de agua obtenidas de la cuenca. El menor porcentaje de lecturas vinculadas a microbiomas intestinales humanos en las muestras fue del $2,32 \cdot 10^{-4}\%$, en un sitio con bajo grado de disturbios perteneciente al arroyo Las Piedras. En cambio, la muestra que presentó mayor porcentaje de lecturas de origen fecal ($9,94 \cdot 10^{-2}\%$) pertenece a un sitio de muestreo a escasa distancia y aguas abajo del anterior, pero ubicado en la interfaz de la urbanización periférica y carente de red cloacal. Se realizó también un análisis de partición de la varianza de la riqueza y abundancia relativa de ASVs de origen fecal en la cuenca en el que se observó que las condiciones ecológicas de los hábitats locales son aquellas que más contribuyeron a explicar la varianza global, con una contribución del 16% en forma pura y del 29% en forma conjunta con los factores espaciales, mientras que los parámetros urbanos sólo explicaron en un 9% en su contribución pura. Además, se realizó una búsqueda bibliográfica de géneros y especies bacterianas potencialmente patógenas, los cuales fueron posteriormente identificados en las asignaciones taxonómicas del bacteriplancton. Se observó un porcentaje de detección en las muestras que sigue un patrón semejante al observado tras el análisis de las bacterias de origen fecal previo. Asimismo, la contribución global de las condiciones del hábitat fueron las que mayormente explicaron la distribución de taxones potencialmente patógenos en la cuenca (48%), con una contribución menor por parte de los factores urbanos (27%). En esta línea, se observó también que tanto la riqueza como el porcentaje de lecturas de ASVs fecales, de ASVs potencialmente patógenos, de ASVs asignados al género *Escherichia/Shigella* y la cuantificación por conteo de *E. coli* correlacionan en forma positiva y significativa entre sí, lo cual puede indicar su aporte y/o modulaciones comunes en estos ambientes, a la vez que confirma el rol indicador del riesgo sanitario que cualquiera de estas determinaciones posee sobre las otras. Además, se detectó de manera ubicua la presencia tanto de bacterias asociadas a la contaminación fecal como bacterias potencialmente patógenas en el bacteriplancton de todos los tramos analizados de la cuenca. Para nuestro conocimiento, este es el primer trabajo que analiza a la comunidad bacteriana de una red hidrológica urbana en toda su extensión por secuenciación masiva de ADN ambiental, en particular, con foco en la ocurrencia y distribución de bacterias alóctonas de origen fecal y potencialmente patógenas. Esto resume en un aporte de importancia ecológica y

sanitaria significativo en la elaboración de estrategias de saneamiento y mitigación de la contaminación de los sistemas lóticos urbanos.

2.2. Materiales y métodos

2.2.1 Secuenciación de ADN ambiental y procesamiento bioinformático

Con el objetivo de caracterizar la estructura de la comunidad del bacteriplancton en la cuenca, se tomaron muestras de agua subsuperficial en cada ubicación intra-sitio (42 muestras en total), en botellas estériles de 1 litro de volumen. Se realizó un filtrado secuencial de las muestras, primero por una membrana de 3 μm y luego por un filtro de policarbonato de 0,22 μm de poro, y se almacenaron las muestras a -80°C hasta su procesamiento. El protocolo de extracción y purificación del ADN ambiental procariota contenido en los filtros se basó en el empleo de buffer de lisis CTAB y mezcla de alcoholes cloroformo y alcohol isoamílico para su purificación. Posteriormente se secuenciaron las regiones hipervariables V3-V4 del gen codificador del ARNr 16S a partir del ADN ambiental extraído (cebadores 341F y 805R) mediante la plataforma Miseq de Illumina (lectura emparejadas 2 x 300)^{92,93}.

El procesamiento bioinformático de limpieza de las secuencias crudas se condujo empleando el paquete DADA2 en el entorno R⁹⁴. La asignación taxonómica de las secuencias se realizó empleando la base de datos SILVA como referencia. El procesamiento involucró el filtrado de calidad de las secuencias crudas y el corte de sus extremos usando los parámetros truncLen c = (260, 190) y maxEE = c (2, 5), la fusión de secuencias pareadas, la eliminación de quimeras y la asignación taxonómica de las variantes de secuenciación (ASV, del inglés *amplicon sequence variants*). Se eliminaron las secuencias no deseadas: cloroplastos, mitocondrias, eucariotas y arqueas (dataset de 19937 ASVs). Se hicieron análisis de rarefacción en los que se observó que todas las muestras llegaron al estado asintótico, lo que indica que la profundidad de secuenciación fue suficiente para representar la diversidad y riqueza del bacteriplancton. Se hicieron análisis de rarefacción para las 42 muestras a partir de la matriz obtenida. Las curvas resultantes demostraron que todas las muestras llegaron al estado asintótico (Figura 2.1), lo que indica que la profundidad de secuenciación fue suficiente para representar la diversidad y riqueza del bacteriplancton. Por último, se retuvieron sólo los ASVs que poseían más de 10 lecturas y que se encontraban al menos en dos muestras. El número final ASVs fue de 5867. Luego, las muestras se normalizaron al número de lecturas de la muestra con menos lecturas (Cmin) usando el flujo de trabajo SRS⁹⁵. Los análisis descriptivos sobre la tabla de ASV se realizaron empleando los paquetes *Phyloseq* y

Biostrings en R^{96,97}. Se evaluó la eficiencia de muestreo a través de curvas de acumulación de especies utilizando la función *specaccum* en el paquete *vegan* en R⁹⁸.

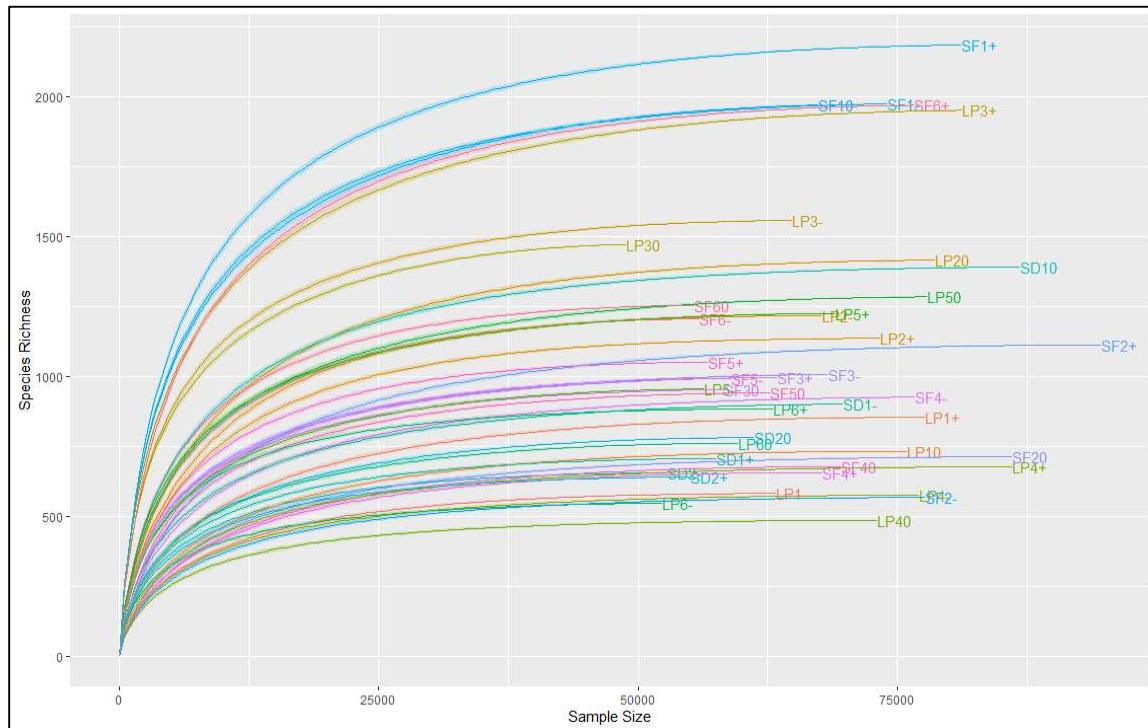


Figura 2.1: Curvas de rarefacción por muestra.

2.2.2 Identificación del core de ASVs a partir de curvas de rango y abundancia

La detección del núcleo o *core* de taxones bacterianos de la red hidrológica se realizó mediante un análisis de curvas de rango y abundancia empleando la tabla de ASV normalizada con la totalidad de las muestras⁹⁹. Para ello, se evaluó la frecuencia de detección de los ASVs en las muestras, en términos presencia/ausencia y de su abundancia relativa, información que sirvió de insumo para el armado de un ranking que combina ambos parámetros, ordenando los ASVs desde los más ubicuos y abundantes, a los más locales y escasos. Luego se obtuvo una matriz de similitud de Bray-Curtis con la totalidad del *dataset* y se evaluó progresivamente la contribución de los ASVs ordenados en el ranking en explicar la diversidad beta. El umbral para la inclusión de taxones fue un incremento final de la similitud de Bray Curtis igual o mayor al 2%, obteniendo así un *core* de 364 ASVs.

2.2.3 Determinación de secuencias potencialmente patógenas

En primer lugar, se identificaron los géneros potencialmente patógenos de humanos a partir de bases de datos de instituciones de referencia internacional en la investigación de patógenos (NIH, 2019; NIAID-*emerging pathogens*), así como de trabajos publicados

que ampliaran el listado de grupos bacterianos asociados a infecciones comunes, oportunistas, enfermedades emergentes, nosocomiales, etc.¹⁰⁰⁻¹⁰⁶. De esta búsqueda bibliográfica, se identificaron 53 géneros potencialmente patógenos para humanos. A partir de la información colectada, se rastrearon los géneros potencialmente patógenos en el *dataset* de nuestra red hidrológica, formando un subconjunto de lecturas pertenecientes a los mismos. Posteriormente, dentro de cada género potencialmente patógeno encontrado en las muestras de la cuenca, se identificaron 174 especies descriptas como patógenas, también en base a la bibliografía específica. De estas especies se descargaron secuencias ARNr 16S provenientes de cepas de referencia, preferentemente de la *American Type Culture Collection* (ATCC), y se evaluó la identidad entre las secuencias de los ASVs contenidos dentro de los géneros potencialmente patógenos y las secuencias de referencia de especies patógenas mediante BLASTn. Los ASVs fueron identificados como PLSs (del inglés, *Pathogen-Like Sequences Species*) cuando la identidad resultante fue igual o superior al 99%. A partir del *dataset* resultante se calcularon los parámetros ecológicos de la comunidad.

2.2.4 Determinación de secuencias vinculadas a microbiomas fecales humanos del AMBA

Con el propósito de identificar taxones bacterianos asociados a la contaminación fecal específicos de la región de estudio se empleó un *dataset* de 30 microbiomas fecales humanos del Área Metropolitana de Buenos Aires publicado por Belforte *et al.* (2019)⁹¹. La población muestral consta de personas sanas y sin sobrepeso, con igual representación de sexos y edad media de $35.23 \pm 8,26$ años. Las secuencias disponibles por los autores constan de la secuenciación del gen del ARN ribosomal 16S (subunidades V3-V4) a partir de muestras fecales, empleando los cebadores 341F y 805R (plataforma Miseq, Illumina; lectura *paired-end* 2 x 300). De este set de muestras se buscó un *core* de variantes de origen bacteriano basado también en un análisis de curvas de rango y abundancia, el cual fue de 316 ASVs. Posteriormente, se intersectaron las secuencias 16S ARNr de la cuenca con las correspondientes a los ASVs *core* fecal (criterio de similitud del 100%), creando el *subset* de ASVs de la cuenca asociados a contaminación fecal. A partir del mismo, se calcularon los parámetros ecológicos de la comunidad.

2.2.5 Procesamiento estadístico

2.2.5.1 Parámetros ecológicos de la comunidad y partición aditiva de la diversidad

Se analizaron los parámetros de diversidad de la comunidad fitoplanctónica en las muestras (42) a partir del set de 5867 ASVs. Los índices incluidos en análisis fueron la riqueza de ASVs, la equitatividad de Pielou y la dominancia de Simpson los cuales fueron calculados mediante el paquete *microbiome* para R^{107,108}. Esta misma metodología se empleó para el cálculo de la riqueza de ASVs fecales y PLSSs detectados en la cuenca. La asociación entre la riqueza de ASVs, la equitatividad de Pielou, la dominancia de Simpson y el índice de heterogeneidad ambiental (metodología explicitada en la Materiales y Métodos del Capítulo 1 de esta Tesis) fue analizada mediante un análisis de correlaciones de Pearson. La significancia de los coeficientes de correlación (r de Pearson) se evaluó con la prueba de correlación de muestras pareadas ajustada para comparaciones múltiples (método de Holm; 999 permutaciones) empleando los paquete *corrplot* (versión 0.84) y *permute* (versión 0.9-5) para R⁸⁹.

La diversidad de ASVs se particionó en la riqueza local de ASVs (número de ASVs distintos) en cada muestra analizada (α), el recambio de ASVs entre muestras dentro de cada sitio (β_1), y el recambio de ASVs entre sitios (β_2) usando la función *adipart* en el paquete *vegan*⁹⁸. Los valores de diversidad observados se compararon con los valores aleatorios esperados obtenidos por modelos nulos (método de *r2dtable*, con 999 simulaciones)¹⁰⁹, que proporcionan una base para las inferencias sobre los impulsores de la diversidad de especies a través de escalas.

2.2.5.2 Elaboración de predictores espaciales

Se diseñaron Mapas de Autovectores Asimétricos (AEMs, del inglés *Asymmetric Eigenvector Maps*) para modelar los procesos espaciales direccionales subyacentes a las singularidades de la red de arroyos^{110,111}. Estos AEMs se basaron en la distribución geográfica de los sitios de muestreo (14 sitios de muestreo con 3 ubicaciones cada una a lo largo de un gradiente longitudinal de 50 m), y la matriz de conectividad entre los sitios se construyó considerando el flujo río arriba-río abajo y una distancia máxima de vecindad espacial de 5 km entre los sitios. Los AEMs también incluyeron vectores de peso obtenidos de la matriz de conectividad, que determinaron la fuerza de conexión entre muestras según su distancia geográfica. De los autovectores obtenidos se seleccionaron aquellos con autocorrelación espacial positiva o negativa significativas (prueba de Moran de dos lados; 999 permutaciones; $p < 0,05$) (Figura 2.2)¹¹². Estos procedimientos se

llevaron a cabo utilizando los paquetes *adespatial* (v 0.3-8), *ade4* (v 1.7-15), *sp* (v 1.3-2) y *spdep* (v 1.1-3) en R.

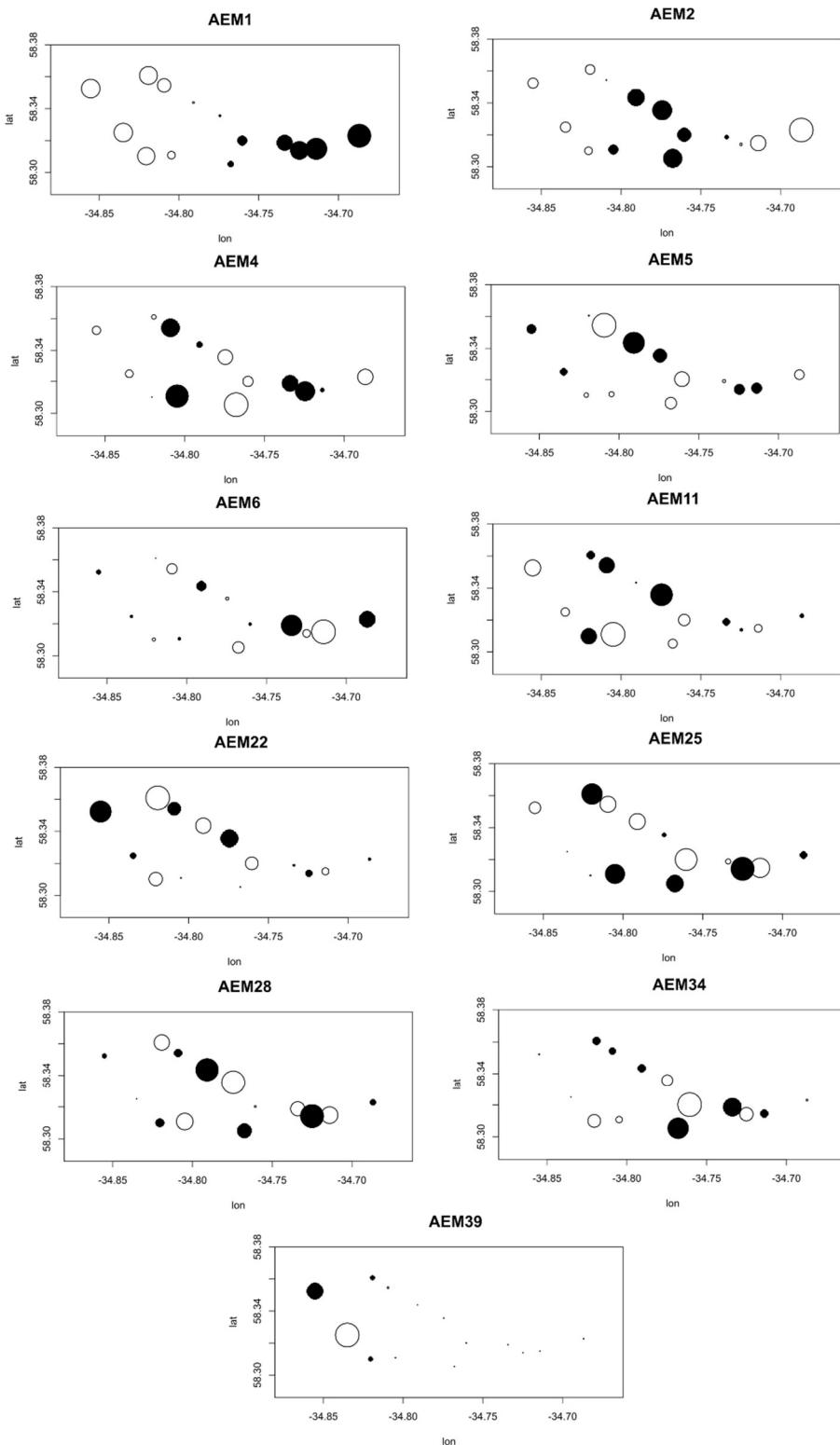


Figura 2.2: AEMs significativos para autocorrelación como resultado de la prueba de Moran de dos lados (999 permutaciones; $p < 0,05$).

2.2.5.3 Análisis de partición de la varianza

Se realizaron distintos análisis de partición de la varianza para descomponer los efectos puros y compartidos del hábitat local, la infraestructura urbana y los predictores espaciales en la determinación de los patrones de a) la riqueza de ASVs en la cuenca, b) la equitatividad en la composición de ASVs; c) la riqueza y abundancia relativa de ASVs asociados a bacterias fecales, y d) la riqueza de ASVs asociados a PLSSs. La proporción en la que cada componente contribuyó a explicar la varianza de las variables se analizó mediante el coeficiente de determinación ajustado (R^2a).

En todos los casos, la matriz de predictores del hábitat local se compuso de las siguientes variables: cobertura porcentual de plantas macrófitas, pH, OD, DIN, SRP, turbidez, DOC, profundidad y velocidad de flujo medias; la matriz de predictores de infraestructura urbana se compuso de las siguientes variables: densidad de cloacas, densidad de pozos sépticos, superficie impermeabilizada y densidad de drenaje; y la matriz de factores espaciales fue distinta para cada variable respuesta con el objetivo de reducir la multicolinealidad detectada en las respectivas particiones de la varianza. La elección de los factores espaciales en cada análisis se realizó mediante un procedimiento de selección directa utilizando la función *forward.sel* desde el paquete *adespatial* para R y reteniendo a los AEMs significativamente asociados a cada variable respuesta^{98,113}. Para la partición de (a) la riqueza de ASVs como variables respuesta se emplearon los AEMs 1, 11, 22 y 34; para (b) la equitatividad de Pielou como variable respuesta los AEMs 2, 4, 11 y 22; para (c) la riqueza y abundancia relativa de ASVs asociados a bacterias fecales como matriz respuesta se emplearon los AEMs 1, 4, 11, 22 y 39; y para (d) la riqueza de ASVs asociados a PLSSs como variable respuesta los AEMs 1, 6, 11, 22 y 34.

La significancia global de cada modelo empleado y las variables explicativas se evaluó a través distintos análisis pRDA, los cuales se realizaron bajo distintas condiciones de restricción de acuerdo con los resultados obtenidos en la partición de la varianza. En términos generales, las distintas condiciones de restricción se emplearon para evaluar las significancias sobre a) los efectos puros de cada matriz explicativa por separado, o b) los efectos conjuntos entre matrices, esto es la suma de los efectos puros de una matriz y en combinación con otra de las matrices explicativas. En el caso de evaluar las significancias sobre los efectos puros de una matriz (a) el análisis pRDA se realizó condicionando a la matriz de interés por el resto de los conjuntos de predictores. Este tipo de pRDA se realizó para analizar la significancia de las matrices sobre la equitatividad de Pielou de los ASVs

de la cuenca (Figura 2.9), la riqueza y abundancia relativa de ASVs fecales (Figura 2.13) y adicionalmente para evaluar la significancia de los predictores espaciales en la riqueza de ASVs de la cuenca (Figura 2.9) y en la riqueza de ASVs PLSs (Figura 2.15). Por otro lado, para evaluar la significancia en los efectos compartidos de una matriz junto a otra (b), el análisis pRDA se realizó condicionando a la matriz explicativa que contiene a la varianza pura sólo por la matriz que no participa en la compartición. Este tipo de pRDA se realizó para analizar la significancia de las matrices del hábitat local y de infraestructuras en explicar la riqueza de ASVs de la cuenca (Figura 2.9) y de ASVs PLSs (Figura 2.15), en ambos casos, con contribuciones compartidas con la matriz de predictores espaciales. Adicionalmente, el tipo de pRDA empleado en cada caso fue reflejado gráficamente con el área coloreada en los respectivos diagramas de Venn. Los resultados de los pRDA de efectos puros de la matriz espacial en cada uno de los distintos análisis mencionados se enlistan en la Tabla A2.2 en el anexo de este capítulo. En consideración de nuestro diseño de muestreo de tipo anidado y la estructura multinivel de las variables explicativas, en todos los casos se probó la significancia estadística de cada pRDA utilizando una prueba de permutación de Monte Carlo restringida de los residuos del modelo completo (9999 permutaciones)¹¹⁴. Adicionalmente, se evaluó tanto la normalidad como la falta de estructura en los residuos. Los procedimientos estadísticos se condujeron empleando los paquetes *vegan* (versión 2.5-6), *qgraph* (versión 1.6.5) y *permute* (versión 0.9-5) en R.

2.2.5.4 Redes de co-ocurrencia

Con el propósito de evaluar modulaciones en las interacciones específicas entre taxones bacterianos según el grado de disturbio detectado en las locaciones se construyeron redes de co-ocurrencia basadas en la metodología de Layeghifard (2018)¹¹⁵. Brevemente, tomando como base el ordenamiento de los sitios de muestreo que emerge del PCA en el que se incluyen las variables ambientales, biológicas e infraestructurales (Figura 1.10, Capítulo 1 de esta tesis) la matriz general de ASVs se subdividió en 3 matrices agrupando a los sitios según su grado de disturbio: disturbios bajos (LP1, LP2 y LP3), medios (SF1, SF2, SF3 y SF6) y elevados (LP4, LP5, LP5, SF4, SF5, SD1 y SD2). Con cada una de las matrices se implementó el algoritmo *SparCC* basado en los coeficientes de correlación de Pearson¹¹⁶. Para eliminar las correlaciones inciertas y débiles, el valor de corte empleado fue 0,3. Además, se detectaron grupos de nodos estrechamente relacionados o vecindarios empleando el algoritmo *walktrap*¹¹⁷. Se calcularon el número total de nodos

y de vínculos por red, así como la densidad de vínculos promedio por nodo y el índice de modularidad y el coeficiente de clusterización de cada red. El coeficiente de clusterización o agrupamiento, también conocido como transitividad, mide la probabilidad de que los nodos adyacentes de un determinado nodo estén conectados. En tanto, las redes con coeficientes de clusterización altos son más cohesivas que las redes con coeficientes más bajos. Por su parte, la modularidad es una medida de la fuerza en la división de una red en grupos o módulos. La alta modularidad indica que la red tiene conexiones densas dentro de ciertos grupos de nodos y escasas conexiones entre estos grupos. Todas las métricas de la red se obtuvieron utilizando el paquete *igraph* R. Los resultados se condensan en la Tabla 2.2.

2.2.5.5 Correlaciones múltiples

Se analizaron mediante correlaciones de Pearson las asociaciones entre los parámetros ecológicos de la comunidad bacteriana y el índice de heterogeneidad ambiental, y entre las riquezas de ASVs fecales y potencialmente patógenos (PLSs), sus abundancias relativas, la abundancia relativa del ASV asignado taxonómicamente a *E. coli*-*S. boydii*,-*S. dysenteriae*,-*S. flexneri* -*S. sonnei* (ASV 52), el logaritmo de la abundancia de *E. coli*, y parámetros de infraestructura urbana y del hábitat local (fisicoquímica e hidráulicas) se analizaron mediante correlaciones de Pearson. La significancia de los coeficientes de correlación (r de Pearson) se evaluó con la prueba de correlación de muestras pareadas ajustada para comparaciones múltiples (método de Holm; 999 permutaciones) empleando los paquete *corrplot* (versión 0.84) y *permute* (versión 0.9-5) para R⁸⁹.

2.3. Resultados

2.3.1 Estructura taxonómica global de la comunidad bacteriana en la cuenca

La composición relativa global del bacteriplancton se encontró dominada por los filos *Proteobacteria* (47,2%), *Bacteroidota* (20,1%), *Campilobacterota* (10,8%), *Firmicutes* (9,8%), *Patescibacteria* (7,1%) y *Actinobacteriota* (3,2%) (Figura 2.3). Dentro del filo *Proteobacteria*, los ASVs del género C39 contribuyeron con una abundancia relativa global del 31%, siendo el género más abundante. En particular, sólo el ASV 1 perteneciente a C39 aporta el 16% de la abundancia global. El segundo género en relevancia dentro de *Proteobacteria* es *Dechloromonas* con una abundancia relativa mucho menor: 3% respecto del total de lecturas. La abundancia relativa global contribuida por el grupo de las bacterias *Bacteroidota* se compone principalmente del género *Cloacibacterium*, con un 16% de las lecturas y el género *Prevotella* con un 2% del total. Por su parte, los géneros *Arcobacter* (6%) y *Pseudoarcobacter* (2%) son los más representados dentro de las *Campilobacterota*, mientras que el género *Anaerosinus* (2%) lo es dentro del filo *Firmicutes*.

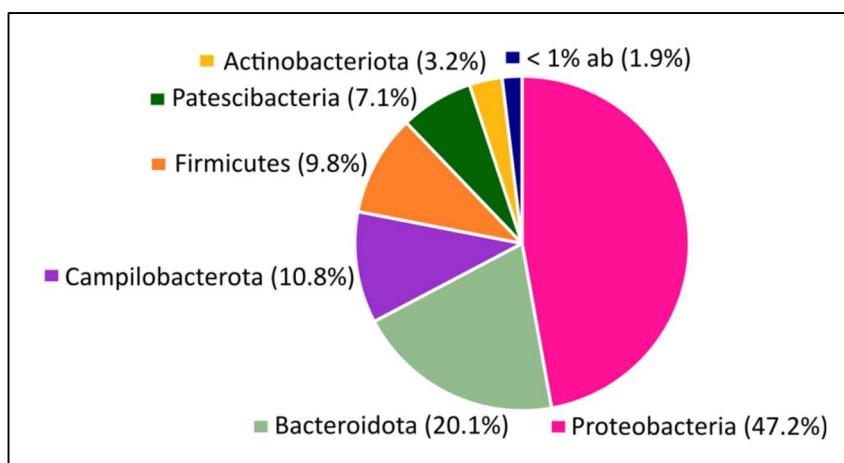


Figura 2.3: Composición relativa global de la comunidad de bacterias a nivel de *phylum*

El análisis de la composición relativa de las muestras evidencia que los filos *Proteobacteria* y *Bacteroidota* son los dominantes en la mayoría de ellas (Figura 2.4). Las bacterias de los filos *Firmicutes* y *Campilobacterota* conforman una parte importante de la comunidad bacteriana a partir de sitios en la cuenca media y alcanzan sus máximos relativos en sitios de la cuenca baja (sitios LP6 y SD1, respectivamente). Los taxones del filo *Actinobacterota* se detectan casi exclusivamente en los sitios peri-urbanos de la cuenca alta del arroyo Las Piedras, con una abundancia relativa próxima al 25% en LP2. El filo *Patescibacteria* fue detectado en la totalidad de la cuenca, revelando máximos de

abundancia en el sitio LP3, tras la inserción del arroyo Las Piedras en zona urbana, y en el sitio SF1, luego de que el arroyo San Francisco emerja del entubamiento.

Con el propósito de identificar los taxones bacterianos nucleares a la red hidrológica se llevó a cabo una curva de rango y abundancia, a partir de la cual se consideraron taxones *core* aquellos con una ocurrencia y relevancia entre muestras a partir de cierto umbral determinado que enfatiza su representatividad (ver metodología). De la matriz inicial de ASVs, ya curada y estandarizada (5867 ASVs), se obtuvieron 364 ASVs *core* cumpliendo con los umbrales propuestos (Figura 2.5), esto implica a un 6% de las variantes de secuenciación. Es de destacar que sólo esta pequeña fracción de ASVs concentra el 78% de la abundancia relativa respecto de la totalidad, lo que indica que la red hidrológica se halla mayormente dominada por pocos grupos bacterianos.

De analizar los taxones seleccionados, se observa que los primeros 10 ASVs *core* coinciden con los géneros bacterianos principales previamente mencionados (Tabla 2.1). Estos ASVs, que abarcan el 45% de la abundancia relativa global, fueron asignados taxonómicamente a los géneros *C39* (f. *Proteobacteria*), *Cloacibacterium* (f. *Bacteroidota*), *Dechloromonas* (f. *Proteobacteria*), *Arcobacter* (f. *Campylobacterota*), *Anaerobius* (f. *Firmicutes*) y *Aurantimicrobium* (f. *Actinobacteriota*). Además, cuatro de estos ASVs *core* principales fueron asignados al género *C39*, el cual condensa el 31% de las lecturas totales, lo que confirma su dominancia en la red hidrológica. Los géneros *Cloacibacterium* y *Arcobacter* le siguen en representatividad, con el 13% y 6% de las lecturas, respectivamente.

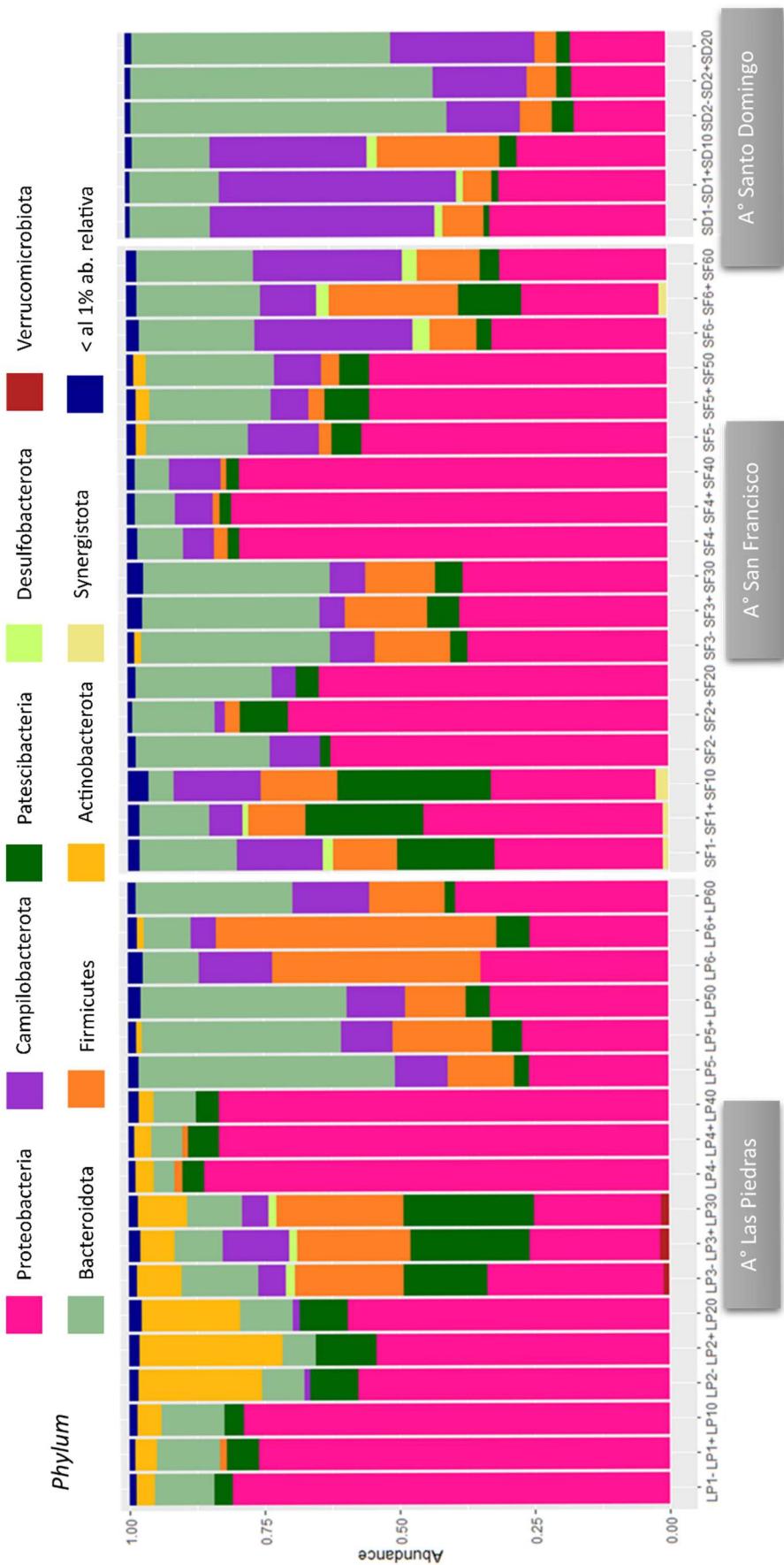


Figura 2.4: Composición s a nivel de *phylum* por muestra (42 en total), agrupadas por arroyo.

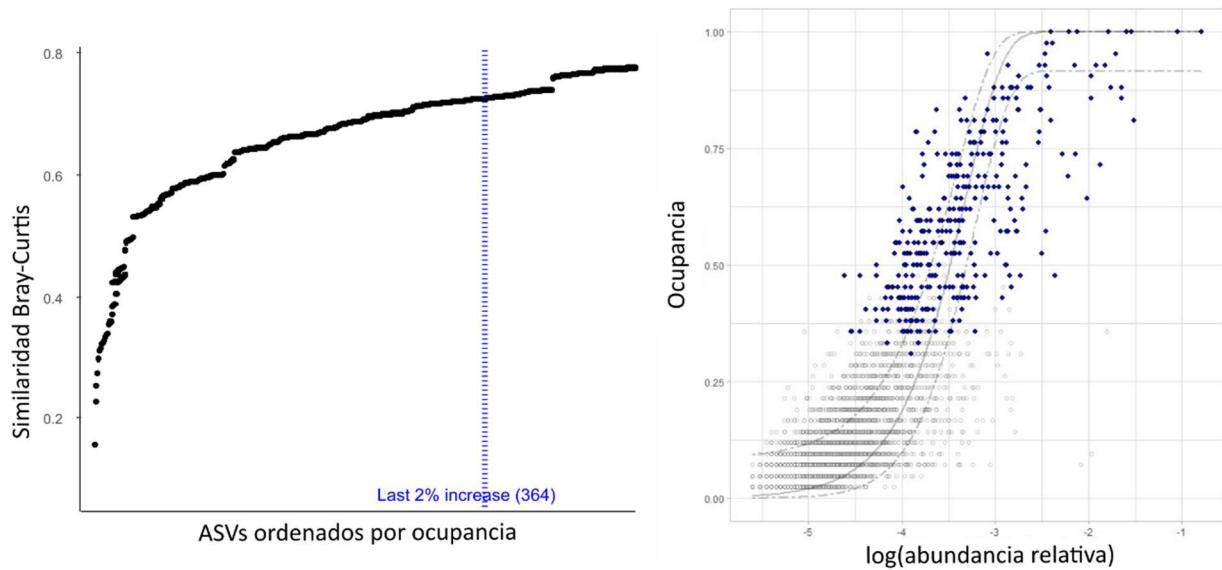


Figura 2.5: Izq: Ordenamiento de ASVs según su contribución a la disimilitud de Bray-Curtis y punto de corte. Der: Curva de ocupancia-abundancia+Modelo neutral con 95% de confianza. Los puntos azules representan a los 364 ASVs del *core* de taxones bacterianos.

Tabla 2.1: Listado de ASVs principales identificados en el *core* de taxones.

ASV_ID	Reino	Filo	C1ase	Orden	Familia	Género
ASV1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	<i>C39</i>
ASV2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Weeksellaceae	<i>Cloacibacterium</i>
ASV3	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	<i>C39</i>
ASV4	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	<i>Dechloromonas</i>
ASV5	Bacteria	Campylobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae	<i>Arcobacter</i>
ASV6	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	<i>C39</i>
ASV7	Bacteria	Campylobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae	<i>Arcobacter</i>
ASV8	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Sporomusaceae	<i>Anaerosinus</i>
ASV9	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	<i>C39</i>
ASV10	Bacteria	Actinobacteriota	Actinobacteria	Micrococcales	Microbacteriaceae	<i>Aurantimicrobium</i>

2.3.2 Análisis de la distribución espacial de la comunidad bacteriana

La riqueza de ASVs bacterianos observada por sitio reveló una gran variabilidad a lo largo de la red hidrológica (Figura 2.6, Tabla A2.1), la cual *a priori* no se observa regida por patrones evidentes que den cuenta de procesos hidrológicos y/o geográficos a escala global, aunque los sitios con una riqueza taxonómica más elevada pertenecen a tramos de la cuenca alta de los arroyos San Francisco y Las Piedras, SF1 (1356 ASVs) y LP3 (1113 ASVs), respectivamente. Estos sitios fueron destacados previamente al mencionar los picos de abundancia identificados para el filo *Patescibacteria*. Por su parte, la equitatividad de Pielou en estos puntos destacados mostró sus máximos en comparación al resto de la cuenca (Figura 2.6, Tabla A2.1), lo que indica que, si bien hay mayor riqueza de variantes microbianas, no se establecen dominancias, sino por contrario, la representatividad en la comunidad del bacteriplancton se halla repartida entre los taxones presentes. Como contraparte, la riqueza de ASVs mostró sus mínimos en los sitios LP1 y LP4 del arroyo Las Piedras pertenecientes a la cuenca alta y media. En LP1, cabecera del arroyo, se detectaron sólo 468 ASVs distintos, mientras que en LP4 se detectaron 432 ASVs. Estas dos locaciones están entre las que menor equitatividad de Pielou presentan, lo que sugiere que hay menos ASVs y que se establecen fuertes dominancias en la comunidad de bacterias planctónicas.

En congruencia con la descripción anterior, las correlaciones múltiples entre los indicadores de diversidad ecológica promedio de la comunidad del bacteriplancton indican una fuerte correlación positiva y significativa entre la riqueza y la equitatividad de Pielou de los ASVs (r de Pearson=0,828; $p<0,05$) (Figura 2.7). A la vez, la riqueza y la dominancia de Simpson correlacionan negativamente de manera significativa (r de Pearson=-0,596, $P<0,05$). Esto indica que los sitios con mayor riqueza de taxones bacterianos presentan una estructura de comunidad más uniformemente.

Por otro lado, el índice de heterogeneidad ambiental calculado para cada sitio, según su diversidad en las condiciones ecológicas internas, muestra una correlación positiva (aunque no significativa) entre la riqueza de ASVs y la heterogeneidad de los sitios (r de Pearson=0,343). Esto sugiere que conforme aumenta la disponibilidad de nichos dentro de los sitios, la riqueza de bacterias respondería también tendiendo a un aumento, pero que el grado heterogeneidad captada no sería un factor clave en estructurar la riqueza en la cuenca.

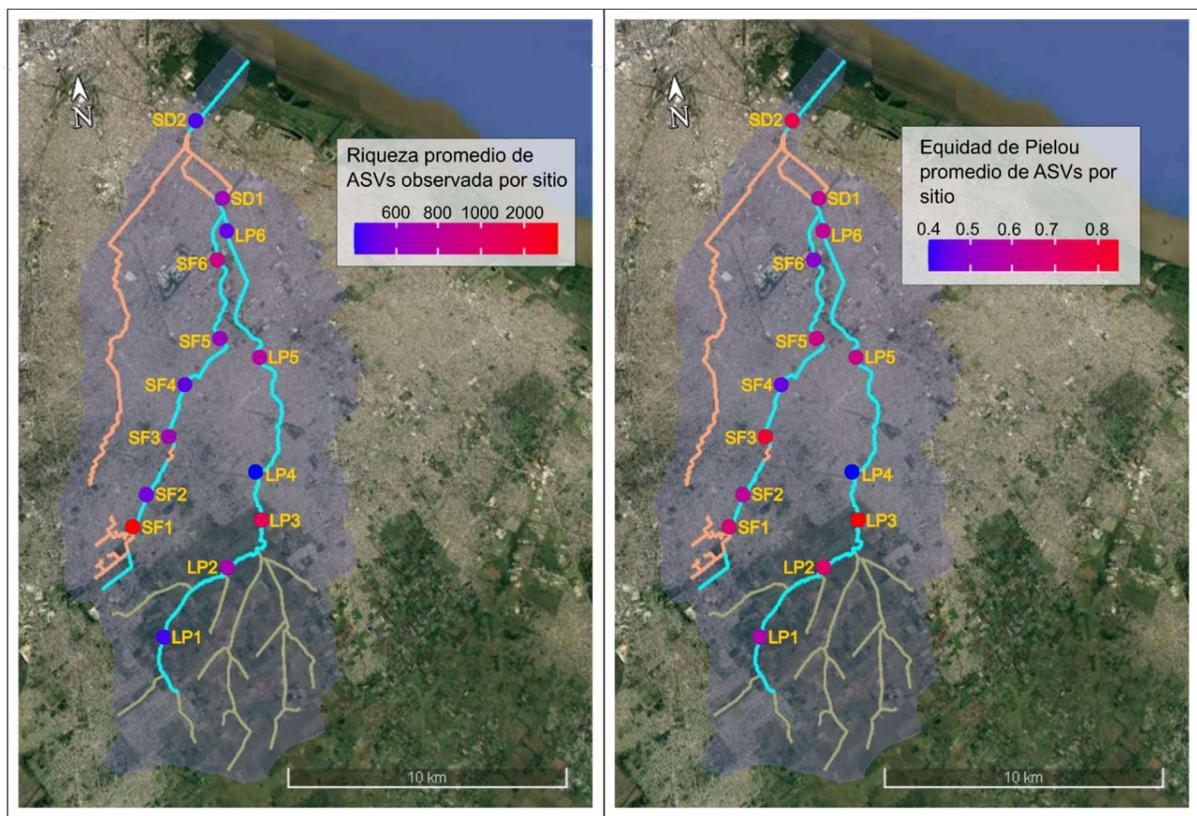


Figura 2.6: Distribución de la riqueza (izquierda) y la equitatividad de Pielou (derecha) promedio de ASVs en los sitios de muestreo.

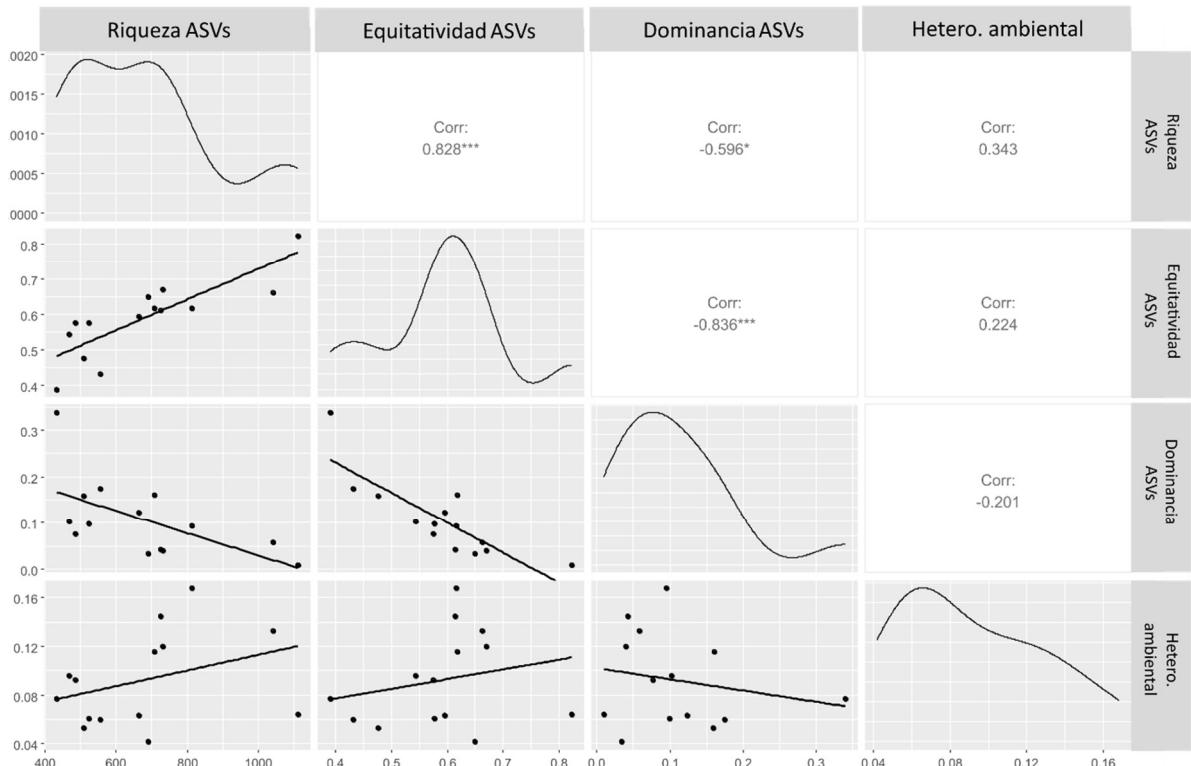


Figura 2.7: Correlaciones entre parámetros ecológicos de la comunidad bacteriana y el índice de heterogeneidad ambiental. Significancias verificadas con permutaciones (nperm=999): ***p<0.001, **p<0.01, *p<0.05.

Por otro lado, se realizó una partición aditiva de la diversidad (riqueza de ASVs) (Figura 2.8). Se observó que el mayor componente de la diversidad regional está dado por el recambio de ASVs entre sitios (β_2), con el 79,5% de contribución a la riqueza de la cuenca. En segundo y tercer lugar se encuentran las contribuciones locales intra-sitio (α), con un 12%, seguida por la contribución del recambio de ASVs entre muestras intra-sitio (β_1), del 8,3%. Comparando esta estructura de la diversidad con la asociada al modelo nulo, tanto la diversidad α (de cada muestra) como la β_2 (inter-sitio) resultaron significativamente mayores a la esperadas, mientras que la diversidad β_1 (intra-sitio) fue menor a la esperada. La mayor variabilidad en la composición de especies entre los sitios de muestreo, dado que la diversidad β_2 fue mayor a la esperada, podría deberse a la presencia de barreras físicas o a diferencias en los ambientes locales que condicionen la distribución de las variantes bacterianas. Por otro lado, la baja diversidad β_1 podría deberse a la existencia de factores locales que limitan la diversidad de especies dentro de cada sitio de muestreo, como la competencia intra-específica, o la homogeneización en las condiciones ambientales intra-sitio, tal como se dedujo en el capítulo previo (Figura 1.12).

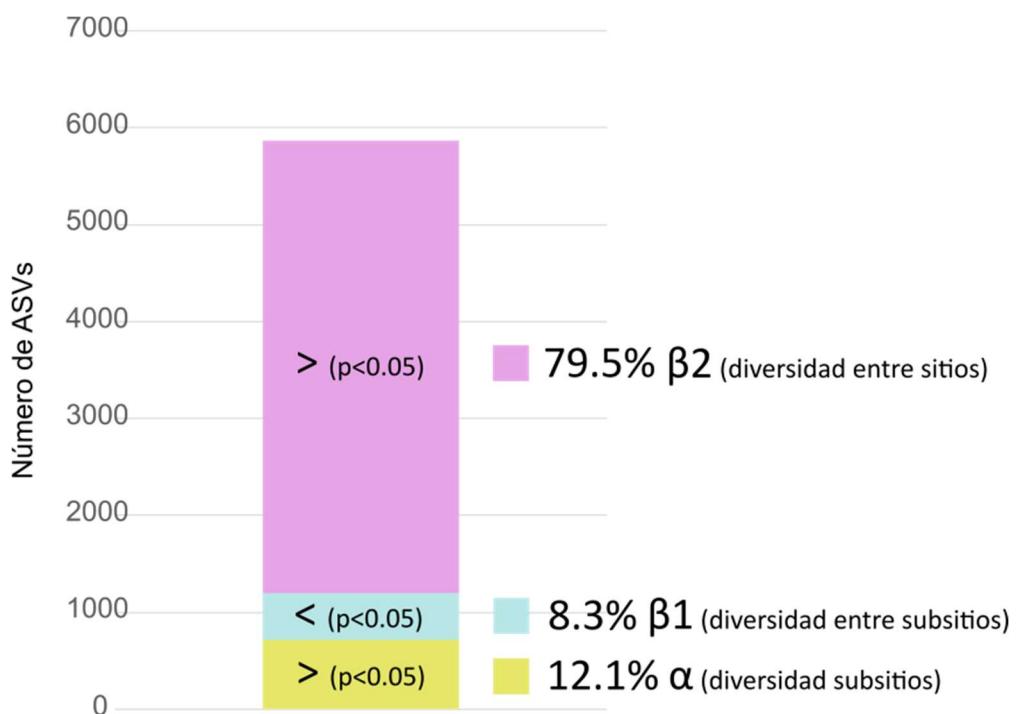


Figura 2.8: Partición aditiva de la diversidad de ASVs en la cuenca.

Entendiendo que la variabilidad específica observada en la comunidad en términos de riqueza se vería mayormente condicionada por la diversidad de condiciones ambientales e infraestructurales que afectan conjuntamente a cada tramo de la red hidrológica, se buscó particionar la contribución parcial de los parámetros del hábitat local e infraestructurales sobre la varianza de la riqueza y la equitatividad de Pielou del bacteriplancton (Figura 2.9). Adicionalmente, se incluyó en análisis a la matriz de factores espaciales para absorber la varianza asociada estrictamente a procesos hidrológicos de autocorrelación espacial. Se observó que las características del ambiente local, en comparación con los condicionantes de infraestructura, son los contribuyentes principales en modular tanto la riqueza como la equitatividad de los ASVs, contribuyendo con un aporte global del 42% sobre la riqueza y del 36% sobre la equitatividad de Pielou. La contribución de las condiciones del hábitat sobre la riqueza fue significativa en la fracción conjunta que agrupa lo contribuido en forma pura y lo contribuido junto a los factores espaciales ($F_{9,28}=3.35$, $P<0,05$; área coloreada en violeta del diagrama correspondiente en la Figura 2.8). De esta fracción se destaca una asociación positiva y significativa del DIN ($F_{1,28}=7.19$, $P<0,05$) y negativa de la cobertura de macrófitas ($F_{1,28}=5.72$, $P<0,05$) con el eje RDA1 significativo ($F_{1,36}=38.73$, $P<0,05$), sugiriendo que la riqueza bacteriana es mayor en sitios disturbados que reúnen concentraciones elevadas de contaminantes como el DIN y menor presencia de plantas acuáticas. Por su parte, las condiciones del hábitat local reflejaron un aporte puro y significativo del 18% sobre la equitatividad de Pielou ($F_{9,24}=14.45$, $P<0,05$). Dentro de esta fracción, la velocidad media se asoció en forma positiva y significativamente ($F_{1,24}=24.83$, $P<0,05$) con el eje significativo RDA 1 ($F_{1,32}=173.4$, $P<0,05$), mientras que el OD se asoció en forma negativa y significativa con este eje ($F_{1,24}=11.14$, $P<0,05$).

La intensidad en la urbanización, valorada a partir del set de parámetros de infraestructura, contribuyó con un aporte global del 6% sobre la partición de la riqueza de ASVs, el cual no presentó significancia. Por su parte, la contribución de este set sobre la equitatividad de Pielou fue del 19%, la cual se descompone en una importante contribución conjunta junto a las condiciones del hábitat local y un aporte puro y significativo del 7% ($F_{4,24}=11.02$, $P<0,05$) (Figura 2.9, diagrama de la derecha). El eje significativo RDA1 ($F_{1,27}=49.6$, $P<0,05$) para la contribución pura se relacionó negativa y significativamente con todas las variables de infraestructura: densidades de pozos sépticos ($F_{1,24}=17.11$, $P<0,05$), de cloacas ($F_{1,24}=15.31$, $P<0,05$), de drenaje ($F_{1,24}=5.34$,

$P < 0,05$) y con la superficie impermeabilizada ($F_{1,24} = 8.51$, $P < 0,05$). La interpretación de los resultados de esta fracción indica que la mayor presencia de infraestructuras favorecería procesos de dominancia en los ensambles bacterianos.

Se destaca también el importante aporte de los factores espaciales direccionales en ambas particiones de la varianza, con R^2 ajustados que superan el 50% de las respectivas contribuciones globales (Figura 2.9). Sólo en el caso de la riqueza de ASVs como variable respuesta se halló explicación compartida de esta matriz junto a los parámetros del hábitat local (22%), lo que hace referencia a un aporte espacialmente estructurado por parte de las condiciones ambientales. Un detalle completo de estos resultados se adjunta en la Tabla A2.2 en el anexo de este capítulo.

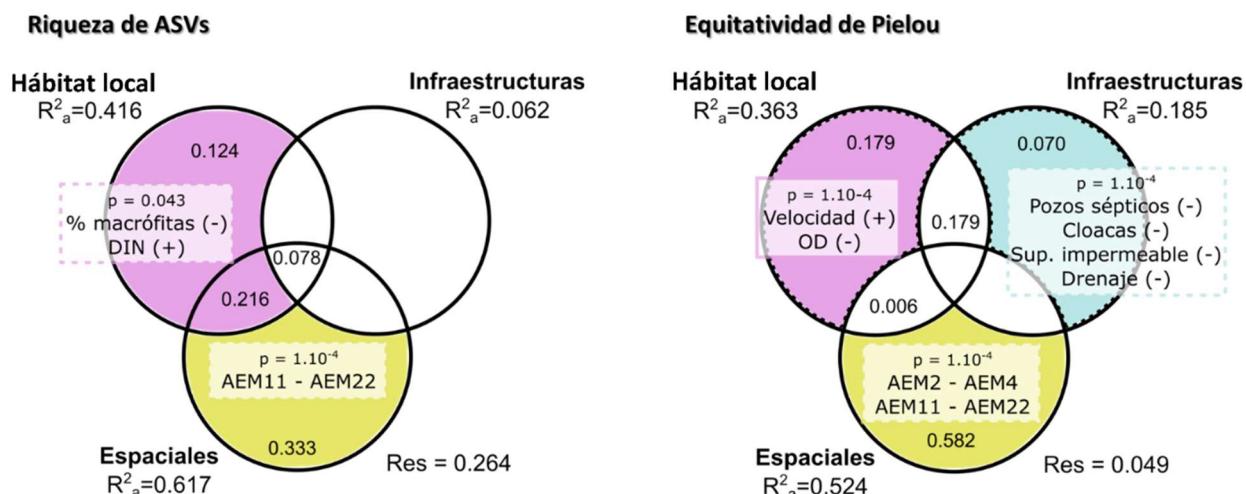


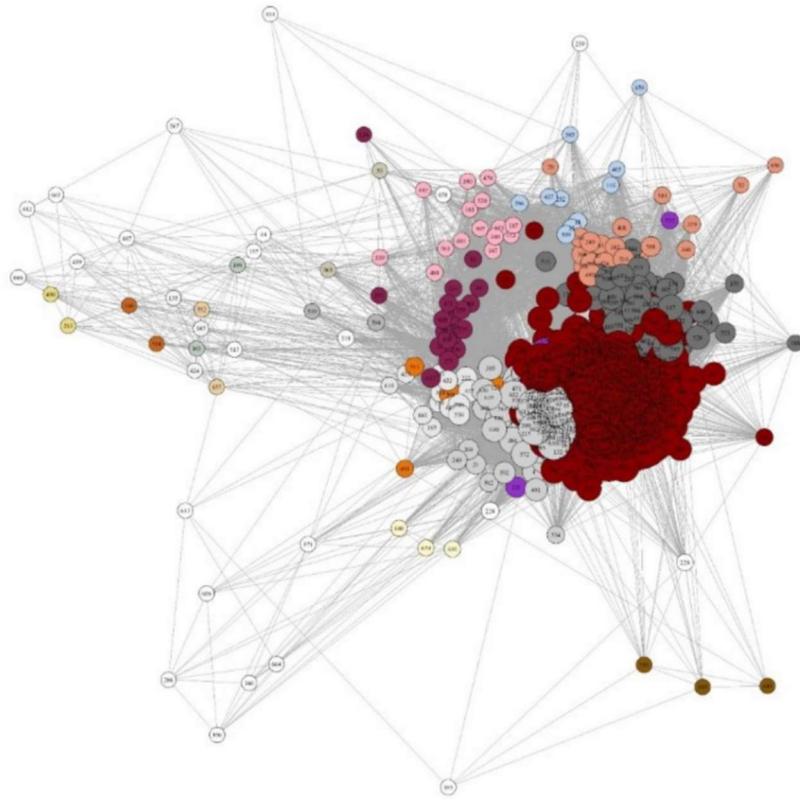
Figura 2.9: Particiones de la varianza de la riqueza de ASVs y de la equitatividad de Pielou entre los conjuntos de predictores del microhábitat, infraestructura urbana y factores espaciales (AEMS). Las contribuciones calculadas de cada pRDA se reflejan en los valores de R^2 ajustado incluidos, mientras que las condiciones de restricción de cada análisis son reflejadas a partir de la coloración alternada de las distintas áreas puras y compartidas de las respectivas gráficas de Venn.

2.3.3 Redes de co-ocurrencia de la comunidad bacteriana

Las redes de co-ocurrencia de especies son aproximaciones gráficas que ponen de manifiesto interacciones inter-específicas basadas en un criterio de coexistencia y de correlación en sus abundancias. Con el objetivo de profundizar en dichas interacciones entre taxones de la comunidad del bacteriplancton de una cuenca urbana en función del grado de disturbio ambiental, se subdividió a la matriz general de ASVs en 3 matrices agrupando a los sitios según el criterio descripto en la metodología: disturbios bajos (LP1, LP2 y LP3), medios (SF1, SF2, SF3 y SF6) y elevados (LP4, LP5, LP5, SF4, SF5, SD1 y SD2), y se estudiaron sus respectivas redes (Figura 2.10).

Bajo condiciones ecológicas de menor impacto (Figura 2.10.a; sitios con mayor cobertura de macrófitas y menor presión de infraestructuras urbanas) la red de co-ocurrencia se ordenó en un agrupamiento principal que abarca a 410 nodos (ASVs), en otros siete agrupamientos periféricos de menor cantidad (el segundo agrupamiento en tamaño tiene sólo 75 ASVs), y finalmente en una serie de nodos inconexos. En términos generales, la red se aprecia cohesiva en torno al agrupamiento central. En este sentido, la densidad de vínculos, la cual relativiza la totalidad de las interacciones subyacentes a la cantidad de nodos, fue de 0,57 (Tabla 2.2). Por su parte, en condiciones de disturbios medios (Figura 2.10.b; rangos intermedios de parámetros de infraestructura), la red evidenció una mayor fragmentación en cuatro grandes agrupamientos de 290, 256, 210 y 145 ASVs. La densidad de vínculos establecidos disminuyó a 0,20 (Tabla 2.2). Por último, la red de coocurrencia basada en la comunidad de los sitios con elevado grado de disturbio (Figura 2.10.c; presión elevada de infraestructuras, menor cobertura de macrófitas y mayor contaminación orgánica) se compuso de un agrupamiento de mayor tamaño (200 ASVs) y de ocho agrupamientos menores (cada uno con menos de 100 ASVs). La densidad de los vínculos establecidos fue la menor de las tres redes analizadas (0,06). Además, en condiciones de bajo disturbio se observó el mayor coeficiente de agrupamiento (0,86) y la menor modularidad (0,03) (Tabla 2.2), en comparación con los parámetros obtenidos para las redes de disturbios medios (coef. agrupamiento=0,59; modularidad=0,34) y elevados (coef. agrupamiento=0,50; modularidad=0,32). Esto sugiere que en los sitios con condiciones de menor presión urbana y mayor integridad ambiental local relativa, la comunidad se compone en forma más cohesiva, propiedad que disminuye conforme se intensifican los disturbios y aumenta la modularidad y la fragmentación en la interacciones globales, con ordenamientos intra-grupos más densos y conexiones inter-grupos escasas.

a) Disturbios bajos (LP1, LP2 y LP3)



b) Disturbios medios (SF1, SF2, SF3 y SF6)

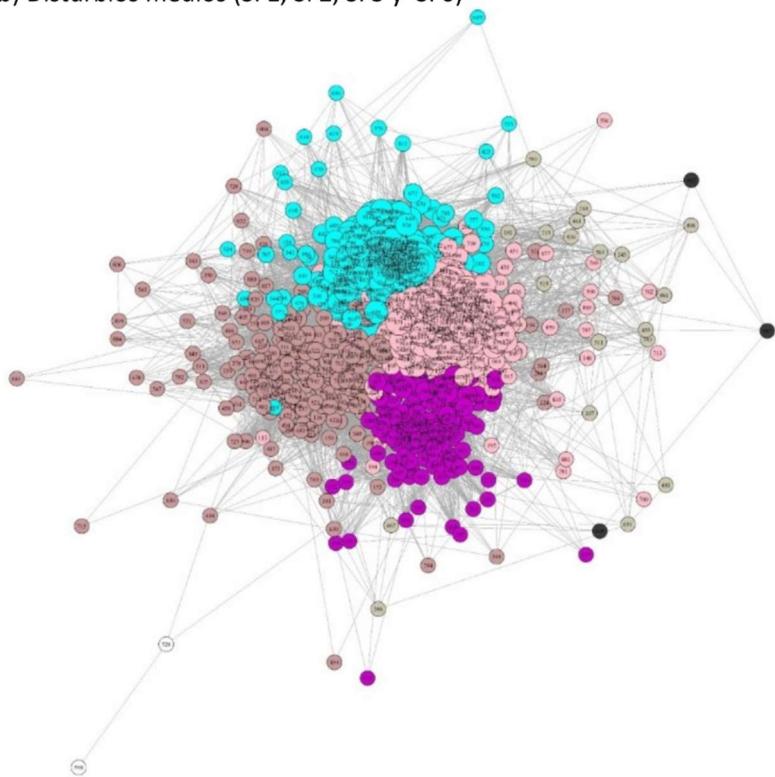


Figura 2.10: Redes de coocurrencia entre ASVs empleando agrupamientos de sitios según grado de disturbio – bajos (a), medios (b) y elevados (c; siguiente página).

c) Disturbios elevados (SF4, SF5, LP4, LP5, LP6, SD1 y SD2)

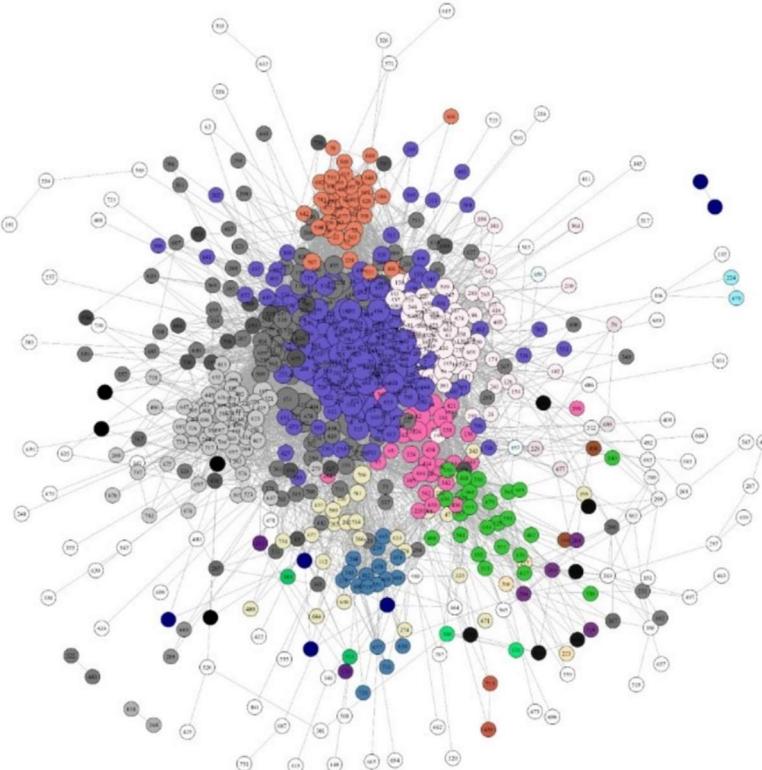


Figura 2.10: Redes de coocurrencia entre ASVs empleando agrupamientos de sitios según su grado de disturbio – bajos (a), medios (b) y elevados (c).

Tabla 2.2: Métricas asociadas a las redes de co-ocurrencia entre ASVs según el grado de disturbio

Parámetros de redes	Redes según grado de disturbios		
	Bajos	Medios	Elevados
Nodos	667	925	746
Vínculos	127199	83506	16679
Modularidad	0,04	0,34	0,32
Coeficiente de clusterización	0,86	0,59	0,51
Densidad de vínculos	0,57	0,20	0,60

2.3.4 Identificación de bacterias de origen fecal en la cuenca

Un aspecto interesante y poco estudiado es la ubicuidad de bacterias de origen fecal en una cuenca urbana. En ese sentido, para indagar sobre la distribución de taxones bacterianos asociados a la contaminación de origen fecal, se aprovechó la disponibilidad de un set de datos de 16S ARNr de 30 microbiomas fecales pertenecientes a personas sanas de la Ciudad Autónomas de Buenos Aires⁹¹. De dicha base se retuvo un *subset* de 316 taxones *core* presentes en las muestras de origen fecal en base a un análisis de curvas de rango y abundancia.

Una vez identificado el *core* de taxones de microbiomas fecales de la región, se rastrearon sus secuencias en el *dataset* de muestras de la red hidrológica con un criterio de homología en las secuencias del 100%, y se construyó un *subset* de la cuenca basado en esta intersección. Como resultado de esta intersección, 139 de los 316 taxones fecales fueron detectados en muestras de agua del presente estudio, representando un promedio de lecturas del 2,30% en el *dataset* original. La muestra con menor cantidad de lecturas asociadas a taxones fecales pertenece al sitio LP2 de la cuenca alta del arroyo Las Piedras ($2,32 \cdot 10^{-4}\%$) (Figura 2.11). Llamativamente, el sitio contiguo aguas abajo sobre el mismo arroyo (LP3) fue el sitio en el que se detectó la muestra con mayor representatividad de estas lecturas ($9,94 \cdot 10^{-2}\%$). Tal como se analizó en el capítulo 1 de esta Tesis, este sitio se corresponde con el inicio de una zona de elevada densidad demográfica dispuesta en urbanizaciones no planificadas y, por lo tanto, con infraestructuras sanitarias insuficientes.

Los filos de origen fecal más abundantes en el bacteriplancton de la cuenca (Figura 2.11) fueron *Bacteroidota*, con un rango de abundancia relativa en las muestras entre 21% (SF2) y 73% (SF4), *Proteobacteria*, no detectado en LP2 y con una abundancia del 70% en SF2; *Verrucomicrobiota*, tampoco detectado en LP2 y con su máximo de abundancia en LP1 (22%), *Firmicutes*, con el 0,06% en SF2 y 53% en LP3; y, por último, el filo *Actinobacteriota*, no detectado en numerosos sitios como LP2, LP6 y SF2, y con una abundancia máxima detectada en LP4.

Los ASVs asociados a la contaminación fecal se agruparon en 19 órdenes, máxima categoría taxonómica en el que la totalidad de los ASVs alcanza asignación, y 51 géneros. En la Tabla 2.3 se enlistan los principales órdenes bacterianos fecales detectados, su distribución en la cuenca y sus características ambientales principales basadas en bibliografía preexistente. Se observa que los órdenes abundantes *Bacteroidales* (47% de abundancia relativa promedio), *Oscillospirales* (16%), *Lachnospirales* (9%) y *Verrucomicrobiales* (6%) tienen una distribución ubicua en los sitios muestreados, mientras que los órdenes *Enterobacteriales* (10%) y *Lactobacillales* (4%) se identifican sólo en tramos particulares de la red.

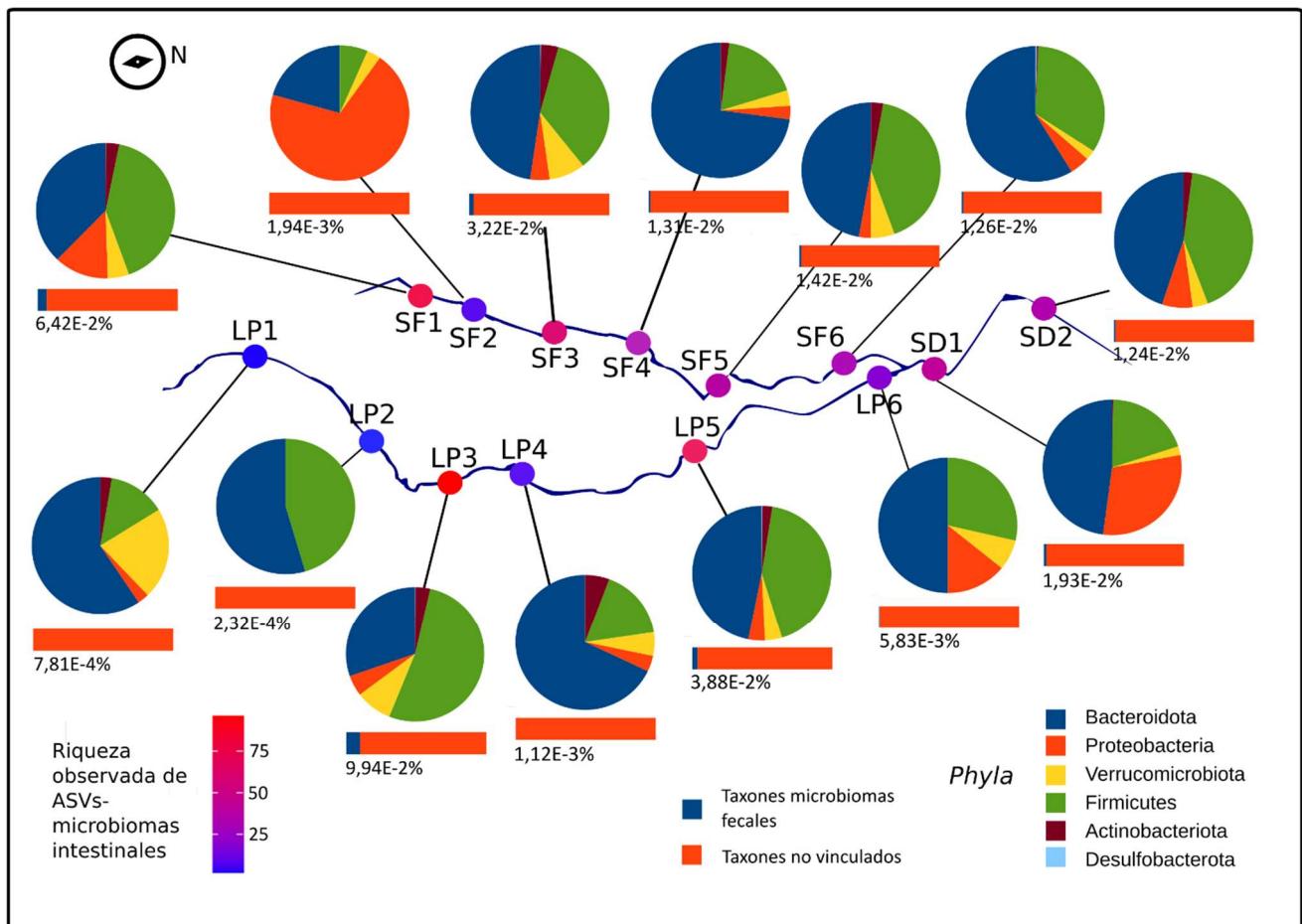


Figura 2.11: Resumen de resultados de la detección de bacterias fecales en la cuenca. Sobre cada sitio se observa la riqueza de ASVs fecales detectada, en un gradiente de color del azul al rojo. Adicionalmente, se esquematiza la composición relativa de los filos de origen fecal de cada sitio (gráficas circulares) y el porcentaje de lecturas del total de cada sitio que intersectaron con el *core* de microbiomas fecales del AMBA (barras debajo de gráficas circulares).

La riqueza de ASVs fecales tuvo un mínimo promedio de 2 ASVs en las muestras de LP2 y un máximo promedio de 96 ASVs entre las muestras de LP3 (Figura 2.11). La distribución espacial de la riqueza presenta valores medios en los sitios aguas abajo del arroyo San Francisco, al igual que en el resto de la cuenca baja. Contrariamente, la cuenca alta del arroyo San Francisco y la mayoría de los sitios de la cuenca alta y media del Arroyo Las Piedras muestran valores dispares de riqueza entre sitios conectados por el cauce, lo que es indicio de controles locales modulando la diversidad específica de la contaminación de origen fecal.

Orden	Distribución en sitios de muestreo	Características ambientales según bibliografía	Referencia
Bacteroidales (filo Bacteroidetes)	Ubicua	Microbiota intestinal humana	¹¹⁸
Lachnospirales (filo Bacillota)	Ubicua	Microbiota intestinal humana	¹¹⁹
Oscillospirales (filo Bacillota)	Ubicua	Microbiota intestinal humana (más asociado a poblaciones rurales)	^{120,121}
Verrucomicrobiales (filo Verrucomicrobiota)	Ubicua	Microbiota intestinal humana	¹²²
Enterobacteriales (filo Proteobacteria)	↑SF2 y 84 en cuenca baja	Microbiota intestinal humana. Incluye taxones patógenos y con resistencia antibiótica	^{122,123}
Lactobacillales (filo Firmicutes)	↑LP2 (sitio con escasas lecturas en general)	Microbiota intestinal humana	¹²²

Tabla 2.3: Listado de ordenes bacterianos principales asociados a la contaminación fecal que fueron detectados en los sitios de muestreo. Se incluye su distribución entre el set muestras de la cuenca y su nicho asociado según la bibliografía consultada.

Se analizó la partición de la varianza de la riqueza específica y de la abundancia relativa de ASVs de bacterias fecales detectadas a lo largo de la red hidrológica, en base a las condiciones del hábitat local, la intensidad de la urbanización (parámetros de infraestructura) y los factores espaciales específicos, para la captación de la varianza asociada a fenómenos de autocorrelación espacial (AEMs) (Figura 2.12). El set de predictores del hábitat local conformado por la cobertura de plantas macrófitas, el pH, OD, DIN, SRP, turbidez, DOC, profundidad y velocidad de flujo medias contribuyó a explicar las variables respuesta en un 16% en forma pura y significativa ($F_{9,23} = 10.43$, $P < 0,05$) y en un 29% en forma conjunta con los factores espaciales. Dentro de la contribución pura, la velocidad media ($F_{1,23} = 4.89$, $P < 0,05$) se asoció en forma positiva y significativa con el eje significativo RDA1 ($F_{1,30} = 115.31$, $P < 0,05$), mientras que la

cobertura de macrófitas y la profundidad media se asociaron en forma significativa y negativa con el mismo (respectivamente, $F_{1,23}=4.51$ y $F_{1,23}=17.74$, $P <0,05$).

Las infraestructuras urbanas (densidades de cloacas, pozos sépticos, drenaje y porcentaje de superficie impermeabilizada) realizaron un aporte significativo mucho menor a la varianza, con una contribución pura del 9% ($F_{4,23}=10.43$, $P <0,05$). Sin embargo, ninguno de los parámetros resultó significativo. Por su parte, los AEMs basados en la autocorrelación espacial positiva y negativa (AEMs 1, 4, 11, 22 y 39) abarcaron una contribución pura y significativa del 51% ($F_{5,23}=48.35$, $P <0,05$). De estos, los AEM 22 y 39, ambos de grano fino, resultaron significativos (respectivamente, $F_{1,23}=22.48$ y $F_{1,23}=3.64$, $P <0,05$) (ver detalle de AEMs en Figura 2.2), evidenciando fuertes modulaciones locales sobre las variables respuesta.

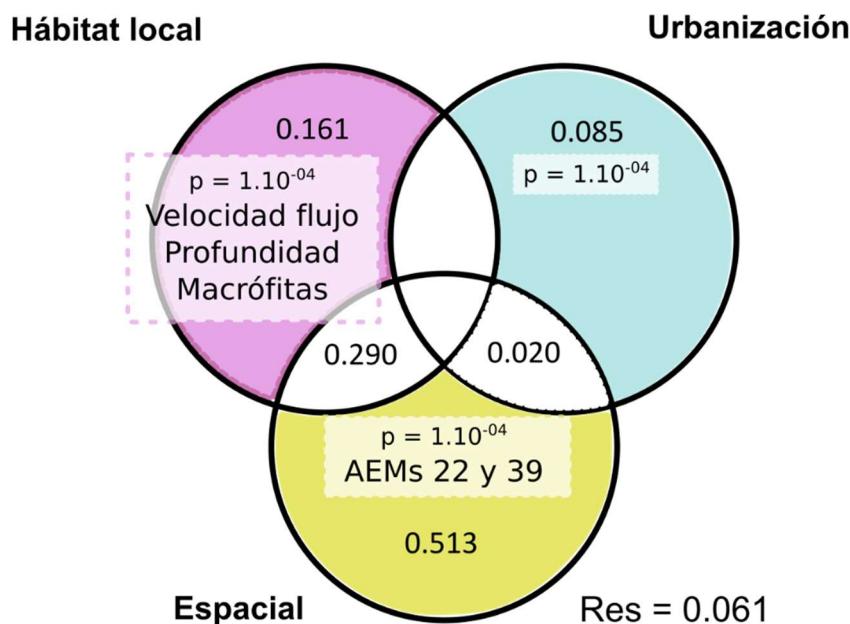


Figura 2.12: Partición de la varianza de la riqueza y abundancia relativa de ASVs fecales entre los conjuntos de predictores del microhábitat, infraestructura urbana y factores espaciales (AEMS). Las contribuciones globales calculadas de cada set se reflejan en los valores de R^2 ajustado incluidos. Mientras que las condiciones de restricción de cada análisis son reflejadas a partir de la coloración alternada de las distintas áreas puras y compartidas de las respectivas gráficas de Venn.

2.3.5 Rastreo de taxones potencialmente patógenos

La identificación de bacterias con probable origen fecal humano no refiere explícitamente a su patogenicidad, ni tampoco abarca posibles patógenos provenientes de otros orígenes. Con el objetivo de indagar específicamente en dicha cuestión, en primer lugar, se realizó una búsqueda bibliográfica en la que se identificaron 53 géneros bacterianos potencialmente patógenos para humanos, para luego analizar su ocurrencia en las muestras de la red hidrológica. Se detectaron 26 géneros, siendo los siguientes los que

tienen mayor cantidad relativa de lecturas en la totalidad del proyecto: *Arcobacter* (58% de abundancia relativa), *Erysipelothrix* (15% de abundancia relativa), *Pseudomonas* (9% de abundancia relativa), *Acinetobacter* (7% de abundancia relativa) y *Aeromonas* (6% de abundancia relativa) (Figura 2.13 y Tabla A2.3 en el anexo de este capítulo). Un total de 293 ASVs se hallaron asignados a los géneros detectados, lo que implicó un promedio del 10.04% de lecturas en las muestras.

Posteriormente se validó la identidad de secuencia de los 293 ASVs contenidos en los 26 géneros potencialmente patógenos con secuencias 16S ARNr de cepas de referencias de especies con patogenicidad conocida (BlastN > 99% de identidad entre secuencias). Entre las 174 especies patógenas contenidas dentro de los géneros potencialmente patógenos detectados, 26 se hallaron en nuestro *dataset* (Tabla A2.3 del anexo de este capítulo). La estructura de la distribución espacial de las secuencias asociadas a patógenos a nivel de especie (PLSs; del inglés *Pathogen-like sequences at species level*) mostró sus mínimos en sitios de la cuenca alta del arroyo Las Piedras (sitios LP1 y LP2) y sus máximos de detección en la cuenca baja (SD1 y SD2) (Figura 2.13). El promedio de lecturas asociadas a las especies patógenas fue del 6.68%. En los sitios LP1 y LP2 se detectaron las más bajas cantidades de lecturas PLSs (0.19% y 0.22% promedio, respectivamente) y en el sitio SD1 se observó el máximo (19.1% promedio). Las principales PLSs halladas corresponden a las especies *Arcobacter cryaerophilus*, *Arcobacter butzleri*, *Aeromonas veronii* y *E. coli*-*S. boydii/dysenteriae/flexneri/sonnei*, las cuales están clínicamente vinculadas a enfermedades de carácter gastrointestinal y diarreico, y asociadas a la contaminación fecal (Tabla A2.3 del anexo de este capítulo). Es de destacar que los ASVs con identidad de secuencia mayor al 99% con *Arcobacter cryaerophilus* conforman el 52% de las lecturas dentro del *subset* de secuencias asociadas a géneros potencialmente patógenos. ASVs con identidad de secuencia con *Arcobacter butzleri* y *Aeromonas veronii* reúnen el 3% de las secuencias dentro de este *subset*, y ASVs con identidad de secuencia con *E. coli*-*S. boydii/dysenteriae/flexneri/sonnei* conforman el 2% (los ASVs que superaron el criterio del 99% de identidad con la cepa de referencia de *E. coli* también lo hicieron con *S. boydii*, *S. dysenteriae*, *S. flexneri* y *S. sonnei* por lo que se informan como un grupo único). La riqueza de PLSs se distribuye con valores intermedios en la porción media de la red y en casi la totalidad de los sitios sobre el arroyo San Francisco, a excepción del primer sitio aguas arriba del mismo, que posee la mayor abundancia promedio de PLSs máxima reportada, con 29 taxones patógenos (Figura 2.13). Por su

parte, la naciente del arroyo Las Piedras posee los valores mínimos, con sólo 5 ASVs asociados a PLSs en el sitio LP1 y 6 ASVs en LP2. Los sitios SD1 y SD2 sobre el arroyo Santo Domingo, que poseen la mayor cantidad de lecturas por muestra asociadas a secuencias patogénicas, presentan las mayores riquezas de PLSs, con 25 y 16 ASVs respectivamente.

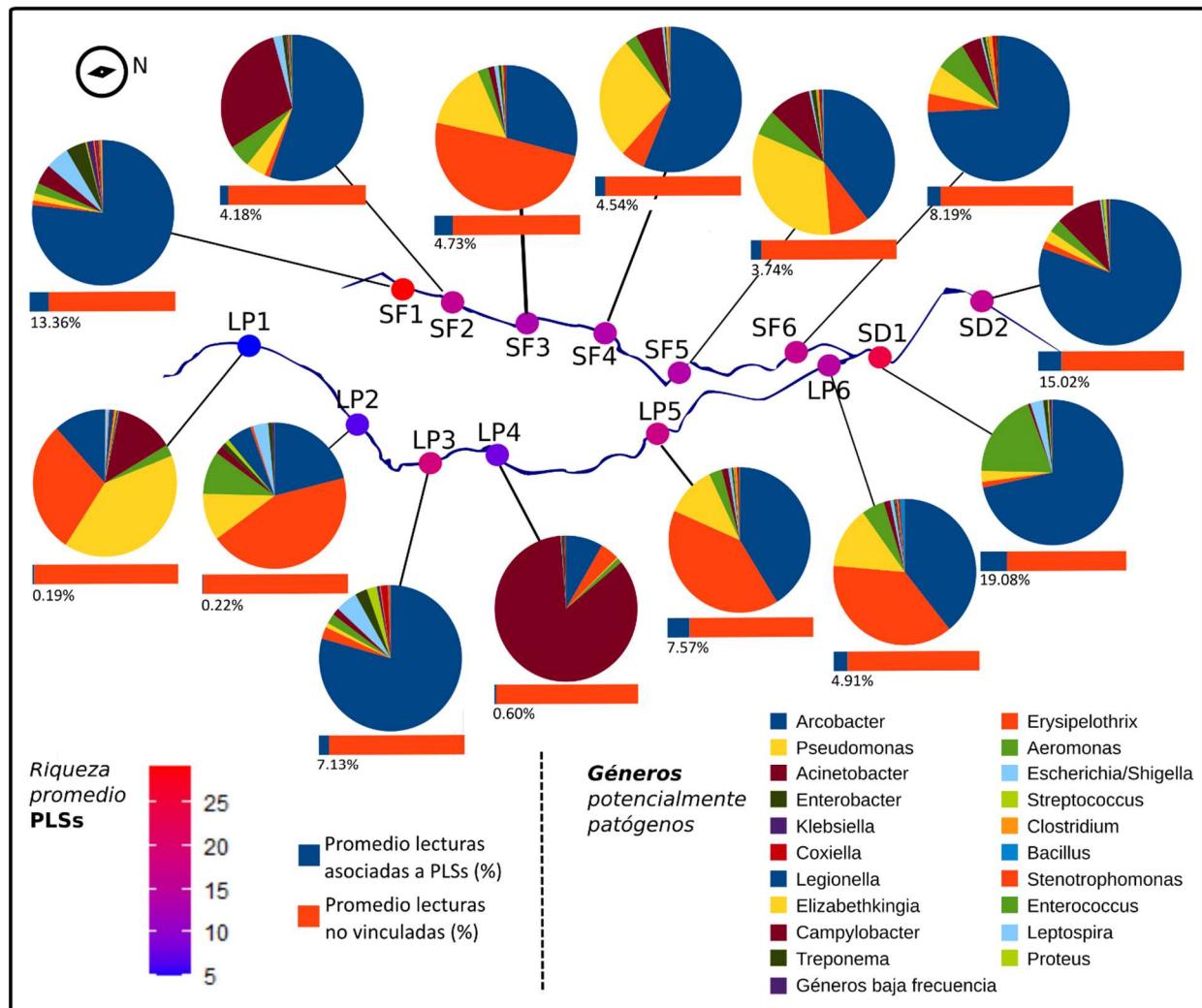


Figura 2.13: Resumen de resultados de la detección de taxones potencialmente patógenos en la cuenca. Sobre cada sitio se indica la riqueza de PLSs (secuencias asociadas a especies potencialmente patógenos) detectada, en un gradiente de color del azul al rojo, y el porcentaje de lecturas del total de cada sitio que intersectaron con las secuencias PLSs. Adicionalmente, se esquematiza la composición relativa de los géneros potencialmente patógenos de cada sitio (gráficas circulares).

Haciendo un análisis global de la distribución de taxones potencialmente patógenos en la cuenca, evaluamos la contribución de las condiciones del hábitat local, las infraestructuras urbanas y los factores espaciales sobre la distribución espacial de la riqueza de PLSs a partir de un análisis de partición de la varianza. Se identificó un aporte principal del 48% por parte de los parámetros del hábitat local, del cual un 27% es explicado junto a los parámetros espaciales (Figura 2.14). Por su parte, los parámetros de

urbanización contribuyeron a explicar un 8% de la varianza en la riqueza de PLSs en forma pura. Se detectó a la vez una covarianza entre los tres sets de predictores del 22%, lo que sugiere importantes efectos anidados y/o con algún grado de dependencia entre los mismos. Por último, los factores espaciales contribuyeron a la varianza con un 14% puros.

Del análisis de redundancia parcial respecto de la riqueza de PLSs se obtuvo que la contribución por parte de las condiciones del hábitat es significativa ($F_{9,28}=4.14$, $P<0,05$), destacándose una asociación positiva con el DIN ($F_{1,28}=6.66$, $P<0.05$), la turbidez ($F_{1,28}=3.29$, $P<0.05$), y negativa con la velocidad de flujo ($F_{1,28}=4.14$, $P<0.05$), sugiriendo que la riqueza de bacterias potencialmente patógenas es mayor en sitios con elevada contaminación orgánica, turbidez y aguas calmas. Por su parte, el aporte de las variables infraestructurales o urbanas, controlando por la contribución de las variables del hábitat local, también resultó significativo ($F_{9,28}=3.00$, $P<0,05$), destacándose una asociación positiva con la densidad de pozos sépticos ($F_{1,28}=6.80$, $P<0.05$) y de hogares conectados a cloacas ($F_{1,28}=4.81$, $P<0.05$), indicando como era esperable que a mayor intensidad de urbanización se ve favorecida una mayor ocurrencia de taxones bacterianos potencialmente patógenos.

La contribución por parte de los factores espaciales a esta variable respuesta también fue significativa ($F_{5,23}=3.8$, $P<0,05$) y un mayor detalle de estos resultados se adjunta en la tabla A2.2 del anexo de este capítulo.

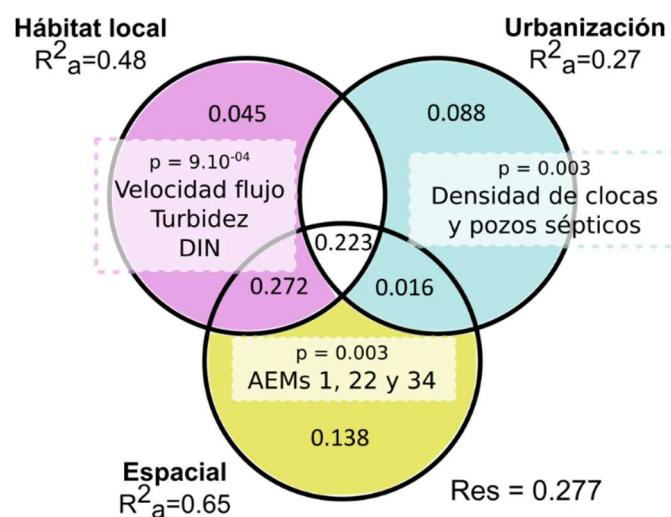


Figura 2.14: Partición de la varianza de la riqueza de PLSs entre los conjuntos de predictores del microhábitat, infraestructura urbana y factores espaciales (AEMS). Las contribuciones calculadas de cada fracción se reflejan en los valores de $R^2_{ajustado}$ incluidos. Las variables significativas incluidas son las obtenidas de distintos pRDA bajo las condiciones de restricción reflejadas a partir de la coloración alternada de las distintas áreas puras y compartidas de la gráfica de Venn.

2.3.6 Asociaciones entre las distintas comunidades bacterianas analizadas

Se analizaron en forma conjunta los *datasets* empleados a lo largo de este capítulo con el propósito de identificar ASVs comunes a los mismos. Se observó una importante intersección de variantes microbianas entre el *core* de taxones presentes en la red hidrológica y los taxones *core* de microbiomas de origen fecal humano de la región AMBA (34 ASVs comunes), así como con los taxones potencialmente patógenos (30 ASVs comunes) (Figura 2.15). La matriz de ASVs de microbiomas fecales humanos presentó 4 ASVs compartidos con la matriz de ASVs pertenecientes a géneros potencialmente patógenos. El único ASV común a las tres matrices en análisis fue asignado al género *Escherichia/Shigella* empleando la base de datos de SILVA y coincide en un porcentaje de identidad mayor al 99% con las especies *E. coli*-*S. boydii*-*S. dysenteriae*-*S. flexneri* -*S. sonnei*. El grado de resolución taxonómica de la región V3-V4 del gen 16S ARNr no permite su diferenciación entre estas especies empleando la base de datos SILVA. Adicionalmente, se destacan los géneros bacterianos a los que pertenecen los ASVs que fueron comunes tras realizar los distintos cruzamientos matriciales (Figura 2.15).

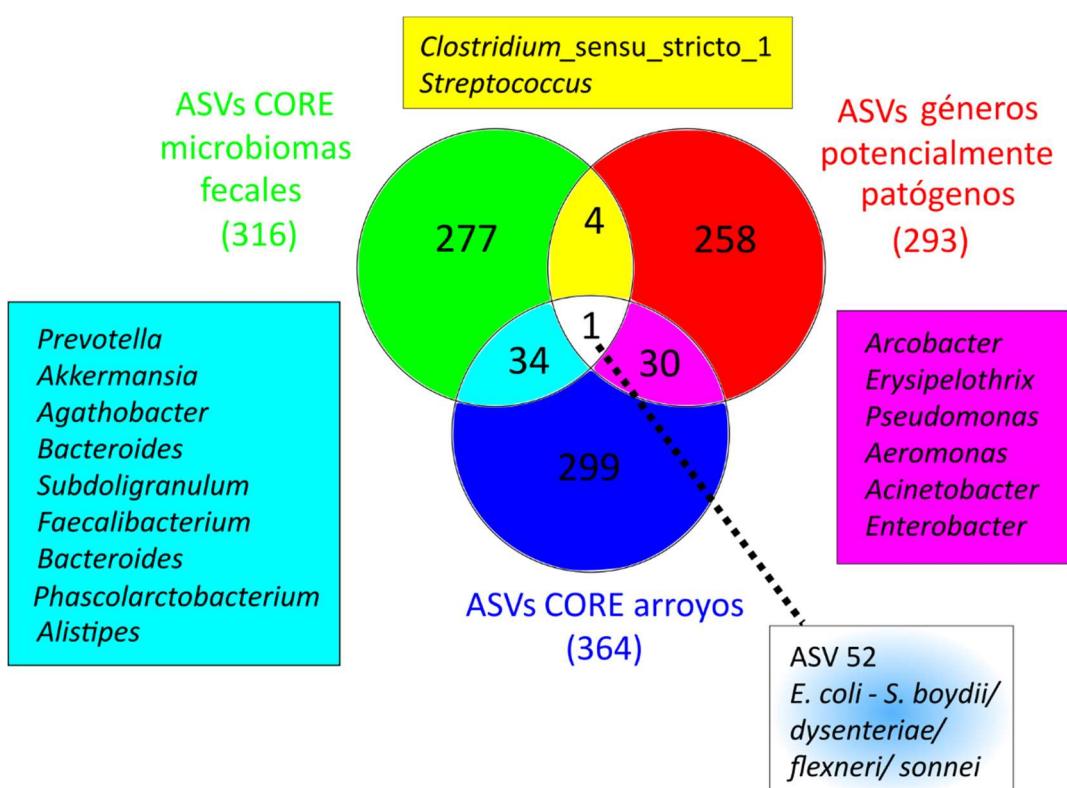


Figura 2.15: Resumen de intersección entre ASVs *core* de la red hidrológica, ASVs *core* de microbiomas fecales del AMBA y ASVs vinculados a géneros potencialmente patógenos en los arroyos. Para cada intersección se incluye un recuadro, siguiendo la clave de color, con los principales taxones compartidos.

Por último, y como era de esperarse, el logaritmo de la abundancia de *E. coli* correlacionó positivamente con los conteos de lecturas del ASV 52, así como también con respecto a la abundancia y riqueza de bacterias potencialmente patógenas (Figura 2.16). A su vez, las métricas asociadas a la matriz de microbiomas fecales humanos y a las PLSS evidenciaron una fuerte asociación positiva, lo que podría sugerir que una parte de los taxones patógenos del bacteriplancton podrían tener su origen en la contaminación fecal humana.

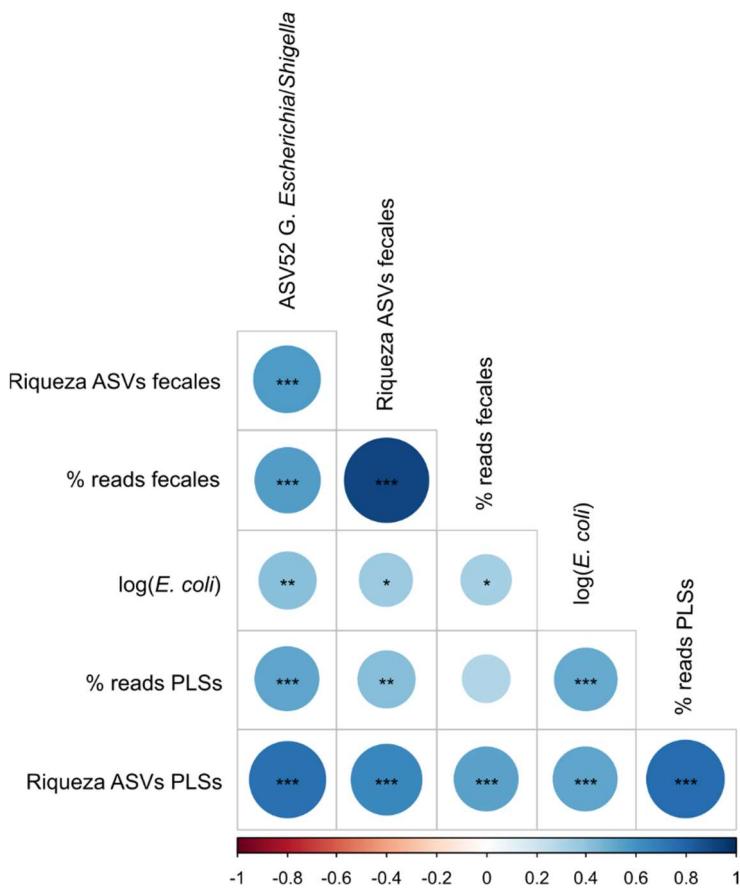


Figura 2.16: Correlaciones de Pearson entre la riqueza y porcentaje de lecturas de bacterias fecales, riqueza y porcentaje de lecturas de PLSSs, cantidad de lecturas del ASV 52 (G. *Escherichia/Shigella*) y el logaritmo del conteo de *E. coli*. Significancias verificadas con permutaciones (nperm=999): ***p<0.001, **p<0.01, *p<0.05.

2.4. Anexo

Tabla A2.1: Rasgos ecológicos de la comunidad del bacteriplancton basados en las matrices de ASVs obtenidas. Los resultados se muestran por sitio de muestreo en la Figura 2.6 (promedio entre ubicaciones internas).

Punto de muestreo	Riqueza observada	Equitatividad de Pielou
LP1	468	0.54
LP2	732	0.67
LP3	1113	0.82
LP4	432	0.39
LP5	812	0.62
LP6	523	0.58
SF1	1041	0.66
SF2	554	0.43
SF3	708	0.62
SF4	508	0.48
SF5	664	0.59
SF6	726	0.61
SD1	688	0.65
SD2	484	0.58

Tabla A2.2: Análisis RDA parcial de efectos puros de la matriz espacial. La contribución de cada conjunto de predictores espaciales se evaluó para cada una de las variables respuesta a través de un análisis pRDA condicionado por los conjuntos de parámetros de infraestructura ambiental y urbana. A) la riqueza de ASVs como variable de respuesta; B) La equitatividad de Pielou como variable de respuesta; C) La riqueza y abundancia relativa de taxones fecales como matriz respuesta; y D) La riqueza de PLSS como variable respuesta. AEM, mapa asimétrico de autovectores.

A)	Valor F	Significancia (P<0.05)
Matriz de predictores espaciales (modelo global)	$F_{4,24}= 9.81$	*
AEM1	$F_{1,24}= 2.50$	
AEM11	$F_{1,24}= 2.93$	*
AEM22	$F_{1,24}= 19.68$	*
AEM34	$F_{1,24}= 0.94$	
Eje RDA1	$F_{1,24}= 44.16$	*
B)	Valor F	Significancia (P<0.05)
Matriz de predictores espaciales (modelo global)	$F_{4,24}= 84.66$	*
AEM2	$F_{1,24}= 14.13$	*
AEM4	$F_{1,24}= 18.89$	*
AEM11	$F_{1,24}= 22.71$	*
AEM22	$F_{1,24}= 31.14$	*
Eje RDA1	$F_{1,24}= 380.95$	*
C)	Valor F	Significancia (P<0.05)
Matriz de predictores espaciales (modelo global)	$F_{5,23}= 48.35$	*
AEM22	$F_{1,23}= 22.48$	*
AEM39	$F_{1,23}= 3.64$	*
Eje RDA1	$F_{1,26}= 272.30$	*
D)	Valor F	Significancia (P<0.05)
Matriz de predictores espaciales (modelo global)	$F_{5,23}= 3.80$	*
AEM1	$F_{1,23}= 2.40$	*
AEM22	$F_{1,23}= 5.27$	*
AEM34	$F_{1,23}= 2.36$	*
Eje RDA1	$F_{1,27}= 22.31$	*

Tabla A2.3: Resumen de géneros y especies potencialmente patógenas (PLSs) rastreados, presencia/ausencia en el *dataset* de la cuenca San Francisco-Las Piedras-Santo Domingo, patologías asociadas y cita bibliográfica por especie bacteriana, e ID NCBI a la secuencia 16S ARNr de la cepa de referencia ATCC con la que se compararon los ASVs para la determinación de PLSs.

Nº Género	Géneros	P/A género	Nº Espe- cie	Especies pa- togénicas	P/A espe- cies (>99%id)	Enfermedades asociadas	ID NCBI	Fuente
1	<i>Acinetobacter</i>	P	1	<i>A. haemolyti- cus</i>	P	Bloody diarrhea	NR 026207	Cui, 2017
			2	<i>A. johnsonii</i>	P	Nosocomial	NR 044975	Cui, 2017
			3	<i>A. junii</i>	P	Septicemia	NR 026208	Cui, 2017
			4	<i>A. baumannii</i>	A		NR 119358	Jones, 2008
2	<i>Actinobacillus</i>	A	5	<i>A. hominis</i>		Sinusitis, bronchopneumonia, and meningitis		NIH, 2019
			6	<i>A. ureae</i>				NIH, 2019
3	<i>Aeromonas</i>	P	7	<i>A. hydrophila</i>	P	Gastroenteritis	NR 043638	Cui, 2017
			8	<i>A. veronii</i>	P	Septicemia	NR 102789	Cui, 2017
4	<i>Arcanobacterium</i>	A	9	<i>A. haemolyti- cum</i>		Ulcerative skin lesions		NIH, 2019
5	<i>Arcobacter</i>	P	10	<i>A. butzleri</i>	P	Diarrhoea, Abdominal pain	NR 074573	Cui, 2017
			11	<i>A. cryaerophilus</i>	P	Diarrhea	NR 025905.1	Cui, 2017
6	<i>Bacillus</i>	P	12	<i>B. anthracis</i>	A	Anthrax	NR 041248	Cui, 2017
			13	<i>B. cereus</i>	A	Nausea, vomiting and diarrhea	NR 074540	Cui, 2017
7	<i>Bartonella</i>	A	14	<i>B. henselae</i>		Cat-scratch disease		Cui, 2017
			15	<i>B. bacillifor- mis</i>				Jones, 2008
			16	<i>B. elizabethae</i>				Jones, 2008
			17	<i>B. quintana</i>				NIH, 2019
			18	<i>B. vinso- nii subsp. ber- khoffii</i>				NIH, 2019; Cheslock, 2019
			19	<i>B. vinso- nii subsp. aru- pensis</i>				NIH, 2019; Cheslock, 2019
8	<i>Bordetella</i>	A	20	<i>B. bronchiseptica</i>		Bronchitis		Cui, 2017
			21	<i>B. pertussis</i>		Whooping cough		Cui, 2017
9	<i>Borrelia</i>	A	22	<i>B. burgdorferi</i>		Lyme disease		Cui, 2017
			23	<i>B. recurrentis</i>				NIH, 2019
			24	<i>B. mayonii</i>				NIAID Emerging Infectious

							NIAID Emerging Infectious
			25	<i>B. miyamotoi</i>			
10	<i>Brucella</i>	A	26	<i>B. melitensis</i>		Brucellosis	Cui, 2017
			27	<i>B. suis</i>			Cui, 2017
			28	<i>B. abortus</i>			NIH, 2019
			29	<i>B. canis</i>			NIH, 2019
			30	<i>B. suis</i>			NIH, 2019
11	<i>Burkholderia</i>	A	31	<i>B. mallei</i>		Glanders	Cui, 2017
			32	<i>B. pseudomallei</i>		Melioidosis	Cui, 2017
			33	<i>B. cepacia</i>			Jones, 2008
12	<i>Campylobacter</i>	P	34	<i>C. jejuni</i>	A	Diarrhea, campylobacteriosis	NR 041834 Cui, 2017
			35	<i>C. fetus</i>	A		NR 118514.1 NIH, 2019
			36	<i>C. coli</i>	A		NR 118511.1 NIH, 2019
13	<i>Chlamydia</i>	A	37	<i>C. pneumoniae</i>		Pneumonia	Cui, 2017
			38	<i>C. psittaci</i>		Psittacosis	Cui, 2017
			39	<i>C. trachomatis</i>		Trachoma	Cui, 2017
14	<i>Chlamydophila</i>	A	40	<i>C. pneumoniae</i>		Pneumonia	Cui, 2017
15	<i>Clostridium</i>	P	41	<i>C. baratii</i>	P	Clostridial disease	NR 029229 Cui, 2017
			42	<i>C. botulinum</i>	A	Botulism	NR 076243.1 Cui, 2017
			43	<i>C. chauvoei</i>	A		NR 026013.1 NIH, 2019
			44	<i>C. difficile</i>	A	Clostridial disease	NR 113132.1 Cui, 2017
			45	<i>C. haemolyticum</i>	A		NR 024749.1 NIH, 2019
			46	<i>C. histolytica</i>	A	Clostridial disease	NR 104889 Cui, 2017
			47	<i>C. novyi</i>	A		NR 040855.1 NIH, 2019
			48	<i>C. perfringens</i>	A	Gas gangrene, Enterotoxemia	NR 112169.1 Cui, 2017
			49	<i>C. septicum</i>	A		NR 026020.1 NIH, 2019
			50	<i>C. sporogenes</i>	A	Clostridial disease	NR 029231 Cui, 2017
			51	<i>C. tetani</i>	A	Tetanus	NR 029260 Cui, 2017
16	<i>Corynebacterium</i>	A	52	<i>C. diphtheriae</i>		Diphtheria	Cui, 2017
			53	<i>C. jeikeium</i>			Cui, 2017
			54	<i>C. pseudotuberculosis</i>			NIH, 2019
			55	<i>C. renale</i>			NIH, 2019
17	<i>Coxiella</i>	P	56	<i>C. burnetii</i>	A	Q fever	NR 104916.1 Jones, 2008
18	<i>Dermatophilus</i>	A	57	<i>D. congolensis</i>			NIH, 2019
19	<i>Edwardsiella</i>	P	58	<i>E. tarda</i>	A		NR 024770.1 NIH, 2019
20	<i>Ehrlichia</i>	A	59	<i>E. ruminantium</i>		Heartwater disease	Cui, 2017
21	<i>Elizabethkingia</i>	P	60	<i>E. anophelis</i>	A		NR 116021.1 McArthur, 2020
			61	<i>E. meningoseptica</i>	A		NR 042267.1 McArthur, 2020

			62	<i>E. miricola</i>	P			NR 036862.1	McArthur, 2020
22	<i>Enterobacter</i>	P	63	<i>E. cloacae</i>	P			NR 102794	Cui, 2017
23	<i>Enterococcus</i>	P	64	<i>E. faecalis</i>	P	Nosocomial infections	NR 115765	Cui, 2017	
			65	<i>E. faecium</i>	A		NR 115764	Cui, 2017	
24	<i>Erysipelothrix</i>	P	66	<i>E. rhusiopathiae</i>	A	Acute septicemia	NR 040837	Cui, 2017	
25	<i>Escherichia/Shigella</i>	P	67	<i>E. coli</i>	P	Enterohemorrhagic diarrhea	NR 114042.1	Cui, 2017	
			68	<i>S. boydii</i>	P	Shigellosis	NR 104901.1	Cui, 2017	
			69	<i>S. dysenteriae</i>	P		NR 026332.1	Cui, 2017	
			70	<i>S. flexneri</i>	P		NR 026331.1	Cui, 2017	
			71	<i>S. sonnei</i>	P	Shigellosis	NR 104826.1	Cui, 2017	
26	<i>Francisella</i>	A	72	<i>F. tularensis</i>		Tularemia			Cui, 2017
27	<i>Haemophilus</i>	A	73	<i>H. ducreyi</i>		Chancroid			Cui, 2017
			74	<i>H. influenzae</i>		Sepsis and bacterial meningitis			Cui, 2017
28	<i>Helicobacter</i>	A	75	<i>H. pylori</i>		Helicobacter Infections			Cui, 2017
			76	<i>H. cinaedi</i>					Jones, 2008
			77	<i>H. fennelliae</i>					Jones, 2008
29	<i>Klebsiella</i>	P		<i>K. pneumoniae</i>	P	Pneumonia, soft tissue infections			
			78	<i>K. aerogenes</i>	P	Opportunistic infections	NR 118556	Cui, 2017	
			79						
30	<i>Legionella</i>	P	80	<i>L. pneumophila</i>	A	Legionnaire's Disease	NR 041742	Cui, 2017	
31	<i>Leptospira</i>	P	81	<i>L. interrogans</i>	A	Leptospirosis	M71241.1	Cui, 2017	
			82	<i>L. fainei</i>	A		NR 043049.1	Jones, 2008	
			83	<i>L. weilii</i>	A		NR 119300.1	Jones, 2008	
32	<i>Listeria</i>	A	84	<i>L. monocytogenes</i>	A	Listeriosis			Cui, 2017
33	<i>Moraxella</i>	P	85	<i>M. catarrhalis</i>	A		NR 118775.1	NIH, 2019	
			86	<i>M. lacunata</i>	A		NR 114416.1	NIH, 2019	
34	<i>Mycobacterium</i>	A	87	<i>M. abscessus</i>					Jones, 2008
			88	<i>M. aurum</i>		Bacteraemia, keratitis			Cui, 2017
			89	<i>M. asiaticum</i>					NIH, 2019
			90	<i>M. avium</i>		M. avium-intracellulare Infection			Cui, 2017
			91	<i>M. brumae</i>		Bacteraemia			Cui, 2017
			92	<i>M. bovis</i>					NIH, 2019
			93	<i>M. celatum</i>					Jones, 2008
			94	<i>M. chelonae</i>					NIH, 2019
			95	<i>M. conspiciuum</i>					Jones, 2008
			96	<i>M. farcinogenes</i>		Prosthesis infection			Cui, 2017
			97	<i>M. fortuitum</i>					NIH, 2019
			98	<i>M. genavese</i>					Jones, 2008
			99	<i>M. gordonaiae</i>					Jones, 2008

			100	<i>M. haemophilum</i>					Jones, 2008
			101	<i>M. malmoense</i>					NIH, 2019
			102	<i>M. marinum</i>					NIH, 2019
			103	<i>M. morio-kaense</i>		Pneumonia			Cui, 2017
			104	<i>M. leprae</i>		Leprosy			Cui, 2017
			105	<i>M. kansasii</i>					NIH, 2019
			106	<i>M. paratuberculosis</i>					NIH, 2019
			107	<i>M. shimoidei</i>					Jones, 2008
			108	<i>M. simiae</i>					NIH, 2019
			109	<i>M. scrofulaceum</i>					NIH, 2019
			110	<i>M. szulgai</i>					NIH, 2019
			111	<i>M. tuberculosis</i>					Jones, 2008
			112	<i>M. ulcerans</i>					NIH, 2019
			113	<i>M. xenopi</i>					NIH, 2019
35	<i>Mycoplasma</i>	A	114	<i>M. bovis</i>					Cui, 2017
			115	<i>M. fermentans</i>					Jones, 2008
			116	<i>M. pneumoniae</i>		Pelvic inflammatory diseases			Cui, 2017
			117	<i>M. pneumoniae</i>		Atypical pneumonia and other respiratory disorders			Cui, 2017
36	<i>Neisseria</i>	A	118	<i>N. gonorrhoeae</i>					NIH, 2019
			119	<i>N. meningitidis</i>		Meningococcal Infections			Cui, 2017
			120	<i>N. weaveri</i>					Jones, 2008
37	<i>Nocardia</i>	A	121	<i>N. asteroides</i>		Nocardiosis			NIH, 2019
			122	<i>N. brasiliensis</i>					NIH, 2019
			123	<i>N. nova</i>					Cui, 2017
			124	<i>N. otitidisca-viarum</i>					NIH, 2019
			125	<i>N. transvalensis</i>					NIH, 2019
			126	<i>N. veterana</i>					Jones, 2008
38	<i>Orientia</i>	A	127	<i>O. tsutsugamushi</i>					Jones, 2008
39	<i>Pasteurella</i>	A	128	<i>P. multocida</i>					NIH, 2019
40	<i>Proteus</i>	P	129	<i>P. mirabilis</i>	P		NR_114419.1		Bereket, 2012
41	<i>Pseudomonas</i>	P	130	<i>P. aeruginosa</i>	A	Pseudomonas infection	NR_114471.1	Cui, 2017	
			131	<i>P. alcaligenes</i>	P	Bloodstream Infection	NR_114472.1	Cui, 2017	
42	<i>Rhodococcus</i>	A	132	<i>R. equi</i>					Jones, 2008
43	<i>Rickettsia</i>	A	133	<i>R. akari</i>					NIH, 2019
			134	<i>R. australis</i>					NIH, 2019
			135	<i>R. canada</i>					NIH, 2019
			136	<i>R. conorii</i>		Boutonneuse Fever			Cui, 2017

			137	<i>R. prowazekii</i>		Epidemic Typhus fever		Cui, 2017
			138	<i>R. rickettsii</i>		Rocky mountain spotted fever		Cui, 2017
			139	<i>R. siberica</i>				NIH, 2019
			140	<i>R. typhi</i>		Typhus, Endemic Flea-Borne		Cui, 2017
44	<i>Salmonella</i>	A	141	<i>S. arizona</i>				NIH, 2019
			142	<i>S. choleraesuis</i>				NIH, 2019
			143	<i>S. enterica</i>		Typhoid Fever, Acute gastroenteritis		Cui, 2017
			144	<i>S. enteritidis</i>				NIH, 2019
			145	<i>S. gallinarum-pullorum</i>				NIH, 2019
			146	<i>S. meleagridis</i>				NIH, 2019
			147	<i>S. paratyphi</i>				NIH, 2019
			148	<i>S. typhi</i>				NIH, 2019
			149	<i>S. typhimurium</i>				NIH, 2019
			150	<i>S. marcescens</i>	A	Nosocomial infections	NR 114043	Cui, 2017
45	<i>Serratia</i>	P	151	<i>S. odorifera</i>	A		NR 114157.1	Jones, 2008
47	<i>Staphylococcus</i>	P	152	<i>S. aureus</i>	P	Coagulase-positive staphylococcal infections	NR 115606.1	Cui, 2017
			153	<i>S. epidermidis</i>	P	Infections of implanted prostheses	NR 113957.1	Cui, 2017
			154	<i>S. haemolyticus</i>	A		NR 036955.1	Jones, 2008
			155	<i>S. lugdunensis</i>	A		NR 024668.1	Jones, 2008
			156	<i>S. saprophyticus</i>	A	Cystitis in women	NR 074999	Cui, 2017
			157	<i>S. schleiferi</i>	A		NR 037009.1	Jones, 2008
48	<i>Stenotrophomonas</i>	P	158	<i>St. maltophilia</i>	P	Immunosuppressed and nosocomial disease/common environmental	NR 112030.1	Jones, 2008
49	<i>Streptococcus</i>	P	159	<i>S. agalactiae</i>	A	Invasive neonatal disease	NR 115728.1	Cui, 2017
			160	<i>S. pneumoniae</i>	P	Pneumococcal infectious	NR 115239.1	Cui, 2017
			161	<i>S. pyogenes</i>	A	Opportunistic infections with septicemia and pneumonia	NR 112088.1	Cui, 2017
			162	<i>S. halichoeri</i>	A		NR 029025.1	Shakir, 2021
			163	<i>S. iniae</i>	A		NR 025148.1	Jones, 2008
50	<i>Treponema</i>	P	164	<i>Treponema pallidum</i>	A	Sífilis	M88726.1	Jones, 2008
51	<i>Tropheryma</i>	A	165	<i>Tropheryma whippelii</i>		Enfermedad de Whipple		Jones, 2008
52	<i>Vibrio</i>	P	166	<i>V. cholerae</i>	P	Diarrheal disease, cholera	NR 119051	Cui, 2017
			167	<i>V. fluvialis</i>	A		NR 119053.1	Jones, 2008
			168	<i>V. metschnikovii</i>	A		NR 114220.1	Jones, 2008
			169	<i>V. mimicus</i>	A		NR 029259.1	Jones, 2008

			170	<i>V. para-haemolyticus</i>	A		NR_119058.1	Jones, 2008
			171	<i>V. vulnificus</i>	A		NR_117907.1	Jones, 2008
53	<i>Yersinia</i>	P	172	<i>Y. enterocolitica</i>	P	Enterocolitis	NR_104903.1	Cui, 2017
			173	<i>Y. pestis</i>	A		NR_025160	Cui, 2017
			174	<i>Y. pseudotuberculosis</i>	A		NR_025158	Cui, 2017

El Capítulo 3 fue retirado a pedido del autor

Capítulo 4

Caracterización de rasgos asociados a fenómenos de persistencia ambiental en aislamientos de *E. coli*

Capítulo 4

Caracterización de rasgos asociados a fenómenos de persistencia ambiental en aislamientos de *E. coli*

4.1. Resumen

Bajo la hipótesis de que los distintos grupos filogenéticos de *E. coli* (A, B1, B2, D, E, F/G) exhiben características fenotípicas diferenciales que contribuyen a la persistencia ambiental extendida por parte de algunos de ellos, en particular de B1, se llevó a cabo un estudio experimental para explorar la capacidad diferencial de reproducirse y de formar estructuras de resistencia, como los biofilms, en condiciones controladas. Para ello, se realizaron ensayos en placas *multiwell* sobre aislamientos obtenidos de la cuenca en estudio, utilizando agua del mismo origen como medio para su cultivo. Los ensayos de supervivencia se realizaron incubando los distintos aislamientos en el agua de la cuenca esterilizada por filtrado (9.88 mg/L DOC; 14.9 mg/L DIN; 1.6 mg/L SRP), monitoreando la absorbancia (570nm) cada hora mediante espectrofotometría. El biofilm formado tras 48hs de incubación, también en agua de la cuenca esterilizada por filtrado, se cuantificó por tinción con cristal violeta. Se evaluaron las diferencias entre los filogrupos para cada variable mediante modelos lineales generales mixtos. En los ensayos de supervivencia (n=68) sólo los grupos A y B1 presentaron tasas de crecimiento positivas en agua de la cuenca (A: 0.32 ± 0.29 día $^{-1}$; B1: 0.16 ± 0.24 día $^{-1}$). La formación de biofilm (n=63) fue significativamente mayor en B1 (*post-hoc Tukey*, P<0.05), aunque la mayoría de los grupos filogenéticos evaluados demostró capacidad de formación de este tipo de estructuras. Con los resultados obtenidos se realizó un agrupamiento jerárquico de los aislamientos, identificándose 4 grupos: los grupos 1 y 2 con tasas de crecimiento positivas y nula formación de biofilm (mayormente A), el grupo 3 con elevada formación de biofilm y tasas de crecimiento bajas (mayormente B1) y el grupo 4 con crecimiento negativo y formación de biofilm nula (enriquecido en D). Estos resultados apoyan la hipótesis de persistencia de *E. coli* en ambientes acuáticos degradados, ya que indican tendencias de crecimiento positivas y capacidad formadora de biofilm bajo condiciones experimentales similares a las de la cuenca, en particular para los filogrupos A y B1, sugiriendo posibles estrategias de persistencia diferenciales para dichos grupos.

4.2. Materiales y métodos

4.2.1 Ensayo de supervivencia

Para evaluar experimentalmente la capacidad de crecimiento de los distintos grupos intraespecíficos de *E. coli* en contextos fisicoquímicos análogos a los representados en la cuenca de estudio, se diseñó un ensayo de supervivencia basado en el cultivo de aislamientos pertenecientes a los distintos grupos filogenéticos en aguas provenientes de la cuenca en estudio. Se utilizaron aislamientos de *E. coli*, pertenecientes a los grupos filogenéticos (A, B1, B2, D, E, F/G), recuperados a partir de muestras de agua de los sitios de muestreo y conservados a -80°C en glicerol analítico al 25%.

Para los ensayos realizados se obtuvieron 5 L agua de la cuenca del sitio de muestreo SF2, locación que reúne condiciones intermedias de disturbio antropogénico (ver figura 1.5), la cual fue mantenida en frío de heladera hasta su esterilización por filtrado (0,22 μm). El agua filtrada se caracterizó a partir de la concentración de sus nutrientes solubles principales (9.88 mg/L DOC; 14.9 mg N-NO₃⁻/L DIN; 1.6 mg P-PO₄³⁻/L SRP según la metodología desarrollada en el capítulo 1 de esta tesis), y posteriormente fue alicuotada y preservada a -20°C.

Previo a cada ensayo, los aislamientos de *E. coli* se reactivaron cultivándose en caldo líquido LB a 36°C y agitación constante (150 rpm), por al menos 8 horas. Posteriormente, las células se centrifugaron (5000 rpm durante 10 minutos) y resuspendieron en agua de la cuenca filtrada 2 veces. Las cantidades de cultivo de *E. coli* reactivado y medio de cultivo (agua de la cuenca estéril) empleados en cada resuspensión siguieron la proporción 1:2. Posteriormente, 150 μl de cada suspensión bacteriana, perteneciente a un mismo aislamiento, se distribuyó en 6 pocillos de una placa. En cada placa *multiwell* se ensayaron un máximo de 9 aislamientos, sin utilizarse la máxima capacidad disponible de pocillos para evitar contaminaciones cruzadas. Los ensayos se realizaron a 25°C (temperatura del agua asociada con la estación de verano) con agitación constante a 120 rpm. En el tiempo inicial y cada 60 minutos (por 8 horas) se midió la DO₅₉₅ (densidad óptica a 595nm) en un equipo lector de placas (Biorad 680). En cada placa se incluyeron también 6 pocillos de cada medio de cultivo estéril y sin inóculo como controles negativos. En paralelo, en cada evento experimental se verificó la cinética bacteriana de los aislamientos en caldo LB y de esta manera se buscó confirmar el comportamiento exponencial esperable de *E. coli* en este medio de cultivo para el rango temporal del ensayo.

4.2.2 Análisis de los ensayos de supervivencia

A partir de los datos de absorbancia experimental obtenidos, se procedió a descontar la absorbancia basal de los blancos respectivos a cada tiempo experimental. De esta manera, se intenta despejar de eventuales modulaciones espurias en las lecturas de absorbancia por parte del equipamiento. Adicionalmente se evaluó en forma independiente la progresión de la absorbancia de los blancos experimentales, considerando siempre una esperable leve dispersión en las respuestas, pero siempre de menor orden respecto a la evolución en la respuesta de las unidades experimentales. Con los datos resultantes se evaluó el ajuste a un modelo de regresión lineal. Se tomó como criterio mínimo de ajuste un R^2 de 0.55. En los casos que hubiese observaciones que se desviaban más de 2.5 desviaciones estándar del valor predicho, dichas observaciones fueron consideradas *outliers*. Se mantuvieron en el análisis series temporales con al menos 5 observaciones (los aislamientos con comportamientos erráticos fueron repetidos para obtener datos robustos). De cada serie temporal se obtuvo la pendiente del modelo lineal como parámetro cinético descriptivo de la performance de cada aislamiento (Figura 4.1). En la Tabla 4.1 se detalla la cantidad de aislamientos de cada grupo filogenético de *E. coli* caracterizados por este ensayo e incluidos en análisis. Para su posterior tratamiento estadístico, los datos de las pendientes obtenidas fueron multiplicados por un factor fijo para obtener una tasa cinética diaria (factor multiplicativo: 1440).

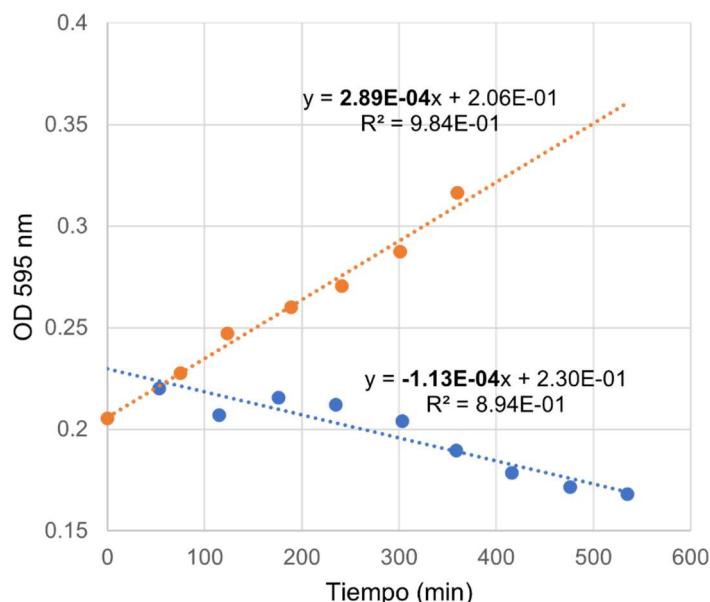


Figura 4.1: Ejemplos de curvas obtenidas evidenciando crecimiento (naranja) o decrecimiento (azul) en el tiempo experimental. En el eje de abscisas el tiempo en minutos; en el eje de ordenadas la densidad óptica a 595 nm (blancos descontados). De cada curva se explicita su recta y ajuste lineal (R^2). En negrita se destaca la pendiente empleada para la obtención de la tasa cinética diaria.

Tabla 4.1: Resumen de cantidad de aislamientos evaluados en los ensayos de supervivencia bacteriana y de formación de biofilm por filogrupos de *E. coli*

	A	B1	B2	D	E	F/G	N total por ensayo
Ensayo de supervivencia	26	23	3	11	4	1	68
Ensayo de formación de biofilm	24	21	1	11	4	2	63

4.2.3 *Ensayo de formación de biofilm*

Se evaluó la capacidad formadora de biofilm de aislamientos de *E. coli* siguiendo la metodología de Reisner *et al.*¹³², con algunas modificaciones. Tras la reactivación de los aislamientos bajo las mismas condiciones que en el método previo, y con el objetivo de estandarizar las condiciones iniciales, se utilizaron suspensiones de bacterias a una DO₅₉₅ de 0,1. A partir del volumen necesario de inóculo para cada ensayo y con el fin de remover el LB utilizado como medio de cultivo, las bacterias se centrifugaron a 5000 rpm durante 10 minutos y se resuspendieron en solución fisiológica estéril (NaCl 0,9%). Este paso fue repetido 1 vez más, y las bacterias resultantes fueron resuspendidas en el volumen de agua de la cuenca esterilizada por filtrado apropiado para el ensayo. Se distribuyeron 200 µl de las suspensiones de cada aislamiento sobre 4 pocillos al azar de una placa estéril de 96 pocillos de fondo plano y sin tratamiento de adhesión. En cada placa se incluyeron 4 pocillos con agua de la cuenca estéril como controles negativos. Las placas se incubaron durante 48 horas a 25°C en cámara húmeda y sin agitación.

Después de la incubación, se trasvasó el sobrenadante de cada pocillo a una placa de fondo plano no estéril y se midió la absorbancia de las células planctónicas (parámetro APL, del inglés *absorbance of planctonic cells*). Los pocillos se lavaron dos veces con 200 µl de solución salina estéril para eliminar todas las células bacterianas no adherentes. Las células adherentes restantes se tiñeron añadiendo 160 µl de violeta cristal (0,1% m/v). Despues de 15 minutos de tinción, se eliminó el exceso de violeta cristal de los pocillos, y se lavaron los pocillos dos veces con 200 µl de solución salina (0,9%). Se añadió etanol 95%-acetona 25% a cada pocillo para destear las células bacterianas adheridas. Luego de 5 minutos, se cuantificó la cantidad de violeta cristal en el etanol midiendo la absorbancia a 570 nm (parámetro ACV, del inglés *absorbance of crystal violet*). La absorbancia de los pocillos control (agua de la cuenca sin inóculo) son el parámetro CW

(del inglés *control well*). Finalmente, se utilizó el Índice de Formación de Biofilm Específico (SBF, del inglés *Specific Biofilm Formation Index*) de cada aislamiento, calculado siguiendo la ecuación: ACV-CW/APL. Se evaluó la dispersión entre réplicas y en los casos con una dispersión menor al 5% se promediaron los valores de absorbancia para el cálculo del SBF. Cada aislamiento fue evaluado de esta manera en al menos 2 eventos independientes. En la Tabla 4.1 se detalla la cantidad de aislamientos de cada grupo filogenético de *E. coli* caracterizados por este ensayo e incluidos en análisis.

4.2.4 Análisis estadístico general

Para evaluar cambios significativos en la tasa de crecimiento diario y el SBF según la identidad filogenética de los aislamientos evaluados, se realizaron sendos modelos lineales mixtos para cada variable respuesta, en los que se incluyó al filogrupo como variable explicativa fija y la placa en la que se realizó cada ensayo como factor aleatorio, y para considerar la heterocedasticidad encontrada se modeló la varianza mediante la función *varIdent*¹³³. Las variables respuesta fueron previamente estandarizadas por la función *decostand* de *vegan* (método *standardize*)⁹⁸. Además, se verificaron los supuestos de los modelos utilizados, incluyendo la normalidad de la variable de efecto aleatorio. A partir de cada modelo se realizó un ANOVA para evaluar la significancia de la variable filogrupo sobre dicha variable respuesta y contrastes *post hoc* de Tukey para evaluar diferencias significativas según el filogrupo ($P<0.05$). Los modelos, verificación de supuestos y contrastes se realizaron empleando los paquetes *nlme* (versión 3.1-160), *stats* (versión 4.2.2), *emmeans* (versión 1.8.5) y *lsmeans* (versión 2.30.0) en R¹³⁴⁻¹³⁷.

Complementariamente, con las tasas de supervivencia diaria y el SBF resultantes se realizó un análisis de agrupamiento jerárquico de los aislamientos para visualizar el ordenamiento de éstos según ambas caracterizaciones experimentales. Para ello, las variables respuesta fueron previamente estandarizadas por la función *decostand* de *vegan* (método *range*)⁹⁸. El agrupamiento jerárquico de los aislamientos se realizó a partir de una matriz de disimilitud computada a través de distancias euclídeas según la función *hclust* del paquete *stats* de R¹³⁵ (versión 4.2.2); el mapa de color y el dendrograma se elaboraron empleando el paquete *gplots* en R (versión 3.1.3)¹³⁸.

Por último, con las tasas de supervivencia diaria y el SBF como factores explicativos y la identidad filogenética como variable diana (categórica) se diseñó un modelo de árbol de clasificación y regresión (CART, del inglés *Classification and Regression Tree*). El

propósito de este análisis fue el de obtener una clasificación de los filogrupos a partir del comportamiento experimental de los aislamientos respectivos. Para ello se emplearon los paquetes *rpart* y *rpart.plot* en R¹³⁹.

4.3. Resultados

4.3.1 Ensayos de supervivencia bacteriana

El análisis de la variación en la tasa de crecimiento en agua de la cuenca obtenidas experimentalmente para cada grupo filogenético evidenció una tendencia de crecimiento positivo por parte de los aislamientos del filogrupo A (media $0.32 \pm 0.29 \text{ día}^{-1}$), y en segundo lugar por parte de los aislamientos B1 (media $0.16 \pm 0.24 \text{ día}^{-1}$) (Figura 4.2a). Contrariamente, el resto de los grupos filogenéticos analizados no mostraron tendencias claras: el conjunto de aislamientos pertenecientes al grupo filogenético D revela una gran dispersión alrededor de una tasa de crecimiento nula (media $0.04 \pm 0.26 \text{ día}^{-1}$) y los filogrupos B2, E y F/G tienen un n muestral pequeño, lo cual no nos permitió observar un comportamiento concluyente. Es notable la elevada variabilidad general en las respuestas experimentales.

Además, se observó en el ANOVA realizado que, a pesar de que el crecimiento estimado por parte de los filogrupos A y B1 fue positivo y distinto de cero (respectivos intervalos de confianza; IC 95%: 0.045-0.41 y 0.017-0.37 día^{-1}), el conjunto de las tasas de supervivencia diarias no logra ser explicado en forma estadísticamente significativa cuando se emplea la filiación filogenética como variable explicativa ($F_{5,50}=1.34$; $P=0.26$) (Figura 4.2b).

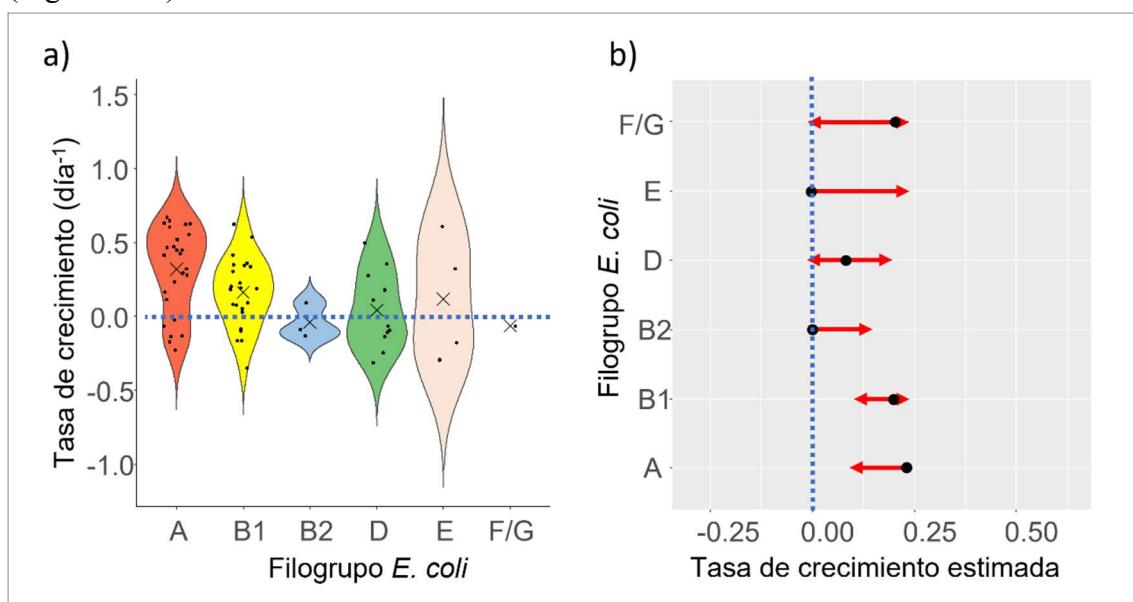


Figura 4.2: a) Gráfica de violín con la tasa de crecimiento obtenida experimentalmente por filogrupos de *E. coli*. La línea punteada horizontal delimita el cero de la gráfica, mientras que la cruz dentro de cada conjunto representa el valor medio del mismo; b) Comparación gráfica del resultado del modelado de la varianza de la tasa diaria por filogrupos. Cada punto negro representa la media marginal estimada, mientras que las flechas rojas permiten realizar la comparación entre grupos a partir de la incertidumbre asociada a la estimación. Líneas rojas superpuestas indican la falta de significancia entre pares de comparaciones. La línea punteada vertical delimita el cero de la gráfica.

4.3.2 Ensayo de formación de biofilm

A partir de los ensayos de formación de biofilm realizados, se observó que los aislamientos provenientes del filogrupo B1 tienen mayor capacidad de formación de biofilm (SBF promedio= 1.35 ± 1.17) (Figura 4.3a). Los aislamientos pertenecientes a los filogrupos A (0.57 ± 0.65) y D (0.38 ± 0.28) también mostraron tendencias positivas, aunque de menor magnitud, en cuanto a este rasgo. Análogamente a lo informado para los ensayos cinéticos de crecimiento bacteriano, los pocos aislamientos B2, E y F/G no permiten sacar conclusiones claras, aunque en general los valores sugieren un índice positivo.

El ANOVA realizado evidenció un efecto significativo del filogrupo en explicar la variabilidad encontrada en la capacidad de formación de biofilm ($F_{5,48}=4.54$; $P<0.05$). Adicionalmente, se realizaron comparaciones *a posteriori* entre filogrupos (Test de Tukey), evidenciando una diferencia significativa en la capacidad de formación de biofilm entre el filogrupo B1 en contraste con los grupos A, D y F/G ($P<0.05$; Figura 4.3b). En todos los casos B1 tiene una capacidad de formación de biofilm mayor que estos filogrupos. El resto de los filogrupos no presentaron diferencias significativas entre sí.

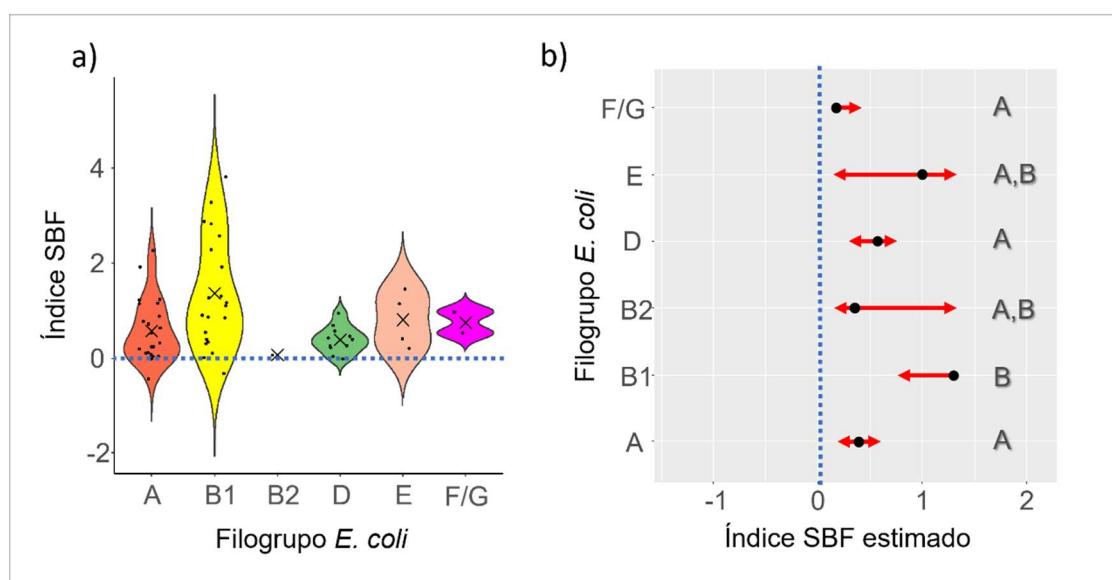


Figura 4.3: a) Gráfica de violín con Índices de Formación de Biofilm Específicos (SBF) experimentales por filogrupo de *E. coli*. La línea punteada horizontal delimita el cero de la gráfica, mientras que la cruz dentro de cada conjunto representa el valor medio del mismo; b) Comparación gráfica de contrastes de Tukey referido al modelo lineal de SBF. Cada punto negro representa la media marginal estimada, mientras que las flechas rojas permiten realizar la comparación entre grupos a partir de la incertidumbre asociada a la estimación. Líneas rojas superpuestas indican la falta de significancia entre pares de comparaciones. Las letras A y B sobre la gráfica expresan las comparaciones *a posteriori* significativas entre grupos filogenéticos.

4.3.3 Análisis integral del comportamiento experimental de los filogrupos

Con el objetivo de valorar de manera integral rasgos complementarios como los obtenidos asociables a fenómenos de persistencia ambiental extendida, se practicaron distintos enfoques estadísticos para analizar la respuesta de los distintos grupos filogenéticos evaluados de *E. coli*. Por un lado, se empleó como estrategia el empleo de un análisis de agrupamiento jerárquico en el que se buscó que los aislamientos se ordenen según su comportamiento con ambas variables experimentales (Figura 4.4). A partir de dicho análisis se obtuvieron 4 agrupamientos: un primer grupo enriquecido en aislamientos pertenecientes a B1, con tasas de crecimiento positivas de valor intermedio, y elevado índice de formación de biofilm; un segundo grupo, dominado por el filogrupo D, con tasas de supervivencia e índice de formación específico de biofilm bajos; y dos agrupamientos restantes, dominados ambos por aislamientos pertenecientes al filogrupo A, con baja capacidad formadora de biofilm, pero con tasas de crecimiento positivas intermedias y elevadas en los respectivos agrupamientos.

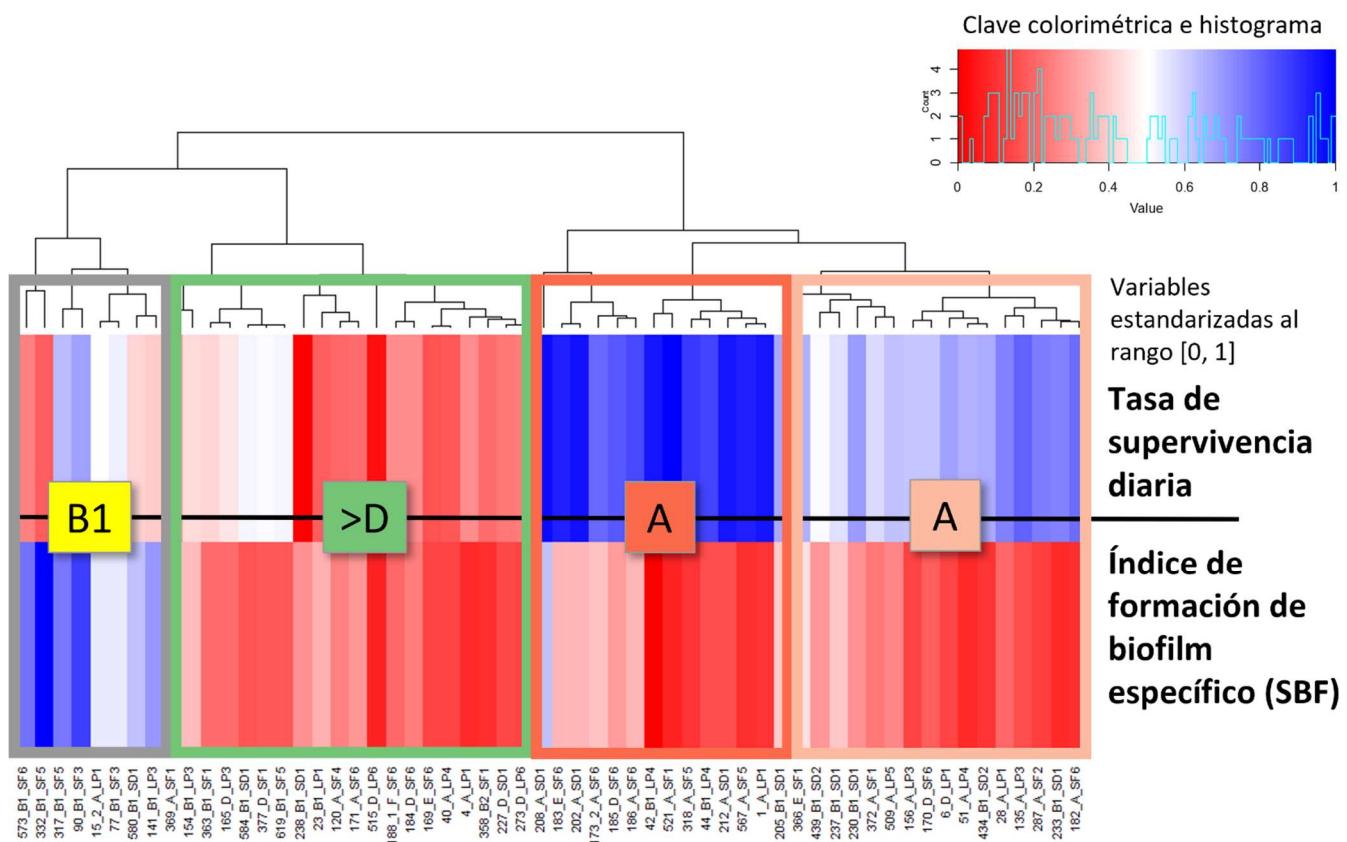


Figura 4.4: Agrupamiento jerárquico de los aislamientos de *E. coli* según su comportamiento conjunto en el ensayo de supervivencia (fracción superior de la gráfica; variable respuesta=Tasa de crecimiento diaria) y el ensayo de formación de biofilm (fracción inferior de la gráfica; variable respuesta=Índice de formación de biofilm específico). El eje X cada aislamiento se encuentra identificado por su ID, filiación filogenética y sitio de aislamiento. De cada agrupamiento resultante se resalta el grupo filogenético más abundante.

Complementariamente, se diseñó un modelo de árbol de decisión basado en los resultados experimentales de supervivencia y de formación de biofilm como variables discriminantes, y la filiación filogenética de los aislamientos como variable diana o a discriminar. El algoritmo empleado permite discriminar los valores umbrales de las variables explicativas que permiten ordenar de manera óptima la probabilidad de que un aislamiento pertenezca a un dado filogrupo (Figura 4.5). Según el modelo obtenido, tasas de supervivencia mayores a 0.42 discriminan a un primer grupo de aislamientos asignados al filogrupo A, independientemente de su capacidad formadora de biofilm (27.6 % de los aislamientos están dentro del mismo). Tasas menores a este umbral conllevan a una segunda instancia de discriminación basada en el índice SBF: aislamientos con índices mayores a 0.81 son discriminados como B1 (31% de los aislamientos). Por debajo de este umbral, una mayor tasa de supervivencia (> 0.18) agrupa a aislamientos del filogrupo A (19% de los aislamientos), mientras que por debajo de este se agrupan los aislamientos pertenecientes al filogrupo D (22.4%).

En resumen, los enfoques estadísticos practicados dan cuenta de que la integración de ambas variables fenotípicas permitió distinguir con mayor detalle el comportamiento de distintos grupos filogenéticos, siendo los filogrupos A y B1 los que presentaron rasgos compatibles con fenómenos de persistencia ambiental. Sin embargo, esta aptitud no sería homóloga entre estos grupos, ya que la persistencia ambiental del filogrupo A, a la luz de estos resultados, podría vincularse más a una aptitud de persistir, incluso multiplicarse, en la columna de agua de sistemas hidrológicos con elevada carga orgánica, mientras que la permanencia ambiental extendida de B1 estaría más asociada a su mayor aptitud formadora de biofilms sobre superficies colonizables.

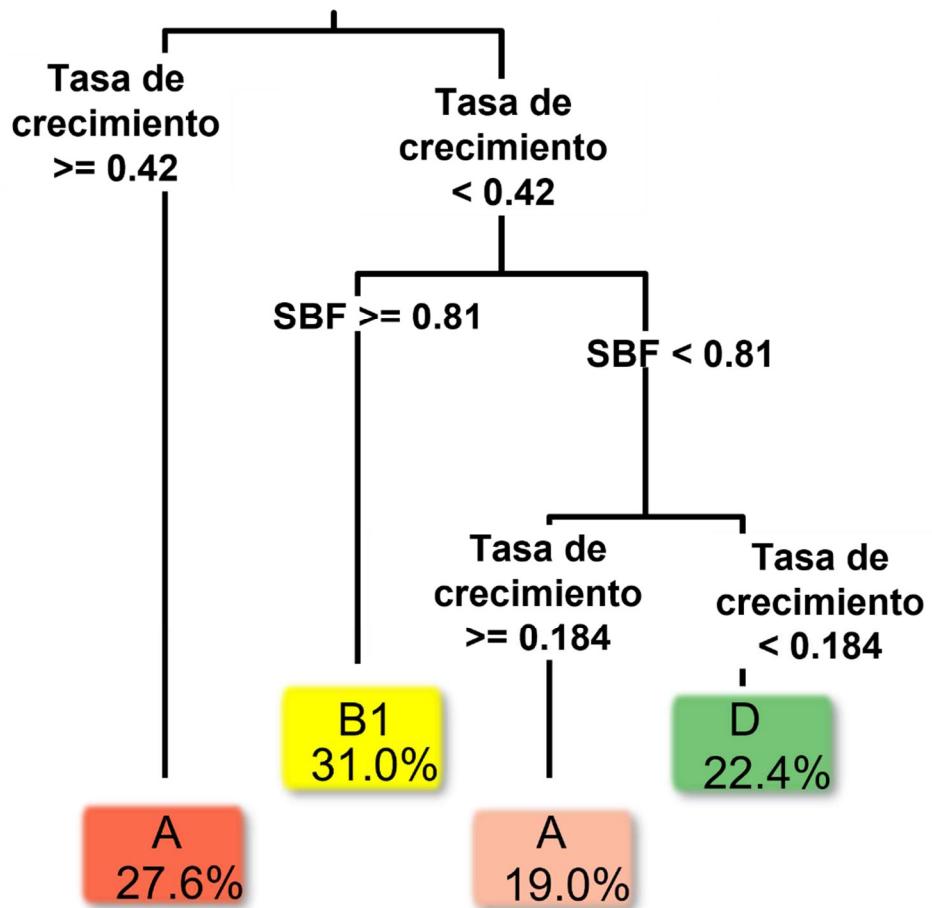


Figura 4.5: Visualización del modelo de árbol de regresión en el que se expresa la probabilidad de que un aislamiento pertenezca a alguno de los grupos filogenéticos evaluados con base a su Tasa de crecimiento diaria y al Índice de formación de biofilm específico como variables discriminantes o explicativas.

Capítulo 5

Estudio genómico comparativo de aislamientos de clado críptico
IV de origen ambiental y comensal

Capítulo 5

Estudio genómico comparativo de aislamientos de clado críptico IV de origen ambiental y comensal

5.1. Resumen

Los clados crípticos son grupos taxonómicos filogenéticamente cercanos a *Escherichia coli* pero que, a diferencia de esta especie, no se asocian con el comensalismo entérico de animales homeotermos, sino que estarían adaptados a la vida libre, lo que se conoce como la "hipótesis ambiental". Con el objetivo de identificar genes que permiten a estos grupos sobrevivir en ambientes secundarios, se secuenció el genoma de un aislamiento del clado críptico IV recuperado de la cuenca en estudio (genoma 339_SF). Este aislamiento es el primero de su linaje en ser reportado en ambientes secundarios de Sudamérica. Además, se analizaron genomas públicos de otros aislamientos del clado críptico IV recuperados tanto del ambiente como de materia fecal (comensalismo). Se evaluó la hipótesis de que los aislamientos del clado críptico IV poseen menos factores de virulencia y resistencia antibiótica que los aislamientos de *E. coli*. Tras secuenciar y analizar el genoma 339_SF, se identificaron 52 factores genéticos asociados a rasgos de virulencia, de los cuales 3 están asociados a perfiles de patogenicidad típicos de *E. coli*, y 2 factores que confieren resistencia antibiótica. Estos factores son compartidos por la mayoría de los genomas de aislamientos del clado críptico IV. A su vez, se evaluó la hipótesis de que los aislamientos del clado críptico IV, pero de distinto origen (ambiental o comensal), tienen un enriquecimiento diferencial de genes que les permiten adaptarse a sus respectivos nichos. Para poner a prueba esta hipótesis, generamos una filogenia con los genomas disponibles basada en SNPs y realizamos un análisis de enriquecimiento diferencial de genes funcionales según el origen. Este análisis permitió identificar numerosos genes que podrían vincularse a una mayor aptitud para habitar ambientes naturales. Consideramos que este análisis es un aporte valioso para la identificación de genes que contribuyen potencialmente en facilitar fenómenos de persistencia ambiental extendida en el género *Escherichia*.

5.2. Materiales y métodos

2.1 Extracción de ADN genómico y secuenciación

La extracción de ADN genómico se practicó sobre un aislamiento ambiental del clado críptico IV obtenido del sitio SF5 del arroyo San Francisco (Figura 3.3 del capítulo 3). La metodología de aislamiento, conservación y caracterización molecular se especificó en los materiales y métodos del capítulo 3 de esta Tesis. La extracción del ADN se inició con la incubación del *pellet* bacteriano en un buffer de extracción integrado por EDTA salino ($C_{10}H_{16}N_2O_8$ 0.01 M y NaCl 0.15 M; pH = 8.0) y las enzimas proteinasa K y ribonucleasa¹⁴⁰. La limpieza del ADN se hizo con lavados secuenciales de cloroformo:alcohol isoamílico (24:1). El ADN se precipitó con isopropanol, y se resuspendió en buffer TE suave (Tris 1 mM y EDTA 0.1 mM a pH = 7.0). Se evaluó la integridad del ADN mediante electroforesis en gel de agarosa 1% (Figura 5.1). Se estimó la masa del material extraído mediante la inclusión de un *ladder* de masa en la corrida electroforética (*MassRuler™ DNA Ladder Mix*). La concentración también se estimó por *Qubit* (*dsDNA Quantitation, broad range*). La secuenciación del genoma completo o WGS (del inglés, *Whole Genome Sequencing*) se realizó en la plataforma Novaseq de Illumina (lecturas paired-end 2 x 150). Las lecturas tuvieron un Q30 de 91,99%, puntaje que indica una secuenciación de excelente calidad, y una profundidad de cobertura estimada de 890X.

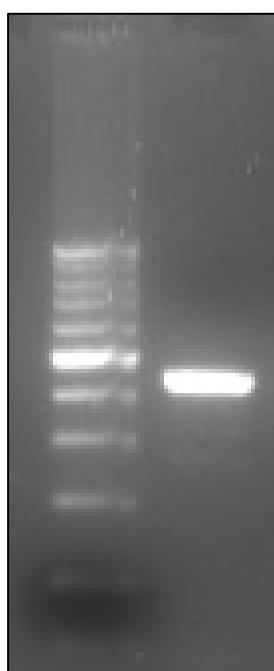


Figura 5.1: Fotografía del gel de agarosa 1%. La calle de la izquierda se sembró con 10 μ l del *MassRuler™ DNA Ladder Mix*. La calle de la derecha se sembró con 10 μ l de la purificación de ADN del genoma 339_SF.

5.2.2 Procesamiento bioinformático

5.2.2.1 Ensamble de novo del genoma, control de calidad y anotación

Tras controlar la calidad de las lecturas mediante la herramienta *FastQC* y evaluar posibles contaminaciones con el algoritmo *Kraken2*, se procedió al ensamblado de las lecturas^{141,142}. Para ello se empleó el algoritmo *Unicycler* disponible en la plataforma de acceso abierto *Galaxy* (Versión 1.0.1). La calidad del ensamblado obtenido se evaluó con la herramienta *Quast* (ver el detalle de este reporte en la Tabla 5.1)¹⁴³⁻¹⁴⁵. La anotación del genoma se realizó con el programa *Prokka*, seleccionando un tamaño mínimo en los *contigs* a anotar de 300 pares de bases y una anotación específica a genes del género *Escherichia*. La búsqueda de factores genómicos de interés se realizó en *ABRicate* (<https://bio.tools/ABRicate>) empleando las bases de datos *VFDB* para factores de virulencia, las bases de datos *NCBI* y *Resfinder* para factores de resistencia antibiótica, y la base de datos *Plasmidfinder* para plásmidos¹⁴⁶⁻¹⁴⁸. Se tomaron aquellos factores con una cobertura de la secuencia mayor al 99% y una identidad mínima del 85%. Entre los factores de virulencia se puso especial atención a la presencia de los genes *stx1*, *stx2*, *eaeH*, *bfp*, *stp*, *st*, *astA*, *lt*, *stgA*, *stgB*, *stgC*, *aggR* y *hlyA* por tratarse de factores de asociados a distintos perfiles de patogenicidad típicos de *E. coli*^{64,149-151}.

5.2.2.2 Pangenoma y filogenia

Se rastrearon genomas ensamblados pertenecientes al clado críptico IV (44 genomas), y a los clados crípticos I (2 genomas), II (1 genoma), III (2 genomas) y V (2 genomas). Además, se usaron genomas de las especies *Escherichia albertii* (1 genoma), *Escherichia fergusonii* (1 genoma) y *Salmonella typhi* (1 genoma). Además, se incorporaron en el análisis el genoma de referencia de *E. coli* (la cepa K-12), y 8 genomas de *E. coli* patogénicas. En total se analizaron 64 genomas (la Tabla A5.1 en el anexo de este capítulo resume la información de acceso a los genomas). Todos los genomas se analizaron con el mismo protocolo de anotación y búsqueda de factores (virulencia/resistencia antibiótica/plásmidos) usados para el aislamiento 339_SF. A partir de los archivos de anotación obtenidos se procedió a la búsqueda del pangenoma y los genes *core* mediante el programa *Roary*¹⁵². Se rastrearon SNPs en los genes *core* con la herramienta *SNP-site*, cuyo resultado fue insumo para el armado de la filogenia (ver resultados en Tabla 5.2). La filogenia se basó en un modelo de sustitución de nucleótidos empleando el método de máxima verosimilitud con un *bootstrap* igual a 1000 mediante el programa *IQ-TREE*¹⁵³.

Por último, el árbol resultante se enraizó en un punto medio (*midpoint*) mediante el software *FigTree*.

5.2.2.3 Enriquecimiento de genes funcionales en clados crípticos IV según origen de aislamiento

Se diseñó una matriz de presencia/ausencia de genes funcionales (base UniProtKB) en los genomas del clado críptico IV (44 genomas), que se empleó para realizar un análisis de agrupación jerárquica. Los genes detectados en un único genoma o en la totalidad de ellos fueron descartados para el análisis de agrupación jerárquica.

Posteriormente, el *set* de genomas fue dividido en dos *subsets* según el ecotipo de origen hospedador/ambiental (Figura 5.2) y se analizó la cantidad de veces que cada gen se detectó en los respectivos grupos (los genes repetidos se contabilizaron como una detección única). Con estos datos se calculó la frecuencia de detección de cada gen funcional en ambos *subsets*. El criterio empleado para asociar a los genes como enriquecidos en un ecotipo se basó en las frecuencias resultantes, considerando para su asociación a uno u otro grupo un umbral mínimo del 80% de detección en uno de los *subsets* y, conjuntamente, un porcentaje menor al 80% de detección en el *subset* complementario. Tras identificar genes asociados a los ecotipos, se realizó una búsqueda bibliográfica con el propósito de asignar rasgos funcionales a los mismos.

5.3. Resultados

5.3.1 Análisis genómico del clado críptico IV hallado en la cuenca de estudio

Tras identificar al aislamiento 339_SF de nuestra colección como perteneciente al clado críptico (CC) IV se procedió a la secuenciación de su genoma. Para nuestro conocimiento, no hay registro previo de detección y secuenciación de un aislamiento de este clado críptico que haya sido recuperado del ambiente natural en el continente Sudamericano.

El ADN se secuenció en un equipo Novaseq (*Illumina*) y se generaron 29.677.758 lecturas (4,5 Gigabases). El contenido de GC en las lecturas fue del 50,86%, lo que es coherente con lo esperado para un genoma del género *Escherichia*. El ensamblado de lecturas produjo un total de 29 *contigs* (Tabla 5.1). En los *contigs* obtenidos se identificaron 4283 regiones codificantes. Además, se identificaron 52 factores de virulencia, de los cuales tres (*astA*, *stgB* y *stgC*) se encuentran asociados a perfiles de virulencia típicos de *E. coli*^{154,155}. Otro factor encontrado es *sat*, que está asociado también a *E. coli* patogénicas (*E. coli* enteroaggregativa). También se encontraron numerosas variantes del gen precursor de fimbrias tipo 1 asociado a *E. coli* uropatogénicas (*fimB*, *fimC*, *fimG*, *fimD*, *fimI*)¹⁵⁶, un factor asociado al secuestro de hierro, *entS*, con presuntos efectos negativos en los huéspedes¹⁵⁷, factores que codifican para componentes de la vía general de secreción de gram negativas (*gspG*, *gspH*, *gspI* y *gspL*)¹⁵⁸, un factor clave en la síntesis de la matriz polimérica extracelular en *E. coli* llamado *KpsD*¹⁵⁹, y los factores *fepA*, *fepB* y *fepD*, y *chuX*, que son factores de virulencia ubicuos en *E. coli*^{160,161}.

Empleando la base de datos *Resfinder* se encontró al gen de resistencia a antibióticos *mdf(A)* (cobertura del 100% e identidad del 91,73%). Este gen es ubicuo en *E. coli* y se sabe que confiere un amplio espectro de resistencias a distintos antibióticos. Empleando la base de datos NCBI se encontró al gen *blaEC-15* (cobertura del 100% e identidad del 90,21%), que está asociado a la resistencia a cefalosporinas. Además, se detectaron 4 plásmidos: *IncHIIb(CIT)_1_pNDM-CIT* (cobertura: 100% e identidad: 93,68), *IncHII_1* (cobertura: 100% e identidad: 92,86%), *Col8282_1* (cobertura: 99,03% e identidad: 89,27%) y *ColRNAI_1* (cobertura: 98,46% e identidad: 80,62%).

Tabla 5.1: Reporte sobre la calidad de ensamble del genoma del *CC IV* secuenciado (339_SF).

	339_SF
Nro. de contigs	29
Contig más largo (pb)	1109346
Largo total (pb)	4637586
GC (%)	50.54
N50	536699
N75	266954
L50	3
L75	6

5.3.2 Relaciones filogenéticas entre el genoma 339_SF y otros miembros del clado criptico IV

Se analizó la ubicación del aislamiento del *CC IV* en un árbol hecho con otros genomas del género *Escherichia*. Se incluyeron en el árbol genomas del *CC IV* de origen ambiental (10 genomas) y de origen comensal humano (14 genomas) y no humano (20 genomas), provenientes de distintas regiones geográficas del mundo (ver *origen* y *región* de genomas en Figura 5.2). Además, se incluyeron genomas de aislamientos de los *CCs I* (2 genomas), *II* (1 genoma), *III* (2 genomas) y *V* (2 genomas) y los genomas de *E. coli* K-12, *E. albertii* y *E. fergusonii*. También se incluyeron genomas de referencia de los principales grupos patogénicos de *E. coli*: EPEC (*E. coli* enteropatogénica), ETEC (*E. coli* enterotoxigénica), EIEC (*E. coli* enteroinvasiva), STEC (*E. coli* productora de la toxina Shiga), EAEC (*E. coli* enteroagregativa), EHEC (*E. coli* enterohemorrágica), UPEC (*E. coli* uropatogénica) y NMEC (*E. coli* asociada a la meningitis neonatal). Como grupo externo se incluyó al genoma de *Salmonella typhi*. Luego de anotar los genes en todos estos genomas, se observó que existen 501 genes *core* (genes que poseen > 99% de prevalencia en los genomas) (Tabla 5.2). Los genes compartidos fueron alineados y comparados para analizar SNPs. Se detectaron 108348 SNPs, que fueron utilizados para construir la filogenia.

Tabla 5.2: Resumen de los tipos de genes según su prevalencia en los genomas anotados del *CC IV* y otros grupos.

Tipos de genes	Porcentaje de prevalencia en los genomas	Cantidad de genes por tipo
Genes <i>core</i>	99% <= genomas <= 100%	501
Genes <i>soft core</i>	95% <= genomas < 99%	1604
Genes <i>shell</i>	15% <= genomas < 95%	2493
Genes <i>cloud</i>	0% <= genomas < 15%	18485
Total de genes	0% <= genomas <= 100 %	23083

La estructura general de la filogenia obtenida es coherente con lo esperado (Figura 5.2). Los *CC IV* se agruparon entre sí conformando un grupo monofilético. Por su parte, los genomas de los *CC II*, *III* y *V* se ubicaron como grupos hermanos al *CC IV*, mientras que el *CC I* se ubicó junto a los genomas agrupados de *E. coli* (*E. coli* K-12 y genomas de grupos patogénicos). Se destaca que el genoma de *E. fergusonii* se ubicó junto a los genomas de *E. coli*, en tanto que *E. albertii* se ubicó en forma externa a todos los genomas antes mencionados. El genoma de *Salmonella typhi*, por su parte, se ubicó externo a todos los pertenecientes al género *Escherichia*. En la filogenia se puede ver que 6 de los genomas más cercanos al del aislamiento 339_SF son también de origen ambiental. Esto indica que podrían existir características genéticas comunes en estos aislamientos. Además, estos genomas presentan bajas cantidades de factores de resistencia antibiótica (2 genes, *mdf(A)_1* y *blaEC-15*) y de virulencia (3 genes, *astA*, *stgB* y *stgC*).

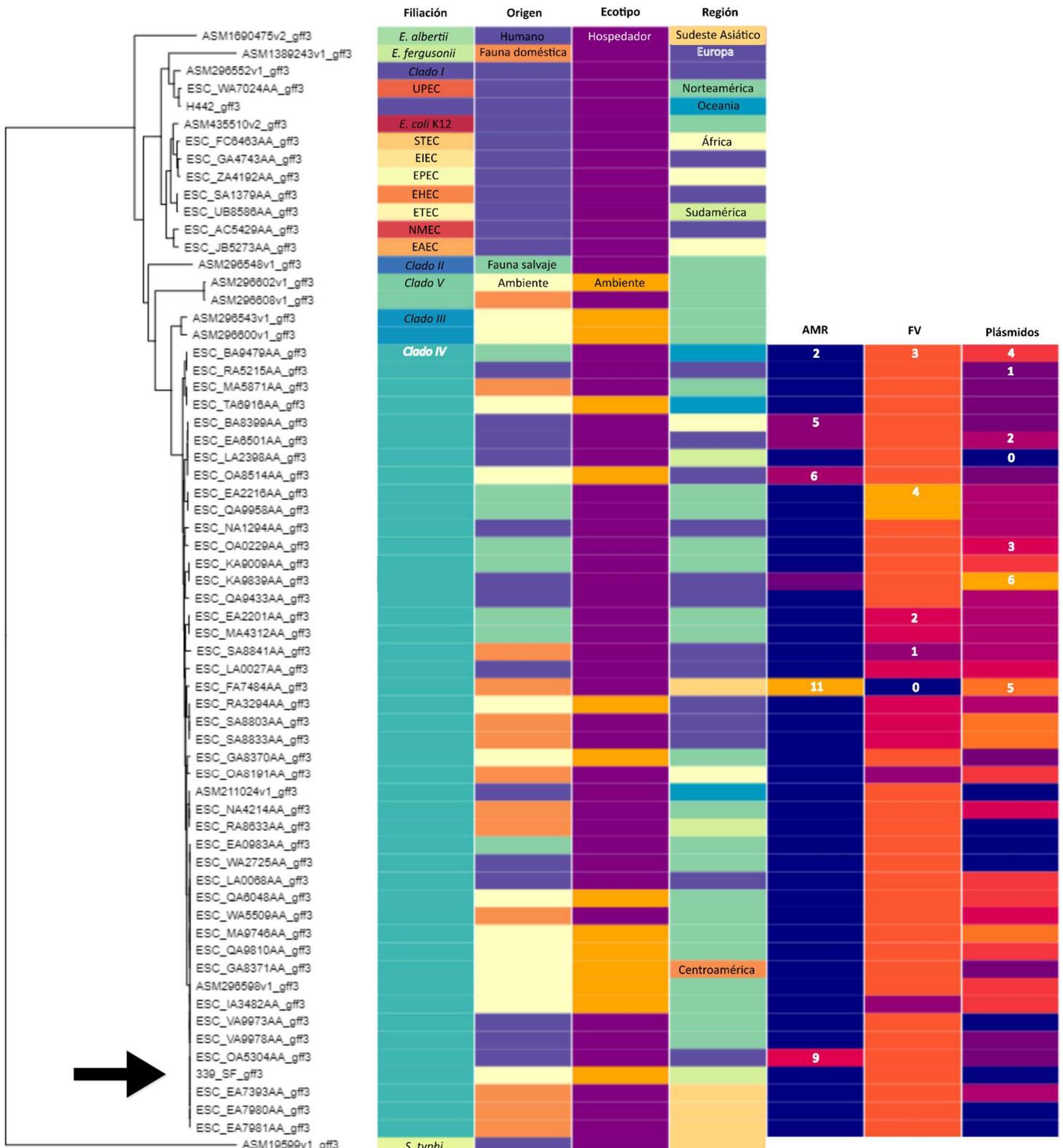


Figura 5.2: Izq) Filogenia construida a partir de SNPs en genes *core* de genomas del clado críptico IV y otros grupos. El aislamiento secuenciado en este trabajo (339_SF) se indica con una flecha. En las primeras columnas se indica el origen, ecotipo y región del taxón en clave colorimétrica. También se muestra la cantidad de factores de resistencia antibiótica (AMR), la cantidad de factores de virulencia asociados a perfiles patogénicos principales de *E. coli* (FV) y la cantidad de plásmidos.

5.3.3 Análisis genómico comparativo de aislamientos del clado criptico IV

Con la anotación génica realizada sobre los genomas disponibles asignados al *CC IV* (44 genomas) se elaboró una matriz de presencia/ausencia de genes, la cual constó inicialmente de 3524 genes con funciones validadas experimentalmente o con función putativa (según la base de datos *UniProtKB*). Tras eliminar genes detectados en un único genoma o presentes en todos los genomas, la matriz se condensó en 870 entradas. A partir de esta matriz se realizó una agrupación jerárquica basada en correlaciones de Pearson positivas entre genomas (Figura 5.3). El dendrograma resultante ordena a los genomas del *CC IV* en dos agrupamientos principales, de 18 y 27 genomas. Este agrupamiento es similar al observado en la filogenia de SNPs de los genes *core* (Figura 5.2). El genoma 339_SF se ubicó en el grupo de 18 genomas y este grupo es el que posee la mayor proporción de aislamientos de origen ambiental (7 de los 18 genomas, versus 4 de los 27). Se observó adicionalmente que 13 de estos 18 genomas presentan una correlación positiva fuerte, incluyendo en este clúster de correlación a 5 genomas ambientales, de los cuales uno es el 339_SF (esquina superior derecha, Figura 5.3).

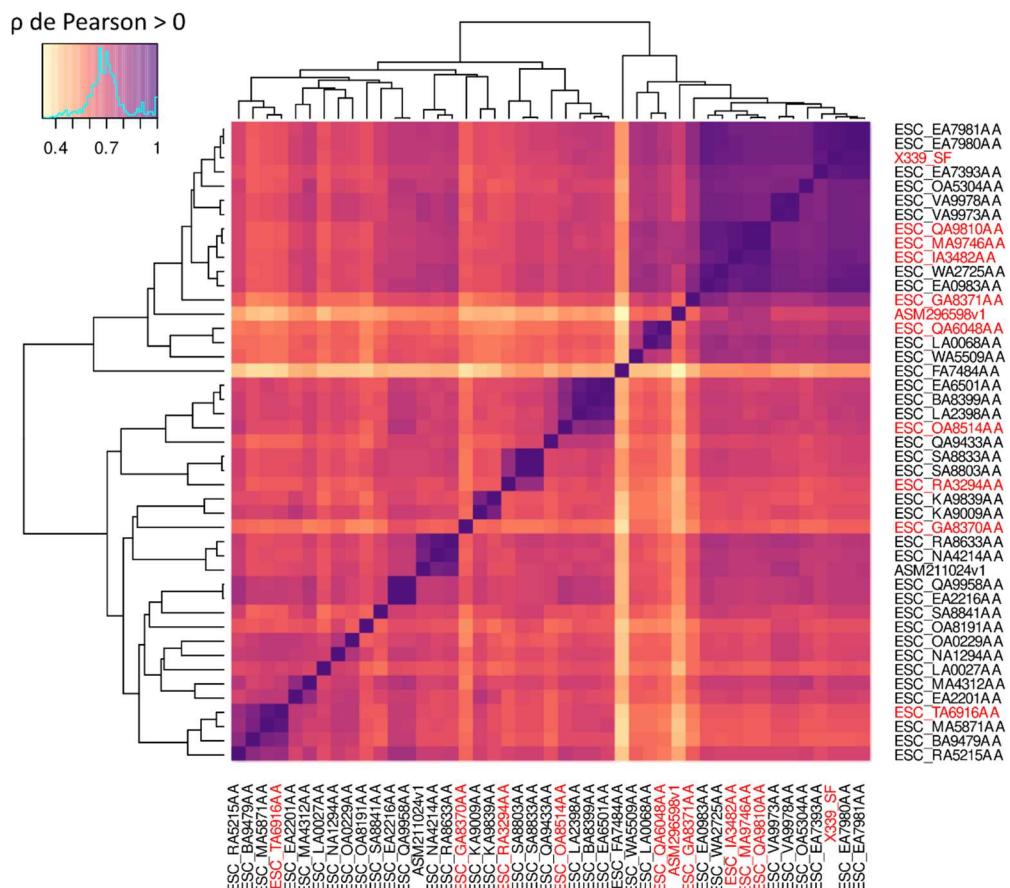


Figura 5.3: Agrupamiento jerárquico de los coeficientes de correlación de Pearson determinados a partir de la presencia/ausencia de genes funcionales en los 44 genomas disponibles del clado críptico IV. Los colores del mapa de calor indican el coeficiente de correlación de Pearson entre los genomas. El ecotipo de origen se indica con el color de la etiqueta: rojo para genomas de origen ambiental y negro para el origen comensal.

En paralelo, se indagó en la búsqueda de genes que pudieran estar más representados o *enriquecidos* en los genomas de *CCs IV* según su procedencia de aislamiento, distinguiendo entre el origen ambiental o ligado al comensalismo entérico de animales.

Se encontraron 24 genes representados con elevada probabilidad (>80%) entre los genomas de origen ambiental, y 3 genes Enriquecidos entre los de origen comensal (Tabla A5.2 del anexo de este capítulo). De los 24 genes Enriquecidos en el origen ambiental, 12 fueron previamente descriptos para *E. coli*, mientras que los 12 restantes fueron identificados en genomas de referencia pertenecientes a otras especies bacterianas. Los genes Enriquecidos en genomas de origen ambiental se pueden vincular a distintas funciones biológicas. Entre los 24 genes, se identificó a un grupo de factores asociados a la respuesta a estrés y a daños en el ADN: *UmuC* y *UmuD* reparan daño al ADN y participan en la respuesta SOS, *yrecE* y *ybcO* son nucleasas que intervienen en la reparación del ADN, *dnaK* tiene actividad de chaperona ante eventos de estrés y *kilR* codifica para un inhibidor de la división celular en respuesta a antibióticos. Otros genes Enriquecidos en genomas ambientales se relacionan con las infecciones virales: *cas6f* y *csy3* pertenecen al sistema de defensa CRISPR, el gen *hpaIIM* codifica para una enzima de restricción de *Haemophilus parainfluenzae* y el gen *intQ* ha sido relacionado con la integración de fagos al genoma huésped. A la vez, se encontró al gen *ehicA* el cual codifica para un componente de un sistema toxina-antitoxina tipo II en *E. coli*. También se identificaron genes vinculados al metabolismo energético bacteriano como *ahr*, cuya proteína es partícipe del metabolismo de lípidos, y *hypBA1*, relacionado con el metabolismo de hidratos de carbono. Otro grupo de genes está asociado a la estructura de membranas y al transporte transmembrana (*yidK*, *csbX*, *ompC*, *yidI* y *yknY*). Además, se encontró al gen *fliC*, que es parte del flagelo bacteriano en *Salmonella dublin*, y al gen *gpFI*, que codifica para una chaperona del fago λ . Además, los genes *bepC*, que confieren resistencia antibiótica a sustancias de naturaleza hidrofóbica (descriptos en *Brucella suis*), e *icsA*, el cual codifica para una proteína esencial para la adhesión y virulencia bacteriana (descripta en *Shigella flexneri*).

Por otro lado, dentro de los 3 genes Enriquecidos en los genomas de procedencia comensal se encontró a *rrrD*, que codifica para un factor con actividad hidrolítica de paredes celulares bacterianas y a *tolA*, que codifica para una proteína del sistema Tol-Pal (este gen es un factor importante para la división celular y para la integridad de la membrana externa). Estos dos genes fueron descriptos en *E. coli*. El tercer gen, llamado *prfA*, fue

descripto para *Mycobacterium tuberculosis* y codifica para una proteína intermediaria en la biosíntesis de proteínas.

5.4. Anexo

Tabla A5.1: Información de acceso y generalidades de los genomas empleados en este análisis.

ID	NCBI BioSample ID	Bio Project ID	Filiación	Origen	Ecotipo	Región	País
ASM296562v1	CI_SAMN05597624	PRJNA230969	Clado I	Humano	Comensal	Europa	Dinamarca
H442	CI_SAMEA6968677	PRJEB38885	Clado I	Humano	Comensal	Oceanía	Australia
ASM296548v1	CII_SAMN08395604	PRJNA230969	Clado II	Fauna salvaje	Comensal	Norteamérica	EE. UU.
ASM296543v1	CIII_SAMN05440288	PRJNA230969	Clado III	Ambiente	Ambiente	Norteamérica	EE.UU.
ASM296600v1	CIII_SAMN08395599	PRJNA230969	Clado III	Ambiente	Ambiente	Norteamérica	EE.UU.
ASM211024v1	CIV_SAMN06392855	PRJNA39025	Clado IV	Humano	Comensal	Oceanía	Australia
ASM296598v1	CIV_SAMN08395601	PRJNA230969	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC BA8399AA	CIV_SAMN02687529	PRJNA233818	Clado IV	Humano	Comensal	África	Tanzania
ESC BA9479AA	CIV_SAMEA1888904	PRJEB2879	Clado IV	Fauna salvaje	Comensal	Oceanía	Australia
ESC EA0983AA	CIV_SAMN04992183	PRJNA230969	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC EA2201AA	CIV_SAMN04992208	PRJNA230969	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC EA2216AA	CIV_SAMN04992377	PRJNA230969	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC EA6501AA	CIV_SAMEA3918377	PRJEB12888	Clado IV	Humano	Comensal	Europa	Reino Unido
ESC EA7393AA	CIV_SAMEA3981847	PRJEB12887	Clado IV	Fauna doméstica	Comensal	SE Asiático	Vietnam
ESC EA7980AA	CIV_SAMEA3980610	PRJEB12887	Clado IV	Fauna doméstica	Comensal	SE Asiático	Vietnam
ESC EA7981AA	CIV_SAMEA3980596	PRJEB12887	Clado IV	Fauna doméstica	Comensal	SE Asiático	Vietnam
ESC FA7484AA	CIV_SAMEA4061316	PRJEB12887	Clado IV	Fauna doméstica	Comensal	SE Asiático	Vietnam
ESC GA8370AA	CIV_SAMN02469845	PRJNA224116	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC GA8371AA	CIV_SAMN02469833	PRJNA59765	Clado IV	Ambiente	Ambiente	Centroamérica	Puerto Rico
ESC IA3482AA	CIV_SAMN08395601	PRJNA230969	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC KA9009AA	CIV_SAMN10485904	PRJNA357722	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC KA9839AA	CIV_SAMEA104370298	PRJEB23294	Clado IV	Humano	Comensal	Europa	NA
ESC LA0027AA	CIV_SAMEA104370058	PRJEB23294	Clado IV	Humano	Comensal	Europa	NA
ESC LA0068AA	CIV_SAMEA104370000	PRJEB23294	Clado IV	Humano	Comensal	Europa	NA
ESC LA2398AA	CIV_SAMN09832945	PRJNA486009	Clado IV	Humano	Comensal	Sudamérica	Ecuador
ESC MA4312AA	CIV_SAMN11478079	PRJNA230969	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC MA5871AA	CIV_SAMN11640643	PRJNA324565	Clado IV	Fauna doméstica	Comensal	Norteamérica	EE.UU.
ESC MA9746AA	CIV_SAMN11918111	PRJNA230969	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC NA1294AA	CIV_SAMEA4607767	PRJEB23294	Clado IV	Humano	Comensal	Europa	NA
ESC NA4214AA	CIV_SAMN12221183	PRJNA292667	Clado IV	Fauna doméstica	Comensal	Norteamérica	EE.UU.
ESC OA0229AA	CIV_SAMN12722938	PRJNA230969	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC OA5304AA	CIV_SAMEA5953375	PRJEB34275	Clado IV	Humano	Comensal	Europa	Países bajos
ESC OA8191AA	CIV_SAMN13262684	PRJNA293225	Clado IV	Fauna doméstica	Comensal	África	Nigeria
ESC OA8514AA	CIV_SAMN12429450	PRJNA558173	Clado IV	Ambiente	Ambiente	Europa	Suiza
ESC QA6048AA	CIV_SAMN14140199	PRJNA283914	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC QA9433AA	CIV_SAMEA7040573	PRJEB39252	Clado IV	Humano	Comensal	Europa	Francia
ESC QA9810AA	CIV_SAMN14503430	PRJNA357722	Clado IV	Ambiente	Ambiente	Norteamérica	EE.UU.
ESC QA9958AA	CIV_SAMN14503470	PRJNA357722	Clado IV	Fauna salvaje	Comensal	Norteamérica	EE.UU.
ESC RA3294AA	CIV_SAMN11124973	PRJNA526980	Clado IV	Ambiente	Ambiente	Europa	Alemania
ESC RA5215AA	CIV_SAMN15052611	PRJNA592140	Clado IV	Humano	Comensal	Europa	Países bajos
ESC RA8633AA	CIV_SAMN15515979	PRJNA633873	Clado IV	Fauna doméstica	Comensal	Sudamérica	Perú
ESC SA8803AA	CIV_SAMN20959669	PRJNA757375	Clado IV	Fauna doméstica	Comensal	Europa	Reino Unido
ESC SA8833AA	CIV_SAMN20959706	PRJNA757375	Clado IV	Fauna doméstica	Comensal	Europa	Reino Unido

ESC SA8841AA	CIV SAMN20959692	PRJNA757375	Clado IV	Fauna doméstica	Comensal	Europa	Reino Unido
ESC TA6916AA	CIV SAMN12996506	PRJNA576546	Clado IV	Ambiente	Ambiente	Oceanía	Nueva Zelanda
ESC VA9973AA	CIV SAMN20065249	PRJNA230968	Clado IV	Humano	Comensal	Norteamérica	EE.UU.
ESC VA9978AA	CIV SAMN20065245	PRJNA230968	Clado IV	Humano	Comensal	Norteamérica	EE.UU.
ESC WA2725AA	CIV SAMN20600830	PRJNA218110	Clado IV	Humano	Comensal	Norteamérica	EE.UU.
ESC WA5509AA	CIV SAMN21217652	PRJNA510384	Clado IV	Fauna doméstica	Comensal	Norteamérica	EE.UU.
ASM296602v1	CV SAMN08395600	PRJNA230969	Clado V	Ambiente	Ambiente	Norteamérica	EE.UU.
ASM296608v1	CV SAMN04279488	PRJNA230969	Clado V	Fauna doméstica	Comensal	Norteamérica	EE.UU.
ASM435510v2	Ecoli K12 BW25113	PRJNA525908	E. coli K12	Laboratorio	Domestic	Norteamérica	EE.UU.
ESC JB5273AA	Ecoli SAMEA7160185	PRJEB39739	EAEC	Humano	Comensal	África	Sudáfrica
ESC SA1379AA	Ecoli SAMN09207722	PRJNA248792	EHEC	Humano	Comensal	Europa	Reino Unido
ESC GA4743AA	Ecoli SAMN03612246	PRJNA243331	EIEC	Humano	Comensal	Europa	Italia
ESC ZA4192AA	Ecoli SAMEA112667364	PRJEB39739	EPEC	Humano	Comensal	África	Sudáfrica
ASM1690475v2	Ealber SAMN17525958	PRJNA693666	Escherichia albertii	Humano	Comensal	SE asiático	China
ASM1389243v1	Efergu SAMN15148162	PRJNA605147	Escherichia fergusonii	Fauna doméstica	Comensal	Europa	Reino Unido
ESC UB8586AA	Ecoli 303191	NA	ETEC	Humano	Comensal	Sudamérica	Paraguay
ESC AC5429AA	Ecoli KMB-1032	NA	NMEC	Humano	Comensal	Europa	Eslovenia
ASM19599v1	Styphi SAMEA1705914	PRJNA236	Salmonella typhi	Humano	Comensal	SE asiático	Vietnam
ESC FC6463AA	Ecoli SAMEA112667359	PRJEB39739	STEC	Humano	Comensal	África	Sudáfrica
ESC WA7024AA	Ecoli SAMN09405878	PRJNA475142	UPEC	Humano	Comensal	Norteamérica	EE.UU.

Tabla A5.2: Listado de genes detectados en mayor abundancia relativa según su procedencia de aislamiento o ecotipo. Los genes asociados al origen ambiental se colorean en verde, mientras que en rojo aquellos que se asociaron al comensalismo. Se incluye el nombre de la proteína que codifica, organismo en el cual se describió y su función biológica atribuida basada en el registro de la base de datos UniProtKB.

UniProtKB ID	% origen ambiente	% origen comensal	Gen	Proteína	Organismo	Función biológica UniProtK
P0AG11	100.00%	76.47%	<i>umuD</i>	Proteína UmuD	<i>Escherichia coli</i> (cepa K12)	Daño en el ADN; reparación de ADN; mutagénesis SOS; respuesta SOS ¹⁶²
P15446	100.00%	76.47%	<i>hpaII</i>	Metiltransferasa Tipo II M.Hpall	<i>Haemophilus parainfluenzae</i>	Sistema de restricción ¹⁶³
P76168	100.00%	70.59%	<i>intQ</i>	Proteína defectuosa putativa IntQ	<i>Escherichia coli</i> (cepa K12)	Integración de ADN; recombinación de ADN; integración del genoma viral; entrada del virus en la célula huésped
Q8G0Y6	100.00%	79.41%	<i>bepC</i>	Proteína de salida de la membrana externa BepC	<i>Brucella suis</i> biovar 1 (cepa 1330)	Resistencia antibiótica; Transporte ¹⁶⁴
P76106	90.91%	67.65%	<i>hicA</i>	Probable mRNA interferasa HicA	<i>Escherichia coli</i> (cepa K12)	Respuesta al estrés; sistema toxina-antitoxina ¹⁶⁵

Q2YIQ4	90.91%	67.65%	<i>eryD</i>	Proteína reguladora del catabolismo del eritritol EryD	<i>Brucella abortus</i> (cepa 2308)	Regulación de transcripción ¹⁶⁶
P15032	90.91%	44.12%	<i>recE</i>	Exodesoxiribonucleasa 8	<i>Escherichia coli</i> (cepa K12)	Exonucleasa; hidrolasa; nucleasa
P04152	90.91%	73.53%	<i>umuC</i>	Proteína UmuC	<i>Escherichia coli</i> (cepa K12)	Daño en el ADN; reparación de ADN; mutagénesis SOS; respuesta SOS ¹⁶²
Q9I747	90.91%	76.47%	<i>hcp1</i>	Proteína hcp1	<i>Pseudomonas aeruginosa</i> (cepa ATCC 15692)	Región extracelular; Unión entre proteínas idénticas
Q7BCK4	90.91%	76.47%	<i>icsA</i>	Autotransportador de membrana exterior IcsA	<i>Shigella flexneri</i>	Adhesión celular; Virulencia ¹⁶⁷
P31448	90.91%	76.47%	<i>yidK</i>	Symporter no caracterizado YidK	<i>Escherichia coli</i> (cepa K12)	Transporte de sodio; simporte
O05390	90.91%	67.65%	<i>csbX</i>	Permeasa de alfa-cetoglutarato	<i>Bacillus subtilis</i> (cepa 168)	Transporte
P68661	90.91%	70.59%	<i>ybcO</i>	Nucleasa putativa YbcO	<i>Escherichia coli</i> (cepa K12)	Hidrolasa; nucleasa ¹⁶⁸
P31446	81.82%	70.59%	<i>yidI</i>	Proteína de membrana interna Yidl	<i>Escherichia coli</i> (cepa K12)	Membrana interna celular; Proteína de membrana multipaso
Q02TE1	81.82%	73.53%	<i>gpFI</i>	Supuesta proteína de la vaina de la cola principal del profago	<i>Pseudomonas aeruginosa</i> (cepa UCBPP-PA14)	Supuesta proteína de la vaina de la cola principal del profago ¹⁶⁹
Q8CVW1	81.82%	47.06%	<i>ompC</i>	Porina C de la membrana externa	<i>Escherichia coli</i> O6:H1 (cepa ATCC 700928)	Transporte de iones ¹⁷⁰
P27250	81.82%	61.76%	<i>ahr</i>	Aldehído reductasa Ahr	<i>Escherichia coli</i> (cepa K12)	Metabolismo de ácidos grasos; Metabolismo de lípidos ¹⁷¹

B7J282	81.82%	70.59%	<i>dnaK</i>	Proteína chaperona DnaK	<i>Borrelia burgdorferi</i> (cepa ZS7)	Respuesta al estrés
Q06971	81.82%	58.82%	<i>fliC</i>	Flagelina	<i>Salmonella dublin</i>	Componente celular; flagelo bacteriano
Q1RE32	81.82%	70.59%	<i>cas6f</i>	Endonucleasa asociada a CRISPR Cas6/Csy4	<i>Escherichia coli</i> (cepa UTI89)	Defensa antiviral ¹⁷²
O31711	81.82%	47.06%	<i>yknY</i>	Proteína ATPasa parte de un transportador ABC no caracterizado	<i>Bacillus subtilis</i> (cepa 168)	Transporte transmembrana
E8MGH8	81.82%	73.53%	<i>hypBA1</i>	Beta-L-arabinofurano-sidasa terminal no reductora	<i>Bifidobacterium longum</i> subsp. <i>longum</i> (cepa ATCC 15707)	Metabolismo de los carbohidratos; Degradación de polisacáridos ¹⁷³
P38393	81.82%	79.41%	<i>kilR</i>	Killing protein KilR	<i>Escherichia coli</i> (cepa K12)	Ciclo celular; regulación negativa del ciclo celular; respuesta al antibiótico ¹⁷⁴
Q6D0W6	81.82%	67.65%	<i>csy3</i>	Proteína asociada a CRISPR CsY3	<i>Pectobacterium atrosepticum</i> (cepa ATCC BAA-672)	Defensa antiviral ¹⁷⁵
P78285	72.73%	94.12%	<i>rrrD</i>	Lisozima RrrD	<i>Escherichia coli</i> (cepa K12)	Antimicrobiano; enzima bacteriolítica; glucosidasa; hidrolasa ¹⁷⁶
P9WHG3	72.73%	91.18%	<i>prfA</i>	Factor de liberación de cadena peptídica 1	<i>Mycobacterium tuberculosis</i> (cepa ATCC 25618)	Biosíntesis de proteínas
P19934	72.73%	82.35%	<i>tolA</i>	Proteína del sistema Tol-Pal TolA	<i>Escherichia coli</i> (cepa K12)	Ciclo celular; División celular ¹⁷⁷

**Discusión general fue retirada a pedido del
autor**

**Conclusiones generales fue retirado a pedido
del autor**

Bibliografía

Bibliografía

1. Elmqvist, T. *The urban planet: Knowledge towards sustainable cities*. (Cambridge University Press, 2018).
2. Alberti, M. *Cities that think like planets: complexity, resilience, and innovation in hybrid ecosystems*. (University of Washington Press, 2016).
3. Paul, M. J. & Meyer, J. L. Streams in the urban landscape. *Annu. Rev. Ecol. Syst.* 32(1), 333–365 (2001).
4. Walsh, C. J. *et al.* The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society*, 24(3), 706-723 (2005).
5. Pennino, M. J., Kaushal, S. S., Beaulieu, J. J., Mayer, P. M. & Arango, C. P. Effects of urban stream burial on nitrogen uptake and ecosystem metabolism: implications for watershed nitrogen and carbon fluxes. *Biogeochemistry* 121, 247–269 (2014).
6. Kaushal, S. S. *et al.* Freshwater salinization syndrome on a continental scale. *Proc. Natl. Acad. Sci. U. S. A.* 115, E574–E583 (2018).
7. Mainali, J. & Chang, H. Landscape and anthropogenic factors affecting spatial patterns of water quality trends in a large river basin, South Korea. *J. Hydrol.* 564, 26–40 (2018).
8. Booth, D. B., Roy, A. H., Smith, B. & Capps, K. A. Global perspectives on the urban stream syndrome. *Freshw. Sci.* 35, 412–420 (2016).
9. Alberti, M. *et al.* The Complexity of Urban Eco-evolutionary Dynamics. *Bioscience* 1–22 (2020).
10. Bernhardt, E. S., Band, L. E., Walsh, C. J. & Berke, P. E. Understanding, managing, and minimizing urban impacts on surface water nitrogen loading. *Ann. N. Y. Acad. Sci.* 1134, 61–96 (2008).
11. Hosen, J. D., McDonough, O. T., Febria, C. M. & Palmer, M. A. Dissolved organic matter quality and bioavailability changes across an urbanization gradient in headwater streams. *Environ. Sci. Technol.* 48, 7817–7824 (2014).
12. Hatt, B. E., Fletcher, T. D., Walsh, C. J. & Taylor, S. L. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. *Environ. Manage.* 34, 112–124 (2004).
13. Smith, R. M., Kaushal, S. S., Beaulieu, J. J., Pennino, M. J. & Welty, C. Influence of infrastructure on water quality and greenhouse gas dynamics in urban streams.

Biogeosciences 14, 2831–2849 (2017).

- 14. Grimm, N. B. *et al.* The changing landscape: Ecosystem responses to urbanization and pollution across climatic and societal gradients. *Front. Ecol. Environ.* 6, 264–272 (2008).
- 15. Violin, C. R. *et al.* Effects of urbanization and urban stream restoration on the physical and biological structure of stream ecosystems. *Ecol. Appl.* 21, 1932–1949 (2011).
- 16. Peipoch, M., Brauns, M., Hauer, F. R., Weitere, M. & Valett, H. M. Ecological Simplification: Human Influences on Riverscape Complexity. *Bioscience* 65, 1057–1065 (2015).
- 17. Levi, P. S. *et al.* Macrophyte Complexity Controls Nutrient Uptake in Lowland Streams. *Ecosystems* 18, 914-931 (2015).
- 18. Allan, J. D., Castillo, M. M., & Capps, K. A. *Stream ecology: structure and function of running waters.* (Springer Nature, 2021).
- 19. Kalloul, S. *et al.* Source contributions to heavy metal fluxes into the Loukous estuary (Moroccan Atlantic Coast). *J. Coast. Res.* 28, 174–183 (2012).
- 20. Hwang, H. M., Fiala, M. J., Park, D. & Wade, T. L. Review of pollutants in urban road dust and stormwater runoff: part 1. Heavy metals released from vehicles. *Int. J. Urban Sci.* 20, 334–360 (2016).
- 21. Kot-Wasik, A., Dąbrowska, D. & Namieśnik, J. The importance of degradation in the fate of selected organic compounds in the environment. Part I. General considerations. *Polish J. Environ. Stud.* 13, 607–616 (2004).
- 22. Tripodi, M., Cueto, G. & Suárez, O. Evaluation of changes in spatial pattern of metals pollution in surface water using multivariate analysis: case study an urban basin from Buenos Aires, Argentina. *Urban Water J.* 18, 725–732 (2021).
- 23. Isaak, D. J. *et al.* Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews: Water*, 1(3), 277-294 (2014).
- 24. Bernhardt, E. S. *et al.* The metabolic regimes of flowing waters. *Limnol. Oceanogr.* 63, S99–S118 (2018).
- 25. Graham, E. B. *et al.* Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? *Front. Microbiol.* 7, 214 (2016).
- 26. Bardgett, R. D. & Van Der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nat.* 515, 505–511 (2014).

27. Nemergut, D. R., Shade, A., Violle, C., Stein, L. Y. & Nie, M. When, where and how does microbial community composition matter? *Frontiers in microbiology*, 5, 497. (2014).
28. Wang, P. *et al.* Shift in bacterioplankton diversity and structure: Influence of anthropogenic disturbances along the Yarlung Tsangpo River on the Tibetan Plateau, China. *Sci. Rep.* 7, 1–11 (2017).
29. McLellan, S. L., Fisher, J. C. & Newton, R. J. The microbiome of urban waters. *Int. Microbiol.* 18, 141–149 (2015).
30. Richards, S., Paterson, E., Withers, P. J. A. & Stutter, M. Septic tank discharges as multi-pollutant hotspots in catchments. *Sci. Total Environ.* 542, 854–863 (2016).
31. Sowah, R. A., Habteselassie, M. Y., Radcliffe, D. E., Bauske, E. & Risse, M. Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia, United States. *Water Res.* 108, 330–338 (2017).
32. Ly, D. K. & Chui, T. F. M. Modeling sewage leakage to surrounding groundwater and stormwater drains. *Water Sci. Technol.* 66, 2659–2665 (2012).
33. Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* 38, 1–40 (2014).
34. Newton, R. J. & McLellan, S. L. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. *Front. Microbiol.* 6, 1–13 (2015).
35. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4516–4522 (2011).
36. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, 1–11 (2013).
37. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egoscue, J. J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 1–6 (2017).
38. Jang, J. *et al.* Environmental Escherichia coli: ecology and public health implications—a review. *J. Appl. Microbiol.* 123, 570–581 (2017).
39. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic Escherichia coli. *Nat. Rev. Microbiol.* 19, 37–54 (2021).

40. Rochelle-Newall, E., Nguyen, T. M. H., Le, T. P. Q., Sengtaheuanghoun, O. & Ribolzi, O. A short review of fecal indicator bacteria in tropical aquatic ecosystems: Knowledge gaps and future directions. *Front. Microbiol.* 6, 1–15 (2015).
41. Mackowiak, M., Leifels, M., Hamza, I. A., Jurzik, L. & Wingender, J. Distribution of *Escherichia coli*, coliphages and enteric viruses in water, epilithic biofilms and sediments of an urban river in Germany. *Sci. Total Environ.* 626, 650–659 (2018).
42. Lee, C. M. *et al.* Persistence of fecal indicator bacteria in Santa Monica Bay beach sediments. *Water Res.* 40, 2593–2602 (2006).
43. Ishii, S., Ksoll, W. B., Hicks, R. E. & Sadowsky, M. J. Presence and growth of naturalized *Escherichia coli* in temperate soils from lake superior watersheds. *Appl. Environ. Microbiol.* 72, 612–621 (2006).
44. Surbeck, C. Q., Jiang, S. C. & Grant, S. B. Ecological control of fecal indicator bacteria in an urban stream. *Environ. Sci. Technol.* 44, 631–637 (2010).
45. Jaureguy, F. *et al.* Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9, 560 (2008).
46. Clermont, O., Gordon, D. & Denamur, E. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 161, 980–988 (2015).
47. Clermont, O. *et al.* Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. *Environ. Microbiol.* 21, 3107–3117 (2019).
48. Ratajczak, M. *et al.* Influence of hydrological conditions on the *Escherichia coli* population structure in the water of a creek on a rural watershed. *BMC Microbiol.* 10, (2010).
49. Johnson, J. R. *et al.* Phylogenetic backgrounds and virulenceassociated traits of *Escherichia coli* isolates from surface waters and diverse animals in Minnesota and Wisconsin. *Appl. Environ. Microbiol.* 83, 1–33 (2017).
50. Petit, F. *et al.* Change in the structure of *Escherichia coli* population and the pattern of virulence genes along a rural aquatic continuum. *Front. Microbiol.* 8, 1–14 (2017).
51. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29, 592–599 (2015).
52. Méric, G., Kemsley, E. K., Falush, D., Saggers, E. J. & Lucchini, S. Phylogenetic

distribution of traits associated with plant colonization in *Escherichia coli*. *Environ. Microbiol.* 15, 487–501 (2013).

- 53. Berthe, T., Ratajczak, M., Clermont, O., Denamur, E. & Petit, F. Evidence for coexistence of distinct *escherichia coli* populations in various aquatic environments and their survival in estuary water. *Appl. Environ. Microbiol.* 79, 4684–4693 (2013).
- 54. Elsas, J. D. Van, Semenov, A. V, Costa, R. & Trevors, J. T. Survival of *Escherichia coli* in the environment : fundamental and public health aspects. *ISME J.* 5, 173–183 (2010).
- 55. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5, (2009).
- 56. Touchon, M. *et al.* Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLOS Genet.* 16, e1008866 (2020).
- 57. Ingle, D. J. *et al.* Biofilm formation by and thermal niche and virulence characteristics of *Escherichia* spp. *Appl. Environ. Microbiol.* 77, 2695–2700 (2011).
- 58. Sharma, G. *et al.* *Escherichia coli* biofilm: development and therapeutic strategies. *J. Appl. Microbiol.* 121, 309–319 (2016).
- 59. Murphy, H. M., Payne, S. J. & Gagnon, G. A. Sequential UV- and chlorine-based disinfection to mitigate *Escherichia coli* in drinking water biofilms. *Water Res.* 42, 2083–2092 (2008).
- 60. Holden, N., Pritchard, L. & Toth, I. Colonization outwith the colon: Plants as an alternative environmental reservoir for human pathogenic enterobacteria: Review article. *FEMS Microbiol. Rev.* 33, 689–703 (2009).
- 61. Walk, S. T. The ‘Cryptic’ *Escherichia*. *EcoSal Plus* 6, (2015).
- 62. Steinsland, H., Lacher, D. W., Sommerfelt, H. & Whittam, T. S. Ancestral lineages of human enterotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* 48, 2916–2924 (2010).
- 63. Stecher, B. & Hardt, W. D. Mechanisms controlling pathogen colonization of the gut. *Curr. Opin. Microbiol.* 14, 82–91 (2011).
- 64. Acosta-Dibarrat, J. *et al.* Characterization of commensal *Escherichia coli* isolates from slaughtered sheep in Mexico. *J. Infect. Dev. Ctries.* 15, 1755–1760 (2021).
- 65. Becattini, S., Taur, Y. & Pamer, E. G. Antibiotic-Induced Changes in the Intestinal Microbiota and Disease. *Trends Mol. Med.* 22, 458–478 (2016).

66. Jernberg, C., Löfmark, S., Edlund, C. & Jansson, J. K. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* 156, 3216–3223 (2010).
67. Ingle, D. J. *et al.* Biofilm formation by and thermal niche and virulence characteristics of *Escherichia* spp. *Appl. Environ. Microbiol.* 77, 2695–2700 (2011).
68. Vignaroli, C. *et al.* Adhesion of marine cryptic *Escherichia* isolates to human intestinal epithelial cells. *ISME J.* 9, 508–515 (2015).
69. Luo, C. *et al.* Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc. Natl. Acad. Sci.* 108, 7200–7205 (2011).
70. Clermont, O., Gordon, D. M., Brisse, S., Walk, S. T. & Denamur, E. Characterization of the cryptic *Escherichia* lineages: Rapid identification and prevalence. *Environ. Microbiol.* 13, 2468–2477 (2011).
71. Censo nacional de población, hogares y viviendas 2022 : resultados provisionales. Instituto Nacional de Estadística y Censos (INDEC), 2023).
72. Cirelli, A. F. & Ojeda, C. Wastewater management in Greater Buenos Aires, Argentina. *Desalination* 218, 52–61 (2008).
73. Graziano, M., Giorgi, A. & Feijoó, C. Multiple stressors and social-ecological traps in Pampean streams (Argentina): A conceptual model. *Sci. Total Environ.* 142785 (2020).
74. Elordi, M. L., Lerner, J. E. C. & Porta, A. Evaluacion del impacto antropico sobre la calidad del agua del arroyo Las Piedras, Quilmes, Buenos Aires, Argentina. *Acta Bioquim. Clin. Latinoam.* 50, 669–677 (2016).
75. Capítulo, A. R., Gómez, N., Giorgi, A. & Feijoó, C. Global changes in pampean lowland streams (Argentina): Implications for biodiversity and functioning. *Hydrobiologia* 657, 53–70 (2010).
76. Aliaga, V. S., Ferrelli, F., Alberdi-Algarañaz, E. D., Bohn, V. Y., & Piccolo, M. C. Distribution and variability of precipitation in the Pampean Region, Argentina. *Cuad. Investig. Geográfica* 42(1), 261–280 (2016).
77. Cabrera, A. Regiones fitogeográficas Argentinas. *Encycl. argentina Agric. y Jard.* 2, 1–85 (1976).
78. Gatti, I. A. *et al.* Anticipating the Flood. Community-based cartography for disaster flood events in Argentina. *Proc. ICA* 2, 1–8 (2019).

79. Re, M., Kazimierski, L. D., Garcia, P. E., Ortiz, N. E. & Lagos, M. Assessment of crowdsourced social media data and numerical modelling as complementary tools for urban flood mitigation. *Hydrological Sciences Journal* 67, 1295–1308 (2022).

80. Gomez, B. M., Diaz, N. S., Saraceno, M., Sinistro, R. & Graziano, M. Co-production of just nature-based solutions to mitigate the impact of domestic effluents on stream water quality in an informal urbanization in Latin America: Diagnosis and a pilot study co-production of just nature-based solutions. En proceso de publicación.

81. Efron, S. T., Aquino, J., de Cabo, L., dos Santos Afonso, M. & Graziano, M. Evaluación de la capacidad de auto-depuración de un arroyo urbano y el uso de macrófitas nativas como estrategia de restauración. *Biol. Acuática* 30, 275–285 (2014).

82. Taller de Aguas. Facultad de Ciencias Exactas y Naturales; Universidad de Buenos Aires. Determinación de la calidad del agua del Arroyo San Francisco. (2010).

83. Instituto Nacional de Estadística y Censos. *Censo Nacional de Población, Hogares y Viviendas 2010: Censo del Bicentenario. 1a Ed.* (2012).

84. Gordon, N. D., McMahon, T. A., Finlayson, B. L., Gippel, C. J., & Nathan, R. J. Stream hydrology: an introduction for ecologists. *John Wiley and Sons.* (2004).

85. LLC, G. Google Earth Pro. (2022).

86. Congedo Pg, L. Adapting to Climate Change in Coastal Dar es Salaam Project Ref. EC Grant Contract No 2010/254-773 Semi-Automatic Classification Plugin for QGIS. (2013).

87. Marker AFH, Nusch A, R. H. y R. B. The measurement of photosynthetic pigments in freshwaters and standardization of methods: conclusions and recommendations. *Arch. für Hydrobiol. Beihefte. Ergebnisse der Limnol.* 14, 91–106 (1980).

88. Wilkinson, L. & Friendly, M. History corner the history of the cluster heat map. *Am. Stat.* 63, 179–184 (2009).

89. Wei, T. *et al.* Visualization of a Correlation Matrix. *Statistician* 56, 316–324 (2017).

90. EPA. Recreational Water Quality Criteria. *U. S. Environ. Prot. Agency* 1–69 (2012).

91. Belforte, F. S. *et al.* Getting to know the gut microbial diversity of metropolitan buenos aires inhabitants. *Front. Microbiol.* 10, 1–10 (2019).

92. Herlemann, D. P. R. *et al.* Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579 (2011).

93. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624 (2012).

94. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583 (2016).

95. Beule, L. & Karlovsky, P. Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. *PeerJ* 8, e9593 (2020).

96. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, (2013).

97. H Pagès, P Aboyoun, R Gentleman, S. D. Biostrings: Efficient manipulation of biological strings. *Bioconductor* (2019).

98. Oksanen, J. Vegan: ecological diversity. *R Packag. Version 2.4-4* 1, 11 (2017).

99. Shade, A. & Stopnisek, N. Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr. Opin. Microbiol.* 49, 50–58 (2019).

100. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* 451, 990–993 (2008).

101. Cui, Q., Fang, T., Huang, Y., Dong, P. & Wang, H. Evaluation of bacterial pathogen diversity, abundance and health risks in urban recreational water by amplicon next-generation sequencing and quantitative PCR. *J. Environ. Sci. (China)* 57, 137–149 (2017).

102. Nshimyimana, J. P., Freedman, A. J. E., Shanahan, P., Chua, L. C. H. & Thompson, J. R. Variation of Bacterial Communities with Water Quality in an Urban Tropical Catchment. *Environ. Sci. Technol.* 51, 5591–5601 (2017).

103. NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules. *NIH Guidel.* 2, 1–7 (2019).

104. Shakir, S. M. *et al.* Clinical laboratory perspective on streptococcus halichoeri, an unusual nonhemolytic, lancefield group b streptococcus causing human infections. *Emerg. Infect. Dis.* 27, 1309–1316 (2021).

105. Bereket, W. *et al.* Update on bacterial nosocomial infections. *Eur. Rev. Med. Pharmacol. Sci.* 16, 1039–1044 (2020).

106. Cheslock, M. A. & Embers, M. E. Human bartonellosis: An underappreciated

public health problem? *Trop. Med. Infect. Dis.* 4, (2019).

107. Magurran, A. E. Measuring biological diversity. Environmental and Ecological Statistics (Blackwell Publishing Ltd, 2004).
108. Lahti, L., & Shetty, S. Introduction to the microbiome R package. Bioconductor 2018. (2018).
109. Crist, T. O., Veech, J. A., Gering, J. C. & Summerville, K. S. Partitioning Species Diversity across Landscapes and Regions: A Hierarchical Analysis of α , β , and γ Diversity. *Am. Nat.* 162, 734–743 (2003).
110. Blanchet, F. G., Legendre, P., Maranger, R., Monti, D. & Pepin, P. Modelling the effect of directional spatial ecological processes at different scales. *Oecologia* 166, 357–368 (2011).
111. Borcard, D., Gillet, F. & Legendre, P. Numerical ecology with R. (Springer, 2018).
112. Bivand, R. S. & Wong, D. W. S. Comparing implementations of global and local indicators of spatial association. *Test* 27, 716–748 (2018).
113. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* 89, 2623–2632 (2008).
114. Simpson, G. Restricted permutations; using the permute package. *R.Adu.Org.Za* 1–9 (2007).
115. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling Interactions in the Microbiome: A Network Perspective. *Trends Microbiol.* 25, 217–228 (2017).
116. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* 8, 1–11 (2012).
117. Csardi, G., & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Syst.* 1695, 1–9 (2006).
118. Nshimyimana, J. P., Cruz, M. C., Thompson, R. J. & Wuertz, S. Bacteroidales markers for microbial source tracking in Southeast Asia. *Water research*, 118, 239–248. (2017).
119. Vacca, M. *et al.* The Controversial Role of Human Gut Lachnospiraceae. *Microorg.* 8, 573 (2020).
120. Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S. & Kyrpides, N. C. New insights from uncultivated genomes of the global human gut microbiome. *Nat.* 2019 5687753 568, 505–510 (2019).
121. Sun, S. *et al.* Loss of Novel Diversity in Human Gut Microbiota Associated with

Ongoing Urbanization in China. *mSystems* 7, (2022).

- 122. Rajilić-Stojanović, M., Smidt, H. & De Vos, W. M. Diversity of the human gastrointestinal tract microbiota revisited. *Environ. Microbiol.* 9, 2125–2136 (2007).
- 123. Amaretti, A. *et al.* Antibiotic resistance, virulence factors, phenotyping, and genotyping of non-escherichia coli enterobacterales from the gut microbiota of healthy subjects. *Int. J. Mol. Sci.* 21, 1–14 (2020).
- 124. Clermont, O., Bonacorsi, S., Bingen, E. & Bonacorsi, P. Rapid and Simple Determination of the Escherichia coli Phylogenetic Group Rapid and Simple Determination of the Escherichia coli Phylogenetic Group. *Appl. Environ. Microbiol.* 66, 4555–4558 (2000).
- 125. Clermont, O., Christenson, J. K., Denamur, E. & Gordon, D. M. The Clermont Escherichia coli phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental microbiology reports*, 5(1), 58-65 (2012).
- 126. Lescat, M. *et al.* Commensal Escherichia coli strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environmental microbiology reports*, 5(1), 49-57 (2012).
- 127. Clermont, O. *et al.* Evidence for a human-specific Escherichia coli clone. *Environ. Microbiol.* 10, 1000–1006 (2008).
- 128. Robinaugh, Donald J., Alexander J. Millner, and R. J. M. Identifying highly influential nodes in the complicated grief network. *J. Abnorm. Psychol.* (2016).
- 129. Lepš, J. & Šmilauer, P. Multivariate analysis of ecological data using CANOCO. (Cambridge University Press, 2003).
- 130. Peres-Neto, P. R., Legendre, P. L., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87, 2614–2625 (2006).
- 131. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280 (2001).
- 132. Reisner, A., Krogfelt, K. A., Klein, B. M. & Zechner, E. L. In Vitro Biofilm Formation of Commensal and Pathogenic. *Society* 188, 3572–3581 (2006).
- 133. Laird, N. M. & Ware, J. H. Random-Effects Models for Longitudinal Data. *Biometrics* 38, 963–974 (1982).
- 134. Jose, P., Douglas, B., Saikat, D. & Deepayan, S. Linear and nonlinear mixed effects

models. (2009).

135. Bolar, K. Interactive Document for Working with Basic Statistical Analysis. (2019).
136. Lenth, R., Love, J. & Herve, M. Estimated Marginal Means, aka Least-Squares Means. *R J.* 1–57 (2018).
137. Russell Lenth & Jonathon Love. Least-Squares Means. *Am. Stat.* 34, 216–221 (2017).
138. Warnes, A. G. R. *et al.* Various R Programming Tools for Plotting Data. (2016).
139. Therneau, T., Atkinson, B., R. B. The rpart package. (2018).
140. Salvà-Serra, F. *et al.* A protocol for extraction and purification of high-quality and quantity bacterial DNA applicable for genome sequencing: a modified version of the Marmur procedure. (2018).
141. de Sena Brandine, G. & Smith, A. D. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Research* 8, 1874 (2019).
142. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 1–13 (2019).
143. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13, 1–22 (2017).
144. Jalili, V. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* 48, W395–W402 (2021).
145. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075 (2013).
146. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genomics* 8, 1–10 (2022).
147. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 50, D912–D917 (2022).
148. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903 (2014).
149. Gangiredla, J. *et al.* Draft Genome Sequences of *Escherichia albertii*, *Escherichia*.

Microbiol. Genome Announc. 5, 18–19 (2018).

- 150. Masters, N., Wiegand, A., Ahmed, W. & Katouli, M. Escherichia coli virulence genes profile of surface waters as an indicator of water quality. *Water Res.* 45, 6321–6333 (2011).
- 151. Ishii, S., Meyer, K. P. & Sadowsky, M. J. Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of Escherichia coli strains from diverse human and animal sources. *Appl. Environ. Microbiol.* 73, 5703–5710 (2007).
- 152. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693 (2015).
- 153. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274 (2015).
- 154. Savarino, S. J. *et al.* Enteropathogenic Escherichia coli heat-stable enterotoxin is not restricted to enteropathogenic *E. coli*. *J. Infect. Dis.* 173, 1019–1022 (1996).
- 155. Lymberopoulos, M. H. *et al.* Characterization of Stg fimbriae from an avian pathogenic Escherichia coli O78:K80 Strain and assessment of their contribution to colonization of the chicken respiratory tract. *J. Bacteriol.* 188, 6449–6459 (2006).
- 156. Müller, C. M. *et al.* Type 1 fimbriae, a colonization factor of uropathogenic Escherichia coli, are controlled by the metabolic sensor CRP-cAMP. *PLoS Pathog.* 5, (2009).
- 157. Bleuel, C. *et al.* TolC is involved in enterobactin efflux across the outer membrane of Escherichia coli. *J. Bacteriol.* 187, 6701–6707 (2005).
- 158. Py, B., Loiseau, L. & Barras, F. An inner membrane platform in the type II secretion machinery of Gram-negative bacteria. *EMBO Rep.* 2, 244–248 (2001).
- 159. Russo, T. A., Wenderoth, S., Carlino, U. B., Merrick, J. M. & Lesse, A. J. Identification, genomic organization, and analysis of the group III capsular polysaccharide genes kpsD, kpsM, kpsT, and kpsE from an extraintestinal isolate of Escherichia coli (CP9, O4/K54/H5). *J. Bacteriol.* 180, 338–349 (1998).
- 160. Ozenberger, B. A., Schrotte Nahlik, M. & McIntosh, M. A. Genetic organization of multiple fep genes encoding ferric enterobactin transport functions in Escherichia coli. *J. Bacteriol.* 169, 3638–3646 (1987).
- 161. García-Aljaro, C., Ballesté, E., Muniesa, M. & Jofre, J. Determination of

crAssphage in water samples and applicability for tracking human faecal pollution. *Microb. Biotechnol.* 10, 1775–1780 (2017).

162. Tang, M. *et al.* Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* 404, 1014–1018 (2000).

163. Roberts, R. J. *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812 (2003).

164. Posadas, D. M. *et al.* The TolC homologue of *Brucella suis* is involved in resistance to antimicrobial compounds and virulence. *Infect. Immun.* 75, 379–389 (2007).

165. Jørgensen, M. G., Pandey, D. P., Jaskolska, M. & Gerdes, K. HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J. Bacteriol.* 191, 1191–1199 (2009).

166. Sangari, F. J., Agüero, J. & García-Lobo, J. M. The genes for erythritol catabolism are organized as an inducible operon in *Brucella abortus*. *Microbiology* 146, 487–495 (2000).

167. Bernardini, M. L., Mounier, J., D’Hauteville, H., Coquis-Rondon, M. & Sansonetti, P. J. Identification of icsA, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc. Natl. Acad. Sci. U. S. A.* 86, 3867–3871 (1989).

168. Kaur, G. & Subramanian, S. Classification of the treble CLEF zinc finger: Noteworthy lessons for structure and function evolution. *Sci. Rep.* 6, 1–9 (2016).

169. Lee, D. G. *et al.* Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* 7, (2006).

170. Lou, H. *et al.* Altered antibiotic transport in OmpC mutants isolated from a series of clinical strains of multi-drug resistant *E. coli*. *PLoS One* 6, (2011).

171. Rodriguez, G. M. & Atsumi, S. Isobutyraldehyde production from *Escherichia coli* by removing aldehyde reductase activity. *Microb. Cell Fact.* 11, 1 (2012).

172. Cady, K. C. & O’Toole, G. A. Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.* 193, 3433–3445 (2011).

173. Fujita, K., Takashi, Y., Obuchi, E., Kitahara, K. & Suganuma, T. Characterization of a novel β -l-arabinofuranosidase in *bifidobacterium longum*: Functional elucidation of a duf1680 protein family member. *J. Biol. Chem.* 289, 5240–5249 (2014).

174. Conter, A., Bouché, J. P. & Dassain, M. Identification of a new inhibitor of essential division gene *ftsZ* as the *kil* gene of defective prophage *Rac*. *J. Bacteriol.* 178, 5100–5104 (1996).

175. Vercoe, R. B. *et al.* Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genet.* 9, (2013).

176. Srividhya, K. V. & Krishnaswamy, S. Sub classification and targeted characterization of prophage-encoded two-component cell lysis cassette. *J. Biosci.* 32, 979–990 (2007).

177. Webster, R. E. The *tol* gene products and the import of macromolecules into *Escherichia coli*. *Mol. Microbiol.* 5, 1005–1011 (1991).

178. Gabriel Basílico, Marcos Brito, Gabriela Iglesias, L. de C. Problemas ambientales recurrentes en cuencas hídricas del AMBA, Argentina: tres casos de estudio. *Congr. Interam. Ing. Y Ambient.* (2021).

179. Castané, P. M. *et al.* Spatial and temporal trends of physicochemical parameters in the water of the Reconquista river (Buenos Aires, Argentina). *Environ. Monit. Assess.* 117, 135–144 (2006).

180. Elordi, M. L. Tesis doctoral: "Microbiología ambiental: estudio de patógenos asociados a enfermedades hídricas en arroyos urbanos bonaerenses". Universidad Nacional de La Plata (2016).

181. Menone, M. L. *et al.* Calidad del agua y niveles guía para la protección de la biodiversidad acuática. Interacción entre ciencia y gestión. *Ecol. Austral* 32, 245–257 (2021).

182. HYTSA Estudios y Proyectos S.A. Niveles Guía (NG) para la Cuenca del Plata. (2001).

183. Steele, M. K., McDowell, W. H. & Aitkenhead-Peterson, J. A. Chemistry of Urban, Suburban, and Rural Surface Waters. *Urban Ecosyst. Ecol. agronomymo*, 297–339 (2010).

184. Ramírez, A., Rosas, K. G., Luago, A. E. & Ramos-González, O. M. Spatio-temporal variation in stream water chemistry in a tropical urban watershed. *Ecol. Soc.* 19, 45 (2014).

185. Coleman, J. E. *et al.* Canadian water quality guidelines for the protection of aquatic life: zinc. In: Canadian environmental quality guidelines, 1999. Canadian Council of Ministers of the Environment, Winnipeg, MB. *Environ. Sci. Technol.* 22, 1–13

(2018).

- 186. Office of Water Policy and Technical Guidance on Interpretation and Implementation of Aquatic Life Metals Criteria. (1997).
- 187. CCME. Canadian Environmental Quality Guidelines: Canadian Water Quality Guidelines for the Protection of Aquatic Life - Cadmium. Winnipeg: Canadian Council of Ministers of the Environment. (2014).
- 188. Canadian Council of Ministers of the Environment. Canadian Water Quality Guidelines for the Protection of Aquatic Life: AMMONIA. *Can. Environ. Qual. Guidel.* 8 (2010).
- 189. Peijnenburg, W. J. G. M., Posthuma, L., Eijsackers, H. J. P. & Allen, H. E. A conceptual framework for implementation of bioavailability of metals for environmental management purposes. *Ecotoxicol. Environ. Saf.* 37, 163–172 (1997).
- 190. Rey, N., Casaux, R., & Colautti, D. C. El síndrome del arroyo urbano y su impacto sobre la biodiversidad acuática en cuencas urbanas de la Argentina. *Interciencia* 38, 163–170 (2013).
- 191. Ibekwe, A. M., Ma, J. & Murinda, S. E. Bacterial community composition and structure in an Urban River impacted by different pollutant sources. *Sci. Total Environ.* 566–567, 1176–1185 (2016).
- 192. Staley, C. *et al.* Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River. *Sci. Total Environ.* 505, 435–445 (2015).
- 193. Zhang, H. *et al.* Water bacterial and fungal community compositions associated with urban lakes, Xi'an, China. *Int. J. Environ. Res. Public Health* 15, (2018).
- 194. Wang, F. *et al.* Fates of antibiotic resistance genes during upgrading process of a municipal wastewater treatment plant in southwest China. *Chem. Eng. J.* 437, 135187 (2022).
- 195. McLellan, S. L. & Roguet, A. The unexpected habitat in sewer pipes for the propagation of microbial communities and their imprint on urban waters. *Curr. Opin. Biotechnol.* 57, 34–41 (2019).
- 196. Roguet, A., Newton, R. J., Eren, A. M. & McLellan, S. L. Guts of the Urban Ecosystem: Microbial Ecology of Sewer Infrastructure. *mSystems* (2022).
- 197. Montoya, J. M., Pimm, S. L. & Solé, R. V. Ecological networks and their fragility. *Nature* 442, 259–264 (2006).

198. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19891–19896 (2007).

199. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8577–8582 (2006).

200. Hosen, J. D., Febria, C. M., Crump, B. C. & Palmer, M. A. Watershed Urbanization Linked to Differences in Stream Bacterial Community Composition. *Frontiers in microbiology*, 8, 1452. (2017).

201. Newton, R. J. *et al.* Sewage Reflects the Microbiomes of Human Populations. *MBio* 6, 1–9 (2015).

202. Newton, R. J., Bootsma, M. J., Morrison, H. G., Sogin, M. L. & McLellan, S. L. A Microbial Signature Approach to Identify Fecal Pollution in the Waters Off an Urbanized Coast of Lake Michigan. *Microb. Ecol.* (2013).

203. Peipoch, M., Brauns, M., Hauer, F. R., Weitere, M. & Valett, H. M. Ecological Simplification: Human Influences on Riverscape Complexity. *Bioscience* 65, 1057–1065 (2015).

204. Stoppe, N. de C. *et al.* Worldwide phylogenetic group patterns of *Escherichia coli* from commensal human and wastewater treatment plant isolates. *Front. Microbiol.* 8, (2017).

205. Kleinheinz, G. *et al.* Effect of aquatic macrophytes on the survival of *Escherichia coli* in a laboratory microcosm. *Lake Reserv. Manag.* 25, 149–154 (2009).

206. Escobar-Páramo, P. *et al.* Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl. Environ. Microbiol.* 70, 5698–5700 (2004).

207. Walk, S. T., Alm, E. W., Calhoun, L. M., Mladonicky, J. M. & Whittam, T. S. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. 9, 2274–2288 (2007).

208. Corzo-Ariyama, H. A. *et al.* Phylogroups, pathotypes, biofilm formation and antimicrobial resistance of *Escherichia coli* isolates in farms and packing facilities of tomato, jalapeño pepper and cantaloupe from Northern Mexico. *Int. J. Food Microbiol.* 290, 96–104 (2019).

209. Walk, S. T. *et al.* Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* 75, 6534–6544 (2009).

210. van der Putten, B. & Mende, D. R. *Escherichia ruysiae* sp. nov., isolated from an international traveller. *International Journal of Systematic and Evolutionary Microbiology*, 71(2). (2019).

211. Chaudhuri, R. R. & Henderson, I. R. Infection , Genetics and Evolution The evolution of the *Escherichia coli* phylogeny. *Infect. Genet. Evol.* 12, 214–226 (2012).
212. Clermont, O., Christenson, J. K., Denamur, E. & Gordon, D. M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* 5, 58–65 (2013).
213. Uddin, T. M. *et al.* Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *J. Infect. Public Health* 14, 1750–1766 (2021).
214. Casadevall, A. & Pirofski, L. A. Host-pathogen interactions: Basic concepts of microbial commensalism, colonization, infection, and disease. *Infect. Immun.* 68, 6511–6518 (2000).
215. Tegha, G. *et al.* Genomic epidemiology of *escherichia coli* isolates from a tertiary referral center in lilongwe, Malawi. *Microb. Genomics* 7, 1–12 (2021).
216. Martinez, J. L. Environmental pollution by antibiotics and by antibiotic resistance determinants. *Environ. Pollut.* 157, 2893–2902 (2009).
217. Vrede, K. Nutrient and temperature limitation of bacterioplankton growth in temperate lakes. *Microb. Ecol.* 49, 245–256 (2005).
218. Jiang, L. & Patel, S. N. Community assembly in the presence of disturbance: A microcosm experiment. *Ecology* 89, 1931–1940 (2008).
219. Van Melderen, L. & De Bast, M. S. Bacterial toxin-Antitoxin systems: More than selfish entities? *PLoS Genet.* 5, (2009).
220. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science (80-.)* 356, (2017).

Financiamiento

- Subsidio a Unidades Ejecutoras 2016 del CONICET (22920160100122CO/IEGEBA-CONICET).
- Premio Tonolli 2019 – Otorgado por la Sociedad Internacional de Limnología (*International Society of Limnology, SIL*).