



OPEN

Unraveling the ecological processes modulating the population structure of *Escherichia coli* in a highly polluted urban stream network

Martín Saraceno^{1,2}, Sebastián Gómez Lugo², Nicolás Ortiz³, Bárbara M. Gómez³, Carmen A. Sabio y García², Nicolás Frankel^{2,4} & Martín Graziano^{1,2}✉

Escherichia coli dynamics in urban watersheds are affected by a complex balance among external inputs, niche modulation and genetic variability. To explore the ecological processes influencing *E. coli* spatial patterns, we analyzed its abundance and phylogenetic structure in water samples from a stream network with heterogeneous urban infrastructure and environmental conditions. Our results showed that environmental and infrastructure variables, such as macrophyte coverage, DIN and sewerage density, mostly explained *E. coli* abundance. Moreover, main generalist phylogroups A and B1 were found in high proportion, which, together with an observed negative relationship between *E. coli* abundance and phylogroup diversity, suggests that their dominance might be due to competitive exclusion. Lower frequency phylogroups were associated with sites of higher ecological disturbance, mainly involving simplified habitats, higher drainage infrastructure and septic tank density. In addition to the strong negative relationship between phylogroup diversity and dominance, the occurrence of these phylogroups would be associated with increased facilitated dispersal. Nutrients also contributed to explaining phylogroup distribution. Our study proposes the differential contribution of distinct ecological processes to the patterns of *E. coli* in an urban watershed, which is useful for the monitoring and management of fecal pollution.

Escherichia coli, one of the most frequently found pathogens in urban waters, has historically been used as a proxy for recent microbiological pollution, mainly due to its straightforward and low-cost detection methods¹. There is consistent evidence that some *E. coli* strains may persist and even grow in secondary habitats such as fresh and marine waters and sediments^{2–6}. The ecological processes influencing the population structure of *E. coli* in urban environments are unclear; its ubiquity result from a complex balance involving external inputs, ecological conditions and niche processes acting on aquatic habitats, while affecting its intraspecific genetic variability^{7–9}. Therefore, a better understanding of the ecological processes underlying the environmental occurrence of *E. coli* is a matter of methodological, sanitary and management concern^{9,10}.

E. coli intra-specific variability has been grouped into eight major phylogroups based on genetic analyses: A, B1, B2, C, D, E, F and G^{11–13}. Although laboratory and pathogenic *E. coli* strains have been thoroughly described, the environmental ecological niches associated with phylogroups remain largely unknown. The abundance pattern of each phylogroup seems to vary among surface freshwater systems, but some of them (A and B1) are consistently predominant and prevalent while others are less frequent and associated to urban areas (e.g., D and F)^{14–16}. According to the coexistence theory, the over-representation of some phylogroups in the environment can be ascribed to competitive fitness differences (e.g., in resource exploitation) that stabilize their coexistence¹⁷. Experimental microcosm studies support the notion that the most abundant *E. coli* phylogroups show greater capacity of environmental survival. For example, strains from phylogroup B1 were found to persist longer and

¹Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEB), CONICET - Universidad de Buenos Aires, 1428 Buenos Aires, Argentina. ²Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina. ³Instituto Nacional del Agua, 1804 Ezeiza, Argentina. ⁴Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), CONICET - Universidad de Buenos Aires, 1428 Buenos Aires, Argentina. ✉email: marting@ege.fcen.uba.ar

to tolerate lower temperatures than the remaining phylogroups¹⁸. Moreover, phenotypic analysis revealed that phylogroup B1 strains were more likely to exhibit traits indicative of a higher ability to colonize aquatic plants (and therefore to persist in freshwater secondary habitats), while A and B2 phenotypes were linked to an animal-associated lifestyle¹⁹. *Escherichia* strains closely related to *E. coli* (i.e., *cryptic clades* I to V), were also associated with long-term environmental persistence^{3,20,21}. Altogether, this strongly suggests that niche processes such as competitive exclusion and niche partitioning may contribute to shaping the spatial distribution of the environmental population structure of *E. coli*. However, there is no *in-situ* evidence supporting the role of niche processes.

Several studies focused on bacterial community in urban streams have revealed the impact of urbanization on population structure, highlighting the need to consider watersheds as part of a complex social-ecological-technological system^{22–24}. Within this framework, ecological modulations may be understood as a balance between local features of the ecosystems, demographic pressures and built-in infrastructure. In this sense, urban infrastructures, such as stormwater networks and roads, facilitate the transfer of pollutants by surface runoff, acting as links between human activity in land areas of the watershed and water bodies, turning the latter into sinks²⁵. Indeed, the presence of drainage infrastructures and impervious surfaces was positively associated with concentrations of dissolved organic carbon (DOC), soluble reactive phosphorus (SRP), total phosphorus (TP), ammonium, nitrates and nitrites, and electrical conductivity in urban streams^{26–28}. Some of these compounds can be used as nutrients by aquatic bacteria^{8,29}. Thus, heavily impacted urban water bodies with high availability of these nutrients and other suitable environmental conditions (e.g., relatively high temperatures and a wide range of colonizable niches) may become hot habitats for *E. coli* strains showing long-term persistence^{30–33}.

Previous studies have suggested that environmental persistence of *E. coli* in urban freshwater systems is influenced by abiotic factors such as soil moisture, temperature and nutrient availability^{7,34}. Furthermore, it was shown that high levels of soluble nutrients (e.g., DOC and phosphorus) have a positive influence on growth and survival of *E. coli* in the water column of urban streams⁸. Moreover, the input of black and grey waters into streams entails the dispersal of allochthonous bacteria, representing one of the major drivers of shifts in the microbial community composition of these systems^{35–37}. Septic tanks are also known to contribute to non-point pollution through leakage, acting as a source of pathogens, dissolved nutrients and metals to groundwater and nearby streams^{38,39}. Likewise, sanitary systems such as sewer networks, if present, cause pollution by leakage and subsequent exfiltration to water bodies^{39,40}. Hence, ecological niche conditions and human-facilitated dispersal processes may also play an important role in modulating *E. coli* population structure in urban surface waters.

In this work we analyzed the ecological processes affecting the spatial structure of *E. coli* in a heavily polluted urban stream network of Buenos Aires province, Argentina. The watershed was characterized as a social-ecological and technological system in order to analyze the contributions of local urban infrastructure and environmental conditions. Spatial processes structuring *E. coli* populations that were not explained by the latter were assessed by including asymmetric eigenvector maps. We hypothesize that the density of input sources to streams and the features of the ecological habitat determine the spatial patterns of abundance and phylogenetic composition of *E. coli*. In this context, we predict that (1) *E. coli* abundance will be positively associated with the density of different urban input sources and higher nutrient availability, (2) that the occurrence of phylogroup B1 will be associated with higher nutrient availability and aquatic vegetation coverage and (3) that the occurrence of low-frequency *E. coli* phylogroups (D and F) will be positively associated with locations of higher urbanization. Finally, we discuss our results in relation to the different ecological processes that could influence the dynamics of *E. coli* and their implications for urban water management.

Methods and materials

Study area and sampling procedure. The watershed of San Francisco, Las Piedras and Santo Domingo streams is located in the Pampean plain, in the south of the highly urbanized Buenos Aires Metropolitan Area (AMBA), surrounding Buenos Aires city (Fig. 1). This watershed encompasses a total surface area of approximately 160 km², with a longitudinal extension of about 23 km. The streams draining this area are affected by multiple anthropogenic stressors⁴¹. The watershed and stream network have been highly modified by rectifications, channel incisions and some streams were partially piped; one of the latter (Las Perdices stream) was even relocated into this watershed. In addition, urban growth in the AMBA was not accompanied by investments in sanitary infrastructure and as a result, local water bodies receive the impact of untreated sewage and domestic effluents^{42–44}.

We selected 14 sampling sites across the watershed representing a broad range of urban characteristics related to population density and coverage of sanitary services. Data were obtained from the Census Data Center⁴⁵. The sampling campaign was conducted in summer, the season with the most favorable mean temperature for *E. coli* survival, during 4 consecutive days in February 2018. A dry period of at least 4 days was considered before the sampling campaign and no rainfall were recorded during its duration. Sites across the stream network were visited randomly at morning-noon to avoid spatio-temporal dependence. Habitat heterogeneity was assessed along a 50 m-longitudinal transect established along the main watercourse at each sampling site, where sub-samples were taken at 0 m, 25 m and 50 m. On each sub-sampling site, pH, dissolved oxygen (DO), conductivity and temperature (T) were measured *in situ* from the sub-surface layer of the water column, using a Lutron WA-2017SD multiparameter field sensor. Local hydraulic and habitat parameters of the sampled stream section were assessed in terms of physical dimensions (width and depth), flow velocity and macrophyte coverage. Flow velocity was determined using the flotation method by triplicate and water flow from physical and flow velocity parameters⁴⁶. Mean values of flow velocity, depth and water flow were calculated per site. Macrophyte coverage was estimated according to the Braun-Blanquet methodology, by establishing a 0.5 m width plot perpendicular to the flow and accounting for emerged and floating-leaved macrophytes⁴⁷. Water samples from each sub-sampling site were

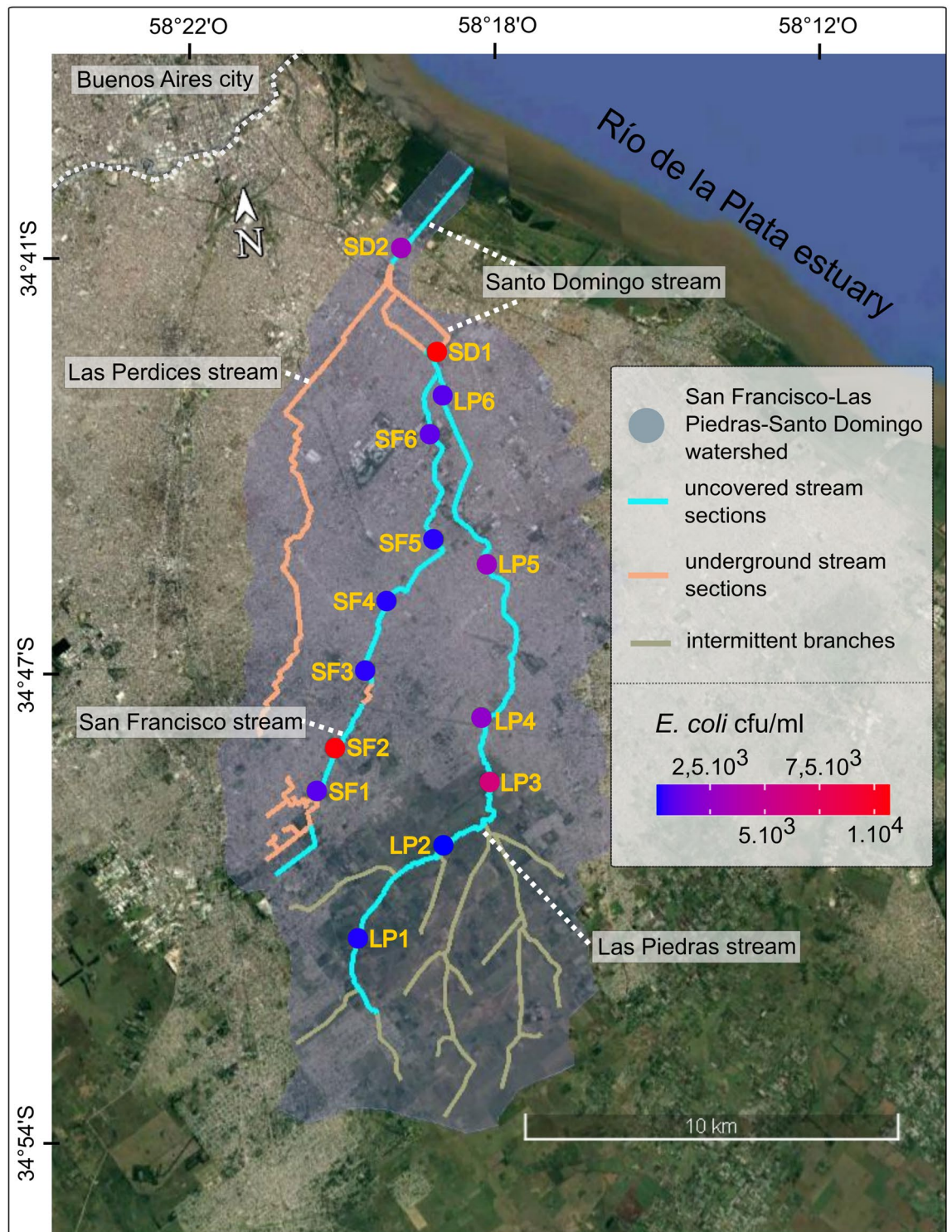


Figure 1. Location of the San Francisco-Las Piedras-Santo Domingo watershed and *E. coli* abundance across sampled sites. The studied watershed is located on the southern side of the Metropolitan Area of Buenos Aires. Spatial distribution of mean *E. coli* abundance (cfu/ml) at the sampled sites (dots) is color-coded in a blue-red gradient. LP, Las Piedras stream; SF, San Francisco stream; SD, Santo Domingo stream. Base map created in Google Earth V 7.3.3.7786. (October 5, 2020). Buenos Aires Metropolitan Area, Argentina. 34°46'47.32"S, 58°19'27.95"O, Eye alt 47.5 km. Maxar Technologies 2020. <http://www.earth.google.com>. Graphical edition in Inkscape V 0.92.4. Inkscape Project (2020). <https://inkscape.org>.

collected in clean 250-ml bottles for the analysis of dissolved nutrients and other analytes and in sterile 810-ml plastic bags (Hixwer) for measuring microbiological parameters. Water samples were taken from the sub-surface layer and kept cold until processing in the laboratory.

Physicochemical characterization. Upon arrival at the laboratory, turbidity was measured with a turbidimeter HACH 2100P. Samples were filtered simultaneously for further quantification of soluble analytes. First, samples were filtered through a cellulose nitrate filter of 0.47 μm -pore diameter (Sartorius) for subsequent analysis of soluble reactive phosphorus (SRP), ammonium and nitrates. Then, they were passed through a polycarbonate filter of 0.2 μm -pore diameter (Merck Millipore) for dissolved organic carbon (DOC), soluble iron and chloride. SRP was determined by the molybdovanadate method, ammonium by the salicylate method and nitrate plus nitrite by the cadmium reduction method⁴⁸. Measurements were made by colorimetric analysis with a spectrophotometer Hach® DR 2800 within 48 h of sample collection; limits of quantification (LOQ) were 0.1 mg/L for SRP and nitrates and 0.4 mg/L for ammonium. Ammonia and nitrate plus nitrite were considered together as dissolved inorganic nitrogen (DIN). For DOC determination, filtered samples were acidified with high-quality sulfuric acid (Merck) and measurements were performed with the high-temperature combustion method using a total organic carbon analyzer Shimadzu 5000A; LOQ was 0.1 mg/L⁴⁸. Chlorides were measured by the iodometric titration method; LOQ was 4 mg/L⁴⁸. Finally, filtered samples pre-treated with ultrapure nitric acid (Merck) were analyzed for soluble iron by inductively coupled plasma-mass spectrometry (ICP-MS) using a mass spectrometer Agilent 7500cx; LOQ was 0.3 $\mu\text{g/L}$.

***Escherichia coli* abundance and phylogenetic group annotation.** *Escherichia coli* abundance was determined for each sample (42 samples in total, three per site) by plate counting using Chromocult® Coliform Agar selective medium (MilliporeSigma). Serial dilutions in sterile ultrapure water were performed on each sample in triplicate and plates were cultured for 24 h at 37 °C. After scoring for blue- to purple-colored colonies in the best dilution for each sample, triplicates counts were averaged per sample and relativized to one milliliter of volume. The detection limit was 200 cfu/100 mL.

To obtain the relative phylogenetic composition of *E. coli* per location, in a first step, blue- to purple-colored colonies from the Chromocult® plates from each site were streaked onto Levine E.M.B. (Eosin Methylene Blue) agar plates and cultured for 24 h at 36 °C. The streaking procedure was repeated at least once under the same culture conditions to ensure the purity of each isolate. Isolates yielding typical responses for *E. coli* on all media were designated *E. coli*, while those exhibiting no typical phenotypic responses in any of the media were further screened using a colony PCR method with primers directed to *E. coli* essential *trpA* gene⁴⁹. A total of 327 isolates were reliably assigned to *E. coli*, ranging from 18 to 33 isolates per sampled site. Isolates were randomly selected from the sub-sampling plates (see Table S1, Supplementary Appendix A1 for total counts per site). Phylogenetic annotation within the *E. coli* intraspecific groups was then performed using the Clermont's multiplex PCR method¹³. The closely related phylogroups G and F were grouped together because this strategy is unable to differentiate one from the other¹². Selected isolates were grown overnight in MacConkey broth at 37 °C with continuous shaking and used as templates in the PCR assays. The first round of phylogroup assignment was performed through the amplification of the *araA*, *chuA*, *yjaA* and *TspE4.C2* genes. A remarkable advantage of this method is that at least one of the targeted genes is guaranteed to be amplified, thus allowing a positive control of species identity. Additional primers were used to identify phylogroups E and C if needed, by detection of *arpA* and *trpA* genes, respectively. A double PCR method, based on *aes* and *chuA* allele-specific amplification, was employed to assign an *Escherichia* strain a cryptic lineage membership^{50,51}. All PCR reactions were carried out in a 20 μL -total volume containing 2 μL of templates and 10 μL of 2X GoTaq® Green Master Mix (Promega) following the published literature^{13,51,52}. The complete assignment of isolates to phylogroups per site is shown in Table S1 (Supplementary Appendix A1).

Demographic, hydraulic and sanitary infrastructure parameters. A 500-m radius circular area was established around each site to survey the total number of dwellings and residents and the total number of dwellings with sanitary sewers, septic tanks and potable water, using open data provided by national agencies⁴⁵. The impervious surface coverage was calculated through GIS image processing (QGIS v. 2.18.20) coupled with the Semi-automatic Classification Plugin algorithm, employing Landsat 8 satellite images of the watershed⁵³. This methodology is based on the preliminary creation of a spectral firm by a supervised classification procedure for the generation of regions of interest (ROIs). Each pixel in the ROI was categorized as impervious or pervious area. The trained ROIs were employed with the Spectral Angle Mapper algorithm to characterize the satellite image of the watershed. Impervious surface coverage was first calculated by manual digitalization of about 30 ha distributed homogeneously throughout the basin, obtained from a Google Maps image⁵⁴. This was used to generate polygons representing areas with different percentages of imperviousness. Permeable areas showed a limit of impermeability of 5%. Finally, considering a 500-m radius buffer area for each sampling site, GIS processing tools were used to calculate the areas for the two permeability types, to further express the impervious surface as a percentage within each buffer area. To estimate drainage and road densities, layouts of the pluvial and road networks were provided by local and provincial administrations agencies. Data were intersected with the buffer areas to calculate the length of their respective pluvial and road sections.

Estimation of spatial predictors. Asymmetric Eigenvector Maps (AEMs) were generated to model directional spatial processes underlying streams network singularities^{54–56}. These AEMs were based on the geographical distribution of the sampling sites, and the connectivity matrix among sites was constructed considering the upstream–downstream flux and a maximum spatial neighborhood distance of 5 km between sites. The AEMs

also included weight vectors obtained from the connectivity matrix, which determined the connection strength between samples according to their geographic distance. A two-sided Moran's test (999 permutations; $p < 0.05$) was used to retain AEMs showing significant positive or negative spatial autocorrelations (see Supplementary Fig. S1 for more details)⁵⁷. Finally, *forward selection* procedures were carried out separately for each response variable to retain AEMs significantly associated with them. This yielded 6 and 11 AEMs that were significantly associated with *E. coli* abundance and phylogenetic composition, respectively. The four AEMs gathering the greatest R^2 contribution were employed to avoid multicollinearity in further analysis involving *E. coli* phylogenetic composition. These procedures were conducted using the packages *adespatial* (v 0.3–8), *ade4* (v 1.7–15), *sp* (v 1.3–2) and *spdep* (v 1.1–3) in R. For more details on the methodology for spatial-predictor selection, see Appendix A2 of the Supplementary Material.

Data analysis. We first calculated the phylogroup specific diversity and dominance at each site for an initial characterization of the phylogroup population structure^{58,59}. Phylogroup diversity was obtained by counting the number of different phylogroups present at each site, while Simpson's dominance index (D) was calculated as the weighted arithmetic mean of their proportional abundances⁵⁸. The association between phylogroup diversity, dominance and *E. coli* mean abundance was analyzed using a Pearson's correlation analysis.

Different statistical approaches were applied to analyze relationships among predictors. Initially, two sets of predictor variables were established: an environmental matrix including local physicochemical and habitat conditions and an urban infrastructure set, combining demographic information with different aspects of sanitary infrastructure at the watershed scale. A Principal Component Analysis (PCA) was performed to analyze the association between sites and environmental and urban infrastructure variables⁶⁰. Additionally, a Pearson's correlation matrix with hierarchical clustering was performed among variables. The significance of the correlation coefficients (Pearson's ρ) was evaluated with the paired-samples correlation test adjusted for multiple comparisons (Holm's method)⁶¹. In addition, we analyzed the network of global correlations between pairs of variables showing an absolute Pearson's ρ greater than 0.5. To understand the importance of each parameter in determining the whole network's structure, we estimated associated centrality measures such as Expected Influence (EI), which represents the strength of a node's influence within the network⁶².

A variation partitioning analysis was performed to disentangle the pure and shared effects of the local environment, urban infrastructure and spatial predictors determining the patterns of abundance and phylogroup composition (based on presence-absence data) of *E. coli*⁶³. An adjusted multivariate redundancy statistic (R_a^2) was used to analyze the proportion of variance explained by each component. To reduce the number of variables and to avoid multicollinearity between predictors, a set of multivariate analyses were performed. This was accomplished by identifying several groups of covariate variables by means of a PCA analysis applied to the environmental and infrastructure matrices separately, along with the use of variance inflation factors (for more details on the procedure for variable selection, see Appendix A3 of the Supplementary Material).

For variance partitioning and partial redundancy analyzes, the *E. coli* abundance and the predictor matrices were previously standardized by the *Standardization* method⁶⁴. In accordance with Leps & Smilauer⁶⁵, we preliminarily assessed the use of a linear constrained model (RDA) onto phylogroup composition data, by considering the length of the gradient obtained with a Detrended Correspondence Analysis (DCA; 1st axis length = 1.44). The significance of global models and variables was further evaluated through partial redundancy analysis (pRDA), which was subjected to different conditioning constraints. In the case of the *E. coli* abundance, two different constraints were implemented (1) a pRDA analysis of the environmental or infrastructure matrices, conditioned by each other, and (2) a pRDA analysis of spatial predictors conditioned by the rest of the predictor sets. In contrast, for the occurrence of *E. coli* phylogroups, each predictor matrix was conditioned by the rest of the variables. In all cases, statistical significance of pRDA was tested using a restricted Monte Carlo Permutation Test of the residuals of the full model (9999 permutations), to account for our nested sampling design and the multilevel structure of the explanatory variables⁶⁶. Both normality and lack of structure in residuals were tested. Statistical procedures were implemented with the packages *vegan* (version 2.5–6), *corrplot* (version 0.84), *qgraph* (version 1.6.5) and *permute* (version 0.9–5) in R.

Results

Patterns of *E. coli* abundance associated with environmental and urban infrastructure features.

To gain insight into the ecological processes modulating the population structure of *E. coli*, we assessed the joint and independent contributions of urban features and local stream habitat conditions across 14 stream reaches (locations) distributed throughout a heavily polluted urban stream network with heterogeneous built-in sanitation and drainage infrastructure (Fig. 1). The characterization of urban features included demographic, hydraulic (impervious surfaces, drainage density and road density) and sanitary conditions (sanitary sewers, septic tanks and potable water) (Table 1). Local habitats were characterized for their physicochemical profile (pH, conductivity, temperature, dissolved oxygen, DIN, SRP, DOC, turbidity, chlorides and iron) and for local hydraulic and habitat parameters of the stream section (physical dimensions, water flow, flow velocity and macrophyte coverage) (Table 1). Spatial predictors were also included in the analysis to account for directional and correlated effects (see Appendix A2 of the Supplementary Material). *E. coli* abundance was quantified at each location and the phylogenetic affiliation of isolates was assigned by molecular methods.

The locations surveyed had a wide variability in environmental and urban infrastructure across the hydrological network (Table 1 and Fig. 1). *E. coli* abundance exceeded the level recommended by the US-EPA 2012¹ of 235 cfu/100 mL for recreational freshwaters at all studied sites, which poses a great risk for people in direct contact with the studied streams (see Table S1, Appendix A1 of the Supplementary Material for detailed results). The analysis of urbanization proxies (infrastructure coverage, impervious surface coverage and population density)

	Range	Mean	SD
pH	5.27–8.02	7.36	0.74
Conductivity ($\mu\text{S}/\text{cm}$)	762–1533	1078	189
Temperature ($^{\circ}\text{C}$)	17.4–23.1	19.6	1.7
Dissolved oxygen (mg/L)	2.5–10.0	4.1	2.0
DIN ($\text{mg}/\text{L N-NO}_3^-$)	5.6–20.0	12.4	4.3
SRP ($\text{mg}/\text{L P-PO}_4^{3-}$)	0.8–7.3	2.3	1.7
DOC (mg/L)	6.2–77.9	18.0	18.0
Turbidity (NTU)	6–109	34	33
Chloride (mg/L)	38–107	70	22
Iron ($\mu\text{g}/\text{L}$)	23.5–337	135	91.6
Macrophyte coverage (%)	0–92	38	31
Water flow (m^3/s)	3.0×10^{-2} –3.8	0.9	1.2
Flow velocity (m/s)	3.21×10^{-2} –0.43	0.21	0.12
Depth (cm)	28–80	46	16
Drinking water coverage (%)	0.01–31.1	13.4	12.5
Sanitary sewer density (dwelling/ha)	0.0–16.2	3.5	5.4
Septic tank density (dwelling/ha)	0.1–28.9	14.0	9.9
Impervious surface (%)	5–75	53	27
Drainage density (m/km^2)	0–3.837	1.519	1.235
Road density (m/km^2)	1.248–22.369	13.983	7.323
Population density (Hab./ha)	0.3–135.0	68.7	45.4
<i>E. coli</i> (cfu/ml)	3–10,700	2735	3732

Table 1. Basic statistical summary of environmental (unbolded) and urban infrastructure measured parameters (bolded). DIN, dissolved inorganic nitrogen; SRP, soluble reactive phosphorus; DOC, dissolved organic carbon; SD, standard deviation.

allowed to characterize the watershed as predominantly urban (mean impervious surface of 53%), with a low-urbanized area located in the headwater of Las Piedras stream. The rest of the basin is highly urbanized and values of impervious surface are higher downstream (> 70%). However, the coverage of sanitary infrastructure services, such as drinking water or sewerage, is heterogeneous across the watershed, reflecting differences in the level and quality of urbanization. A PCA revealed a gradient of urban infrastructure along the watershed (Fig. 2a) and grouped locations into distinct clusters based on their features. A first cluster includes locations from the upper section of Las Piedras stream, with higher macrophyte coverage and DIN, together with lower infrastructure coverage. A second cluster groups locations from the upper section of the San Francisco stream, which mainly show intermediate infrastructure coverage. Lastly, a third cluster is composed of locations from densely populated areas with a high development of hydraulic and sanitary infrastructure, which are associated with lower levels of nutrients and high flow velocity. One location in the Santo Domingo stream (SD1) was not grouped in any of the clusters identified, mainly associated with elevated levels of DOC, *E. coli* abundance and turbidity.

A network of global Pearson's correlations was used to explore co-variation patterns between variables (see Figure S4, Supplementary Appendix A4 for the overall significance analysis). The obtained network was analyzed in terms of the influence or relative importance of each variable (see Fig. 2b and Table S4 in Appendix A4 of the Supplementary Material for full-network metrics). Urban infrastructure variables were ordered close to each other by their positive and significant correlations, mainly interacting through the proportion of impervious surface (expected influence, $|EI|=4.80$), drinking water coverage ($|EI|=3.65$) and population density ($|EI|=3.45$). Among environmental variables, flow velocity ($|EI|=2.48$), water flow ($|EI|=2.46$), and with lower relative influence, macrophyte coverage ($|EI|=1.91$) concentrated the links with urban infrastructure, mainly interacting with sanitary sewer density ($|EI|=2.95$) and impervious surface. A second branch of interactions was evidenced between impervious surface and conductivity ($|EI|=2.54$) and chlorides ($|EI|=1.52$). Moreover, pH ($|EI|=4.82$) and DOC ($|EI|=3.60$) represented key variables within the environmental matrix, gathering the largest number of links. Local environmental variables such as DOC, SRP, Iron, conductivity, chlorides and DIN were also clustered in a second positive and significant co-variation group. *E. coli* abundance was significantly and positively associated with nutrients such as SRP and DOC, and with physicochemical conditions such as pH (–) and conductivity (+), suggesting that urbanization parameters may indirectly influence *E. coli* abundance through environmental factors. Finally, a cluster of strong positive and negative significant correlations was found among sanitary sewer density, pH, water flow, turbidity, depth and macrophyte coverage.

Spatial distribution of *E. coli* phylogenetic groups. A total of 327 environmental isolates were characterized based on the Clermont's multiplex PCR method for *E. coli* phylogroup annotation (see Table S1, Supplementary Appendix A1 for the absolute abundances detected per site). We detected most of the phylogenetic groups, except for phylogroup C. Phylogroup A was the most abundant in all locations sampled, with a relative frequency of up to 50% at most sites, followed by phylogroup B1 (Fig. 3). Mean relative frequency was 64% for

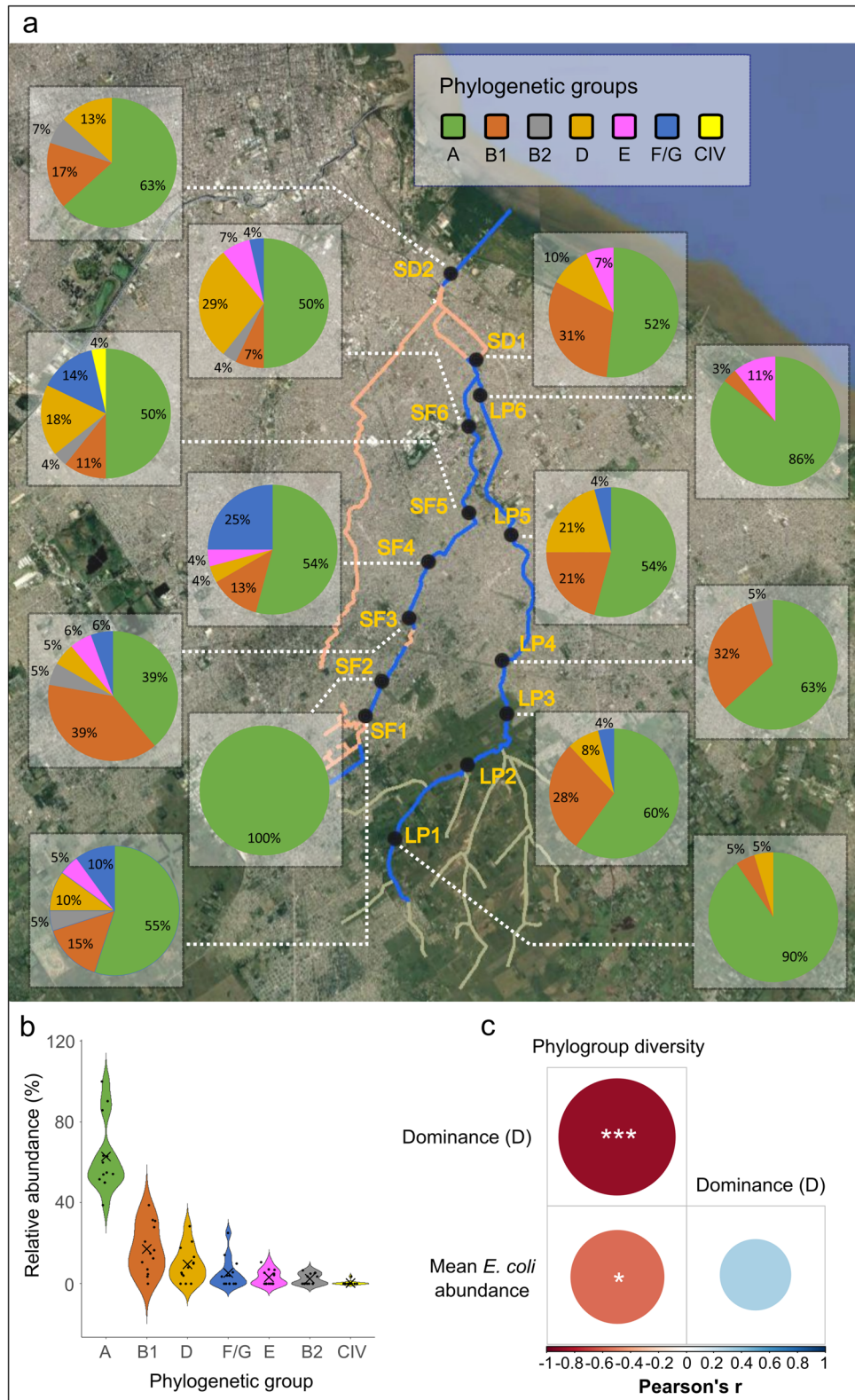


Figure 3. Differences in phylogroup composition of *E. coli* among locations. (a) Spatial distribution of the phylogenetic composition in the sampled locations; pie charts indicating the relative frequency of the phylogenetic groups; (b) Violin plot of the relative abundance of the phylogenetic groups. Filled dots represent the relative abundance at each site; crosses indicate the means; (c) Correlogram of Pearson's correlations between Simpson's dominance index (D), diversity of *E. coli* phylogroups, and mean *E. coli* abundance. Significance levels of Pearson's correlation: * $p < 0.05$, ** $p < 0.001$. Base map created in Google Earth V 7.3.3.7786. (October 5, 2020). Buenos Aires Metropolitan Area, Argentina. 34°46'47.32"S, 58°19'27.95"O, Eye alt 47.5 km. Maxar Technologies 2020. <http://www.earth.google.com>. Data analysis performed in R software⁷³. Graphical edition in Inkscape V 0.92.4. Inkscape Project (2020). <https://inkscape.org>.

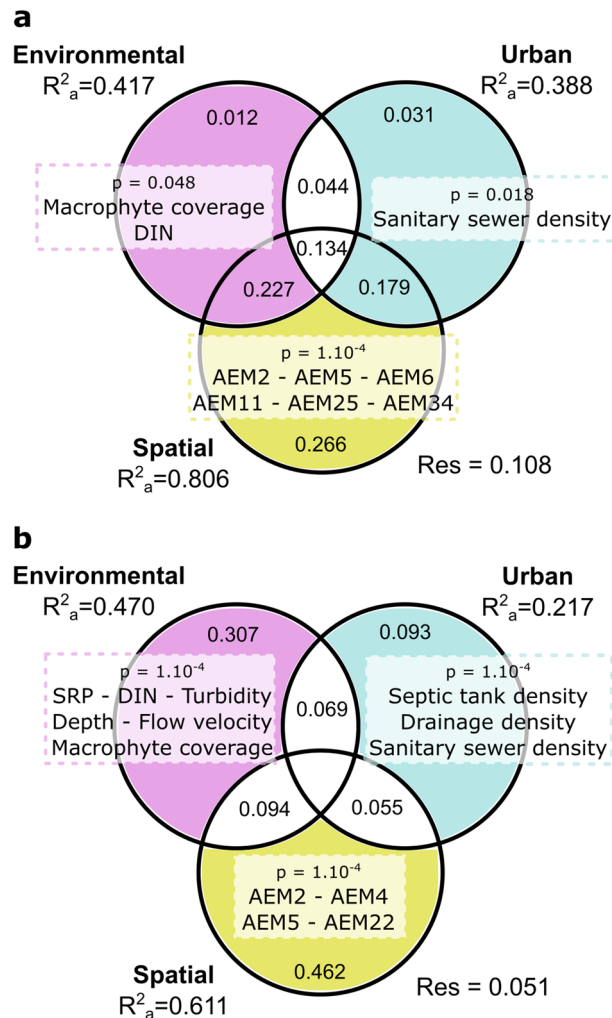


Figure 4. Drivers of *E. coli* population structure. Venn diagrams of the results of variation partitioning analysis of response variables explained by environmental, urban infrastructure and spatial contributing factors: (a) abundance of *E. coli*; (b) phylogenetic composition. The contribution of each predictor set is represented by R^2 adjusted values and p -values, when available; values < 0 are not shown. Significant parameters derived from each pRDA from each predictor matrix are listed in each diagram ($p < 0.05$); color-shaded areas represent the evaluated fraction for each main matrix. DIN, dissolved inorganic nitrogen; SRP, soluble reactive phosphorus; AEM, asymmetric eigenvector map; Res, residuals. Data analysis performed in R software⁷³. Graphical edition in Inkscape V 0.92.4. Inkscape Project (2020). <https://inkscape.org>.

phylogroup A, 16% for B1, 10% for D, 5% for F/G, 3% for E and 2% for B2 (Fig. 3b). Notably, a single isolate of the *cryptic clade* IV (further confirmed by a specific multiplex PCR assay for the identification of cryptic clades within the genus *Escherichia*) was collected in location SF5. To our knowledge, this is the first report of a cryptic clade member in surface waters of South America.

A correlation analysis of community structure metrics applied to *E. coli* phylogenetic composition showed that phylogroup diversity was negatively correlated with total mean abundance of *E. coli* (coefficient of correlation $\rho = -0.56$; $p = 0.05$) (Fig. 3c). In addition, Simpson's dominance index (D) was negatively correlated with phylogroup diversity (coefficient of correlation $\rho = -0.87$; $p = 1.10^{-4}$).

Disentangling the effects of environmental, urban and spatial factors on patterns of *E. coli* abundance and phylogenetic composition.

A variance partitioning analysis (Fig. 4) showed that important independent effects of environmental and urban infrastructure predictors (23% and 18% of the fraction shared with the spatial matrix, respectively) affect the distribution of *E. coli* abundance. At the same time, the three predictor sets together explained 13% of the total variance. The spatial matrix of AEMs showed a strong pure contribution (27%), while environmental and infrastructure matrices exhibited remarkably lower pure contributions ($< 3\%$) (Fig. 4a). In terms of overall influence, 42% of total contribution to the explained variance was related to the environmental matrix and 39% to the urban infrastructure matrix, indicating a similar contribution of both sets of predictors to the observed patterns of *E. coli* abundance. The high contribution of

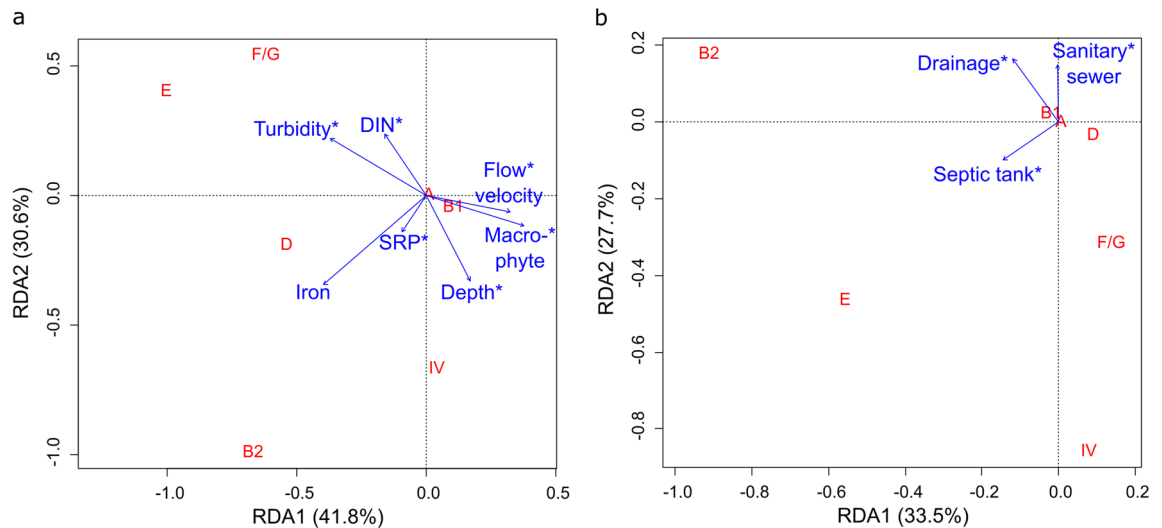


Figure 5. Independent effects of urban infrastructure and ecological habitat conditions affecting the composition of *E. coli* phylogroups. Biplots of partial redundancy analysis relating the occurrence of *E. coli* phylogroups to: (a) environmental and (b) urban infrastructure explanatory factors (scaling = 3). In each case, pRDA was performed controlling for the rest of the variable sets (including spatial effects). Significant variables, assessed using the Monte Carlo restricted permutation test (9999 permutations), are identified with an asterisk. DIN, dissolved inorganic nitrogen; SRP, soluble reactive phosphorus. Data analysis performed in R software⁷³. Graphical edition in Inkscape V 0.92.4. Inkscape Project (2020). <https://inkscape.org>.

spatial factors in shared and pure fractions indicates that *E. coli* abundance has a strong directional spatial structure, potentially reflecting distinct spatially structured processes (i.e., different hydrological, environmental and infrastructure patterns).

To further assess the contribution of spatially structured environmental and urban infrastructure variables, a pRDA was performed, where each explanatory set was conditioned by the other (Fig. 4a). Results showed that both environmental ($F_{7,31} = 4.48$, $P < 0.05$) and urban ($F_{3,31} = 7.38$, $P < 0.05$) matrices were significant in explaining *E. coli* abundance. Within the environmental set, significant variables were macrophyte coverage (score = 0.43; $F_{1,31} = 12.38$, $P < 0.05$) and DIN (score = 0.43; $F_{1,31} = 12.81$, $P < 0.05$). In regard to the urban infrastructure matrix, sanitary sewer density had a significant contribution (score = 0.51; $F_{1,31} = 18.02$, $P < 0.05$). As well, several coarse- and fine-grade AEMs of the spatial matrix had significant pure contributions (Fig. 4a; see more details in Appendix A2 of the Supplementary Material, Table S2).

In contrast to the results obtained for abundance, spatial variability in phylogroup composition at the watershed (based on phylogroup presence-absence data) was mostly explained by pure contributions of the three matrices (Fig. 4b). The contribution of spatial factors was the largest (46%), followed by environmental (31%) and urban infrastructure factors (9%). Partial RDA of environmental variables, controlling for the effect of urban infrastructure and spatial AEMs, was statistically significant ($F_{7,24} = 27.69$, $P < 0.05$). The following variables were significant: DIN ($F_{1,24} = 25.95$, $P < 0.05$), turbidity ($F_{1,24} = 22.99$, $P < 0.05$), SRP ($F_{1,24} = 20.41$, $P < 0.05$), depth ($F_{1,24} = 9.03$, $P < 0.05$), flow velocity ($F_{1,24} = 13.33$, $P < 0.05$) and macrophytes coverage ($F_{1,24} = 7.04$, $P < 0.05$) (Fig. 4b). The first two pRDA axes were also found to be significant: RDA1 accounted for an explained proportion of 42% ($F_{1,26} = 98.55$, $P < 0.05$) and RDA2 of 31% ($F_{1,26} = 72.29$, $P < 0.05$). RDA1 was mainly positively related to macrophyte coverage (score = 0.29), flow velocity (score = 0.25) and depth (score = 0.13), and negatively related to turbidity (score = -0.29) and potential nutrients DIN (score = -0.13) and SRP (score = -0.07) (Fig. 5a), while RDA2 was positively related to DIN (score = 0.20) and turbidity (score = 0.19), and negatively associated to depth (score = -0.28), SRP (score = -0.12) and macrophyte coverage (score = -0.10). Phylogroups D, E, F/G and B2 were strongly and negatively correlated with RDA1 suggesting their association with turbidity and higher nutrients conditions, while B1 was slightly and positively associated with it, thus distinguishing between dominant and less frequent phylogroups.

In addition, partial RDA of urban infrastructure predictors, controlling for the effect of environmental and spatial variables was statistically significant ($F_{3,24} = 17.39$, $P < 0.05$), with a significant association of phylogroup composition with drainage density ($F_{1,24} = 11.49$, $P < 0.05$), septic tanks density ($F_{1,24} = 9.12$, $P < 0.05$) and sanitary sewers density ($F_{1,24} = 7.27$, $P < 0.05$) (Fig. 4b). Significant axis RDA1 ($F_{1,24} = 25.53$, $P < 0.05$), with an explained proportion of 34%, was negatively related to septic tank density (score = -0.29) and drainage density (score = -0.24). Moreover, phylogroups B2, E and, to a lesser extent, B1 were negatively associated with RDA1, while F/G and D were positively associated with this axis (Fig. 5b). Axis RDA2, which accounted for 28% of the explained variance, was also significant ($F_{1,24} = 21.10$, $P < 0.05$) and positively related to the drainage and sanitary sewers densities (scores 0.33 and 0.31, respectively), while septic tanks density was negatively associated to this axis (score = -0.21). Phylogroups F/G and E, and cryptic clade IV and negatively related to RDA2, in contrast to phylogroups B2, D and B1.

Finally, partial RDA of spatial factors, constrained by environmental and urban infrastructure, showed that all the included AEMs (ranging from broad- to fine-scale resolution levels) were significant in explaining spatial variability (see Table S2 and Figure S2 in Appendix A2 of the Supplementary Material for more information).

Discussion

Our study indicates that the spatial structuring of *E. coli* in a highly polluted urban stream network with favorable growth conditions is strongly influenced by the hydraulic and sanitary infrastructure and local ecological features of the habitat. The spatial distribution of *E. coli* abundance and the phylogenetic composition of its population reflect the complex interactions between the socio-technological attributes of the urban environment and the ecological conditions of the stream habitat, supporting our original hypothesis. We found that both urban infrastructure and micro-habitat conditions, which were in a straight relationship with spatial predictors, acted as important drivers of *E. coli* abundance in the streams network studied. Moreover, *E. coli* abundance was negatively correlated with phylogroup diversity. On the other hand, *E. coli* phylogenetic composition was driven by independent effects of spatial, infrastructure and environmental variables, as well as by a strong negative relationship between phylogroup diversity and dominance. We hypothesize that these results reflect the influence of distinct ecological processes on *E. coli* populations, thereby opening future avenues of research in other urban watersheds and additional hydro-climatic conditions.

It has been shown that the effects of geomorphological modifications made to urban streams (channel incisions, rectifications, concrete riverbanks, buried sections) result in a process of ecological simplification and in the predominance of exogenous control over internal equilibrium states, leading to undesirable resilient states^{67,68}. In this vein, we hypothesize that the indivisible effect of infrastructure, environmental and spatial predictors explaining *E. coli* abundance patterns reflects the effect of large watershed-scale processes that cause a significant modification to local habitats, modulating microhabitat features that influence *E. coli* availability. Our hypothesis is supported by the negative co-variation observed between urbanization proxies (e.g., drainage density) and microhabitat features (e.g., macrophyte coverage).

Regarding the independent effects of environmental and urban conditions on *E. coli* abundance, we found a positive link with sanitary sewer density, which is in line with previous studies reporting the contribution of sewer leaks^{68,69}. Furthermore, the association of *E. coli* abundance with macrophyte coverage suggests that its growth may be favored in locations with lower urbanization levels and aquatic vegetation. Indeed, macrophytes are known to be colonizable surfaces for *E. coli*³². Once established, and if nutrient levels are sufficient, bacteria may grow and eventually release cells into freshwater bodies. Surbeck et al.⁸ reported that *E. coli* grows above threshold concentrations of DOC (7 mg/L) and SRP (0.07 mg/L), evidencing its ability to exploit soluble nutrients available in urban freshwaters. The nutrient concentrations detected across the basin were above these values, thus providing favorable environmental conditions for growth. In addition, we found a significant correlation between DOC and the presence of *E. coli*, while other potential nutrients (DIN and Iron) were also spatially correlated with DOC. Altogether, these results validate our first prediction of a positive association of *E. coli* abundance with urban income sources and high nutrient availability.

We found several remarkable aspects in relation to phylogenetic composition. Phylogroups A and B1 showed a co-dominant structure throughout the sampled sites, suggesting their wide ubiquity regardless of differences in environmental conditions. This is in agreement with Petit et al.¹⁶ who reported co-dominance of A and B1 in the water column, with an increase in the frequencies of phylogroups A, D and F in urban locations. With respect to human gut microbiomes in South America, Stoppe⁶⁹ found a predominance of phylogroup A, followed by B2, B1 and D, in an urban population from Brazil, in line with the results of Escobar-Páramo et al.⁷⁰. On this basis, the dominance of the phylogroups detected in the stream network studied here suggests that human fecal contamination is widespread throughout the watershed.

In our study, lower frequency phylogroups (D, E, F/G and B2) were associated with water turbidity, which is positively correlated with water flow and the local density of sanitary sewers around each sampled site, and higher nutrient availability (e.g., SRP and DIN), an association that can be interpreted as a product of common sources of contamination. Our results also suggest that the occurrence of lower frequency phylogroups (B2, D, E, F/G) are mostly associated with recent external contributions from urban drainage infrastructures or septic tank exfiltration. This is consistent with the idea that some of them (such as D or F) occur in highly urbanized locations. On the other hand, B1 was linked to contrasting conditions where aquatic vegetation prevails, which is partially consistent with our second prediction, while there were no strong associations to urban infrastructures. Bearing in mind that B1 can be naturalized in water bodies -thus forming part of the indigenous microbiota- and that this phylogroup is able to colonize surfaces like macrophytes through biofilm development, our results may be interpreted as a positive environmental effect on B1 growth^{18,19,71,72}. Finally, one strain belonging to the cryptic clade IV was isolated from a highly populated and urbanized location with low levels of sewerage infrastructure.

Altogether, our results can be interpreted in terms of different scenarios behind *E. coli* abundance and its phylogenetic composition in urban streams with favorable growth conditions (Fig. 6). A first scenario corresponds to locations with a relatively low level of disturbance, characterized by low infrastructure and stable habitat conditions, including the presence of high macrophyte coverage. These conditions, together with a high availability of nutrients, would promote high rates of bacterial growth and competitive exclusion of ecologically similar phylogroups. In this context, *E. coli* strains with the potential for survival and growth dominate the population structure, thus reducing the survival probability of lower frequency phylogroups. The second scenario involves highly disturbed locations with heavy external inputs. The lack of aquatic vegetation in local habitats and the periodic disturbances produced by drainage networks, may limit bacterial growth and offset competitive exclusion. The phylogenetic composition of *E. coli* populations in these habitats may be influenced by dispersal, facilitated through urban infrastructures and other diffuse sources like septic tanks leakage. Finally, the sanitary

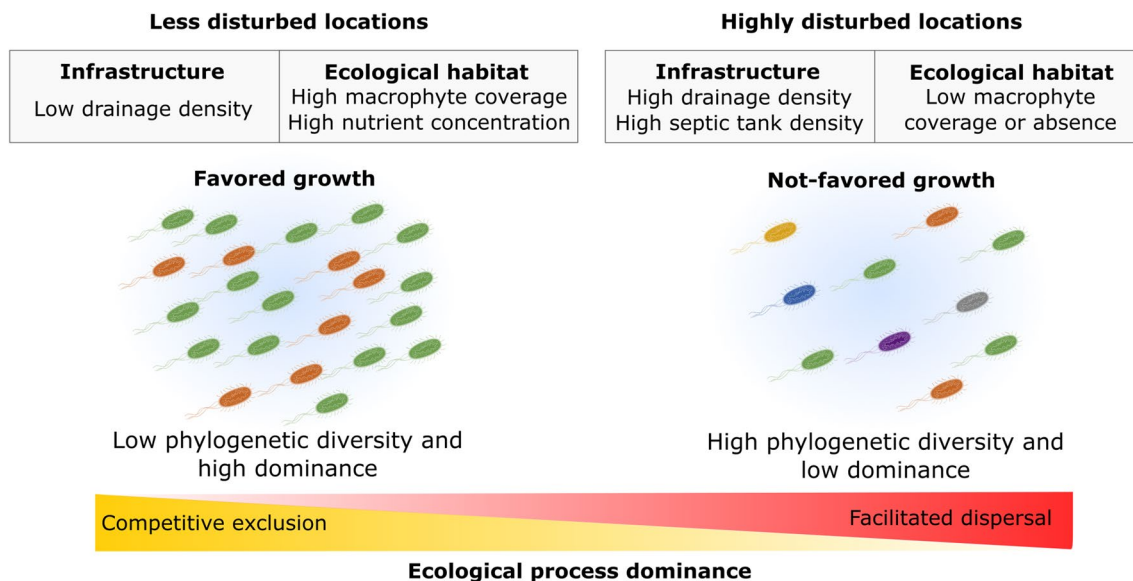


Figure 6. Hypothesized processes affecting the dynamics of *E. coli* population in an urban stream. Schematic diagram of contrasting scenarios along a gradient of environmental and urban conditions with different dominant ecological processes affecting the *E. coli* population. Graphic representation of *E. coli* from Togo picture gallery (Database Center for Life Science). <https://togovt.dbcls.jp/en>. Graphical edition in Inkscape V 0.92.4. Inkscape Project (2020). <https://inkscape.org>.

sewer network was also observed to positively contribute to the *E. coli* abundance, exerting its influence on *E. coli* occurrence irrespective of the environmental disturbance conditions.

Finally, we want to comment on some methodological issues of our analysis. First, we uncovered that the spatial predictors were relevant for explaining *E. coli* abundance and phylogenetic composition (Fig. 4). The selected broad- to fine-scale asymmetric eigenvectors reflect an expectable directional process through the stream network, involving a strong level of autocorrelation between close locations. They also reveal distinct spatially structured processes underlying transport-based (hydrological), spatial environmental or infrastructure patterns. It is worthy to note that the lack of updated information on public sanitary conditions and demographic characteristics may have introduced some bias into the selected AEMs showing pure spatial contributions. Therefore, the usage of spatial predictors allowed us to pull out spatial autocorrelation processes inherent to the structure of stream networks, enabling a more accurate analysis of the influence of local habitats or urban infrastructure features throughout the watershed. Second, the results and hypothetical mechanisms analyzed in this work must be assumed as general trends of the drivers for phylogroup diversity, since they arise from a single point in time, in favorable growth conditions (summer season). The use of presence/absence data due to limitations in sample size also implies a simplification of the real population structure, thus limiting further inferences. The analysis of the spatio-temporal dynamics of *E. coli* in urban several watersheds and the collection of a higher number of *E. coli* isolates will allow us to reach more extensive conclusions, while the present work offers contrastable hypothesis for further evaluation.

To conclude, our results provide information useful for global comparisons of *E. coli* persistence in the environment, while also having important implications for water quality assessment, as they question the use of *E. coli* as a proxy for recent microbiological pollution. In this regard, the genomic analysis of isolates, together with experimental assays may provide useful information on the determinants of *E. coli* environmental persistence and the differential fitness of phylogroups.

Received: 29 December 2020; Accepted: 7 July 2021

Published online: 19 July 2021

References

- 2012 *Recreational Water Quality Criteria*. (U. S. Environmental Protection Agency, 2012).
- Lee, C. M. *et al.* Persistence of fecal indicator bacteria in Santa Monica Bay beach sediments. *Water Res.* **40**, 2593–2602 (2006).
- 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).
- 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).
- Rochelle-Newall, E., Nguyen, T. M. H., Le, T. P. Q., Sengtaheuanghoung, 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).
- Tymensen, L. D. *et al.* Comparative accessory gene fingerprinting of surface water *Escherichia coli* reveals genetically diverse naturalized population. *J. Appl. Microbiol.* **119**, 263–277 (2015).

7. Ishii, S. & Sadowsky, M. J. *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes and environments / JSME* **23**, 101–108 (2008).
8. 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).
9. Jang, J. *et al.* Environmental *Escherichia coli*: Ecology and public health implications—A review. *J. Appl. Microbiol.* **123**(3), 570–581. <https://doi.org/10.1111/jam.13468> (2017).
10. Van Elsas, J. D., 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).
11. Jauregui, F. *et al.* Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* **9**, 560 (2008).
12. 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).
13. 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.* <https://doi.org/10.1111/1758-2229.12019> (2012).
14. 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**, 1–10 (2010).
15. Johnson, J. R. *et al.* Phylogenetic backgrounds and virulence associated traits of *Escherichia coli* isolates from surface waters and diverse animals in Minnesota and Wisconsin. *Appl. Environ. Microbiol.* **83**, 1–33 (2017).
16. 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).
17. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).
18. 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).
19. Méric, G., Kemsley, E. K., Falush, D., Saggars, E. J. & Lucchini, S. Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. *Environ. Microbiol.* **15**, 487–501 (2013).
20. Walk, S. T. The “Cryptic” *Escherichia*. *EcoSal Plus* **6**, 2 (2015).
21. Ingle, D. J. *et al.* Biofilm formation by and thermal niche and virulence characteristics of *Escherichia* spp. *Appl. Environ. Microbiol.* **77**, 2695–2700 (2011).
22. Elmqvist, T. *The Urban Planet: Knowledge Towards Sustainable Cities* (Cambridge University Press, 2018).
23. Hosen, J. D., Febria, C. M., Crump, B. C. & Palmer, M. A. Watershed urbanization linked to differences in stream bacterial community composition. *Front. Microbiol.* **8**, 1–17 (2017).
24. Wang, S.-Y., Sudduth, E. B., Wallenstein, M. D., Wright, J. P. & Bernhardt, E. S. Watershed urbanization alters the composition and function of stream bacterial communities. *PLoS ONE* **6**, e22972 (2011).
25. 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).
26. 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).
27. 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).
28. 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).
29. Handler, N. B., Paytan, A., Higgins, C. P., Luthy, R. G. & Boehm, A. B. Human development is linked to multiple water body impairments along the California coast. *Estuar. Coasts* **29**, 860–870 (2006).
30. Ishii, S. *et al.* Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. *Microbes Environ.* **25**, 8–14 (2010).
31. Whitman, R. L. *et al.* Microbes in beach sands: Integrating environment, ecology and public health. *Rev. Environ. Sci. Biotechnol.* **13**, 329–368 (2014).
32. Kleinheinz, G. *et al.* Effect of aquatic macrophytes on the survival of *Escherichia coli* in a laboratory microcosm. *Lake Reserv. Manage.* **25**, 149–154 (2009).
33. Moreira, S. *et al.* Persistence of *Escherichia coli* in freshwater periphyton: Biofilm-forming capacity as a selective advantage. *FEMS Microbiol. Ecol.* **79**, 608–618 (2012).
34. Pachepsky, Y. A. & Shelton, D. R. *Escherichia coli* and fecal coliforms in freshwater and estuarine sediments. *Crit. Rev. Environ. Sci. Technol.* **41**, 1067–1110 (2011).
35. Walsh, C. J. & Kunapo, J. The importance of upland flow paths in determining urban effects on stream ecosystems. *J. N. Am. Benthol. Soc.* **28**, 977–990 (2009).
36. McLellan, S. L., Fisher, J. C. & Newton, R. J. The microbiome of urban waters. *Int. Microbiol.* **18**, 141–149 (2015).
37. Newton, R. J. *et al.* Sewage reflects the microbiomes of human populations. *MBio* **6**, 1–9 (2015).
38. 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).
39. 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).
40. Ly, D. K. & Chui, T. F. M. Modeling sewage leakage to surrounding groundwater and stormwater drains. *Water Sci. Technol.* **66**, 2659–2665 (2012).
41. Graziano, M., Giorgi, A. & Feijóo, C. Science of the Total Environment Multiple stressors and social-ecological traps in Pampean streams (Argentina): A conceptual model. *Sci. Total Environ.* **765**, 142785 (2020).
42. Graziano, M. *et al.* Fostering urban transformations in Latin America: Lessons around the ecological management of an urban stream in coproduction with a social movement (Buenos Aires, Argentina). *Ecol. Soc.* **24**, 13 (2019).
43. Cirelli, A. F. & Ojeda, C. Wastewater management in Greater Buenos Aires, Argentina. *Desalination* **218**, 52–61 (2008).
44. Elordi, M. L., Lerner, J. E. C. & Porta, A. Evaluación del impacto antrópico sobre la calidad del agua del arroyo Las Piedras, Quilmes, Buenos Aires, Argentina. *Acta Bioquímica Clínica Latinoamericana* **50**, 669–677 (2016).
45. *Censo nacional de población, hogares y viviendas 2010 : censo del Bicentenario : resultados definitivos, Serie B nº 2.* (Instituto Nacional de Estadística y Censos, 2012).
46. Gordon, N. D., McMahon, T. A., Finlayson, B. L., Gippel, C. J. & Nathan, R. J. *Stream Hydrology: An Introduction for Ecologists* (Wiley, 2004).
47. Elosegui, A., Sabater, S. (eds.). *Conceptos y técnicas en ecología fluvial*. 243–251. (Fundación BBVA, 2009)
48. Baird, R. B., Eaton, A. D., Rice, E. W., & Bridgewater, L. (eds.) *Standard methods for the examination of water and wastewater*, **23**. (American Public Health Association, 2017).
49. Clermont, O., Bonacorsi, S., Bingen, E. & Bonacorsi, P. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**, 4555–4558 (2000).
50. 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).

51. Lescat, M. *et al.* Commensal *Escherichia coli* strains in Guiana reveal a high genetic diversity with host-dependant population structure. *Environ. Microbiol. Rep.* **5**, 9–57 (2013).
52. Clermont, O. *et al.* Evidence for a human-specific *Escherichia coli* clone. *Environ. Microbiol.* **10**, 1000–1006 (2008).
53. Congedo, L. Semi-automatic classification plugin for QGIS. *Sapienza Univ.* 1–25 (2013).
54. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008).
55. Borcard, D., Gillet, F. & Legendre, P. *Numerical Ecology with R* (Springer, 2018).
56. Legendre, P., Borcard, D. & Roberts, D. W. Variation partitioning involving orthogonal spatial eigenfunction submodels. *Ecology* **93**, 1234–1240 (2012).
57. Bivand, R. S. & Wong, D. W. S. Comparing implementations of global and local indicators of spatial association. *TEST* **27**, 716–748 (2018).
58. Magurran, A. E. *Measuring Biological Diversity* (Wiley, Hoboken, 2004).
59. Oksanen, J. *et al.* Vegan: Ecological Diversity. *R Project*, **368**. <http://cran.r-project.org> (2013).
60. Wilkinson, L. & Friendly, M. History corner the history of the cluster heat map. *Am. Stat.* **63**, 179–184 (2009).
61. Wei, T. *et al.* Visualization of a correlation matrix. *Statistician* **56**, 316–324 (2017).
62. Robinaugh, D. J., Millner, A. J. & McNally, R. J. Identifying highly influential nodes in the complicated grief network. *J. Abnormal Psychol.* **125**(6), 747 (2016).
63. 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).
64. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280 (2001).
65. Lepš, J. & Šmilauer, P. *Multivariate Analysis of Ecological Data Using CANOCO* (Cambridge University Press, Cambridge, 2003).
66. Simpson, G. Restricted permutations; using the permute package. <http://cran.r-project.org> (2012).
67. Booth, D. B., Roy, A. H., Smith, B. & Capps, K. A. Global perspectives on the urban stream syndrome. *Freshw. Sci.* **35**, 412–420 (2016).
68. Peipoch, M., Brauns, M., Hauer, F. R., Weitere, M. & Valett, H. M. Ecological simplification: Human influences on Riverscape complexity. *Bioscience* **65**, 1057–1065 (2015).
69. Stoppe, N. D. C. *et al.* Worldwide phylogenetic group patterns of *Escherichia coli* from commensal human and wastewater treatment plant isolates. *Front. Microbiol.* **8**, 2512 (2017).
70. Escobar-Páramo, P. *et al.* Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl. Environ. Microbiol.* **70**, 5698–5700 (2004).
71. 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. *Environ. Microbiol.* **9**, 2274–2288 (2007).
72. Touchon, M. *et al.* Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. *PLoS Genet.* **16**, e1008866 (2020).
73. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).

Acknowledgements

This work was financed by grants from CONICET (PUE 22920160100122CO). MS is recipient of a doctoral fellowship from CONICET. We thank Dr. Inés O’Farrell and Dr. Griselda Chaparro for their assistance during the sampling campaign, and one anonymous reviewer for helpful suggestions to improve variance partitioning analysis of phylogroup composition.

Author contributions

M.G. and M.S. conceived the scope and the general design of the study. M.G., M.S., S.G.L. and C.S.G. were involved in the sampling campaign. M.G., M.S., S.G.L., C.S.G. and B.G. carried out analytical determinations. N.F., M.G. and M.S. performed microbiological procedures. N.F., M.S., C.S.G. and S.G.L. performed molecular biology procedures and phylogroup identification. N.O. and M.S. participated in the characterization of the watershed infrastructure. M.G. and M.S. contributed to data and statistical analysis. M.G. and M.S. were responsible for drafting the article and figure design. All other authors provided critical input for the final manuscript. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-94198-1>.

Correspondence and requests for materials should be addressed to M.G.

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

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



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

© The Author(s) 2021