Title page

Characterizing the spatial signal of environmental DNA in river systems using a community ecology approach

Isabel Cantera[1]*, Jean-Baptiste Decotte[3], Tony Dejean[3,4], Jérôme Murienne[1,] Régis Vigouroux[5], Alice Valentini[4] and Sébastien Brosse[1]

[1]Laboratoire Evolution et Diversité Biologique, UMR5174, CNRS, IRD, Université Toulouse III Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse, France.

isa_cantera@hotmail.com, jerome.murienne@univ-tlse3.fr, sebastien.brosse@univ-tlse3.fr

[3]VIGILIFE, 17 rue du Lac Saint-André Savoie Technolac - BP 274, Le Bourget-du-Lac 73375, France. jb.decotte@vigilife.org

[4]SPYGEN, 17 rue du Lac Saint-André Savoie Technolac - BP 274, Le Bourget-du-Lac 73375, France. tony.dejean@spygen.com, alice.valentini@spygen.com

[5]HYDRECO, Laboratoire Environnement de Petit Saut, B.P 823, F-97388, Kourou Cedex, French Guiana. regis.vigouroux@hydrecolab.com

Abstract

Environmental DNA (eDNA) is gaining a growing popularity among scientists but its applicability to biodiversity research and management remain limited in river systems by the lack of knowledge about the spatial extent of the downstream transport of eDNA. Up to now, attempts to measure eDNA detection distance compared known species distributions to eDNA results, limiting therefore studies to a few intensively studied rivers. Here we developed a framework to measure the detection distance of eDNA in rivers based on the comparison of faunas across an increasing range of spatial extents, making it independent from knowledge on species distributions. We hypothesized that under short detection distance the similarity between fish faunas should peak between nearby sites, whereas under long detection distance each site should cumulate species from a large upstream area. Applying this framework to the fish fauna of two large and species rich Neotropical river basins (Maroni and Oyapock), we show that fish eDNA detection distance did not exceed a few kilometers. eDNA hence provided inventories of local species communities. Those results were validated by retrieving the distance decay of species similarity, a general pattern in ecology based on the decline of local species community similarity with spatial distance between them. We finally compared species distribution derived from eDNA to the known distribution of the species based on capture data, and this comparison also confirmed a global match between methods, testifying for a short distance of detection of the fauna by eDNA.

Keywords: Neotropical fishes, metabarcoding, detection distance, spatial patterns, freshwater.

Introduction

In aquatic ecosystems, DNA molecules flowing in the water after being separated from their source organisms can be assigned to taxa using the environmental DNA (eDNA) metabarcoding approach. The method consists on the sampling, extraction, amplification and sequencing of free DNA suspended in the water or linked to suspended particles. Subsequently, the obtained sequences taxonomically assigned to species based on comparisons to a reference molecular database (Taberlet *et al.* 2018). This approach has proven to be particularly efficient at measuring the diversity of fishes (Civade *et al.* 2016; Hänfling *et al.* 2016; Olds *et al.* 2016; Valentini *et al.* 2016; Evans *et al.* 2017; Pont *et al.* 2018; Cantera *et al.* 2019), amphibians (Valentini *et al.* 2016; Lopes *et al.* 2017) and macro-invertebrates (Pri *et al.* 2020) in aquatic ecosystems. Indeed, in those studies, the method provided similar or more complete inventories than traditional methods, such as netting, electrofishing as well as visual and audio records, at the same sites. The great detection performance of the method may be due to its high sensitivity to detect species that are rare and difficult to sample with traditional methods (Ficetola *et al.* 2008; Jerde *et al.* 2011; Goldberg *et al.* 2011). In fact, the species list obtained with one eDNA sampling session was often comparable to cumulative traditional sampling sessions and historical records (Civade *et al.* 2016; Hänfling *et al.* 2016; Nakagawa *et al.* 2018; Pont *et al.* 2018; Cantera et al. 2019). Due to its sampling efficiency to reconstruct whole aquatic communities, eDNA methods are expected to revolutionize survey methods for ecological research and biodiversity monitoring on aquatic ecosystems (Rees *et al.* 2014, 2015; Keck *et al.* 2017; Zinger *et al.* 2020). This is particularly true for large rivers where sampling the fauna can be very difficult, as it is time consuming, costly and limited to specific habitats (Pont *et al.* 2018).

The application of eDNA has been however limited in running waters because the species detection recovered by eDNA sampling is decoupled from the species' physical habitat because of downstream eDNA transport. The network structure of rivers is governed by a continuous and directional flow (McCluney et al. 2014) that can transport eDNA downstream from the DNA source while it degrades (Barnes and Turner 2016). The spatial extent of the downstream transport depends on the degradation rate of eDNA (Barnes and Turner 2016) combined with the discharge rates of the water system (Jane *et al.* 2015). For instance, fast-flow systems may transport eDNA downstream over large distances before its degradation and dilution. Moreover, the identity and density of target species were found to influence eDNA detection distance (Deiner and Altermatt 2014; Pilliod *et al.* 2014; Pont *et al.* 2018). In aquaria and mesocosms, eDNA remained detectable from a few days to a few weeks after its release in the water (see Table 1, Barnes and Turner 2016). Besides being variable, these experimental results are difficult to relate with water velocity in the field to estimate eDNA detection distance, given that natural ecosystems are more complex and local environmental conditions, which influence eDNA degradation rate (Barnes and Turner 2016), are covarying. For instance, Takahara *et al.* (2012) found a significant relationship between eDNA production from common carp and water temperature in a natural lagoon but not in aquaria. Moreover, in experimental designs the DNA of target species are usually in higher concentrations compared to natural ecosystems while DNA concentration was shown to influence degradation rate (Barnes and Turner 2016).

Deiner *et al.* (2016) advocated that eDNA-based inventories represent a spatially integrated measure of biodiversity describing the diversity at the catchment scale and considered that rivers act as conveyor belts for eDNA. Therefore, eDNA collected at one sampling point may

represent the combination of local and upstream diversity from this point. eDNA signal was however shown to exhibit variability in space, as it accurately described the spatial succession of fish communities along a longitudinal gradient from a lake to downstream in French rivers (Civade *et al.* 2016; Pont *et al.* 2018). The assumption that eDNA of entire communities will accumulate downstream indeed omits the phenomenon of DNA degradation and sedimentation once it has been shed from an organism (Barnes and Turner 2016; Pont et al. 2018). The detection distance of eDNA was investigated more explicitly in natural rivers and streams in temperate ecosystems. Although the studies agreed that eDNA from a species can be detected downstream from where the species occurs and that eDNA signal decays over distance, the estimation of detection distance displays a great variability among studies. The introduction of caged animals in streams revealed that eDNA was detected 5 m but not 50 m downstream from caged Idaho giant salamander (Pilliod *et al.* 2014) and up to 1 km downstream from caged brook trout (Wilcox *et al.* 2016). Contrastingly, eDNA from lake species has been found in downstream streams over greater distances. DNA from a lacustrine fish species was detected up to 3 km downstream from the outlet of a lake by Civade *et al.* (2016), and Deiner and Altermatt (2014) detected two lake invertebrate species (a planktonic crustacean and a mollusc) up to 12.3 km downstream from a lake. Nakagawa *et al.* (2018) assessed detection distance by comparing eDNA fish inventories with observational data within intervals from 1 to 10 km upstream from the eDNA sampling site, and showed eDNA inventories were more similar with observational data when comparing with traditional inventories within 6 km upstream from the eDNA sampling site. Pont *et al.* (2018) detected eDNA of Whitefish, an abundant species in Lake Geneva, up to 130 km downstream from the lake, whereas the detection of less abundant species was restricted to a few kilometres, suggesting that abundant species are detected farther downstream than rare ones.

The variability among detection distances may reflect the specific local environmental conditions of the studied water bodies (Jane *et al.* 2015; Barnes and Turner 2016). Local conditions such as acidity, oxygen demand, primary production, water temperature, solar radiation, organic materials and the activity of microorganisms, have been shown to influence eDNA degradation rate (Barnes and Turner 2016; Eichmiller *et al.* 2016; Seymour *et al.* 2018). As all the above-mentioned characteristics as well as discharge rate vary strongly between rivers, the spatial extent of downstream transport of eDNA may be highly dependent of the studied system. For example, by treating eDNA downstream transport as the vertical transfer of fine particulate organic matter from the water column to the riverbed, Pont *et al.* (2018) suggested that detection distance varies with river size, ranging from a few km in small streams to more than 100 km in large rivers.

Describing species communities using eDNA requires characterizing the spatial signal of eDNA and hence determining the spatial extent of an eDNA sample. This constitutes a key step to the use of eDNA in ecological research and biomonitoring. The above-mentioned studies reached this aim because they targeted species or communities with well-known spatial distributions. However, one of the main advantages of eDNA is to provide biodiversity inventories that cannot be performed by traditional methods, as in remote and highly diverse tropical freshwater ecosystems (Cilleros *et al.* 2018; Zinger *et al.* 2020). Here, we propose a methodological framework to characterize the extent of eDNA spatial signal to detect community structure independently from the known distribution of the species. This approach is convenient because there is no need to use prior knowledge of species distribution or site diversity, making it applicable to any taxa or river ecosystem. Along a river course, our framework compares the species similarity between a given site (called focal site) and the sites located upstream, allowing to test for spatial integration distance

of eDNA (Figure 1**A**). If eDNA is transported over long distances, community similarity should be maximal when comparing focal site species diversity to the species diversity combining sites over large upstream spatial extents. In contrast, if eDNA detection distance is short, the species similarity will peak between nearby sites (Figure 1**B**). This framework was tested in the Maroni and Oyapock rivers, two Neotropical rivers differing in their anthropization levels. We then tested the ability of eDNA data to describe a well-established spatial patterns in ecology, the decay of similarity between communities with increasing distance, which postulates that the similarity between local measures of biodiversity should decline with the distance between observations (Soininen *et al.* 2007; Morlon *et al.* 2008). This widely recognised ecological pattern should be retrieved if eDNA detection distance is short, whereas it should be blurred under the hypothesis of a long-distance transport of eDNA. Finally, we compared the distribution range of detected species along the upstream-downstream gradient of the river to the distribution range recorded by capture-based data in the same river. Again, under the hypothesis of a local eDNA detection of the species, the distribution range of species retrieved using eDNA should fall within the known species range, while under the hypothesis of a long-distance transport, species should be detected downstream from their known distribution range.

Materials and methods

Study sites

This study was conducted in two tropical rivers bordering French Guiana (Figure 2) and located in the Northern East of the Amazonian region (*sensu lato,* including Guiana shield and Amazon river drainage). The Maroni river, measuring 612 km from the source to the estuary, separates the West

of French Guiana from Surinam and the Oyapock river (404 km long) which separates the East of French Guiana from Brazil. Both rivers host typical freshwater Amazonian fauna with more than 270 and 190 described fish species in the Maroni and Oyapock, respectively (Le Bail *et al.* 2012). Among these species, 141 and 119 species have been recorded in the main channel of the Maroni and Oyapock rivers, excluding estuarine species (Planquette *et al.* 1996; Keith *et al.* 2000). Despite being located in one of the largest unfragmented rainforest areas in the world, the Maroni river is facing an unprecedented rise of land use changes due mainly to gold-mining activities but also to agriculture and urbanization (Rham *et al.* 2017; Gallay *et al.* 2018). Contrastingly, the Oyapock river remains less impacted, with low human population density and gold mining activities limited to a few tributaries (Gallay *et al.* 2018).

eDNA sampling

We collected environmental DNA samples in 53 sites along the main channel of the two rivers (Figure 2), with 28 sites sampled on the Maroni channel (excluding the estuary) and 22 sites sampled on the Oyapock channel (excluding the estuary). The estuary sections, from the sea to the upstream area influenced by the tide, were not considered in this study, because tidal water movements could bias our results through downstream-upstream movement of eDNA during the rising tide. The sampling was achieved during a 2 weeks field session during the dry season (October- November) in 2017 and 2018 for the Maroni and Oyapock, respectively. In all sites, the river was wider than 20 meters and deeper than one meter (Strahler orders 4 to 8). The distance between adjacent sites ranged from 1.6 to 31.8 km (mean: 13.8 and SE: 1.4) in the Maroni, and from 1.2 to 23.8 km (mean: 10.8 and SE: 1.5) in the Oyapock. In both rivers, sites were sequentially

sampled from downstream to upstream, with 1 to 4 sites sampled per day, according to the travel time between sites.

Following the protocol implemented by Cantera et al. (2019), we filtered two samples of 34 liters of water at each site to collect eDNA. A peristaltic pump (Vampire sampler, Burlke, Germany) and a single-use tubing were used to pump the water into a single-use filtration capsule (VigiDNA 0.45 µm; SPYGEN, le Bourget-du-Lac, France). All the material used to collect water was single use, and therefore replaced between each sample. Single use gloves were also used to avoid contamination. To collect eDNA, the input part of the tubing was placed 10 to 20 centimetres below the surface in areas with high water flow as recommended by Cilleros et al. (2018). Sampling was achieved in turbulent area (rapid hydromorphologic unit) to ensure an optimal homogenization of the eDNA throughout the water column. To avoid DNA contamination among sites, the operator always remained downstream from the filtration area and stayed on emerging rocks. At the end of the filtration, the filtration capsule was emptied of water, filled with 80 mL of CL1 conservation buffer (SPYGEN) and stored in individual sterile plastic bags kept in the dark.

eDNA laboratory and bioinformatic analyses

For DNA extraction, each filtration capsule was agitated for 15 min on an S50 shaker (cat Ingenieurbüro™) at 800 rpm and then emptied into a 50 mL tube before being centrifuged for 15 min at 15,000×g. The supernatant was removed with a sterile pipette, leaving 15 mL of liquid at the bottom of the tube. Subsequently, 33 mL of ethanol and 1.5 mL of 3M sodium acetate were added to each 50 mL tube and stored for at least one night at -20°C. Tubes were centrifuged at 15 000 ×g for 15 min at 6°C, and the supernatants were discarded. After this step, 720 µL of ATL

buffer from the DNeasy Blood & Tissue Extraction Kit (Qiagen) was added. The tubes were then vortexed, and the supernatants were transferred to 2 mL tubes containing 20 µL of proteinase K. The tubes were finally incubated at 56°C for two hours. Afterwards, DNA extraction was performed using NucleoSpin Soil (Macherey-Nagel GmbH & Co., Düren Germany) starting from step six and following the manufacturer's instructions. The elution was performed by adding 100 µL of SE buffer twice. Four negative extraction controls were also performed. They were amplified and sequenced in parallel to the field samples to monitor possible laboratory contaminants. After the DNA extraction, the samples were tested for inhibition by qPCR (Biggs et al. (2015). If the sample was considered inhibited, it was diluted 5-fold before the amplification.

We performed DNA amplifications in a final volume of 25 µL including 1 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA), 10 mM of Tris-HCl, 50 mM of KCl, 2.5 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of "teleo" primers (Valentini *et al.* 2016) and 3 µL of DNA template. We also added human blocking primer for the "teleo" primers with a final concentration of 4 µM and 0.2 µg/µL of bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland) to the mixture. We performed 12 PCR replicates per field sample. The forward and reverse primer tags were identical within each PCR replicate. The PCR mixture was denatured at 95°C for 10 min, followed by 50 cycles of 30 s at 95°C, 30 s at 55°C and 1 min at 72 °C and a final elongation step at 72°C for 7 min. This step was done in a room dedicated to amplified DNA with negative air pressure and physical separation from the DNA extraction rooms (with positive air pressure). We also amplified 14 negative extraction controls and five PCR negatives controls and sequenced them in parallel with the PCR replicates. We pooled the purified PCR products in equal volumes to achieve an expected sequencing depth of 500,000 reads per sample before the

libraries preparation. Ten libraries were prepared using a PCR-free library protocol (https://www.fasteris.com/metafast), at Fasteris facilities (Geneva, Switzerland). Four libraries were sequenced using an Illumina HiSeq 2500 (2x125 bp) (Illumina, San Diego, CA, USA) and the HiSeq SBS Kit v4, three using a MiSeq (2x125 bp) (Illumina, San Diego, CA, USA) and the MiSeq Flow Cell Kit Version3 (Illumina, San Diego, CA, USA) and three using a NextSeq (2x150 bp+8bp) (Illumina, San Diego, CA, USA) and the NextSeq Mid kit (Illumina, San Diego, CA, USA). The libraries ran on the NextSeq were equally distributed in four lanes. Sequencing were performed following the manufacturer's instructions at Fasteris facilities (Geneva, Switzerland).

The sequence reads were analyzed using the OBITools package (Boyer *et al.* 2016) following the protocol described in Valentini et al. (2016). The ecotag program was used for the taxonomic assignment of molecular operational taxonomic units (MOTUs). We kept only the MOTUS with a similarity to our reference database above 98%. Our local reference database represent an updated version of the reference database available in Cilleros et al. (2018), resulting in 251 Guianese fish species. The GenBank nucleotide database was checked but Guianese fishes being poorly referenced (most of the sequences are from Cilleros et al. (2018)), it did not provided additional information in our case. We discarded all MOTUs with a frequency of occurrence below 0.001 per library in each sample, considered as tag-jumps (Schnell *et al.* 2015). These thresholds were empirically determined to clear all reads from the extraction and PCR negative controls included in our global data production procedure as suggested by De Barba et al. (2014) and Taberlet et al. (2018). For the samples sequenced on a NextSeq platform, only species present in at least two lanes were kept. The entire procedure was repeated for each sample, giving rise to two species lists per field site. The two obtained species lists per site were finally pooled to obtain a

single species inventory per site, as Cantera *et al.* (2019) showed that pooling two replicate field samples was sufficient to inventory most of the fauna in highly diverse tropical aquatic ecosystems.

Methodological framework to measure detection distance

To determine the detection distance of eDNA at the community scale, we tested to which extent species similarity changed when comparing the fish fauna detected in each site to the fish fauna detected upstream from the site (Figure 1**A**). First, the fish community detected in a site (hereafter called the focal site) was compared to the fish community detected in the closest upstream site from this focal site ("Sim1" in Figure 1**A)**, accounting for an integration distance corresponding to the watercourse distance between these two sites ("ID1" in Figure 1**A**). Then, the integration distance was increased to the next upstream site ("ID2" in Figure 1**A**) and the species similarity was calculated between the focal site and the pool of species detected in the two upstream sites ("Sim2" in Figure 1**A**). This procedure was repeated while increasing integration distance up to the most upstream sampling site ("Sim 3" and "ID 3" in Figure 1**A**). All the sites except the most upstream site were thus successively considered as a focal site.

Species similarity was measured using the Jaccard similarity index (1- Jaccard dissimilarity; Koleff *et al.* 2003), which varies from one, indicating a complete similarity between the compared faunas, to zero indicating a total independence between the two faunas. The Jaccard similarity index is sensitive to the number of considered species, which mechanically increases when increasing integration distance. For each comparison, we controlled species richness differences between focal sites and the pooled fauna from upstream sites using null models where the number of species detected in the focal site was randomly sampled in the upstream pooled

fauna. This simulation was repeated 1000 times for each upstream fauna, generating 1000 similarity values between each upstream fauna and the focal site. Then, the 1000 similarity values obtained for each comparison were averaged to obtain a single value of similarity for each focal site and the integration distance. In a few cases (14 out of the 378 comparisons for the Maroni and 12 out of the 300 comparisons for the Oyapock), the focal site had higher species richness than the upstream considered site(s). In this particular case, the number of upstream species was randomly sampled in the focal site (1000 iterations, as above), to ensure comparisons between communities having the same species richness. The entire procedure was achieved independently for the Maroni and the Oyapock rivers.

We plotted Jaccard similarities according to integration distances for each river. A decreasing trend would indicate that close communities have the highest similarity in species composition, revealing a short distance of detection of eDNA (top panel of Figure 1**B**). In contrast, an increasing trend would indicate that similarity increases with integration distance, and therefore that the river acts as a long-distance conveyor belt for eDNA (bottom panel of Figure 1**B**). Finally, under an intermediate detection distance, similarity between faunas should peak for intermediate values of integration distance (middle panel of Figure 1**B**). For each river, community similarity was modelled as a function of integration distance using linear regressions.

eDNA data ability to reveal spatial patterns in ecology

We first tested the ability of eDNA data to reveal patterns of decay of similarity in species composition between pairs of sites with increasing geographic distance. Presence/absence matrices of the species detections in each site was used to build a species compositional distance matrix

between pairs of sites, for each river separately with the *vegdist* function. Pairwise compositional distances between sites were calculated using the Jaccard index. The correlation between the species similarity matrix and the matrix of watercourse distances between sites (in km) was tested using the Mantel test with 999 permutations and the Spearman correlation coefficient. This test determines the significance level of the analysed correlation by permuting 999 times the rows and columns of one matrix and comparing the permuted distribution with the value of the observed statistic for the actual data. Considering that the environment has been shown to influence the rate of similarity decay (Nekola and White 1999), differences in the rates of decay in similarity with distance between rivers were tested. To do this, the slopes of the decay relationship found in each river were compared by using a permutation procedure with 999 iterations based on linear regressions of two datasets using the *diffslope* function of the 'simba' R package. This function compares the observed difference between the slopes of the linear regressions from each river with permuted slope differences by randomly interchanging the values between the two river datasets. The p-value is computed as the ratio between the number of cases where the differences in slopes exceed the difference in slope of the actual configuration. For the Maroni river, we compared the obtained pattern to the one retrieved with capture-based methods in five sites sampled annually during the dry season from 2007 to 2016 (Figure 2). Yearly data was combined to provide a broad characterisation of the fish fauna in each site. The fish captures were performed using a standardized gill-net protocol designed by Tejerina-Garro and De Mérona (2001) and used as a routine biodiversity monitoring protocol by the French ministry of Environment, as part of the European Water Framework directive. The 9 years of data per site were pooled to deal with the limited efficiency of gillnets to collect the entire fauna, gill nets being strongly dependent of fish activity and needing repeated sampling effort to provide realistic species community inventories

(Murphy and Willis 1996). We here hypothesise that under a short eDNA detection distance, distance decay of similarity between eDNA and capture based methods should be similar. This was done for the Maroni river only because the main channel of the Oyapock river was only sampled in two sites with the same capture-based protocol.

Second, we compared the spatial upstream-downstream range of each detected species to the known distribution range of the species derived from standardized captures (from the Maroni river only, see above). We used the watercourse distance from the river mouth (in km) of the sampled sites to represent the spatial distribution range along the upstream-downstream gradient. To keep the spatial range comparable between the two sampling methods, we discarded the most upstream sites (S1 to S6, see Figure 2), that were not investigated using gill nets. Moreover, gill net being selective for species (*e.g.* small and/or elongated species are not captured), we restricted this analysis to the species detected by both eDNA and by gill nets (in at least two sites or sampling events). Indeed, the 11 species detected only once over the 9 years' gillnet surveys represent "bycatches" of small species (<5 cm) such as *Characidium zebra* or *Hemigrammus surinamensis*, which are hardly captured using gill-nets. We here predict that under a short eDNA detection distance, species detections using eDNA should not systematically occur downstream from the known distribution range of the species as established by gill-net data.

Results

Fish species were detected in all the eDNA replicates. The number of species per site ranged from 30 to 94 (mean = 57) in the Maroni sites and from 50 to 71 (mean = 62) in the Oyapock sites. We detected a total of 160 species, with 116 and 96 species in the Maroni and Oyapock rivers,

respectively. They account for more than 80% (82.3% and 80.7% for the Maroni and Oyapock, respectively) of the fish species known from the main channel of both river basins, excluding estuarine species (Planquette *et al.* 1996; Le Bail *et al.* 2000). Among the five Maroni sites sampled with capture-based methods (gill-nets), a total of 94 fish species were captured, with species richness ranging from 50 to 66 per site. The comparison of eDNA (22 sites, sites S7 to S28, Figure 2) and gill-nets (5 sites) inventories revealed that 53 species were detected by both methods in the Maroni river.

Similarity values calculated using our theoretical framework were significantly and negatively related with the log-transformed integration distance in both rivers ($p<0.001$, Figure 3). Figure 3 shows that in both rivers, the similarity in species composition is maximal when comparing the species diversity of focal sites to the species diversity in nearby upstream sites. Similarity values showed a steep decline between 0 and 50 km of integration distance. Above 50 km of integration distance, the decline of similarity with distance slowed and saturated towards Jaccard similarities of 0.25 and 0.5 for the Maroni and Oyapock rivers, respectively.

Mantel tests showed that Jaccard similarity matrices were significantly and negatively related to the watercourse distance between sites pairs in both rivers (simulated p-values <0.001, Mantel statistic $R= -0.8$ and -0.7, for Maroni and Oyapock rivers, respectively). As sites became more distant, their corresponding fish communities became more dissimilar in terms of community composition (Figure 4). For both rivers, the similarity decay with distance relationship fitted well without obvious change of the slope of the regression in any distance value. The negative effect of spatial distance on species similarity was significantly higher ($p = 0.001$) in the Maroni river (slope value = -1.255e-03) than in the Oyapock river (slope value = -1.085e-03). Conversely, the distance

decay of similarity was not significant in the sites sampled with gill-nets in the Maroni river (simulated p-value = 0.077, Mantel statistic R = -0.6, Figure 4**A**) despite a trend of similarity decrease with distance between site pairs.

The comparison of the species spatial ranges along the main channel of the Maroni river provided by the eDNA approach and by gill-net captures revealed a global match of the species distribution between methods (Figure 5). Among the 53 species detected by both methods, 74.5% of the fauna (38 species) was not detected by eDNA downstream from capture-based records. Indeed, distribution ranges were similar between methods for 20 species (39% of the fauna, e.g. *Hoplias aimara*, HAIM; *Eigenmannia virescens*, EVIR; *Brycon falcatus*, BFAL; Figure 5) and for 18 species (35.3% of the fauna) the capture-based method detected the species downstream from the distribution range found using eDNA (*e.g. Moenkhausia georgiae*, MGEO; *Pristobrycon eigenmanni*, SHUM; Figure 5). Only 25.5% of the considered fauna (13 species) were detected using eDNA downstream from the distribution provided by capture-based records (*e.g. Electrophorus electricus*, EELE; *Metaloricaria paucidens*, MPAU; *Geophagus harreri*, GHAR, *Guianacara owroewefi*, GOWR; Figure 5). Furthermore, eDNA and capture-based records retrieved consistent wide distribution ranges along the main channel of the river (*i.e.* species detected in almost all the eDNA and gill-net sites) for 21 species (*e.g. H. aimara*, HAIM; Figure 5). Nevertheless, only six species (*G. harreri*, GHAR; *P. barbatus*, PBAR; *Acnodon oligacanthus*, AOLI; *Pseudoplatystoma fasciatum*, PFAS; *E. electricus*, EELE) had a wide distribution range in eDNA records, but not in capture-based records.

Discussion

The theoretical framework that we developed to measure the spatial extent of eDNA signal of fish communities in running waters did not support the hypothesis of eDNA drift over large downstream distances as suggested by Deiner *et al.* (2016) in temperate European rivers. Here we showed that faunistic similarity was maximal between nearby sites, whereas it declined when considering large integration distances (as illustrated in the top panel of the theoretical figure 1B). Our results therefore support previous studies suggesting a short downstream detection of eDNA between 50 m and 6 km reported in temperate streams (Pilliod *et al.* 2014, Civade *et al.* 2016, Wilcox *et al.* 2016; Nakagawa *et al.* 2018). Nevertheless, we cannot exclude that the eDNA of some species, probably the most abundant ones, can be transported over long distances as shown by Pont *et al.* (2018). In the case that this occurs, it would certainly be limited to a few species in our samples as this potential bias did not blurred the significant decline of faunistic similarity with increasing integration distance. Ours results do not allow to provide a precise measure of eDNA integration distance over the upstream-downstream continuum due to the strong variability of similarity values at short distances. This variability accounts for the local environmental variability that determines niche availability and therefore selects the local fish fauna at local scale (Weiher and Keddy 1999; Jackson et al. 2001; Webb *et al.* 2002). Those local environmental effects probably also explain the steep decline of faunistic similarities at low integration distances for both rivers (less than 50km). Integrating faunistic information over larger scales (more than 50km) may buffer local environmental effects (increasing integration distance increases the probability to consider local environmental conditions similar to those of the focal site), and reflect biogeographic effects responsible for the turnover in species composition across the upstream-downstream continuum (Weiher and Keddy 1999; Webb *et al.* 2002; Cilleros *et al.* 2018).

The limited spatial grain at which the fauna was detected using eDNA, makes this method appropriate to test ecological theories, such as the distance decay of similarity, a general rule in ecology. Distance decay theory predicts a decrease of the similarity between communities with the spatial distance separating them (Nekola and White 1999). This rule has already been verified in river ecosystems where it reflects the dynamic and heterogeneous architecture of river systems that shape community composition at different spatial scales along the river continuum (Muneepeerakul *et al.* 2008). However, the distance decay of similarity has rarely been examined in Neotropical aquatic diversity (Araújo *et al.* 2013), and we here confirm this trend for the two studied rivers. Importantly, the fact that eDNA data was able to reveal distance decay of faunistic dissimilarity patterns reinforces the hypothesis of a short downstream transport of eDNA. Indeed, according to Nekola and White (1999), the grain size, corresponding to the contiguous area over which a single observation is made (corresponding to the eDNA integration distance in the present study), should be smaller than the whole spatial extent of the study. This will allow to capture variation in similarities among comparisons and thus observe distance decay patterns. In our case, this means that the spatial signal of eDNA to detect one fish community need to be smaller than the sampled portion of the river, which is verified for the two rivers, and remains true even when considering only a restricted section of each river. Moreover, the slope of similarity decay differed between the two rivers, with a more marked decline in the Maroni river, suggesting a more marked decrease of environmental similarities between sites and/or more dispersal constraints with distance in this river (Nekola and White 1999; Soininen *et al.* 2007). It can be expected that the decrease of environmental similarities and the increase of dispersal constraints are reinforced by human impacts in downstream sites of the Maroni river resulting in a higher rate of similarity decay with distance. Indeed, the Maroni river faces six fold higher deforestation levels than the Oyapock river

(0.37% *versus* 0.06% of catchment area, Gallay *et al.* 2018). Deforestation was mainly caused by gold-mining activities that deeply impact sediment fluxes in the Maroni River (Gallay *et al.* 2018) and thus strengthen environmental constrains for species (Hammond *et al.* 2007; Brosse *et al.* 2011). This difference between river basins was significantly observed in our eDNA samples, testifying for the relevance of ecological patterns retrieved using eDNA. Furthermore, the distance decay of similarity pattern was more pronounced with eDNA sampling than with capture-based samplings, indicating that the ecological information derived from eDNA was more relevant than that derived from gill-net samples. This can be explained by the known limitations of gill-net samples that are particularly selective in species and habitats (Murphy and Willis 1996) and therefore only provide a partial image of the fauna occurring in each site (Cilleros *et al.* 2018).

Comparing species distribution ranges retrieved from eDNA and capture samples using gill-nets revealed a global match in more than 70% of the species. More importantly, only 25% of the species were detected using eDNA downstream from their distribution range recorded using capture-based methods. Although this could be considered as species detections due to a downstream drift of eDNA, 10 out of these 13 species have an extended spatial distribution in the Maroni drainage basin and are known from the downstream reaches of the Maroni (Planquette et al. 1996; Keith et al. 2000). The absence of those species in the recent gill net captures is probably explained by their low capturability due to their morphological and ecological characteristics (Murphy and Willis 1996). For instance, the anguilliform body shape of *E. electricus* (EELE) makes the species difficult to capture with gill-nets, and the sedentary or territorial behaviour of *M. paucidens* (MPAU), *G. harreri* (GHAR) and *G. owroewefi* (GOWR) or *K. Itany* (KITA) reduces the probability to capture them with passive sampling gears such as gillnets. Conversely, the 20

species having consistent distributions between methods are species having high capture probability, such as large bodied (*e.g. H. aimara*, HAIM), mobile and gregarious species (*e.g. Brycon falcatus*, BFAL; *Hemiodus unimaculatus*, HEUN), or species with spines and ossified fins that easily tangle in the nets (*e.g. Acuchenipterus nuchalis*, ANUC; *Doras micropoeus* DMIC, *Pimelodus ornatus*, PORN).

Conclusions

We here demonstrated that fish inventories achieved using eDNA provide a local inventory of the fauna, making this biodiversity data appropriate to describe spatial patterns of fish communities. The scale of eDNA detection certainly not accounts for the local habitat of the species that has to be measured over a few square meters. Nevertheless, a detection distance encompassing several hundred of meters to a few kilometers appears as a reasonable sample grain size for studying communities or to consider anthropic impacts on fish communities (Tonn 1990; Jackson *et al.* 2001). Detection distance might vary across rivers and biomes where different physical and chemical characteristics of the rivers may affect eDNA persistence in the water (Barnes and Turner 2016; Eichmiller *et al.* 2016; Seymour *et al.* 2018). Nevertheless, the developed framework allows to quantify the detection distance of eDNA irrespective of the river or the taxonomic group considered. This framework is also independent from the current knowledge on species distribution in river systems, which up to now has limited the attempts to measure detection distance in well-studied ecosystems, where comparisons between eDNA species inventories to known species distributions are feasible (*e.g*. Pont et al. 2018). Our framework, based on species communities, can also be used for taxa or river ecosystems lacking molecular reference databases, as Molecular Operational Taxonomic Units (MOTUs; Floyd *et al.* 2002) can be used as surrogates to taxonomic

knowledge in community ecology (Ryberg 2015). Given the lack of ecological knowledge for a substantial part of the world rivers, and the deep incompleteness of local distribution knowledge for the major part of the freshwater fauna of the globe, our framework allows to determine the grain size of eDNA samples in all rivers, a prerequisite to use eDNA as a tool in ecological research and biodiversity management in those understudied areas.

Acknowledgements

This work was funded by DGTM Guyane (project DISTDET), Investissement d'Avenir grant DRIIHM (ANR-11-LABX-0010), through the OHM Oyapock, and the VigiLIFE project. The PAG (Parc Amazonian de Guyane), VigiLIFE and SPYGEN provided field and laboratory support. Investissement d'Avenir projects CEBA (ANR-10-LABX-25-01), TULIP (ANR-10-LABX-0041) and ANR DEBIT (ANR-17-CE02-0007-01) also provided financial support.

References

Araújo ES, Marques EE, Freitas IS, *et al.* 2013. Changes in distance decay relationships after river regulation: similarity among fish assemblages in a large Amazonian river. *Ecol Freshw Fish* **22**: 543–52.

Barnes MA and Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* **17**: 1–17.

Biggs J, Ewald N, Valentini A, *et al.* 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). *Biological Conservation* **183**: 19–28.

Boyer F, Mercier C, Bonin A, *et al.* 2016. obitools: a unix-inspired software package for DNA metabarcoding. *Molecular ecology resources* **16**: 176–182.

Brosse S, Grenouillet G, Gevrey M, *et al.* 2011. Small-scale gold mining erodes fish assemblage structure in small Neotropical streams. *Biodiversity and Conservation* **20**: 1013–26.

Cantera I, Cilleros K, Valentini A, *et al.* 2019. Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. *Scientific Reports* **9**.

Cilleros K, Valentini A, Allard L, *et al.* 2018. Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*.

Civade R, Dejean T, Valentini A, *et al.* 2016. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PloS one* **11**: e0157366.

De Barba M, Miquel C, Boyer F, *et al.* 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol Ecol Resour* **14**: 306–23.

Deiner K and Altermatt F. 2014. Transport distance of invertebrate environmental DNA in a natural river. *PLoS One* **9**: e88786.

Deiner K, Fronhofer EA, Mächler E, *et al.* 2016. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nature Communications* **7**: 12544.

Eichmiller JJ, Best SE, and Sorensen PW. 2016. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. *Environ Sci Technol* **50**: 1859–67.

Evans NT, Li Y, Renshaw MA, *et al.* 2017. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences* **74**: 1362–74.

Ficetola GF, Miaud C, Pompanon F, and Taberlet P. 2008. Species detection using environmental DNA from water samples. *Biology Letters* **4**: 423–5.

Floyd R, Abebe E, Papert A, and Blaxter M. 2002. Molecular barcodes for soil nematode identification. *Molecular Ecology* **11**: 839–50.

Gallay M, Martinez J, Allo S, *et al.* 2018. Impact of land degradation from mining activities on the sediment fluxes in two large rivers of FRENCH GUIANA. *Land Degrad Dev* **29**: 4323–36.

Goldberg CS, Pilliod DS, Arkle RS, and Waits LP. 2011. Molecular Detection of Vertebrates in Stream Water: A Demonstration Using Rocky Mountain Tailed Frogs and Idaho Giant Salamanders (B Gratwicke, Ed). *PLoS ONE* **6**: e22746.

Hammond DS, Gond V, Thoisy B de, *et al.* 2007. Causes and consequences of a tropical forest gold rush in the Guiana Shield, South America. *AMBIO: A Journal of the Human Environment* **36**: 661–70.

Hänfling B, Lawson Handley L, Read DS, *et al.* 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology* **25**: 3101–19.

Jackson DA, Peres-Neto PR, and Olden JD. 2001. What controls who is where in freshwater fish communities — the roles of biotic, abiotic, and spatial factors. **58**: 14.

Jane SF, Wilcox TM, McKelvey KS, *et al.* 2015. Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources* **15**: 216–27.

Jerde CL, Mahon AR, Chadderton WL, and Lodge DM. 2011. “Sight-unseen” detection of rare aquatic species using environmental DNA: eDNA surveillance of rare aquatic species. *Conservation Letters* **4**: 150–7.

Keck F, Vasselon V, Tapolczai K, *et al.* 2017. Freshwater biomonitoring in the Information Age. *Frontiers in Ecology and the Environment* **15**: 266–74.

Koleff P, Gaston KJ, and Lennon JJ. 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* **72**: 367–382.

Le Bail PY, Covain, R, Jegu, M, *et al.* 2012. Updated checklist of the freshwater and estuarine fishes of French Guiana.

Le Bail PY, Keith P, and Planquette P. 2000. Atlas des poissons d’eau douce de Guyane. Tome 2, fascicule II : Siluriformes. Paris, France: M.N.H.N./S.P.N.

Lopes CM, Sasso T, Valentini A, *et al.* 2017. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. *Molecular Ecology Resources* **17**: 904–14.

McCluney KE, Poff NL, Palmer MA, *et al.* 2014. Riverine macrosystems ecology: sensitivity, resistance, and resilience of whole river basins with human alterations. *Frontiers in Ecology and the Environment* **12**: 48–58.

Morlon H, Chuyong G, Condit R, *et al.* 2008. A general framework for the distance-decay of similarity in ecological communities. *Ecology Letters* **11**: 904–17.

Muneepeerakul R, Bertuzzo E, Lynch HJ, *et al.* 2008. Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. *Nature* **453**: 220–2.

Murphy BR and Willis DW. 1996. Fisheries techniques. American Fisheries Society Bethesda, Maryland.

Nakagawa H, Yamamoto S, Sato Y, *et al.* 2018. Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. *Freshwater Biology* **63**: 569–80.

Nekola JC and White PS. 1999. The distance decay of similarity in biogeography and ecology. *J Biogeography* **26**: 867–78.

Olds BP, Jerde CL, Renshaw MA, *et al.* 2016. Estimating species richness using environmental DNA. *Ecology and Evolution* **6**: 4214–26.

Pilliod DS, Goldberg CS, Arkle RS, and Waits LP. 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources* **14**: 109–16.

Planquette P, Keith, P, and Le Bail PY. 1996. Atlas des poissons d'eau douce de Guyane (tome 1). Paris, France: IEBG - M.N.H.N., INRA, CSP, Min. Env.

Pont D, Rocle M, Valentini A, *et al.* 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports* **8**.

Pri V, Lima ML, Taberlet P, *et al.* 2020. Large-scale monitoring of freshwater bivalves: an eDNA point of view on species distribution and conservation. Preprints.

Rees HC, Gough KC, Middleditch DJ, *et al.* 2015. Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals (E Crispo, Ed). *Journal of Applied Ecology* **52**: 827–31.

Rees HC, Maddison BC, Middleditch DJ, *et al.* 2014. REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology (E Crispo, Ed). *Journal of Applied Ecology* **51**: 1450–9.

Rham M, Thibault P, Shapiro A, *et al.* 2017. Monitoring the impact of gold mining on the forest cover and freshwater in the Guiana Shield.

Ryberg M. 2015. Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. *Molecular Ecology* **24**: 5770–7.

Schnell IB, Bohmann K, and Gilbert MTP. 2015. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* **15**: 1289–303.

Seymour M, Durance I, Cosby BJ, *et al.* 2018. Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Commun Biol* **1**: 4.

Soininen J, McDonald R, and Hillebrand H. 2007. The distance decay of similarity in ecological communities. *Ecography* **30**: 3–12.

Taberlet P, Bonin A, Zinger L, and Coissac E. 2018. Environmental DNA for biodiversity research and monitoring. Oxford, U.K.: Oxford University Press.

Takahara T, Minamoto T, Yamanaka H, *et al.* 2012. Estimation of Fish Biomass Using Environmental DNA (JA Gilbert, Ed). *PLoS ONE* **7**: e35868.

Tejerina-Garro FL and De MéRona B. 2001. Gill net sampling standardisation in large rivers of French Guiana (South America). *Bulletin Français de la Pêche et de la Pisciculture*: 227–40.

Tonn WM. 1990. Climate Change and Fish Communities: A Conceptual Framework. *Transactions of the American Fisheries Society* **119**: 337–52.

Valentini A, Taberlet P, Miaud C, *et al.* 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* **25**: 929–42.

Webb CO, Ackerly DD, McPeek MA, and Donoghue MJ. 2002. Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics* **3**: 475–505.

Weiher E and Keddy P (Eds). 1999. Ecological Assembly Rules: Perspectives, Advances, Retreats. Cambridge: Cambridge University Press.

Wilcox TM, McKelvey KS, Young MK, *et al.* 2016. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. *Biological Conservation* **194**: 209–16.

Zinger L, Donald J, Brosse S, *et al.* 2020. Advances and prospects of environmental DNA in neotropical rainforests. In: Advances in Ecological Research. Elsevier.

Authors' contributions

IC and SB conceived the ideas, designed methodology and led the writing of the manuscript; SB, JBD, JM and RV collected the data; IC analyzed de data; AV and TD conducted the laboratory work and bio-informatic analyses.

Conflict of interest

Teleo primers and the use of the amplified fragment for identifying fish species from environmental samples are patented by the CNRS and the Université Grenoble Alpes. This patent only restricts commercial applications and has no implications for the use of this method by academic researchers. SPYGEN owns a licence for this patent. A.V. and T.D. are research scientists at a private company specialising in the use of eDNA for species detection.

Figures

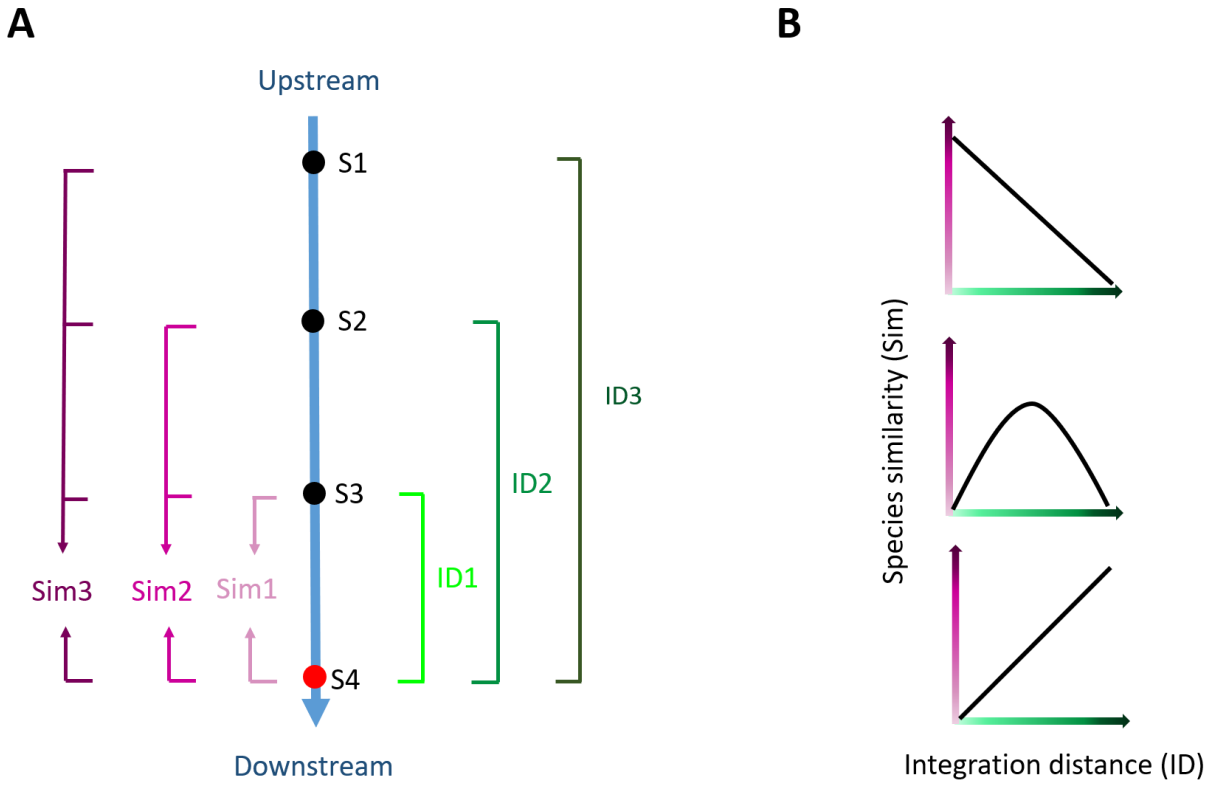

Figure 1: Theoretical framework to characterize the distance of environmental DNA (eDNA) spatial signal along river systems. (**A**) Schematic representation of the measures of species similarity and integration distance to measure detection distance of eDNA. Sites are located linearly along the upstream-downstream gradient of a river (blue arrow). Four eDNA sampling sites (S1 to S4) are represented with dots. The detection distance of eDNA is here measured between the downstream site S4 (in red) and the 3 sites located upstream (in black). ID1 to ID3 indicate the integration distance covering the extent of the upstream area that contributes to eDNA detection. Similarity in species composition (Sim 1 to 3) corresponds to the similarity between the fauna detected between two sites for Sim 1 (similarity between S4 and S3) or between S4 and the combined species inventories of upstream sites, namely S3 and S2 for Sim2, and S3, S2 and S1 for Sim 3 (differences in species richness were controlled using null models, see methods). (**B**) Theoretical representation of the three hypotheses of eDNA detection distance. Top panel represents a short detection distance, with highest similarity between nearby sites. Middle panel represents an intermediate detection distance, with similarity peaking in intermediate integration distance. The bottom panel represents a large detection distance, with similarity increasing with integration distance.

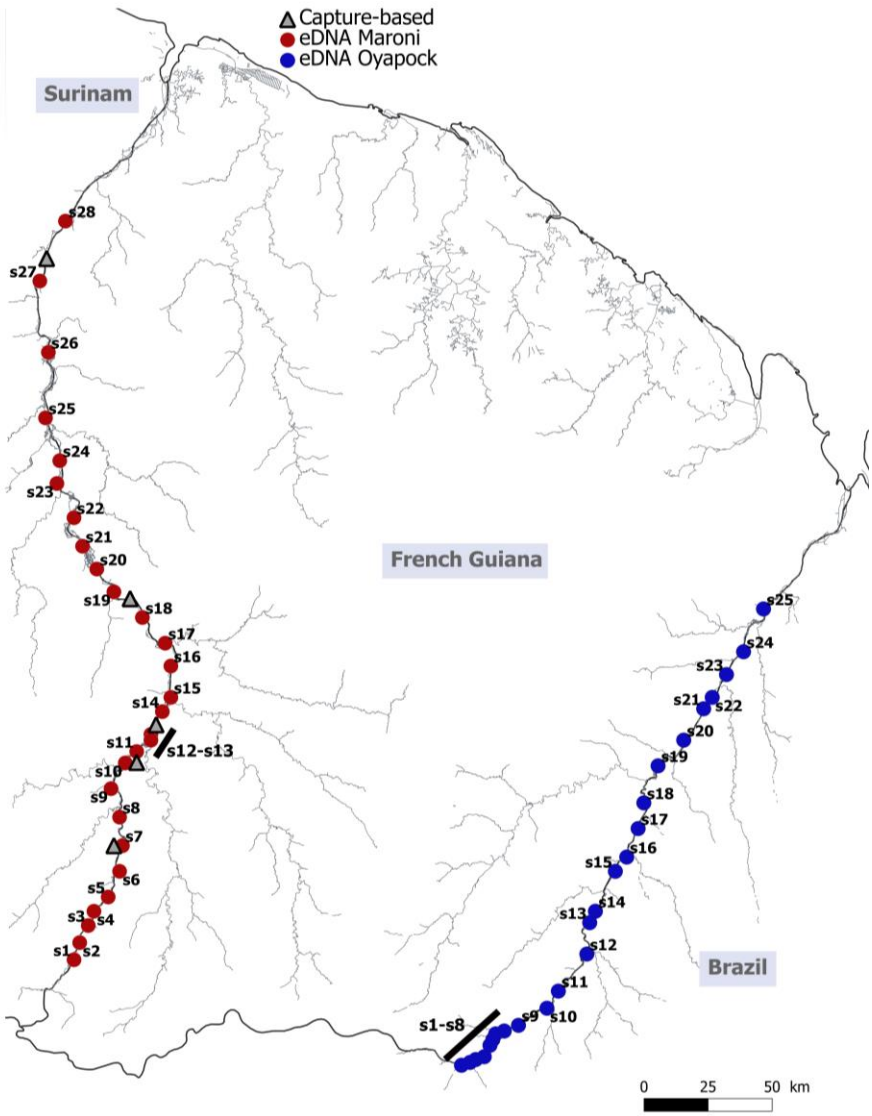

Figure 2: Map of the study area indicating the fish sampling sites in the Maroni (red dots) and Oyapock (blue dots) rivers sampled with eDNA methods. Grey triangles correspond to sites sampled with a capture-based method (gill-nets). Inset map on the top indicate the location of the study area in South America.

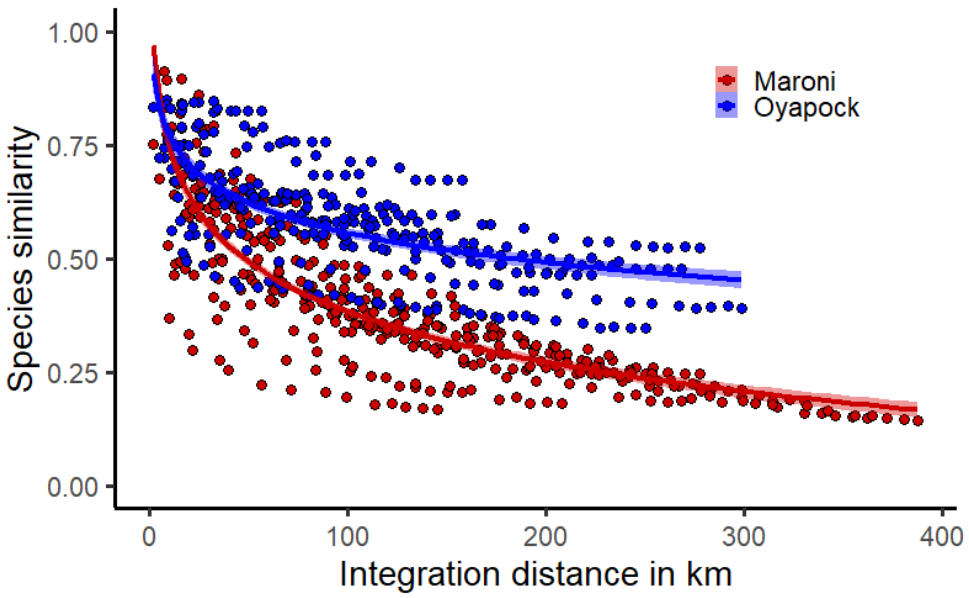

Figure 3: Fish species similarity (Jaccard index) of Maroni (red) and Oyapock (blue) communities plotted against integration distance covering the upstream extent in which species inventories are pooled (see Fig.1 for details). Fitted values (solid lines) and 95% confidence intervals (shaded areas) are derived from a regression model ($p<0.001$ for both rivers and R^2 = 0.7 and 0.46 for the Maroni and Oyapock, respectively).

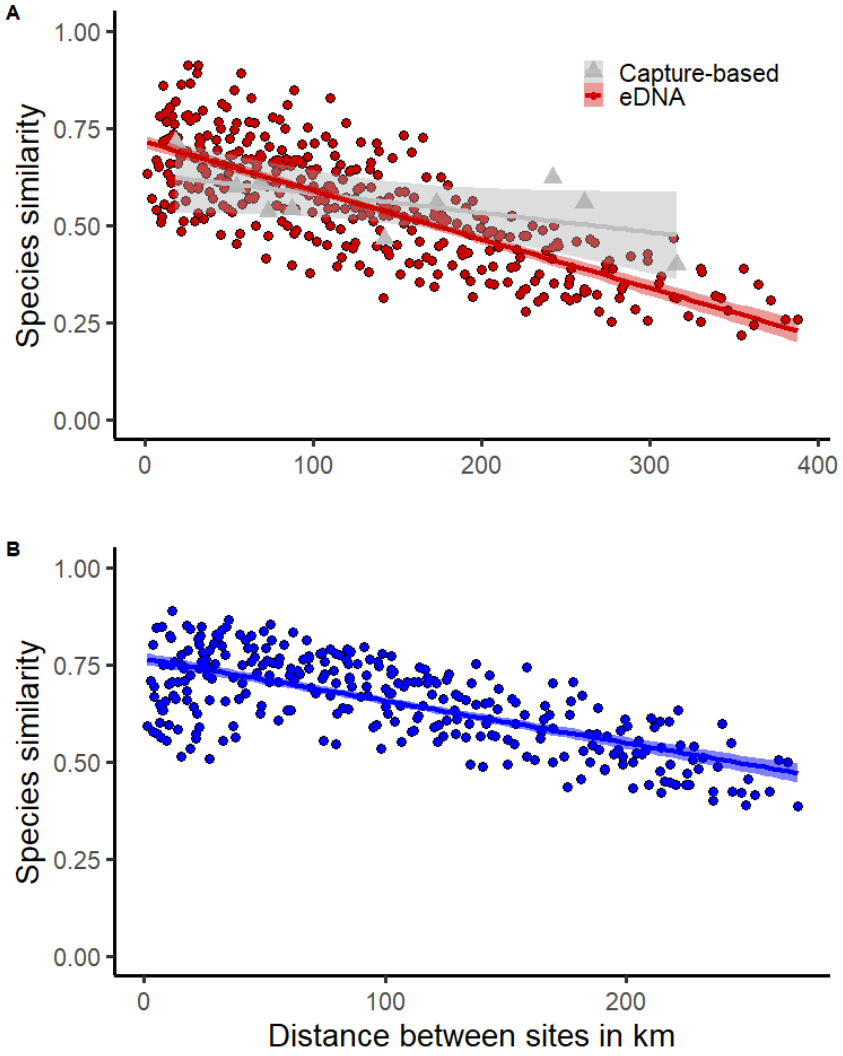

Figure 4: Relationship between fish species similarity (Jaccard index) and watercourse distance between pairs of sites for the (**A**) Maroni and (**B**) Oyapock rivers. Grey triangles on panel (A) are species similarities calculated from capture-based data. The straight lines (solid lines) and 95% confidence intervals (shaded areas) were fitted through linear regression (Mantel statistic R= -0.8 and -0.7, for Maroni and Oyapock rivers, respectively, $p<0.001$)

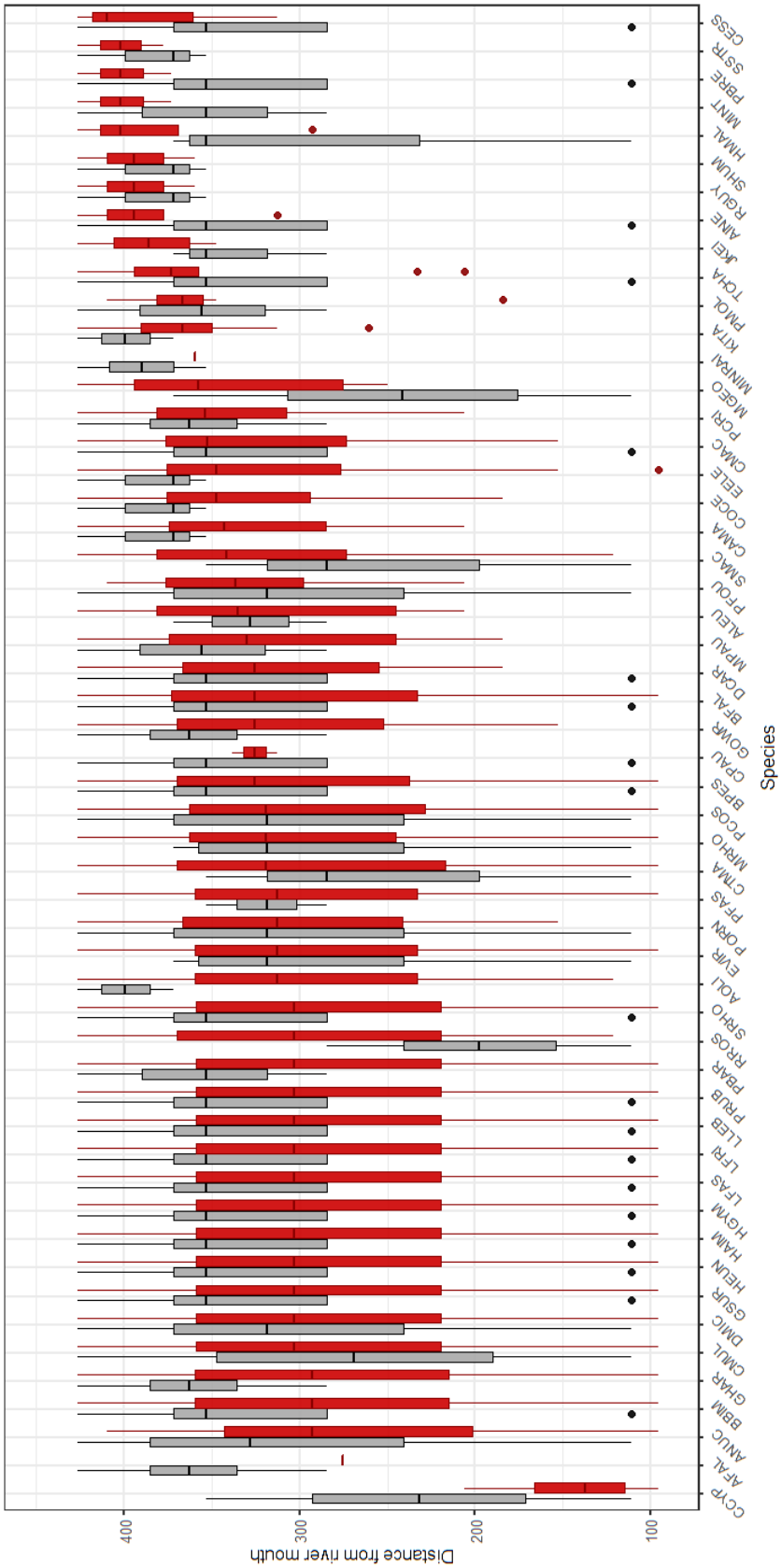

Figure 5: Species distribution ranges along the upstream-downstream gradient of the Maroni river derived from eDNA (red boxplots) and capture-based samples (grey boxplots). Species are ordered according to the mean distance from the river mouth of the sites where they were detected using eDNA (increasing mean from left to right). The species names corresponding to the codes are in Table S1.