eDNA metabarcoding: an effective tool for vertebrate diversity studies in the Colombian Amazon and Orinoco basins

Daniela Martinelli Marín1*, Carlos A. Lasso2 and Susana J. Caballero Gaitan1†

1Laboratorio de Ecología Molecular de Vertebrados Acuáticos (LEMVA), Departamento de Ciencias Biológicas, Universidad de Los Andes, Bogotá, Colombia. 2Centro Colecciones y Gestión de Especies, Dirección de Conocimiento, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia

The development of fast, cost-effective, non-invasive, and efficient sampling alternatives, such as environmental DNA (eDNA), is crucial for understanding the changes in species biodiversity and distributions worldwide, particularly for low abundance, cryptic, and threatened species. This study utilized environmental eDNA to analyze the variety of aquatic, semi-aquatic, and terrestrial vertebrates in the Colombian Amazon and Orinoco basins. The study focused on four main subregions: Bojonawi Natural Reserve and adjacent areas (Vichada Department), Sierra de la Macarena National Park and Tillavá (Meta Department), Puerto Nariño and adjacent areas (Amazonas Department), and the Municipality of Solano (Caquetá Department). A total of 709 operational taxonomic units (OTUs) were identified across all sampling locations. The Orinoco River had the highest number of fish genera (68), while the Guayabero River had the highest number of tetrapod genera (13). New taxonomic records were found for all locations, with the highest number of previously undetected fish diversity being found in the Bita, Orinoco, and Tillavá rivers, compared to traditional surveys. Likewise, the study identified two fish species, four mammal species, and one reptile species as vulnerable. Additionally, four mammal species were identified as endangered, including the giant otter (Pteronura brasiliensis), two subspecies of the Amazon River dolphin (Inia geoffrensis geoffrensis and Inia geoffrensis humboldtiana), and the tucuxi (Sotalia fluviatilis). Standardizing the methodology and improving current DNA sequence databases for the Neotropics is essential to develop future eDNA studies and enhance our understanding of the region’s diversity.

KEYWORDS

eDNA metabarcoding, diversity, richness, vertebrates, Amazon basin, Orinoco basin, Colombia
Introduction

Biodiversity loss on a global scale is a significant environmental issue of this century (Cardinale et al., 2012; Valenti et al., 2016; Johnson et al., 2017; Crossley et al., 2020; DiBattista et al., 2020). Scientific evidence suggests that human activities, such as resource appropriation, habitat fragmentation, non-native species introduction, spread of pathogens, and direct killing of species, are predicting a sixth mass extinction (Barnosky et al., 2011; Ceballos et al., 2015, 2020; Cowie et al., 2022). Traditionally, conservation of fauna and flora has been based on observed local diversity. However, this approach has resulted in significant gaps in worldwide biological information (Pärtel et al., 2011; Delic et al., 2017). There is a relatively new ecological concept called ‘dark diversity’ (Pärtel et al., 2011; Lewis et al., 2016; Fernandes et al., 2019), which refers to the pool of species that are presumed to be present within a certain region. However, traditional sampling methods may miss these species, resulting in them being reported as absent (Moeslund et al., 2017), as diversity encompasses cryptic, rare, and endangered species that are challenging to detect but are crucial for effective ecological management and conservation programs (Rees et al., 2014; Boussarie et al., 2018). Therefore, scientists urgently require the development of rapid, cost-effective, sensitive, and non-invasive molecular tools for species monitoring and conservation (Kelly et al., 2014; Deiner et al., 2017).

Environmental DNA (eDNA) combined with high-throughput sequencing (metabarcoding) is a promising technique for detecting aquatic, semi-aquatic, and terrestrial vertebrates. It has become a powerful complement to traditional surveys, as demonstrated by several studies (Hänfling et al., 2016; Valenti et al., 2016; Sard et al., 2019; Coutant et al., 2021; Dal Pont et al., 2021; Mena et al., 2021; Pawlowski et al., 2021; Jia et al., 2022; Mas-Carrió et al., 2022). Aquatic ecosystems may be affected by various factors, including salinity, temperature, pH, solar radiation, and microorganisms. These factors can contribute to the degradation of eDNA (Thomsen et al., 2012; Barnes et al., 2014; Jo et al., 2017; Saito and Doi, 2021). Therefore, this methodology is believed to produce consistent results with the species present in the current location and time of the environmental sample (Herder et al., 2014; Rees et al., 2014; Thomsen and Willerslev, 2015; Deiner et al., 2017; Seymour, 2019).

Colombia is located in the American tropics, which is one of the most species-rich regions on Earth (Antonelli et al., 2018). It is the third most diverse country, hosting close to 10% of the planet’s biodiversity (Ministry of Environment and Sustainable Development, 2017, 2019; Sistema de Información Ambiental de Colombia, 2020; Convention on Biological Diversity, 2021). Colombia is geographically divided into five regions, two of which are the Amazon and Orinoco. According to the Geographic Institute Agustin Codazzi (2002), the Colombian Amazon region accounts for 6.4% of the total area of the Amazon basin, while the Orinoco River basin represents 35% of the Colombian territory (Correa et al., 2005; Lasso et al., 2010; Universidad Nacional de Colombia, 2013; Lasso et al., 2014; Aldana et al., 2017; Guio and Rojas, 2019; WWF, 2020). These two regions contain a variety of freshwater ecosystems and biomes, which in turn have extraordinary species diversity (Wrona and Reist, 2013). However, these ecosystems are being threatened by a multitude of anthropogenic stressors (Suring, 2020) that critically reduce freshwater biodiversity (Tickner et al., 2020; IUCN, International Union for Conservation of Nature, 2022). Not only is the loss of valuable biological information at stake, but freshwater ecosystems also provide essential goods and services to humans that are now at serious risk (Portocarrero-Aya, 2011; Faghhiinia et al., 2021).

Until now, eDNA has been used in relatively few studies in Colombia. These studies include investigations on marine animals such as fishes, whales, and corals (Juhel et al., 2021; Marques et al., 2021; Stauffer et al., 2021; Polanco et al., 2021b; Mathon et al., 2022), as well as on freshwater, underground, and cave ecosystems in search of vertebrates (Martinelli-Marín et al., 2020; Lozano Mojica and Caballero, 2021; Caballero et al., 2021a, b; Polanco et al., 2021c). However, neither has implemented eDNA in the Amazon and Orinoco basins as a whole. Also, most biodiversity inventorying and monitoring studies in these two regions have relied on traditional, expensive, and time-consuming methodologies (Prieto and Arias, 2007; Lasso et al., 2010; Lasso and Morales-Betancourt, 2017; Lasso et al., 2020). On the other hand, a cost-benefit analysis of eDNA implementation has shown that, at least for the Orinoco Basin, similar or even better results have been obtained with less time, space and economic investment compared to traditional surveys (Martinelli-Marín et al., 2020). This new tool could therefore improve the detection and monitoring of freshwater species by the Colombian environmental authorities in a rapid and cost-effective manner.

To the best of our knowledge, this is the first study to use eDNA and next-generation sequencing to describe the diversity and richness of aquatic, semi-aquatic, and terrestrial vertebrates in the water bodies (rivers, streams, lakes, and lagoons) of four different geographic regions of the Colombian Amazon and Orinoco basins.

Materials and methods

Sample location

Between October 2019 and February 2020, the Aquatic Vertebrate Molecular Ecology Laboratory (LEMVA) at Los Andes University, in collaboration with the Biological Resources Research Institute Alexander von Humboldt, conducted four expeditions to the Colombian Amazon and Orinoco basins. The study examined three Amazon subbasins (Caquetá, Javari, and Loretoyacu rivers) and five Orinoco subbasins (Bita, Guaviare, Meta, the Orinoco main channel, and Vichada), which were divided into four terrestrial subregions: Bojonawi Nature Reserve and adjacent areas, Puerto Nariño and adjacent areas, Sierra de la Macarena National Park, and Tillavá and Solano’s Municipality (Figure 1; Supplementary Material). All field trips were undertaken during low water seasons. During the dry season in the Amazon basin, the Orinoco basin experiences its rainy season and vice versa. As a result, the low water season in the Amazon takes place from May to December, while in the Orinoco it occurs from January to April (Guhl, 2016).
eDNA sampling

25 main locations were sampled using the methodologies employed by the NatureMetrics (2019) laboratory, as well as the studies conducted by Lozano Mojica and Caballero (2021) and Caballero et al. (2021a). At each location, up to seven water subsamples (1L each) were collected in sterile plastic bottles and then transferred to a bucket covered with a plastic bag, using sterile gloves. Prior to use, the plastic bag was decontaminated with 90% ethanol and 10% diluted bleach, and then rinsed with tap water. Water samples were collected at intervals of 10–20 metres along a linear transect using boats for rivers, canoes for lakes and lagoons, and on foot for streams. The coordinates of each sampled point were recorded using a Garmin etrex 12 channel GPS. Once all the subsamples from a location were taken, water samples were filtered using NatureMetrics eDNA collection kits. A 60 ml syringe filled with the collected water was attached to a filter disk with a 0.8 μm pore size. When the filter disk became clogged and water could no longer pass through, we detached the syringe and used a smaller one with a preserving buffer (Longmire’s solution) to protect the filter and prevent DNA degradation (Baker et al., 2018). We stored each filter in an envelope with its corresponding field information and kept them cool in a styrofoam cooler with ice packs.

eDNA extraction, amplification and sequencing

The filters were transported to Nature Metrics in Egham, Surrey, England, where DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit, following the manufacturer’s instructions. The method for disc filters in buffer, with the following modifications, was based on Spens et al. (2016): proteinase K was initially added directly to the filter housing; following incubation, 1 mL of lyse was carried forward for extraction, and all DNeasy Blood and Tissue Kit (Qiagen) reagents were adjusted accordingly. The final elution was in 200 μL. The modifications aim to minimise the risk of contamination and maximise DNA yield. The DNA was purified using the DNeasy PowerClean Pro Cleanup kit to eliminate PCR inhibitors. Subsequently, DNA extracted from each filter was amplified using 12 replicates of the 12S rRNA mitochondrial gene (5’- TAGAACAGGCTCCTCTAG-3’ and 5’-TTAGATACCCCAC TATGC-3’) to target fishes as part of the eDNA survey - Vertebrates pipeline (Riaz et al., 2011; Milan et al., 2020). PCR replicates were combined, and adapters were added at the 5’ end of the primers to complement Illumina Nextera index primers. The amplification mixture for each replicate consisted of 1X DreamTaq PCR Master Mix (Thermo Scientific), 0.4 μM of each of the tailed primers, 1 μL of DNA, and PCR grade water (Thermo Scientific) up to a total reaction volume of 10 μL. All PCRs were performed in the presence of both a negative control and a positive control sample (mock community with a known composition, not expected to occur in Colombia). The PCR conditions comprised an initial denaturation at 95°C for 2 minutes, followed by 10 cycles of 20 seconds at 95°C, a 30-second touchdown annealing step (-0.5°C per cycle) starting at 60°C, and 40 seconds at 72°C, 35 cycles of 20 seconds at 95°C, 30 seconds at 55°C, and 40 seconds at 72°C, and a final elongation step at 72°C for 5 minutes. Amplification success was determined by gel electrophoresis. The amplicons were purified using MagBind TotalPure NGS (Omega Biotek) magnetic beads with a bead to DNA ratio of 0.8:1 to eliminate primer dimers. The purified index PCRs were then quantified using a Qubit high sensitivity kit following
the manufacturer’s protocol. Subsequently, all purified index PCRs were combined into a final library with equal concentrations. Finally, the library was sequenced using an Illumina MiSeq V2 kit at 12 pM with a 10% PhiX spike in. The sequence data underwent processing through a custom bioinformatics pipeline, USEARCH v11, which included quality filtering, dereplication, and taxonomic assignment. Taxonomic assignment was considered successful when there was at least 80% agreement in the overlap. The merged sequences were processed using cutadapt 2.3 (Martin, 2011; Mathon et al., 2021) to remove forward and reverse primers. Only sequences with a trimmed length between 80–120 bp were kept.

Quality filtering was performed to retain sequences with an expected error rate per base of 0.01 or lower. Dereplication was then carried out by sample, retaining only singletons. All unique reads from the samples underwent denoising in a single analysis using UNOISE (Dal Pont et al., 2021). ZOTUs (zero-radius OTUs) with a minimum abundance of 8 were retained and clustered at 99% similarity. An OTU-by-sample table was created by mapping all dereplicated reads for each sample to the OTU representative sequences with an identity threshold of 97%.

Taxonomic information was evaluated for each OTU by conducting sequence similarity searches against the NCBI nt database (GenBank) and PROTAX (Somervuo et al., 2016; Lozano Mojica and Caballero, 2021). Identifications from either source were accepted, and these were consistent at the level at which they were made. Species and genus level assignments were automatically retained if they were supported by unambiguous matches to reference sequences at ≥99% or ≥95%, respectively. Public records from GBIF were used to determine the most likely species present in Colombia when there were multiple equally good matches. This allowed for the resolution of numerous uncertain sequences to the species level. OTUs that were ≥99% similar and had similar co-occurrence patterns were combined using LULU (Frøslev et al., 2017). The OTU table was then filtered to remove low abundance OTUs from each sample (<0.05% or <10 reads). In order to eliminate any erroneous identifications, the sequences corresponding to human, food, and livestock were removed (Supplementary Material). The three most predominant orders in the fish group contained 9 orders, 35 families, 132 genera, and 249 species (Supplementary Material). Although Tillavá is not part of Sierra de la Macarena National Park, it is relatively close (296 km) compared to the other subregions. Therefore, we decided to group them. Furthermore, Tillavá belongs to the same department (Meta) and region (Orinoco) as Sierra de la Macarena National Park. Additionally, we decided to use the taxonomic level of ‘genus’ for all analyses, since it had the highest number of reads and identifications.

R studio (RStudio Team, 2020) (R Project for Statistical Computing, RRID: SCR_001905) version 3.6.0, ‘vegan’ package, was used to calculate alpha and beta diversity for all sampling sites, based on presence/absence (i.e. binary) matrices of genera. Shannon-Weiner and Simpson indices were calculated for alpha diversity (Moreno, 2001; Li et al., 2021). For the Simpson index, low (0.00–0.35), medium (0.36–0.75) and high (0.76–1.00) diversity values were used. For the Shannon index, low (0.1 – 1.5), medium (1.6 – 3.0) and high (3.1 – 4.5) diversity values were used (Magurran, 1988). Beta diversity was then assessed using the Jaccard dissimilarity index (Nakagawa et al., 2018; Wiersma, 2019; Lin et al., 2021) and principal coordinate analysis (PCoA). Similarly, the number of unique genera identified for each site and the genera shared among them were represented by a Euler diagram, both for the fish group and for the tetrapod group.

Results

High-quality vertebrate sequence data were obtained from 31 filters, resulting in a total of 709 OTUs after data curation. The fish group contained 9 orders, 35 families, 132 genera, and 249 species (Supplementary Material), while the tetrapods group contained 22 orders, 32 families, 43 genera, and 49 species (Supplementary Material). The three most predominant orders in the fish group were Siluriformes, Characiformes, and Gymnotiformes, with Characidae and Loricariidae as the leading families. Among the tetrapod group, the orders with the highest prevalence were Anura, Primates, and Rodentia, along with the Psittacidae family.

Orinoco River (RO) had the highest number of fish genera with 67, followed by Tillavá (T) (52), Meta River (RM) (45), Yarumales stream (CY) (43), Bita River (RB) (43), Puerto Nariño (PN) (37), El Silencio lagoon (LES) (33), Verde stream (CV) (31), El Pañuelo lagoon (LEP) (31), Caballococha (PCC) (28), Negro stream (CN) (27), Corre-Tarapoto Redondo lakes (LCT) (25), Don Ricardo stream (CDR) (21), Guayabero River (RG) (18), Caño Cristales River (CC) (11), Caño Paujil River (CP) (11), Tesoro stream (CT) (9) and lastly Caquetá River (RC) (4). In terms of relative abundance of DNA based on percentage of reads, the most abundant genera in Puerto Nariño and adjacent areas were Prochilodus (~38%) and Cyphocharax (~38%), while for the municipality of Solano these were Acestrorhynchus (~22%) and Apteronotus (~18%) (Figure 2). In the Bojonawí Nature Reserve they were Prochilodus (~30%) and Leporinus (~21%), and in the Bioinformatic analyses

Due to the large amount of ichthyological data in comparison to other vertebrates, the results were separated into two groups: fish and tetrapods. Tetrapods included amphibians, birds, mammals, and reptiles, whether they were terrestrial, aquatic, or semi-aquatic.
Sierra de la Macarena National Park and Tillavà they were *Cyphocharax* (~20%) and *Prochilodus* (~13%) (Figure 2).

For the tetrapods, the number of genera found, in order from largest to smallest, was as follows: RG (13), CY (12), LES (8), RO (8), PCC (7), CP (6), RB (5), PN (5), T (5), CV (4), CDR (3), LCT (3), RM (3), CC (2), RC (2), CN (1) and LEP (1). In addition, the most abundant genera were *Phalacrocorax* (birds) (~65%) and *Inia* (mammals) (~33%) for Puerto Nariño and adjacent areas, *Chelus*...
(reptiles) (~41%) and *Steatornis* (birds) (~18%), in the municipality of Solano, *Inia* (mammals) (~41%) and *Peltocephalus* (reptiles) (~21%) for the Bojonawi Nature Reserve and adjacent areas, and *Odocoileus* (mammals) (~21%) and *Bubulcus* (birds) (~14%) in the Sierra de la Macarena National Park and Tillavá, respectively (Figure 3).

**Alpha diversity**

In relation to the fish group, all sites in the Orinoco region showed high diversity according to Simpson’s index, including the Amazonian sites, except for RC, which showed medium diversity (Figure 4B). With regards to Shannon’s index, 10 out of the 13 Orinoco sites had relatively high diversity (CDR, CN, CV, CY, LEP, LES, RB, RM, RO and T), while the remaining three (CC, CT and RG) had medium diversity (Figure 4A). In the case of tetrapods, five locations (CY, LES, RG, RO and T) exhibited relatively high diversity values, while five other locations (CC, CDR, CV, RB and RM) showed medium diversity values, and two locations (CN and LEP) had low diversity values. Conversely, in the Amazon region, three sites (CP, PCC and PN) had high Simpson’s diversity values, while two sites (LCT and RC) had medium values (Figure 4D). The Orinoco region displayed medium diversity (CY, LES, RG, RO and T) and low (CC, CDR, CV, RB, and RM) diversities in the Shannon’s index (Figure 4C). Additionally, CT and LEP had only one genus detected, resulting in a diversity value of $H^* = 0$ (Sen and Grillo, 2018). In contrast, the Amazon region exhibited medium diversity in three locations (CP, PCC and PN) and low diversity in two locations (LCT and RC) (Supplementary Material).

**Beta diversity**

Principal coordinate analysis (PCoA) enabled us to illustrate the ecological distance based on genus diversity. As a result, the fish group was divided into four clusters (Figure 5A). The largest cluster included all locations from the Orinoco basin, except for the lowland streams of Bojonawi Nature Reserve and the Caño Cristales River (CC) for Sierra de la Macarena National Park. The second cluster was formed by Caño Negro (CN) and Caño Verde (CV), while CC formed a third cluster with the two locations of the Caquetá subbasin (Caquetá River and Caño Paujil River), in addition to the third stream of Bojonawi Nature Reserve, the Tesoro stream (CT). The fourth cluster grouped locations of the Loretoaciu River and Javari River subbasins of the Amazon region (PN, LCT and PCC). In the first cluster, the Orinoco and Bita rivers showed the closest proximity as they overlapped. Additionally, these two locations were found to be closer to the Meta River, El Pañuelo lagoon, Tillavá, and Yarumales stream compared to the other locations in the group, such as Don Ricardo stream, Guayabero River, and El Silencio lagoon. The initial two locations were grouped in the upper right quadrant, while El Silencio lagoon was the only location placed in the bottom right quadrant. This indicates a significant difference in the composition of genus diversity within the group.

The study identified three clusters of tetrapods. The first cluster included all locations from Guaviare and Caquetá subbasins, with the Bita River (B) and El Silencio lagoon (LES) grouped together. The second cluster contained the Orinoco and Meta rivers from Bojonawi Nature Reserve, as well as samples from the Loretoaciu River and Javari River subbasins (PN, LCT and PCC). The third cluster consisted
of the fish group, with Caño Negro (CN) and Caño Verde (CV) forming a separate group (Figure 5B).

Taxonomic composition by regions, subregions and locations

A total of 68 fish genera were found exclusively in one of the four sampled subregions. Specifically, 31 genera were detected in the Sierra de la Macarena National Park, 20 in Bojonawi Nature Reserve, 16 in Puerto Nariño and adjacent areas, and one in Solano’s municipality. All sampled subregions shared four genera. Sierra de la Macarena National Park and Bojonawi Nature Reserve had the most genera in common (23), followed by Sierra de la Macarena National Park, Bojonawi Nature Reserve and Puerto Nariño (21). The Amazon basin subregions did not share any genera (Figure 6A; Supplementary Material).

For the tetrapod group, 33 of the 43 genera were found in just one of the four subregions. There were 18 genera in Sierra de la Macarena and Tillavá, seven in Puerto Nariño and adjacent areas, six in the Bojonawi Nature Reserve and adjacent areas, and two in Solano’s municipality. None of the four subregions share genera between them (Figure 6B; Supplementary Material).
Discussion

This study validates the effectiveness of environmental (e)DNA sampling in evaluating the relative taxonomic richness of aquatic, semiaquatic, and terrestrial vertebrates in different macrohabitats, such as rivers, lowland streams, lakes, and lagoons in the Colombian Amazon and Orinoco hydrographic basins. We also identified organisms at the taxonomic levels of families, genera and species,
and allowed the diversity of different communities to be assessed. The results obtained complement and efficiently extend traditional species inventories, although traditional methodologies require greater spatial component and field collection effort compared to eDNA (Fraija-Fernández et al., 2020; Sales et al., 2020b).

The study successfully identified cryptic genera, such as Apteronotus, Eigenmannia, Pellona, Prochilodus, and Sternopygus, which are physically similar but genetically and taxonomically distinct. Additionally, the study detected rare and low-abundance genera that are difficult to observe and not easily detectable in traditional surveys, including Anchoviella, Electrophorus, Pocelia, Synbranchus, and Phreatobius (Ota et al., 2020). Also, the analysis suggests that there are likely more species within the same genus in the Colombian Amazon and Orinoco basins than previously known, including many that are probably new to science. For instance, in the Bojonawí Nature Reserve, seven OTUs of the Prochilodus genus were reported (Supplementary Material), whereas only two species, P. mariae and P. rubrotaeniatus, are recognized in the basin (Donascimento et al., 2021). The high diversity found for this genus in Northern South America (Moyer et al., 2005) could be correlated with potential hybridization and introgression. This likely increases the chances of detecting particular sequences, rather than just a unique sequence for 12s.

As for the genus Electrophorus, only one species is recognised for the Orinoco Basin (E. electricus), but here we detected a second OTU, supporting the study by de Santana et al. (2019), which proposed an unexpected species diversity of electric eels. Additionally, several species of fish that were not previously reported in inventories or studies were detected using environmental DNA (eDNA), particularly in the Bojonawí Nature Reserve and Sierra de la Macarena and Tillavá. The Bita, Orinoco, and Tillavá rivers contain a high percentage of unknown fish diversity, estimated at 75% (65), 35% (113), and 52% (56), respectively (Supplementary Material). Therefore, these results support the need to associate eDNA detection and traditional capture methods to avoid biased estimates of occurrence and obstruct species conservation (Sales et al., 2020b). Additionally, they enable the rapid identification of regions with unexplored diversity to be investigated for future diversity studies in the Orinoco basin.

Furthermore, given the global decline in fisheries and the high cost, logistical demands, and invasiveness of conventional survey methods for stock assessment, eDNA has emerged as an effective tool for assessing methods for stock assessment, Orinoco basin. Diversity to be investigated for future diversity studies in the following genera: Sorubimichtys, Ptyerygoplichthys and Sorubimichthys (Salinas and Agudelo, 2000; Lasso et al., 2011). Therefore, these results suggest that eDNA could be a valuable tool for future fisheries monitoring.

Simultaneously, eDNA enables the identification, surveillance, and protection of threatened species in a more feasible manner (Thomsen et al., 2012; Laramie et al., 2015; Weltz et al., 2017; Loza-Quintana et al., 2021; Plough et al., 2021). The current study identified two fish species as near threatened (Sorubimichthys planiceps and Sorubim lima) and two as vulnerable (Osteoglossum bicirrhosum and Zungaro zungaro) (Mojica et al., 2012). Five vulnerable species were detected in the tetrapods group: the white-lipped peccary (Tayassu pecari), the lowland tapir (Tapiro terrestris), the giant anteater (Myrmecophaga tridactyla), the common woolly monkey (Lagothrix lagotricha), and the Big-headed Amazon River Turtle (Pelocephalus dumerilianus) (Keuroghlian et al., 2013; Miranda et al., 2014; Tortoise & Freshwater Turtle Specialist Group, 1996; Varela et al., 2019; Stevenson et al., 2021). Four endangered species have been identified, including the giant otter (Pteronura brasiliensis), as well as two subspecies of the Amazon River dolphin found in Colombia (Inia geoffrensis geoffrensis and Inia geoffrensis humboldtiana), and the tucuxi (Sotalia fluviatilis) (da Silva et al., 2018, 2020; Groenendijk et al., 2023).

Furthermore, eDNA provides a clear picture of how the distribution of diversity is mainly related to environmental conditions and resources in aquatic and terrestrial systems (Mott, 2010; Wallis et al., 2021). It is expected that sampling locations in the same geographical area would be more similar to each other for biological diversity, compared to locations in different regions (Heino et al., 2015). Most of the fish and tetrapods group data were distributed according to their respective geographic basin in either the Amazon or Orinoco basins, forming two main groups (Figures 5A, B). However, for some taxa, particular combinations resulted in Amazon and Orinoco samples grouping together. A possible explanation for this date back to the late Miocene epoch (approximately 11.6 to 5.3 million years ago) when the eastern Cordillera rose, leading to the emergence of the Macarena massif and the uplift of the Vaupes swell. This event eventually caused the separation of the Orinoco and Amazon systems. These geographical events may have isolated the respective populations, both terrestrial and aquatic, leading to species dispersion among the fragmented Amazon and Orinoco basins (Correa et al., 2005; Mora et al., 2010; Renno et al., 2011; Holbourn et al., 2018; Cala-Cala, 2019; Vargas-Ramirez et al., 2020; Rizo-Fuentes et al., 2021).

It is important to highlight, that eDNA is a relatively new methodology (Peters et al., 2018), therefore limitations remain that must be considered for future studies. Primarily, there is currently no public database that includes sequences from all species described to date, nor a database with divergent sequences between closely related species (e.g., sister species) (Freeland, 2017; Cristescu and Hebert, 2018; Zaiko et al., 2018; Polanco et al., 2021a). This hinders accurate species identification, an issue that has been underlined by previous neotropical eDNA studies (Cilleros et al., 2019; Bevilaqua et al., 2020; Sales et al., 2020b). In the current study, of all the fishes detected, only...
29% were assigned to the species level. This also occurred for lesser extent in the tetrapods (57%). At the same time, misidentification, which arises from error generated as a result of the workflow, could lead to either false positives or false negatives, including false haplotypes (Furlan et al., 2020). As an example, the rough-toothed dolphin (Steno bredanensis) was reported in the results for Caballococha, Amazon, but this organism is a marine species. Phylogenetic studies have identified that the Steno genetic sequence for the 12s ribosomal RNA gene is very similar to that of the Sotalia genus, which is found in the Amazon River by the Tucuxi (Sotalia fluvialitis) (Caballero et al., 2007; Cunha et al., 2011; Mosquera et al., 2015). Another example is the identification of the paújil or black guan (Penelopina nigra) in the Guayabero River, which is not found in Colombia and is found only in Central America (Pineda et al., 2008). However, when evaluating the Penelopinae subfamily phylogenetics, it was found that the Penelopina genus is indeed a sister taxon of the Penelope genus, which is reported in the Sierra de la Macarena National Nature Park with the Spix’s guan (Penelope jacquacu) (Lasso et al., 2018). Therefore, to obtain accurate biological information, especially in megadiverse freshwater systems such as the Neotropics, it is crucial to develop curated and complete molecular DNA databases, including the expansion of currently available 12s and 16s rDNA sequence databases (Bevilaqua et al., 2020; Milan et al., 2020; Jackman et al., 2021).

Furthermore, the effectiveness of organism’s detection (species/genera) between eDNA and traditional sampling varies depending on the analysis approach. As reported in Martinelli-Marin et al. (2020), for the Orinoco River, 223 species were identified by metabarcoding and 128 species by traditional sampling (Lasso et al., 2020), representing a difference of 57% between the two methods. This would also be the case if we consider the whole system (river-floodplain), represented in this case by the Orinoco River and the set of macro-habitats (channels-lagoons) of the floodplain (262 species by metabarcoding vs. 270 species by traditional methods).

However, when the macrohabitats (lagoons, channels and main river) are analyzed at the individual level, there is a large difference between the two methods, with the traditional approach showing a superior efficiency of almost 40%. This pattern is consistent in the other remaining sites, except for the Solano’s municipality, Puerto Nariño and Caballococha, where there were no records for comparison (Table 1). We attribute the difference in ‘efficiency’ between the two methods to multiple aspects. First, eDNA sampling in this study was restricted to the main channel of the rivers. In contrast, the traditional methods included all macrohabitats of the basin, such as main rivers, tributaries, morichals, small ravines and others. Also, there was a significant difference in sampling time, with traditional methods taking weeks, months or even years (historic records) compared to days for eDNA. Finally, eDNA captures a specific sample at a particular time and location, whereas traditional sampling involves using various fishing gears to cover different areas in the water column at different seasons of the year which could contribute to a higher coverage of diversity.

In addition, the study found that the number of tetrapod genera and species was low compared to fishes, which is consistent with previous studies using traditional methods, which reported even lower records (Ruiz et al., 2007; Lasso et al., 2010; Lasso and Morales-Betancourt, 2017; Lasso et al., 2018; Méndez-López et al., 2019; Sánchez et al., 2019; Lasso et al., 2020; Morales-Betancourt and Lasso, 2021). This may be related to the behaviour of the terrestrial animal itself, in particular its opportunistic interactions with water such as swimming, wallowing, salivation during drinking and deposition of urine or faeces (Sales et al., 2020b; Lyet et al., 2021). The results of this eDNA study are consistent with previous research showing that more aquatic, semi-aquatic and nocturnal species were

### Table 1: Comparison of fish species detected by eDNA and traditional sampling in the Amazon and Orinoco basins.

<table>
<thead>
<tr>
<th>Subregion</th>
<th>Location</th>
<th>eDNA</th>
<th>Traditional sampling</th>
<th>Percentage of species detected with eDNA compared with traditional sampling (%)</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bojonawi Nature Reserve and adjacent areas</td>
<td>Bojonawi Nature Reserve (Orinoco River and floodplain)</td>
<td>262</td>
<td>270</td>
<td>97</td>
<td>Lasso et al., 2020</td>
</tr>
<tr>
<td></td>
<td>Orinoco River (main channel)</td>
<td>223</td>
<td>128</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>El Pañuelo lagoon</td>
<td>90</td>
<td>147</td>
<td>61.2</td>
<td>Villa-Navarro et al., 2017</td>
</tr>
<tr>
<td></td>
<td>Streams</td>
<td>119</td>
<td>186</td>
<td>63.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bita River</td>
<td>147</td>
<td>254</td>
<td>57.9</td>
<td></td>
</tr>
<tr>
<td>Sierra de la Macarena National Park and Tillavá</td>
<td>Guayabero River Basin (includes Guayabero River-main channel, EL. Silencio Lagoon, Caño Cristales River, Yarumales stream, Don Ricardo stream)</td>
<td>91</td>
<td>113</td>
<td>83</td>
<td>DoNascimento et al., 2018</td>
</tr>
<tr>
<td></td>
<td>Tillavá River</td>
<td>57</td>
<td>138</td>
<td>41</td>
<td>Morales-Betancourt and Lasso, 2021</td>
</tr>
<tr>
<td>Puerto Nariño and adjacent areas</td>
<td>Correo Tarapoto Redondo lakes</td>
<td>33</td>
<td>169</td>
<td>21</td>
<td>Urbano-Bonilla et al., 2014</td>
</tr>
</tbody>
</table>

Table taken and adapted from Martinelli-Marin et al. (2020).
detected by eDNA than by observational methods: Caenou sp., Hydrochoerus hydrochaeris, L. geoffrensis, M. tridactyla, P. brasilensis, Tamandua tetradactyla, and T. terrestris (Sales et al., 2020a; Coutant et al., 2021). Moreover, the lack of iconic carnivores, such as the jaguar (Panthera onca), cougar (Puma concolor), and ocelot (Leopardus pardalis), in the current study may be due to their ecological characteristics, such as a relatively large home range, dietary preferences, or solitary behavior (Gompper et al., 2006; Harper et al., 2019; Sales et al., 2020a; Broadhurst et al., 2021; Lyet et al., 2021). Alternatively, another possible reason for the low abundance of tetrapods, mainly amphibians, birds and reptiles, may be that we used universal vertebrate primers that amplify a wide range of organisms (Yang et al., 2014); some may have lower primer fidelity and less abundant reads may be too rare to detect. Hence, the identification of specific taxa depends on primers that amplify DNA exclusively from the target taxa and are sensitive enough to detect even small amounts of target DNA from low-quality samples (Macdonald and Sarre, 2017). Moreover, newer studies have employed eDNA metabarcoding with targeted or modified primers and different genetic markers to detect a wider range of amphibian, bird and reptile species (Valentini et al., 2016; Sasso et al., 2017; Wilson et al., 2017; Ushio et al., 2018; Neice and McRae, 2021; Lam et al., 2022). Also, eDNA extraction can be used to target a wide range of environments other than the aquatic ones, which favors the detection of terrestrial organisms (Ficetola et al., 2019). Alternative water sources have been used to obtain environmental DNA, for example, in Sweden, DNA samples for ungulate species identification have been obtained from browsed branches of deciduous trees (Nichols et al., 2012), while in tropical forests, salt licks have been used to detect mammals (Ishige et al., 2017).

Although eDNA is still an exploratory tool, our study demonstrates its effectiveness in the interpretation of taxon richness and diversity of aquatic, semi-aquatic and terrestrial vertebrates in different water bodies of the Colombian Amazon and Orinoco basins. With a simple water sample, we were able to detect organisms that had not been previously identified by traditional surveys which require expensive and time-consuming sampling devices and larger spatial scales. Therefore, we recommend using eDNA as an efficient tool to complement and extend biological inventories in a more rapid and non-invasive manner. However, it is necessary to develop more complete DNA sequence reference databases that include all neotropical species sequences discovered to date. Additionally, it is important to create standardized protocols based on the animal’s habitat and behavior of interest to maximize the benefits of eDNA detection.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://datadryad.org/stash, https://doi.org/10.5061/dryad.15dv41p0d.

Author contributions

DM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. CL: Data curation, Supervision, Validation, Writing – review & editing. SC: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this project was available from a private donation to Universidad de los Andes (project “Conservación del Manatí Antillano (Trichechus manatus) en Colombia y el Caribe uso de nuevas tecnologías como apoyo efectivo en procesos de recuperación de especies amenazadas”) and from the Faculty of Sciences Research Program 2019-2022 Initiative to SC (Programa de Investigación Facultad de Ciencias, Universidad de los Andes).

Acknowledgments

The authors would like to express their gratitude to Juan Esteban Jacome for his valuable assistance in creating the graphs. Additionally, we are thankful for the help provided by C. Romero, J. D. Lozano, L. Baldrich, M. Morales, and B. Castañeda during the field work. We would also like to thank Fundación Omacha for their hospitality at the Bojonawi Nature Reserve and Puerto Nariño.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2024.1409296/full#supplementary-material


Holbourn, A. E., Kuhnt, W., Clemens, S. C., Kochhann, K. G., Johnick, J., Lübbers, J., et al. (2018). Late Miocene climate cooling and intensification of Eocene oceanic anoxic events may have been driven by exchange with the deep ocean. Sci. Rep. 8, 108225. doi: 10.1038/s41598-018-34192-3


Environmental DNA: a new low-cost monitoring tool for pathogens in salmonid mec.16023

Environmental DNA for biomonitoring.

Frontiers in Ecology and Evolution frontiersin.org14

absent species.

580. doi:10.1111/fwb.13094

5 using environmental DNA.

secretive marsh bird.

[Accessed August, 2020].

demography, selection, and coalescence of mitochondrial and nuclear genes in acua


Rodriguez, L. E., Brunbjerg, A. K., Clausen, K. K., Dalby, L., Fløjgaard, C., Juel, A.,


Saito, T., and Doi, H. (2021). A model and simulation of the in


