Assessing freshwater biodiversity through environmental DNA along the Mekong River in Ubon Ratchathani Province, Thailand and from Bolikhamxai to Savannakhet Provinces in Lao P.D.R.

August 2024

Submitted To:
WWF Laos
WWF Thailand

Prepared By:
Jack Eschenroeder
Sinsamout Ounboundisane
Dana Lee

FISHBIO
1617 S. Yosemite Ave.
Oakdale, CA 95361
209.847.6300
www.fishbio.com

Date Submitted:
August, 2024

Table of Contents

Executive Summary

A total of 70 environmental DNA (eDNA) samples were collected between 3 April and 11 April, 2024 across 14 sites (eight sites in Laos and six sites in Thailand). Single-use sample collection kits were used to collect four replicates spanning the width of the river, as well as a single negative control at each site. This eDNA sampling effort was a follow-up to the baseline eDNA monitoring that took place in 2022, which focused on sampling Fish Conservation Zones (FCZs) that were established by phases 1 and 2 of the Lao-Thai Fisheries (LTF) Project. This second iteration of eDNA monitoring included repeated sampling of seven sample sites evaluated by the 2022 study, as well as seven new sample sites, including several representing FCZs established by phase 3 of the LTF Project.

Review and cleaning of the sequence data obtained from the samples resulted in the detection of a total of 106 distinct fish taxa representing 29 families, 64 genera, and 66 species. Note that the term "taxa" is used throughout this report to refer to unique taxonomic assignments of detected DNA sequences, and should be regarded as a conservative estimate of the total, species-level diversity present in the samples. This is a peculiarity of eDNA studies that arises from an inability to perfectly assign all detected sequences to a species-level identity. As such, multiple sequences assigned to the genus, family, or order level represent an unknown number of species. In this particular study, some sequences could only be assigned to the order level (Siluriformes and Osteoglossiformes), and some to the family level (Balitoridae, Cichlidae, Cyprinidae, Gobiidae, Hemiramphidae, Pangasiidae, and Sisoridae). Additional sequences assigned only to the genus level represented 23 distinct genera. Finally, sequences assigned to the species level represented 66 different species. Notably, the total diversity detected in these samples was slightly higher than that reported in the 2022 sequence data (106 versus 93 taxa), and the level of taxonomic resolution was considerably higher (66 species-level IDs in 2024 versus 50 species-level IDs in 2022). A total of 33.3% of the detected fish taxa belonged to the family Cyprinidae. The number of fish taxa detected per site ranged from 16 to 41, with a mean of 26.2. The greatest level of fish diversity was detected in Thapea Village, followed by Kudjub and Kaengsadok. A total of three non-native fish taxa were detected – tilapia (*Oreochromis* sp.), common carp (*Cyprinus* sp.), and silver/bighead carp (*Hypopthalmichtys* sp.). Detections of tilapia and common carp were broadly distributed in nine and six of 14 sites, respectively, whereas silver/bighead carp were detected in only a single location (Kudjub).

Four of the 12 priority target species – lesser tapah (*Wallago attu*), goonch (*Bagarius yarrelli*), Jullien's golden carp (*Probarbus jullieni*), and silver barb (*Barbonymous gonionotus*) – were definitively identified with species-level sequence assignments. DNA sequences that could only be assigned to the genus, family, or order level may have included six additional target species – *Hemisilurus mekongensis*, Asian red tail catfish (*Hemibagrus wyckioides*), Mekong giant catfish (*Pangasianodon gigas*), dog-eating catfish (*Pangasius sanitwongsei*), thick-lipped barb (*Probarbus labeamajor*), and two headed carp (*Bangana behri*) – but lack of species-level assignment and missing genetic reference sequences for some of these species preclude confident determination of whether they were detected or not. Notably, all of the target species identified in

the 2022 were also detected in this sampling effort, with the exception of the family-level stingray detections (Dasyatidae) that occurred in 2022. Stingray eDNA may have been missed in the samples collected this year because either stingrays were not present in the sampled areas, or due to the use of the MiFish primers in this study, which more specifically target bony fish species. However, the MiFish primers provided significantly improved taxonomic resolution and generated a greater number of species-level IDs overall than the primers used in 2022.

Basic comparisons between the eDNA detection data and fisher catch reporting data generated by the LTF Project in spring of 2024 demonstrated that there is imperfect overlap between the species detected by these two methodologies. Specifically, 60 of the 95 genus and species level taxonomic detections in the eDNA data did not occur in the fisher catch reporting data, and 31 of the 63 genus and species level taxonomic detections in the fisher catch reporting data did not occur in the eDNA data. Though caution is warranted in interpreting these results due to differences in the timing and location of the two sampling programs, they suggest that catch reporting and eDNA monitoring may be complimentary to one another. More targeted research to assess this complementarity and how these approaches may best be deployed in tandem is warranted.

This project represents a valuable continuation of molecular monitoring of fish diversity in FCZs established by the LTF project. Further, modification of the sample collection and analysis protocols between the 2022 project and this project provides a valuable opportunity to contribute to understanding of best practices for eDNA studies in the Mekong. Continuing data collection and archiving of both extracted DNA and sequence data will be of great value for monitoring the response of the fish community in the project areas both in response to local management practices and large-scale shifts in the climate and hydrology of the region.

The field crew collecting eDNA samples at Tha Long.

Background

The scope of this project was defined by the stated global project objectives set forth by WWF, which include 1) building freshwater biodiversity baselines for eight river basins in Asia and Africa using eDNA monitoring techniques by 2025, and 2) improving the understanding that targeted stakeholders have of forgotten freshwater biodiversity and providing access to more freshwater biodiversity eDNA data. Beyond these global objectives, specific objectives for WWF's eDNA monitoring include creation of extensive new and transparent freshwater biodiversity datasets; improved understanding of freshwater Asian Species Action Partnership (ASAP), threatened, data deficient, economically and culturally valuable, and invasive non-native species; increased evidence for new and updating Red List Assessments, Key Biodiversity Assessments, and other conservation decision making at the policy and program scale; and increased evidence and understanding of freshwater biodiversity for communities, governments, NGO sector, and private sector decision making. In service of all of these goals, FISHBIO developed and coordinated this eDNA sampling project in a manner that was specifically focused on growing the evidence base for and awareness of freshwater biodiversity, and designed methodologies to ensure the establishment of a comparable baseline of metabarcoding data.

The locations sampled in this study were selected to target existing Fish Conservation Zones (FCZs), which were established with support from WWF through the LTF Project. Sixteen FCZs established by Phase 1 and Phase 2 of this project were evaluated by a similar eDNA study conducted in March and April of 2022. Because Phase 3 of this project was initiated in January of 2023, this follow up study was carried out to collect data on the fish community and the presence of priority species in newly established FCZs, and to serve as a point of comparison with the baseline 2022 data.

Methodology

Field Methods

Study Design and Sampling Protocol

The design of this study was intended to allow for comparison with the eDNA data collected in 2022, while also taking steps to improve species detection and improve confidence in the accuracy of results through the inclusion of additional sample replicates (n = 4 per site) and negative controls (n = 1 per site) in the field. Sites that were included in the 2022 study were prioritized for sampling by this project, with modifications to sampling locations only being made in instances where challenges related to site access precluded revisiting the same location.

All environmental DNA (eDNA) samples for this project were collected using single use kits provided by Jonah Ventures (Boulder, Colorado, USA). These kits each contained a 60mL syringe, a small syringe of Longmire's solution for sample preservation, a pair of gloves, a filter cartridge containing a 5µm filter, and two caps for the filter cartridge. These kits were selected because they do not require specialized equipment, they are simple to use, and samples stabilized with

Longmire's solution do not require refrigeration in the field. Further, the filters are enclosed in a cartridge that reduces potential for contamination, and the single-use nature of the kits eliminates the need for decontamination of equipment.

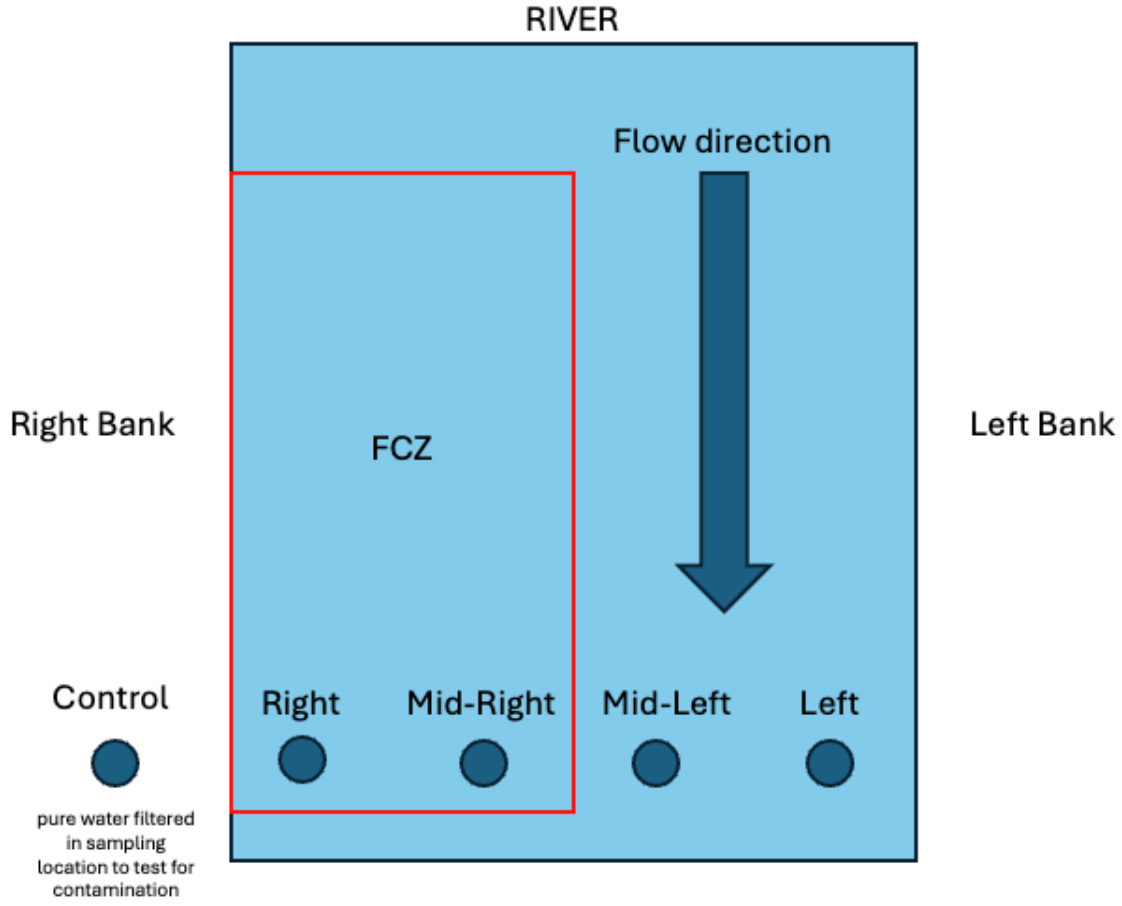


Figure 1. A diagram of the general sampling approach used in each location.

The full field procedure is described in Appendix 1. Briefly, the procedure used to collect a sample involved using the syringe to draw up 60mL of water from just below the surface (i.e., ~5cm deep) at the sample site, affixing the filter disc to the end of the syringe, forcing the water through the filter, removing the disc and drawing up another 60mL of water, and repeating this process until the filter became clogged with sediment and no additional water could be forced through. After filtering as much water as possible, the Longmire's solution was applied to the filter cartridge before capping it and storing it in a secure location protected from sunlight for transport. To limit risk of contamination, care was taken during this procedure to keep sample materials contained in the kit pouch until needed in the sample collection process to limit exposure to potential contaminants, and the individual collecting the sample wore the new gloves contained in each sample kit at each site. Further, samples were collected facing in an upstream direction to ensure that any DNA-containing materials that may have been carried on the boat or on the clothing of the person collecting the sample would not flow into the location where water was being filtered.

Working from one shoreline to the other, a total of four samples spanning the width of the river channel were collected at each site (Figure 1). The collection of samples both in the open water of the mid-channel and near the shoreline on both sides was intended to improve detection of the fish community, as past studies have shown that collection of samples across these two habitats may lead to detection of additional fish taxa (Blackman et al. 2021; Eschenroeder et al. 2024). After collecting the first near-shore sample, the field crews used boats to move across approximately 25% of the width of the river in a straight line and collect a second sample, then another 25% across the width of the river to collect a third sample, before finally collecting the fourth sample near the opposing shore (Figure 1). The samples along the margin of the river were collected approximately 2 meters from shore where possible, although the crew was unable to sample that close in certain location due to restrictions on crossing the national border between Thailand and Lao PDR. A fifth sample was collected at each site by filtering bottled water in the same location. This sample served as a negative control, and was intended to allow for detection of any contamination that may have occurred either during sample collection or during sample transport and storage. The intention of collecting the four in-river samples at the downstream end of the FCZs was to allow for detection of DNA from species present within the FCZ itself, although downstream transport of eDNA means that it is not possible to definitively ascertain whether detected species were present within the FCZ or further upstream.

For each sample collected, the field team recorded a variety of data on field datasheets created for the project (Appendix 1). These data included the date and time that the sample was collected, the name of the site (i.e., the village name), the sample kit code, the total volume of water that was pushed through the filter (in mL), the GPS coordinates (latitude and longitude) where the sample was collected, and the sample location in the river (right bank, right middle, left middle, left bank, or control). Additionally, the crew recorded notes on potential DNA sources in the sample area, including net pens, restaurants, and observed species. These data were all entered into a shared electronic database and were reviewed by FISHBIO staff. In cases of uncertainty, FISHBIO communicated with WWF project staff to obtain the details necessary to correct errors and ensure the quality of the data. Collected samples were labeled with a site code and stored in a cool, dark place in order to reduce the risk of DNA degradation. Samples were transported back to Vientiane, and from there were shipped to the US for analysis.

Laboratory Methods

Sample Processing and Sequencing

Detailed laboratory methodology is provided in Appendix 2. Briefly, DNA metabarcoding employed MiFish primers (Miya et al. 2015), which target the 12S region of the mitochondrial genome and are known to provide genetic resolution of fish taxa to the species level. PCR amplification was performed in replicates of six, none of which were pooled. Each round of PCR included a non-template control to identify any laboratory cross-contamination. Metabarcoding produced hundreds of thousands of sequences, which were processed using a custom bioinformatics pipeline that summarized the number of unique exact sequence variants (ESVs) amplified in each water sample. ESV assignments were based on percent similarity to reference sequences from GenBank, plus additional unpublished sequences from specimens sequenced by

Jonah Ventures. A recursive matching algorithm assigned ESVs to known species according to sequence similarity, and if below a similarity threshold ESVs were assigned to higher taxonomic levels. In most cases, ESVs were designated to species, but genus, family, and order level assignments did occur.

Data Analysis

Data Cleaning and Initial Analyses

Sequence data were provided by Jonah Ventures in csv format, and these data were imported into R statistical software (R Core Team 2024) for filtering, formatting, and analysis. Initial processing steps included identifying any samples that contained no detected sequences, and filtering out sequences belonging to non-fish taxa (e.g., mammals). In addition, a similarity threshold of 97% was applied to the data, meaning that any detected sequences that had less than a 97% match with available reference sequences were removed prior to subsequent analyses. Although there is variation in the thresholds used for filtering of eDNA sequence data, a threshold of 97% is commonly adopted in eDNA studies focused on fish (e.g., Blackman et al. 2021; Evans et al. 2017), and falls within the suggested optimal range for generalist markers (96-99%; Bonin et al. 2021).

Following these initial steps, all taxa identified in the samples were compared to a list of species known to occur in the Mekong (Jerde et al. 2021). This allowed for flagging of detected taxa that are not native to the basin, which may have arisen either from the presence of introduced species (e.g., *Oreochomis* species like tilapia that are known to occur in the study region) or from contamination in the laboratory. Detected sequences belonging to non-native species were reviewed and interpreted based on expert opinion, review of notes on species observations from the field sampling team, and consultation with Lao fisheries biologists.

Detected sequences remaining after these filtering steps were summarized into "molecular operational taxonomic units," or MOTUs for short. These MOTUs represent unique taxonomic identifications based on the detected sequence data. While species level identities are ideal, it is often necessary to consider eDNA sequence data in these more generalized terms, as variation in the taxonomic resolution of the primers used and limitations to existing reference libraries mean that certain sequences may only be reliably identified to higher taxonomic levels such as genus or family. As such, referring to the detected sequences as “species” can be misleading, and instead summaries of detections performed for this project refer to MOTUs, which are hereafter more simply referred to as “taxa.”

Several additional analyses were performed to characterize the diversity captured by the sampling method employed by this study. Basic visualizations were created to evaluate the number of taxa detected per individual sample, the number of taxa detected per site, and the relationship between the number of detected taxa and sample volume (i.e., the amount of water pushed through the collection filter). A basic linear model was used to determine whether there was a significant relationship between sample volume and total taxa detected. In addition, an accumulation curve depicting the relationship between total taxa detected and total samples collected was developed

to evaluate whether the level of sampling intensity applied was sufficient to maximize detection of fish diversity.

Summaries of Diversity and Priority Species Detections

Taxonomic data were pooled by site in order to evaluate species detections and compare detected fish diversity across the sampled locations. These site level detection data were used to construct a presence/absence matrix of all detected taxa across all sites, and additional summaries included total taxa detected at each site, total priority species detected at each site, and non-native taxa detected at each site. Further, stacked bar graphs clustering the detected taxa by family were used to visually compare the taxonomic structure of the fish communities detected in each sampled location.

Similarities and Differences from the 2022 eDNA Study

Sample Processing and Sequencing

The eDNA sample collection protocol employed in the 2022 study relied on the use of a peristaltic pump operated with a battery powered drill, and a large cylindrical filter with a filtration surface area of 500 cm^2. This equipment was used to filter water across a larger area by sampling from a boat being driven in a zip-zag transect throughout the FCZ. In contrast, the study conducted in 2024 relied on the single use kits described above, which were used to collect four replicate samples across the width of the channel at the downstream end of the target FCZ. Although the total volume filtered by the cylindrical filters used in 2022 was considerably larger, there is a tradeoff between larger and smaller filter sizes that must be considered. Filters with larger pore sizes like those used in 2022 allow for filtration of greater volumes (e.g., Durand et al., 2022), but this carries greater risk of sample contamination with PCR inhibitors (Herder et al., 2014), which are abundant in turbid systems (Kumar et al., 2021). Further, when filtering smaller volumes (e.g., < 500 mL), finer pore sizes capture significantly more DNA than filters with larger pore sizes (Jeunen et al., 2019). Although many approaches to eDNA sample collection exist and there is still considerable uncertainty regarding optimization of eDNA sample collection in the Mekong, the best available science suggested that the best option for the application of single-use kits employed in this study was to increase the number of samples rather than seeking to increase the volume filtered per sample. Indeed, sample volumes as low as 100 mL collected with kits similar to those used in this project have been demonstrated to effectively detect biodiversity in turbid, tropical systems if suitable replication is achieved (Blackman et al., 2021).

Another difference between the 2022 sampling and this project was the primers used. The 2022 study relied on the use of the Teleo primers, which sample a hypervariable portion of the 12S region of the mitochondrial genome (Taberlet et al., 2018; Valentini et al., 2016). These primers are commonly used in studies focused on fish diversity. However, the analysis of samples for this project relied on the use of the MiFish primers, which sample a different portion of the 12S region of the mitochondrial genome (Miya et al. 2015). Each of these primers carries advantages and disadvantages. For example, the Teleo primer may provide the ability to detect a broader range of taxa (Polanco-Fernandez et al. 2021), but the MiFish primers are able to provide a greater

taxonomic resolution. Given FISHBIO's success with the use of MiFish primers in past eDNA projects in the Mekong, those primers were selected for use in this project.

Considering these differences in both field and lab protocols employed in 2022 compared with 2024, caution is warranted in drawing direct comparisons between the data obtained from each study. However, the use of variable methods does provide a valuable opportunity to evaluate consistency of results, and potentially glean insight into best practices for eDNA surveys in the Mekong.

Results

Sample Collection

A total of 70 samples were collected, 30 of which were collected by the Thai team across six villages and 40 of which were collected by Lao team across eight villages (Table 1; Figure 2). These samples were comprised of four in-river samples collected across the width of the river at each site, as well as a single negative control collected by filtering pure water at each site. Because the areas covered by the Lao and Thai teams were different than in 2022, not all sites sampled in 2022 were resampled in 2024 (Table 1). In instances where access across the river was not possible by the team, they sampled the next closest downstream FCZ on their side of the national border.

Table 1. Locations where eDNA samples were collected. Four replicates and one negative control were collected in each location, resulting in a total of 70 samples, 14 of which were negative controls.

Date	Village	District	Province	Country	Comparison with 2022 Sampling
3 April 2024	Koum*	Khong Chiam	Ubon Ratchathani	Thailand	New point – extended downstream
	Ta Mui*	Khong Chiam	Ubon Ratchathani	Thailand	New point – extended downstream
4 April 2024	Tha Long*	Khong Chiam	Ubon Ratchathani	Thailand	Same point sampled in 2022
	Kan Tha Kaewn*	Khong Chiam	Ubon Ratchathani	Thailand	Same point sampled in 2022
5 April 2024	Pha Chan*	Pho Sai	Ubon Ratchathani	Thailand	Same point sampled in 2022
	Pak Ka Lang*	Pho Sai	Ubon Ratchathani	Thailand	New point – closed FCZ on Lao side in 2022
8 April 2024	Thaphea	Songkhone	Savannakhet	Laos	Same point sampled in 2022
	Heunhin	Xayphouthong	Savannakhet	Laos	Same point sampled in 2022
9 April 2024	Sivlilay	Xaybouly	Savannakhet	Laos	Same point sampled in 2022
	Kudjub	Nongbok	Khammoaun	Laos	New point – closed FCZ on Thai side in 2022
10 April 2024	Namuang	Thakaek	Khammoaun	Laos	Same point sampled in 2022 (Ban Mueng Kao)
	Poung-Nua	Hinboun	Khammoaun	Laos	New point – closed FCZ on Thai side in 2022
11 April 2024	Boungkuang	Pakkading	Bolixhamxay	Laos	New point – closed FCZ on Thai side in 2022
	Kaengsadok	Paksan	Bolixhamxay	Laos	New point – closed FCZ on Thai side in 2022

*These are new FCZs that were establishing in Phase III of the Lao-Thai Fisheries Project.

Figure 2. A map of the villages where eDNA samples were collected. See Table 1 for site details.

Analyses of Sequence Data

Sequence Detections and Data Cleaning

A total of 1,376 sequences were detected in the 70 samples collected across 14 sites. All sequence and sampling data have been uploaded to the NCBI SRA database (BioProject ID: 1130543) and are publicly available. In total, 12 of the samples yielded no detectable sequences, although 11 of these were negative control samples and therefore were not expected to contain DNA. The only in-river sample that did not contain any detectable DNA was that collected on the right bank at Namuang Village. Of the 1,376 total sequences, 11 could not be identified, and 57 belonged to mammalian taxa, including domestic pigs (*Sus scrofa*), domestic cattle (*Bos* spp.), and humans (*Homo sapiens*). No other non-fish taxa were detected and the remaining 1,308 sequences were all assigned to fish. The filtering out of sequences with <97% match to available genetic reference sequences removed a total of 111 sequences, leaving 1,197 that were retained for subsequent analyses. Of these, six were identified to the class level (Actinopteri), 18 were identified to the order level, 95 to the family level, 171 to the genus level, and 907 to the species level.

These sequences were then compared to the list of native species from the Mekong from Jerde et al. (2021), and potentially erroneous species were flagged. This resulted in the identification of two sequences likely arising from contamination: one belonging to yellow bullhead (*Ameirus natalis*) detected in a sample from Sivilay Village and one belonging to prickly sculpin (*Cottus asper*) detected in a sample from Koum Village. Further, an additional three sequences assigned to the family Cottidae were detected in samples from Koum and Kudjub villages. Fish in the Cottidae family are only found in temperate regions, and both of these species are found in North American streams, and are not likely to have been present in the sampled locations. As such, it was surmised that they were detected due to contamination in the laboratory, and both the species level and Cottidae family level sequences were removed from the dataset prior to subsequent analyses.

Other sequences assigned to non-native fish taxa included many belonging to tilapia species (genus *Oreochromis*), silver/bighead carp (genus *Hypophthalmichthys*), and common carp (genus *Cyprinus*). Because these species are known to have been introduced to the region for aquaculture and have been detected by past eDNA studies (including that conducted by WWF in 2022), these sequences were retained for subsequent analyses.

Several other species level assignments belonged to species not known to occur in the Mekong, but likely arose from limitations in available reference libraries and/or poor resolution among species within certain genera. These included the red algae eater (*Crossocheilus langei*), which is native to Malaysia and Sumatra; the Chiangmai stream goby (*Rhinogobius chiengmaiensis*), which is found only in the Chao Phraya basin in Thailand; and lacustrine goby (*Gobiopterus lacustris*), which is found only in the Philippines. However, all of these genera are represented by different species that do occur in the Mekong – *Rhinogobius mekongianus*, *Crossocheilus oblongus*, *C. cobitis*, *Gobiopterus brachypterus*, and *G. chuno*. Because it could not be definitively determined whether the detections of these species represent novel introductions or simply misidentification of native species, all were retained at the genus level for subsequent analyses (i.e., *Rhinogobius*

sp., *Crossocheilus* sp., and *Gobiopterus* sp.). All other taxonomic IDs were reviewed and deemed to be accurate, leaving a total of 1,190 fish sequences in the final cleaned data set.

Negative Controls

Out of the 14 negative control samples collected (one from each village), only three were found to contain detectable DNA. Two of these contained only sequences belonging to domestic pigs (*Sus scrofa*). These pig sequences were also observed in many of the in-river samples, and likely arose from contamination either during sample collection or during sample transport and storage. The only negative control sample containing DNA belonging to a fish species was that collected at Pong-Nua Village, which was found to contain a sequence that was assigned to goldfin tinfoil barb (*Hypsibarbus malcomi*). Notably, DNA belonging to this species was detected in samples from every village included in the project, although it did not occur in any of the other negative control samples. Given that this is a common species in the sampled region, and because the other negative controls did not indicate widespread contamination across all samples, all detections of this species were retained for subsequent analyses. Overall, the general lack of contamination in the negative control samples greatly improves confidence that detected fish taxa were truly present.

Detected Diversity

Among the samples that yielded detectable DNA, fish sequences detected per sample varied from two to 50, with a mean of 21.71 (Figure 2). Note that in some instances multiple unique sequences were assigned to the same taxa, thus the disparity in counts of taxa detections per site and unique sequence detections per sample. Total volumes of water filtered per sample ranged from 60 to 200 mL, with a mean of 144.73. Comparison of the volume of water filtered with the total number of sequences detected showed no clear pattern (Figure 3), and linear modeling failed to detect any significant relationship between sample volume and total sequences detected.

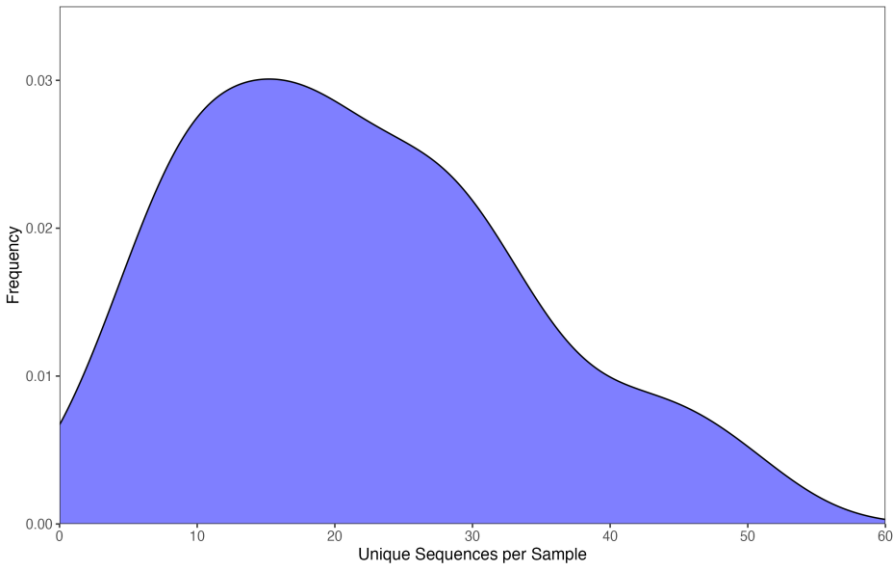


Figure 2. A density curve depicting the number of unique sequences detected per sample. Note that negative controls were excluded from this visualization, and only in-river samples were considered (n = 56).

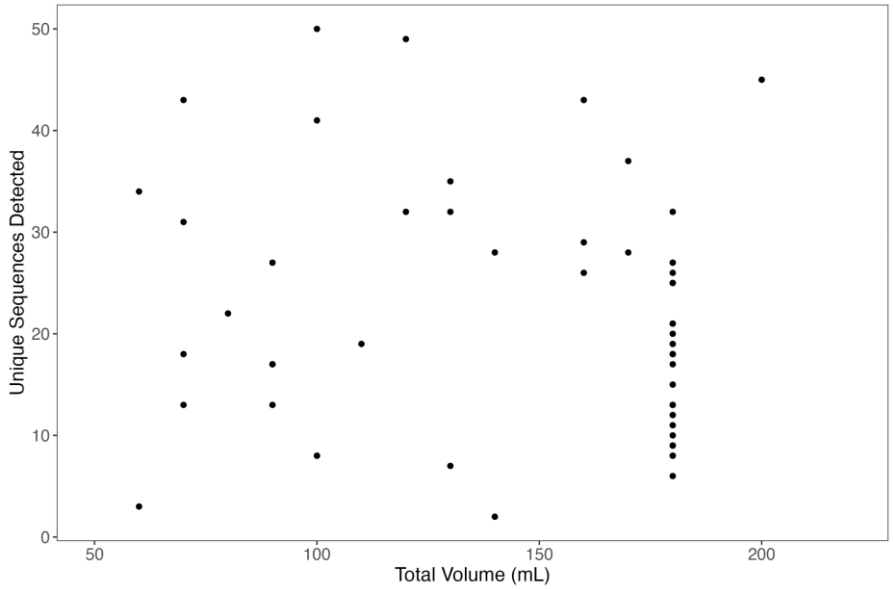


Figure 3. Total sequences detected versus volume of water filtered for each sample. Note that negative controls were excluded from this visualization, and only in-river samples were considered (n = 56).

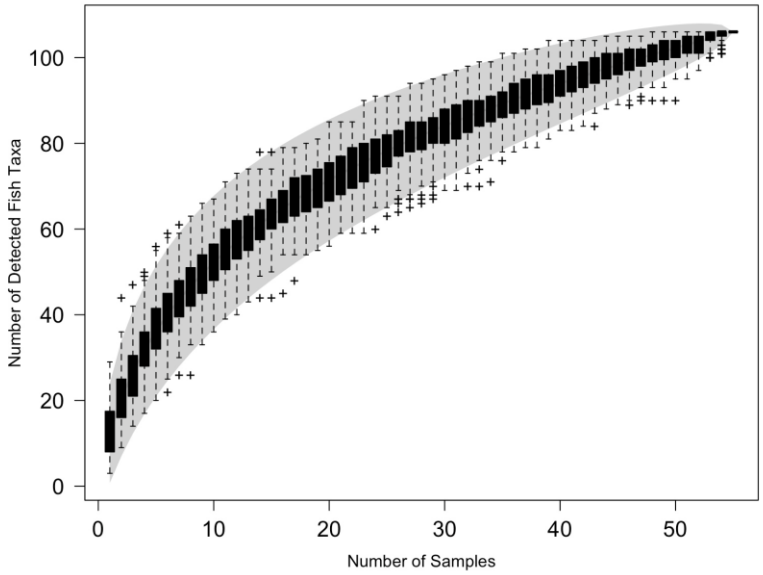


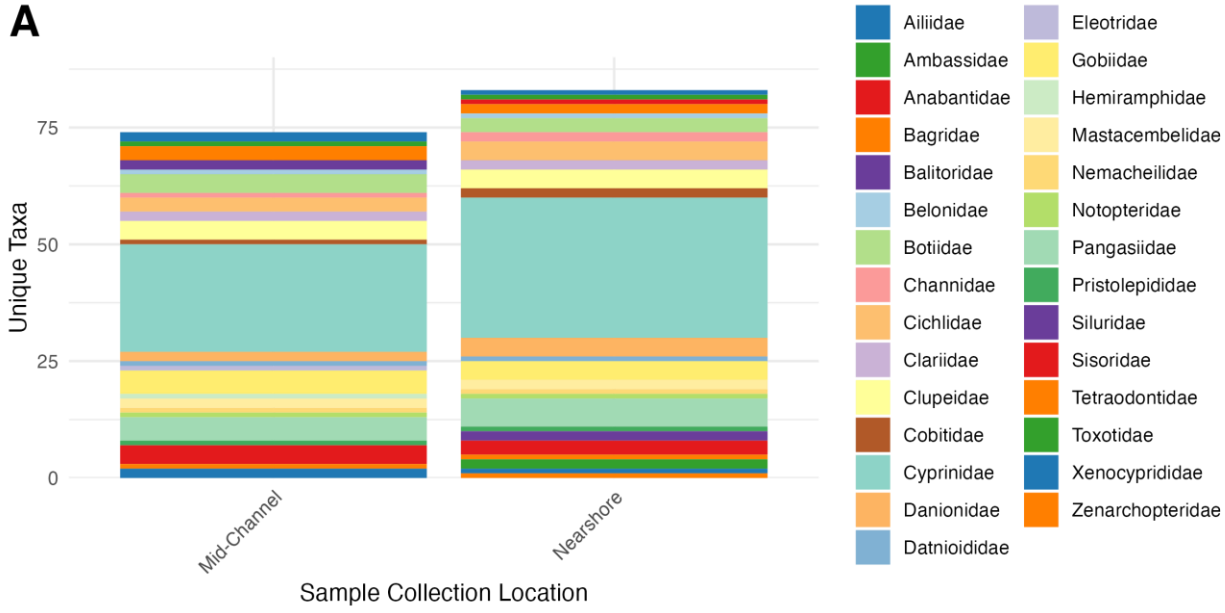
Figure 4. An accumulation curve depicting the number of additional fish taxa detections added with each additional eDNA sample collected. Note that negative controls were excluded from this visualization, and only in-river samples were considered (n = 56).

The accumulation curve indicated a decreasing slope with increasing numbers of samples, indicating diminishing numbers of novel taxa detections with increasing sampling effort (Figure 4). The number of detected taxa continued to rise across all 56 in-river samples, suggesting that additional taxa may have been detected with increased replication. Although the absolute maximum level of detection may not have been achieved, the steep decrease in the slope of the accumulation curve suggests that the number of samples collected for this study was sufficient to

capture the majority of the fish diversity that is detectable with the sampling and laboratory protocols used.

Diversity Detections by Nearshore and Mid-channel Sampling

The collection of four samples spanning the width of the river channel was intended to more fully capture the fish community present in each site, as previous studies have indicated that samples collected near the shoreline may contain eDNA from different taxa compared with samples collected in the middle of the channel (Blackman et al. 2021; Eschenroeder et al. 2024). Comparison of fish taxa detected in the nearshore samples with taxa detected in the mid-channel samples indicated that this was true, as four families (Anabantidae, Toxotidae, Zenarchopteridae, and Siluridae) were only represented by taxa detected in nearshore samples, and three families (Balitoridae, Eleotridae, and Hemiramphidae) were only represented by taxa detected in mid-channel samples (Figure 5A). In terms of total unique taxa detections, 29 taxa were detected only in nearshore samples, and 19 were detected only in mid-channel samples (Figure 5B).



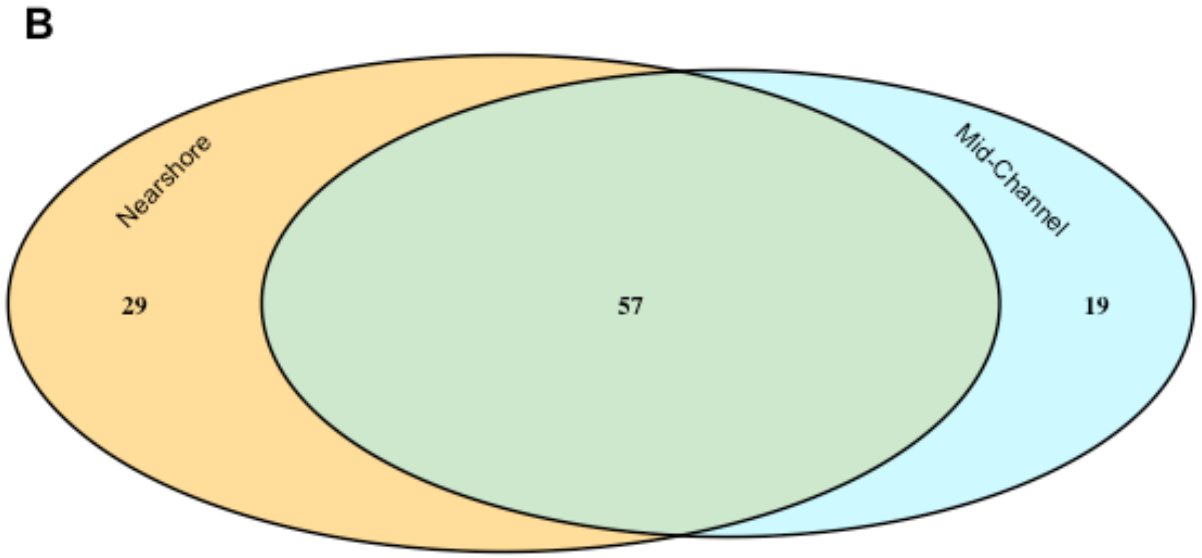


Figure 5. A) Total unique fish taxa detections by sample location (nearshore versus mid-channel), grouped by family. B) Venn diagram showing overlap of unique taxa detected in mid-channel and nearshore samples.

Diversity by Site

The 1,194 sequences in the cleaned dataset contained 486 unique sequence variants, which were assigned to 106 distinct fish taxa representing 29 families, 64 genera, and 66 species. Of the 12 sequences assigned to the order level, 11 belonged to Siluriformes and one belonged to Osteoglossiformes. Among family level assignments, one belonged to Balitoridae, 16 to Cichlidae, 63 to Cyprinidae, one to Gobiidae, one to Hemiramphidae, four to Pangasiidae, and five to Sisoridae. The 171 sequences assigned to the genus level represented 23 genera, and the remaining 906 sequences assigned to the species level represented 66 species (Table 2; Table 3). Notably, the total diversity detected was slightly higher than that reported in the 2022 sequence data (106 versus 93 taxa; Table 2), and the level of taxonomic resolution appeared to be considerably higher (66 species-level IDs in 2024 versus 50 species-level IDs in 2022; Table 3). The vast majority of the detected fish taxa belonged to the family Cyprinidae (33.3%; Figure 6). This pattern was very similar to that observed in 2022, as approximately 30% of taxa detected in eDNA samples from that year belonged to family Cyprinidae.

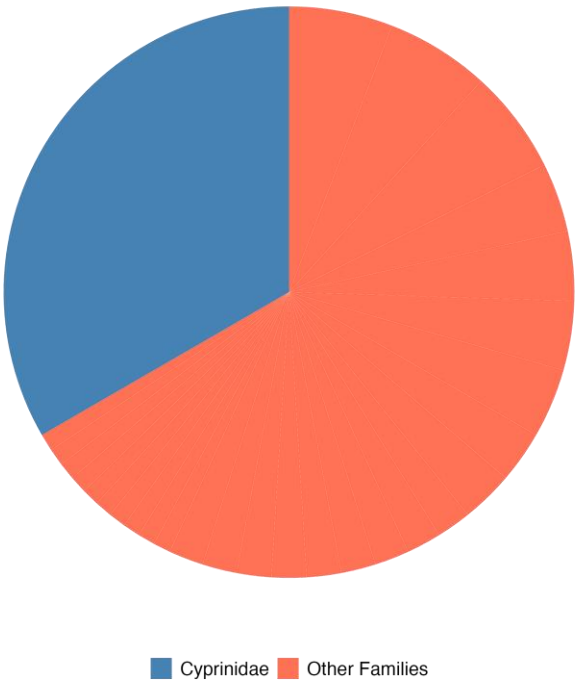


Figure 6. The percentage of taxa within the family Cyprinidae that were detected in the samples.

The number of taxa detected per site ranged from 16 to 41, with a mean of 26.29 (Figure 7). Across all the sample sites, the greatest level of fish diversity was detected in Thapea Village, followed by Kudjub and Kaengsadok (Figure 7). Non-native fish species – particularly tilapia (genus *Oreochromis*) and common carp (*Cyprinus carpio*) – were broadly distributed (Table 4). Tilapia were detected in nine out of 14 sampled locations (Sivilay, Namuang, Kaengsadok, Kudjub, Thapea, Heunhin, Poung-Nua, Boungkuang, and Pak Ka Lang; Table 4), whereas common carp were detected in six out of 14 sampled locations (Sivilay, Namuang, Heunhin, Pha Chan, Kudjub, and Kaengsadok; Table 4). Silver/bighead carp (*Hypophthalmichthys* sp.) were detected in only a single location (Kudjub; Table 4).

Table 3. Detected fish taxaI by sample site. Note that all order and family level assignments fell within the groups represented by the genera and species level assignments with the exception of Hemiramphidae. As such, with the exception of the single sequence assigned to family Hemiramphidae, only detections at the genus or species level are depicted in this table, and all others that could only be assigned to the class, order, or family level (n = 112 sequences) are not included. Native taxa detections are highlighted in green, and non-native taxa detections are highlighted in red. Priority target species and genus-level detections potentially representing priority target species are highlighted in yellow.

Family	Species	Kan Tha Kaewn	Pha Chan	Sivilay	Poung Nua	Koum	Namuang	Ta Mui	Kaengsadok	Kudjub	Thapea	Heunhin	Pak Ka Lang	Tha Long	Boungkuang
Ailiidae	*Clupisoma sinense*	–	–	–	–	–	–	–	X	–	–	–	–	–	–
	Laides longibarbis	X	–	–	X	X	–	–	X	–	–	–	–	–	–
Ambassidae	*Parambassis siamensis*	–	–	–	–	–	–	–	–	X	–	–	–	–	–
	Parambassis sp.	–	X	–	–	–	–	–	–	–	–	–	–	–	–
Anabantidae	*Anabas testudineus*	–	–	–	–	–	–	–	–	–	X	–	–	–	–
Bagridae	*Hemibagrus* sp.	X	X	–	–	–	–	–	–	X	X	X	–	–	X
	Hemibagrus wyckii	–	–	–	–	X	–	–	–	–	X	–	–	–	–
Balitoridae	*Homalopteroides smithi*	X	–	–	–	–	–	–	–	–	–	–	–	–	–
Belonidae	*Xenentodon cancila*	–	X	–	–	–	–	–	–	–	–	X	–	–	–
Botiidae	*Syncrossus helodes*	–	–	–	–	–	–	–	–	–	X	–	–	–	–
	Yasuhikotakia caudipunctata	–	–	–	–	–	–	–	–	–	–	–	X	–	–
	Yasuhikotakia eos	–	–	–	–	–	–	–	–	–	X	–	–	–	–
	Yasuhikotakia lecontei	–	X	–	–	–	–	–	–	–	X	X	–	–	–
	Yasuhikotakia sp.	–	–	–	–	–	–	–	–	–	X	–	–	–	–
	Yasuhikotakia splendida	–	–	–	X	–	–	–	–	–	–	–	–	–	–
Channidae	*Channa micropeltes*	–	–	–	–	X	–	–	X	–	–	–	–	–	–

	Channa striata	–	X	–	–	–	–	–	–	X	–	–	X	–	–
Cichlidae	*Oreochromis niloticus*	–	–	–	–	–	X	–	X	X	–	–	–	–	X
	Oreochromis sp.	–	–	X	X	–	X	–	X	X	X	X	X	–	X
Clariidae	*Clarias macrocephalus*	–	–	X	–	–	–	–	–	–	–	–	X	–	–
	Clarias sp.	–	–	X	–	–	–	–	X	X	–	–	–	–	–
Clupeidae	*Clupeichthys aesarnensis*	–	X	–	X	X	X	–	–	X	–	–	X	–	–
	Clupeichthys perakensis	X	–	–	–	X	–	X	X	X	X	X	–	X	X
	Clupeoides sp.	–	X	–	–	X	–	–	–	–	X	–	X	X	–
	Tenualosa thibaudeaui	–	–	–	–	X	–	–	–	–	–	X	X	X	X
Cobitidae	*Acantopsis ioa*	–	–	–	X	–	–	–	–	–	–	–	–	–	–
	Pangio anguillaris	–	–	–	–	–	–	X	–	–	–	–	–	–	–
	Pangio sp.	–	–	–	–	–	–	–	–	–	–	–	–	–	X
Cyprinidae	*Barbonymus altus*	–	X	X	X	X	X	–	X	X	X	X	X	–	X
	Barbonymus gonionotus	–	X	X	X	X	X	X	X	X	X	X	X	–	–
	Barbonymus schwanefeldii	–	–	–	–	–	–	–	–	X	–	–	–	–	–
	Cirrhinus microlepis	X	–	–	–	–	–	–	–	–	–	–	–	–	–
	Cosmochilus harmandi	X	X	X	X	X	–	X	X	–	X	X	X	X	X
	Crossocheilus sp.	–	–	–	–	X	–	–	–	–	–	–	–	–	–
	Cyclocheilichthys armatus	–	–	X	–	–	–	–	X	X	–	X	–	–	–
	Cyclocheilichthys enoplos	–	X	X	–	–	–	–	–	–	–	–	X	–	–

Cyclocheilichthys repasson	–	–	–	–	–	–	–	–	–	–	–	–	X	X
Cyclocheilichthys sp.	–	X	–	–	–	–	–	–	–	–	–	–	–	–
Cyprinus sp.	–	X	X	–	–	X	–	X	X	–	X	–	–	–
Epalzeorhynchos sp.	–	–	–	–	–	–	X	–	–	–	–	–	–	–
Garra fasciacauda	X	–	X	–	X	–	–	X	–	X	–	X	X	–
Hampala dispar	–	–	–	–	–	–	–	–	–	X	X	X	–	–
Hampala macrolepidota	–	–	–	–	X	–	–	–	–	–	–	–	–	–
Henicorhynchus siamensis	–	–	–	–	–	–	–	–	–	–	–	X	–	–
Hypsibarbus malcolmi	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Labeo chrysophekadion	X	X	X	X	X	X	X	X	–	X	X	X	X	–
Labiobarbus leptocheilus	–	–	–	–	–	–	–	–	–	–	–	–	–	X
Labiobarbus lineatus	–	–	–	–	–	–	–	X	–	–	–	–	–	–
Mekongina erythrospila	–	–	X	X	–	–	X	–	X	–	–	–	–	–
Mystacoleucus ectypus	–	–	–	–	–	–	–	–	X	X	–	–	–	–
Mystacoleucus marginatus	–	–	X	X	–	X	–	X	X	X	X	X	–	–
Mystacoleucus sp.	–	–	X	X	–	–	–	–	X	X	–	–	–	–
Osteochilus melanopleurus	–	–	–	–	–	–	X	–	X	–	–	–	–	–
Osteochilus sp.	–	–	–	–	–	–	–	–	–	X	–	–	–	–
Poropuntius normani	X	–	–	–	X	–	X	–	–	X	X	X	X	–
Probarbus jullieni	–	–	–	–	–	–	–	–	X	X	X	–	–	–

	Puntioplites falcifer	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Puntioplites sp.	X	–	X	X	X	–	X	X	–	–	X	–	–	–
	Scaphognathops bandanensis	X	X	X	X	X	X	X	X	X	X	X	X	X	–
	Sikukia gudgeri	X	–	X	X	X	X	X	X	X	X	X	X	X	X
	Systomus orphoides	–	–	X	–	–	–	–	–	–	–	–	–	–	–
Danionidae	*Opsarius koratensis*	–	–	–	–	X	–	X	–	–	–	–	–	X	–
	Opsarius sp.	–	–	–	–	–	–	–	–	X	–	–	–	–	–
	Raiamas guttatus	–	–	X	X	–	–	–	X	X	–	–	X	–	X
	Rasbora sp.	–	X	–	–	–	–	–	–	–	–	–	X	–	–
Datnioididae	*Datnioides* sp.	–	X	–	–	X	–	X	X	X	X	–	X	X	–
Eleotridae	*Oxyeleotris marmorata*	–	X	–	–	–	–	–	–	–	–	–	–	–	–
Gobiidae	*Eugnathogobius* sp.	–	X	–	–	–	–	–	–	–	–	–	–	–	–
	Gobiopterus lacustris	–	X	–	–	–	–	–	–	–	–	–	–	–	–
	Gobiopterus sp.	–	X	–	–	–	–	–	–	–	–	–	–	–	–
	Papuligobius ocellatus	X	–	–	–	X	–	X	–	–	X	X	–	X	–
	Rhinogobius sp.	–	–	–	–	–	–	–	–	–	X	–	–	–	–
Hemiramphidae	Unknown	–	–	X	–	–	–	–	–	–	–	–	–	–	–
Mastacembelidae	*Macrognathus siamensis*	–	–	–	–	–	–	–	X	–	–	X	–	–	–
	Mastacembelus sp.	–	–	–	X	X	–	–	–	X	X	X	–	X	–
Nemacheilidae	*Schistura* sp.	–	–	–	X	–	–	–	X	X	X	–	–	–	–

Notopteridae	*Chitala ornata*	–	X	X	–	–	X	–	X	X	–	–	–	–	X
Pangasiidae	*Pangasius bocourti*	–	–	–	X	–	–	–	–	X	–	–	–	–	–
	Pangasius macronema	–	X	X	–	X	X	–	X	X	X	–	X	–	–
	Pangasius nasutus	X	–	–	X	–	–	–	X	–	X	–	–	–	–
	Pangasius sp.	–	–	–	–	–	–	–	X	–	–	–	–	–	–
	Pseudolais pleurotaenia	–	–	X	X	–	–	–	–	X	X	–	–	–	–
Pristolepididae	*Pristolepis fasciata*	–	X	–	–	–	–	X	–	–	X	X	–	–	–
Siluridae	*Ompok siluroides*	–	–	–	–	–	–	–	–	X	–	–	–	–	–
	Wallago attu	–	–	–	–	–	–	–	–	X	–	–	–	–	–
Sisoridae	*Bagarius yarrelli*	–	–	–	X	–	–	–	–	–	–	–	–	X	–
	Glyptothorax sp.	–	X	–	–	–	–	–	–	–	X	–	–	–	–
	Glyptothorax trilineatus	–	–	–	–	–	–	–	–	–	X	–	–	–	–
Tetraodontidae	*Pao* sp.	–	–	–	X	–	X	X	–	–	X	–	–	–	–
Toxotidae	*Toxotes chatareus*	–	–	–	–	–	–	–	–	–	–	–	X	–	–
	Toxotes sp.	–	–	–	–	–	–	–	–	–	–	–	X	–	–
Xenocyprididae	*Hypophthalmichthys* sp.	–	–	–	–	–	–	–	–	X	–	–	–	–	–
	Paralaubuca typus	–	X	–	X	–	–	–	–	–	–	–	–	–	–
Zenarchopteridae	*Dermogenys pusilla*	–	X	–	–	–	–	–	–	–	–	–	–	–	–

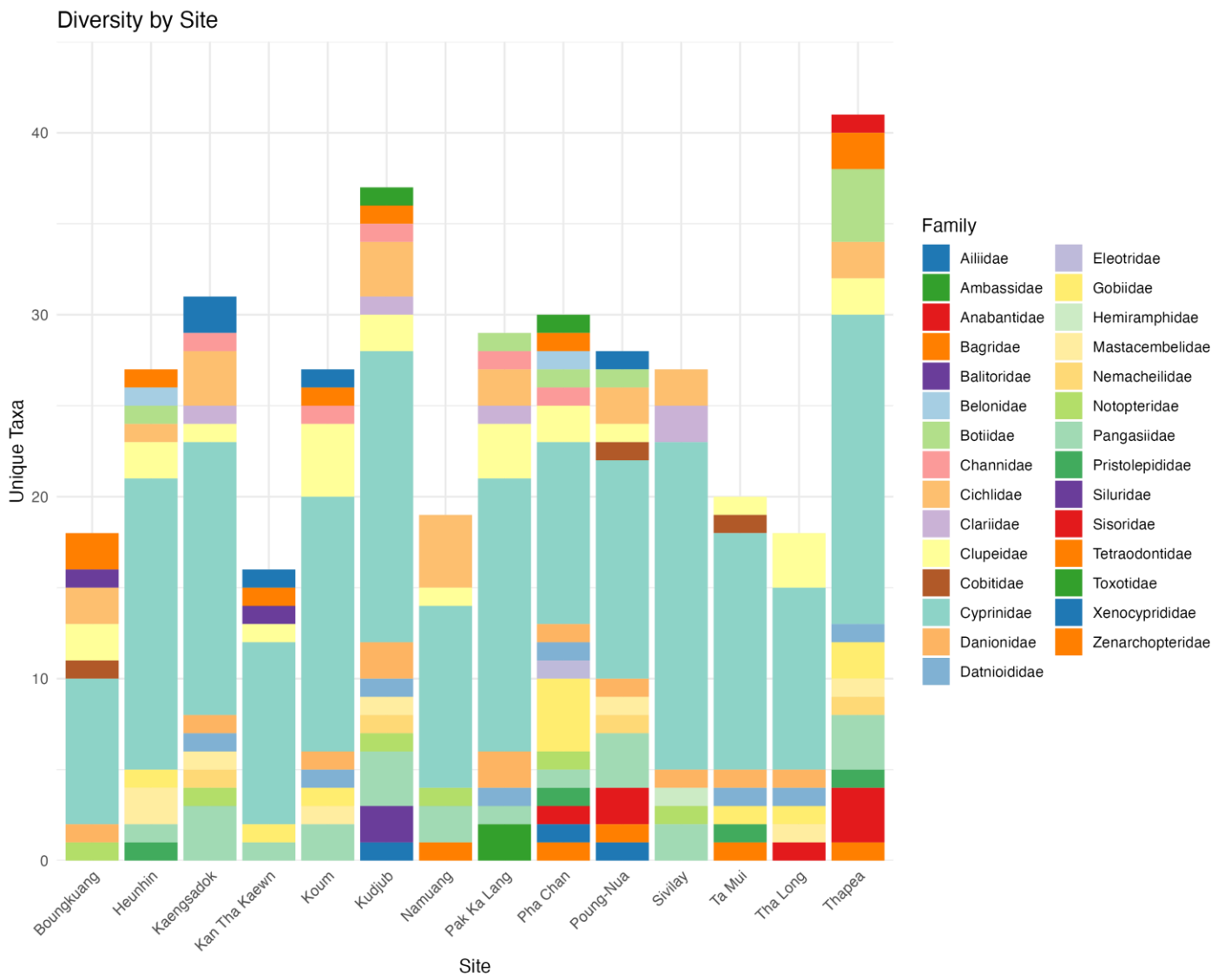


Figure 7. Total unique fish taxa detections by site, grouped by family

Table 3. The 66 species-level identities assigned from the sequence data, arranged by family. Non-native species are highlighted in red.

Family	Species-Level Detections	Total Species-Level IDs
Ailiidae	*Clupisoma sinense, Laides longibarbis*	2
Ambassidae	*Parambassis siamensis*	1
Anabantidae	*Anabas testudineus*	1
Bagridae	*Hemibagrus wyckii*	1
Balitoridae	*Homalopteroides smithi*	1
Belonidae	*Xenentodon cancila*	1
Botiidae	*Syncrossus helodes, Yasuhikotakia caudipunctata, Yasuhikotakia eos, Yasuhikotakia lecontei, Yasuhikotakia splendida*	5
Channidae	*Channa micropeltes, Channa striata*	2
Cichlidae	*Oreochromis niloticus*	1
Clariidae	*Clarias macrocephalus*	1
Clupeidae	*Clupeichthys aesarnensis, Clupeichthys perakensis, Tenualosa thibaudeaui*	3
Cobitidae	*Acantopsis ioa, Pangio anguillaris*	2
Cyprinidae	*Barbonymus altus, Barbonymus gonionotus, Barbonymus schwanefeldii, Cirrhinus microlepis, Cosmochilus harmandi, Cyclocheilichthys armatus, Cyclocheilichthys enoplos, Cyclocheilichthys repasson, Garra fasciacauda, Hampala dispar, Hampala macrolepidota, Henicorhynchus siamensis, Hypsibarbus malcolmi, Labeo chrysophekadion, Labiobarbus leptocheilus, Labiobarbus lineatus, Mekongina erythrospila, Mystacoleucus ectypus, Mystacoleucus marginatus, Osteochilus melanopleurus, Poropuntius normani, Probarbus jullieni, Puntioplites falcifer, Scaphognathops bandanensis, Sikukia gudgeri, Systomus orphoides*	26
Danionidae	*Opsarius koratensis, Raiamas guttatus*	2
Eleotridae	*Oxyeleotris marmorata*	1
Gobiidae	*Gobiopterus lacustris, Papuligobius ocellatus*	2
Mastacembelidae	*Macrognathus siamensis*	1
Notopteridae	*Chitala ornata*	1
Pangasiidae	*Pangasius bocourti, Pangasius macronema, Pangasius nasutus, Pseudolais pleurotaenia*	4
Pristolepididae	*Pristolepis fasciata*	1
Siluridae	*Ompok siluroides, Wallago attu*	2
Sisoridae	*Bagarius yarrelli, Glyptothorax trilineatus*	2
Toxotidae	*Toxotes chatareus*	1
Xenocyprididae	*Paralaubuca typus*	1
Zenarchopteridae	*Dermogenys pusilla*	1
	Total	**66**

Table 4. The non-native species detected in the collected sequence data.

Family	Detected Species	Sites Where Detected
Cichlidae	*Oreochromis* sp.	Sivilay, Namuang, Kaengsadok, Kudjub, Thapea, Heunhin, Poung-Nua, Boungkuang, Pak Ka Lang
	Oreochromis niloticus	Namuang, Kaengsadok, Kudjub
Cyrpinidae	*Cyprinus* sp.	Sivilay, Namuang, Heunhin, Pha Chan, Kudjub, Kaengsadok
Xenocyprinidae	*Hypophthalmichthys* sp.	Kudjub
	Total Detected Taxa = ≥4*	**Total Sites = 10**

*As multiple sequences belonging to non-native taxa could only be assigned to the genus level, the total number of invasive taxa detected may be greater than the four taxonomic assignments, as some may represent multiple species within the same genus.

The detected taxa represent a broad range of diversity across all taxonomic levels from order to species. In general, detected species are typical of those observed in previous studies of Mekong fish diversity using the MiFish primers and single-use sample collection kits (e.g., Eschenroeder et al. 2024). However, some of the detections from this sampling effort are particularly notable. Among these are genus level detections of *Schistura* that occurred in four of the sampled locations. *Schistura* is a genus of small loaches that contains numerous species, and species in this genus alone represent an astonishing 5% of Mekong fish diversity. In previous studies detections of *Schistura* have been rare (Eschenroeder et al. 2024). In fact, no sequences assigned to this genus were detected in the first round of eDNA sampling conducted by WWF for monitoring of the LTF Project in 2022. *Schistura* are a good example of a species that may be easier to detect with eDNA than they are with conventional means (e.g., nets), as their small body size and benthic nature would make them difficult to capture. However, until more reference sequences are made available, it will not be feasible to generate species-level identification of sequences from members of this genus.

There were also some taxa whose absence from the collected sequences was notable. These include giant freshwater whipray (*Urogymnus polylepis*) and Mekong freshwater stingray (*Hemitrygon laosensis*). Whereas the initial sampling in 2022 detected sequences assigned to the family Dasyatidae, which includes these ray species, no identified sequences from the 2024 sampling effort could be attributed to rays. This likely reflects the lack of complete reference sequences for these species for the region of the mitochondrial genome targeted by the MiFish primers, which would preclude assignment of sequences belonging to them.

Priority Species

Of the 12 priority target species, a total of four were definitively identified with species-level sequence assignments (Table 5). DNA sequences belonging two species listed as Vulnerable on the IUCN Red List – goonch (*Bagarius yarrelli*) and lesser tapah (*Wallago attu*) – were detected in Poung-Nua and Kudjub Village, respectively (Table 2). Sequences assigned to the Critically Endangered Jullien's golden carp (*Probarbus jullieni*) were detected in samples from Kudjub, Thapea, and Heunhin villages (Table 2). Silver barb (*Barbonymous gonionotus*), an important food fish, was broadly detected in 11 of the 14 sampled villages (all but Boungkuang, Tha Long, and Kan Tha Kaewn; Table 2).

Detection of sequences assigned to the genus *Hemibagrus* in Kan Tha Kaewn, Pha Chan, Kudjub, Thapea, Heuhin, and Boungkuang villages may have potentially derived from the LTF indicator species Asian redtail catfish (*Hemibagrus wyckioides*). Notably, the only sequence within this genus that was identified to the species level belonged to crystal-eyed catfish (*Hemibagrus wyckii*), and was detected in Koum and Thapea villages. Greater uncertainty surrounds potential detections of Mekong giant catfish (*Pangasianodon gigas*), dog-eating catfish (*Pangasius sanitwongsei*), *Hemisilurus mekongensis*, thick-lipped barb (*Probarbus labeamajor*), and two headed carp (*Bangana behri*). A total of 31 unique sequences that could only be assigned to the family Cyprinidae may have included DNA derived from two headed carp and/or thick-lipped barb, three unique sequences that could only be assigned to family Pangasiidae may have included DNA

derived from Mekong giant catfish and/or dog-eating catfish, and two distinct sequences that could only be assigned to the order Siluriformes may have derived from *Hemisilurus mekongensis*.

Table 5. ASAP and WWF target species. "Unknown" values in the final column indicate that sequences assigned to higher taxonomic levels (family Pangasiidae, family Cyprinidae, and order Siluriformes) may have derived from the target species.

Scientific Name	English Name	Reason for Inclusion	IUCN Status	DNA found in the samples
Pangasianodon gigas	Mekong giant Catfish	ASAP species	CR	Unknown
Pangasius sanitwongsei	Dog-eating catfish	ASAP species	CR	Unknown
Probarbus jullieni	Jullien's golden barb	ASAP species	CR	**Yes**
Hemitrygon laosensis[1]	Mekong stingray	IUCN Redlist	EN	No
Urogymnus polylepis[2]	Giant freshwater whipray	IUCN Redlist	EN	No
Wallago attu	Lesser Tapah	IUCN Redlist	VU	**Yes**
Barbonymus gonionotus	Silver barb	ASAP species	LC	**Yes**
Hemibagrus wyckioides	Asian redtail catfish	LTF Indicator Species	LC	Unknown
Bagarius yarrelli[3]	Goonch	ASAP species	VU	**Yes**
Bangana behri	Two headed carp	ASAP species	VU	Unknown
Probarbus labeamajor[1]	Thick-lipped barb	IUCN Redlist	EN	Unknown
Hemisilurus mekongensis[2]	N/A	LTF indicator species	LC	Unknown

[1]There are no reference sequences for these species on GenBank.
[2]There are only partial reference sequences for these species on GenBank, and they do not cover the 12S region targeted by MiFish.
[3]The taxonomy of the genus has been restructured in recent years, and it is currently believed that three species of *Bagarius* occur in the Mekong Basin (*B. vegrandis*, *B. suchus*, and *B. lica*; Ng and Kottelat, 2021). It is likely that the detected *Bagarius* sequences derived from *B. lica*, but they are reported here as *B. yarrelli* as that is the nomenclature used on the ASAP species list and in the metabarcoding results.

Some of the uncertainty associated with species detection arises from gaps in existing reference databases. Notably, the publicly available genetic reference library used for this project (GenBank) does not include any reference sequences for Mekong stingray (*Hemitrygon laosensis*) nor thick-lipped barb (*Probarbus labeamajor*). In addition, the database does not include any sequences of the 12S region of the mitochondrial genome (the portion targeted by the MiFish primers) for giant freshwater whipray (*Urogymnus polylepis*) nor *Hemisilurus mekongensis*. Should these reference sequences be made available in the future, sequence data archived from this project may be reanalyzed to determine whether sequences belonging to these and other species of interest may be identified. In addition to missing reference sequences, the MiFish primers are known to provide poor taxonomic resolution for certain genera, including *Pangasius* (Jerde et al. 2021), which may have contributed to a lack of species-level taxonomic assignments.

In addition to the 12 priority species for the LTF project (Table 5), DNA sequence data generated by this sampling effort included potential detections of additional ASAP species (Patricio et al. 2023). These include the genus level detection of *Schistura* in four of the sampled villages (Table 2), which may have derived from the ASAP species *Schistura tenura*, a type of loach that is found in Lao PDR. However, this is exceedingly unlikely, as *S. tenura* is only recorded from the Nam Leuk catchment in Laos, which was not sampled by this project. Notably, no publicly available reference sequences currently exist for this species. Another, far more likely ASAP species detection is *Datnioides pulcher*, as sequences assigned to genus *Datnioides* were detected in nine of the sampled villages (Table 2). This species is known to occur in large river channels in the Mekong drainage, and may very well have been present in the mainstem sites sampled for this project. However, detected *Datnioides* sequences may have also derived from the more common *D. undecimradiatus*, and therefore it is not possible to definitively determine whether *D. pulcher* was truly present in the sampled locations.

Comparison with Fisher Catch Reporting Data

Data from fisher catch reporting associated with Phase III of the LTF Project was provided for comparison with the eDNA detection data. The catch reporting data was collected from January through March in five villages: Ban Khamtue, Khanthoungxay, Thakhanxoumxoua, Donmmarkkeua, and Paksamon. Notably, this data covers a larger time period that preceded the eDNA sampling (which took place in April), and covers different villages than those included in the eDNA sampling. Further, it has not been reviewed by local fishers and should be considered preliminary. Despite these limitations, the catch records provide a useful means for basic comparison of the taxa detected by the two sampling approaches.

Because the fisher catch data includes only genus and species level taxonomic identities, we restricted our comparison to just include the genus and species level assignments included in the eDNA data. In plotting the total taxa detected by sampling method and color-coding by family, it becomes clear that eDNA detected a greater number of taxa, which represent a greater amount of diversity at the family, genus, and species levels (Figure 8).

More specific examination of the two datasets revealed that 72 taxa were unique to the eDNA data, and 40 taxa were unique to the fisher catch reporting data. However, variation in the taxonomic level of identification (i.e., genus or species) mean that some of the species level assignments in each dataset fall within genus level assignments in the other dataset. Repeating this comparison with that information reveals that nine species unique to the fisher catch reporting data actually fall under genus-level assignments from the eDNA data, and 12 species from the eDNA data fall under genus-level assignments from the fisher catch reporting data.

Considering this information, a total of 60 of the 95 (~63%) genus and species level taxonomic detections in the eDNA data did not occur in the fisher catch reporting data, and 31 of the 63 (~49%) genus and species level taxonomic detections in the fisher catch reporting data did not occur in the eDNA data. Taken together, these findings suggest that fisher catch reporting and eDNA monitoring capture different components of the fish community, and therefore may be considered complimentary to one another.

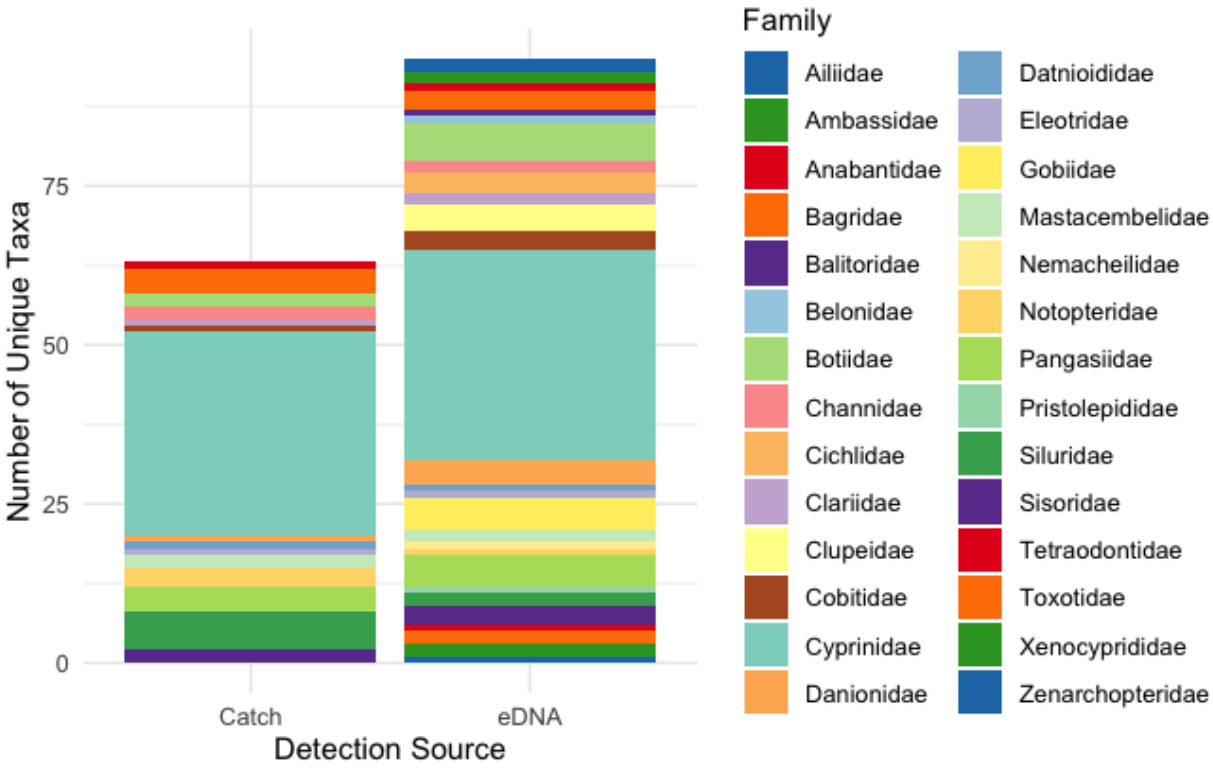


Figure 8. Total detections of unique taxa generated by fisher catch reporting ("Catch") versus eDNA metabarcoding ("eDNA"). Note that eDNA detections only include sequences assigned to the genus or species level, and family and order level assignments are omitted.

Discussion

This study represents a valuable continuation of eDNA-based monitoring in the middle Mekong River, and provides both a useful update on the fish communities inhabiting FCZs established by the LTF Project, as well as insight into variation in the resolution of data provided by varying field sampling and laboratory analysis procedures. Despite sampling fewer sites and a much smaller volume of water, the use of single-use sample collection kits paired with greater sample replication and analysis with MiFish primers resulted in the detection of a slightly greater number of fish taxa in this year as compared to the study conducted in 2022 (106 compared with 93), and a greater number of fish sequences that could be identified to the species level (66 compared with 50). Further, invasive species known to occur in the region that were not detected in the 2022 analysis (*Cyprinus* sp. and *Hypopothalmichthys* sp.) were detected by this project, although an additional invasive species detected in 2022 – African sharptooth catfish (*Clarias gariepinus*) – was not detected by this project. Caution is warranted in drawing direct comparisons between these two datasets given variation in sample replication, sample collection methodology, and taxonomic resolution of primers, but it can be confidently stated that the filtering of considerably smaller volumes of water using four replicates of single-use sample collection kits does not result in

diminished detection of fish diversity, but rather appears to provide equivalent or even greater detection ability.

In general, the identification of similar patterns between the 2022 and 2024 data suggests that both approaches are able to detect key patterns in fish diversity. For example, one of the study sites that was found to contain the highest diversity in 2022 (Thapae Village; n = 56 taxa), was also found to contain the most fish taxa in this study (n = 41 taxa), a trend that is likely a result of this village's proximity to the confluence of the Mekong with the Xe Bang Hieng.

As with all fish diversity monitoring studies, when using eDNA there is a tradeoff between the level of effort and the amount of data obtained. Although sample coverage was limited to 14 discrete locations across many hundreds of river kilometers, the assessment of taxa accumulation across the 56 in-river samples suggests that the level of eDNA sampling applied in this study came very close to maximizing the detection of diversity. Very few additional fish taxa were being detected with each additional sample beyond approximately 45-50 samples. This would indicate that the number of replicates was appropriate, and increasing the number of samples at each site would generate few additional detections.

Further, improvements in available genetic reference libraries would lead to capturing an even more complete picture of the fish community without increasing the intensity of sampling, as availability of reference sequences for more species in the region may allow many of the unknown sequences detected to be assigned to specific taxa and used in the analysis. Based on last sighting surveys (Appendix 3), certain target species that were not detected in the eDNA samples are known to occur in the assessed villages. Although direct comparison of the last sighting data (which covers a different and longer time period) and the eDNA data (which covers a single point in time) is not prudent, the detection of sequences assigned to the genus level may very well represent some of the target species given their apparent commonality based on last sighting surveys, particularly *Hemibagrus* and *Hemisilurus* (Appendix 3). Improved resolution of species within these genera may be achieved through reference library build-out, although this may not be an effective approach for certain genera like *Pangasius* that are known to have limited differentiation in the region targeted by the MiFish primers (Jerde et al. 2021). Similarly, the MiFish primers are more effective at targeting bony fish species and may be less effective at detecting stingrays (Miya et al. 2015), which may explain the lack of stingray detections in the samples despite the apparent wide distribution of *Urogymnus polylepis* and *Hemitrygon laosensis* based on the last sighting surveys (Appendix 3). If projects seek to specifically target certain species in the future, consideration must be given to the availability of reference sequences and the primers that are best suited for distinguishing them, and the approach for detecting as much of the fish community as possible (i.e., MiFish) may not be the best option for detecting a particular species. Importantly, however, any improvements in available libraries may be used to re-analyze the data collected for this project, as well as incorporated into future studies. Further, collected samples could also be analyzed using different primers in the future to target specific types of fish.

Recommendations

Environmental DNA is not a replacement for traditional sampling, as it can provide only data on the presence of species, and is not able to generate information on abundance, size structure, age structure, and other important population demographic factors. However, in the rivers where paired comparisons of traditional and eDNA sampling have been applied, eDNA has been demonstrated to characterize a greater proportion of the total species richness (i.e., the number of species present) in a location compared with traditional sampling, and may be more effective at detecting rare species (Doi et al. 2021; Hallam et al. 2021). Similar studies are underway in the Mekong, and promise to provide greater insight into the complementarity of eDNA and traditional fisheries monitoring in the basin. In addition, the inclusion of fish catch monitoring in the 3rd phase of the LTF Project may serve as an opportunity to conduct a parallel comparative study between traditional and molecular methodologies. Further, the potential value of eDNA sampling in conjunction with occupancy modeling has been demonstrated for monitoring of rare and imperiled species (Martel 2019; Neto et al. 2020; Strickland and Roberts 2019), as well as for detection of invasive species (Erickson et al. 2017; Hunter et al. 2015), and these approaches may be implemented in more targeted studies in the future.

Notably, of the 12 target species included in this study, only seven have available reference sequences for the region of the mitochondrial genome targeted by the MiFish primers. Further, of these seven, species-level identification may not be feasible for some with the MiFish primers (e.g., Pangasius spp.), due to a lack of inter-specific variability in the target region of the genome. To address these limitations, concerted efforts to obtain and publish reference sequences for the entire mitochondrial genomes of species currently missing in public databases are warranted. In addition, analyses to better understand which primers provide the best species-level resolution should be used to inform primer selection based on target species.

Comparison between fisher catch reporting data and eDNA detection data, though somewhat basic, suggests that these two approaches may provide data on different components of the fish community. This is demonstrated through the detection of 31 fish taxa only through fisher catch reporting and the detection of 60 fish taxa only through eDNA. However, as noted above, the comparison between fisher catch reporting and eDNA should be interpreted with caution due to differences in the timing and location between the two sampling programs. To better understand the relationship between fisher reporting data and eDNA data, focused research to specifically evaluate the overlap between the two methods and how they may best be employed in tandem is warranted. The inclusion of robust fisher catch reporting and, if possible, fisheries-independent sampling as part of the LTF Project may provide the data necessary for these evaluations.

Importantly, eDNA sampling confers additional benefits beyond ease of scalability. For instance, unused extracted DNA and/or unused pieces of collected sample filters may be archived as a molecular time capsule of the species present at the time of sample collection, which will serve as a valuable baseline for comparative analysis with future eDNA data as a means of tracking the impacts of climate change, dam development, and other environmental changes on species distributions. Further, collected samples may be reanalyzed at a later date following improvements in available primers (i.e., the markers used in DNA sequencing) or genetic reference libraries, or

reanalyzed using primers that target other taxonomic groups that may be of interest (such as mammals or reptiles). In short, eDNA samples have a wealth of uses that may be leveraged by conservation organizations, researchers, and fisheries managers seeking diverse data to address a myriad of different research questions.

The use of eDNA for monitoring of fish diversity in the Mekong is likely to continue to rapidly expand, as will understanding of associated best practices and its utility as a data collection tool. The two highest priorities for improving the effectiveness of eDNA in the basin are 1) to address gaps in existing reference libraries, with a particular focus on obtaining sequences of priority species, and 2) to conduct targeted studies to evaluate various eDNA field sampling and analysis methodologies to develop protocols that maximize the detection of fish diversity. In particular, robust, strategically designed experimental comparisons of varying sample collection and analysis methodologies would be of great value for improving understanding of best practices for eDNA use in the basin. Numerous projects seeking to address these two objectives are underway, and the protocols used in future eDNA monitoring of the LTF Project may be modified and improved based on the best available science. Continuing data collection and archiving of both extracted DNA and sequence data will be of great value for monitoring the response of the fish community in the project areas both in response to local management practices and large-scale shifts in the climate and hydrology of the region.

Literature Cited

Blackman, R.C., Osathanunkul, M., Brantschen, J., Di Muri, C., Harper, L.R., Mächler, E., Hänfling, B. and Altermatt , F. 2021. Mapping biodiversity hotspots of fish communities in subtropical streams through environmental DNA. *Scientific Reports*. 11: 1–11.

Bonin, A., Guerrieri, A., and Ficetola, G. F. 2023. Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies. *Molecular Ecology Resources*. 23(2): 368-381.

Doi, H., Inui, R., Matsuoka, S., Akamatsu, Y., Goto, M., and Kono, T. 2021. Estimation of biodiversity metrics by environmental DNA metabarcoding compared with visual and capture surveys of river fish communities. *Freshwater Biology*. 66(7): 1257-1266.

Durand, J.D., Simier, M., Tran N.T., Grudpan, C., Chan B., Nguyen B.N.L., Hoang H.D. and Panfi li, J. 2022. Fish diversity along the Mekong River and Delta inferred by environmental DNA in a period of dam building and downstream salinization. *Diversity*. 14, 634.

Erickson, R. A., Merkes, C. M., Jackson, C. A., Goforth, R. R., and Amberg, J. J. 2017. Seasonal trends in eDNA detection and occupancy of bigheaded carps. *Journal of Great Lakes Research*. 43(4): 762-770.

Eschenroeder, J.C., Pilger, T.J., Chea, S., and Hogan, Z. 2024. What's in the water: using environmental DNA metabarcoding to detect fish biodiversity in the Cambodian Mekong. *Cambodian Journal of Natural History*. 2024(1): 2-11.

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., ... and Pfrender, M. E. 2017. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*. 74(9): 1362-1374.

Hallam, J., Clare, E. L., Jones, J. I., and Day, J. J. 2021. Biodiversity assessment across a dynamic riverine system: A comparison of eDNA metabarcoding versus traditional fish surveying methods. *Environmental DNA*. 3(6): 1247-1266.

Herder, J.E., Valentini, A., Bellemain, E., Dejean, T., van Delft, J.J.C.W. and Thomsen, P.F. 2014. Environmental DNA – A Review of the Possible Applications for the Detection of (Invasive) Species. Netherlands Food and Consumer Product Safety Authority, Nijmegen, Netherlands.

Hunter, M. E., Oyler-McCance, S. J., Dorazio, R. M., Fike, J. A., Smith, B. J., Hunter, C. T., ... and Hart, K. M. 2015. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PloS one*. 10(4): e0121655.

Jerde, C.L., Mahon, A.R., Campbell, T., McElroy, M.E., Pin, K., Childress, J.N., Armstrong, M.N., Zehnpfennig, J.R., Kelson, S.J., Koning, A.A., Ngor, P.B., Nuon, V., So, N., Chandra, S. and Hogan, Z.S. 2021. Are genetic reference libraries sufficient for environmental DNA metabarcoding of Mekong River basin fish? *Water*. 13(13): 1767.

Jeunen, G.J., Knapp, M., Spencer, H.G., Taylor, H.R., Lamare, M.D., Stat, M., Bunce, M. and Gemmell, N.J. 2019. Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. Ecology and Evolution. 9: 1323–1335.

Kumar, G., Farrell, E., Reaume, A.M., Eble, J.A. and Gaither, M.R. 2021. One size does not fit all: tuning eDNA protocols for high-and low-turbidity water sampling. Environmental DNA. 4: 167–180.

Martel, C. M. 2019. Using environmental DNA and occupancy modeling to estimate rangewide metapopulation dynamics of the endangered tidewater goby eucyclogobius spp. *Cal Poly Humboldt theses and projects*. 323. https://digitalcommons.humboldt.edu/etd/323

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., and Iwasaki, W. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*. 2(7): 150088.

Neto, J. G. D. S., Sutton, W. B., Spear, S. F., Freake, M. J., Kéry, M., and Schmidt, B. R. 2020. Integrating species distribution and occupancy modeling to study hellbender (*Cryptobranchus alleganiensis*) occurrence based on eDNA surveys. *Biological Conservation*. 251: 108787.

Ng, H.H., & Kottelat, M. (2021). Description of Bagarius vegrandis, a new species of sisorid catfish from Indochina (Actinopterygii: Siluriformes), with notes on the identity of Bagarius bagarius. *Zootaxa*, *4926*(1), zootaxa-4926.

Patricio, H., Ng, N., Baltzer, M., Chao, N., Lyons, T., and Raghavan, R. (2023). A strategic framework to accelerate urgent conservation action for ASAP Freshwater Fishes in Southeast Asia. IUCN SSC Asian Species Action Partnership. Singapore.

Polanco, A., Pellissier, E., Flück, B., Valentini, A., Altermatt, F., Brosse, S., Walser, J.C., Eme, D., Marques, V., Manel, S., Albouy, C., Dejean, T., and Pellissier, L. 2021. Comparing the performance of 12S mitochondrial primers for fish environmental DNA across ecosystems. *Environmental DNA*. 10.1002/edn3.232

Strickland, G. J., and Roberts, J. H. 2019. Utility of eDNA and occupancy models for monitoring an endangered fish across diverse riverine habitats. *Hydrobiologia*. 826(1): 129-144.

Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. 2018. *Environmental DNA: for biodiversity Research and Monitoring*. Oxford university Press, 253 pp.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Brun, M.L., Møller, R.R., Willerslev, E., and Dejean, T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology.* 25: 929-942.

Supplementary Materials

Appendix 1 – Field Sampling Protocol

All environmental DNA (eDNA) samples for this project will be collected using single use kits provided by Jonah Ventures. These kits each contain a 60mL syringe, a small syringe of Longmire's solution for sample preservation, a pair of gloves, a filter disc, and two caps for the filter disc. At each site, the following process will be followed:

1) Open the kit, being careful to keep the contents inside and not allow them to come into contact with potential sources of contamination.
2) Put on the gloves contained in the kit.
3) Take out the 60mL syringe from its packaging.
4) Place the 60mL syringe just below the surface of the river (~5cm deep) and draw up a full 60mL of water.
5) Remove the filter disc from its packaging and lock it onto the end of the 60mL syringe by twisting.
6) Push the water in the syringe through the filter disc (water will drip out the other side).
7) Once all the water is pushed through, remove the filter disc by untwisting, place the syringe back in the river and draw up another 60mL.
8) Reattach the filter disc by twisting it on and push the water through the filter once again.
9) Repeat this process until it becomes too difficult to push water through the filter (i.e., it becomes clogged with sediment); be sure to keep track of the volume of water you have pushed through the filter so it can be recorded.
10) Once no more water can be pushed through the filter, remove it from the syringe, and attach the small syringe of preservative solution by twisting it onto the end.
11) Push the preservative solution through the filter disc.
12) Attach the caps to both sides of the filter disc, ensuring they are screwed on tightly.
13) Place the filter disc in the small pouch labelled with the kit code.
14) Verify all data have been recorded on the field data sheet, and safely store the collected filter out of direct sunlight.

To complete this process, it will be best to have one person holding the eDNA collection kit pouch, one person (the one wearing gloves) collecting the sample, and one person recording the data. When collecting from a boat, do your best to have the boat driver hold the boat in one place facing upstream while you collect the sample. Always collect from upstream of the boat to reduce the risk of contamination. Once one sample is collected, the boat can move across the river in a straight line and collect the next sample. A total of four samples will be collected in the river in each sample site (Figure 1). These should be evenly spaced across the width of the river. The samples along the margin of the river should be collected approximately 2 meters from shore. However, if shore access isn't possible in a sample site then they may be collected further away, just be sure to record this on the datasheet.

A fifth sample will be collected by filtering bottled water in the same location. This fifth sample will serve as a negative control, and will allow us to detect any contamination that may have occurred at the sample site. This is because we expect no DNA to be detected in bottled water, and if DNA is detected we can determine that there was some source of contamination in the field.

To reduce the risk of contamination, be careful to ensure that only the person wearing gloves is touching the equipment inside the sample kit pouch, that the filter disc caps are securely attached, and that the filter is securely stored in the small pouch once it is collected.

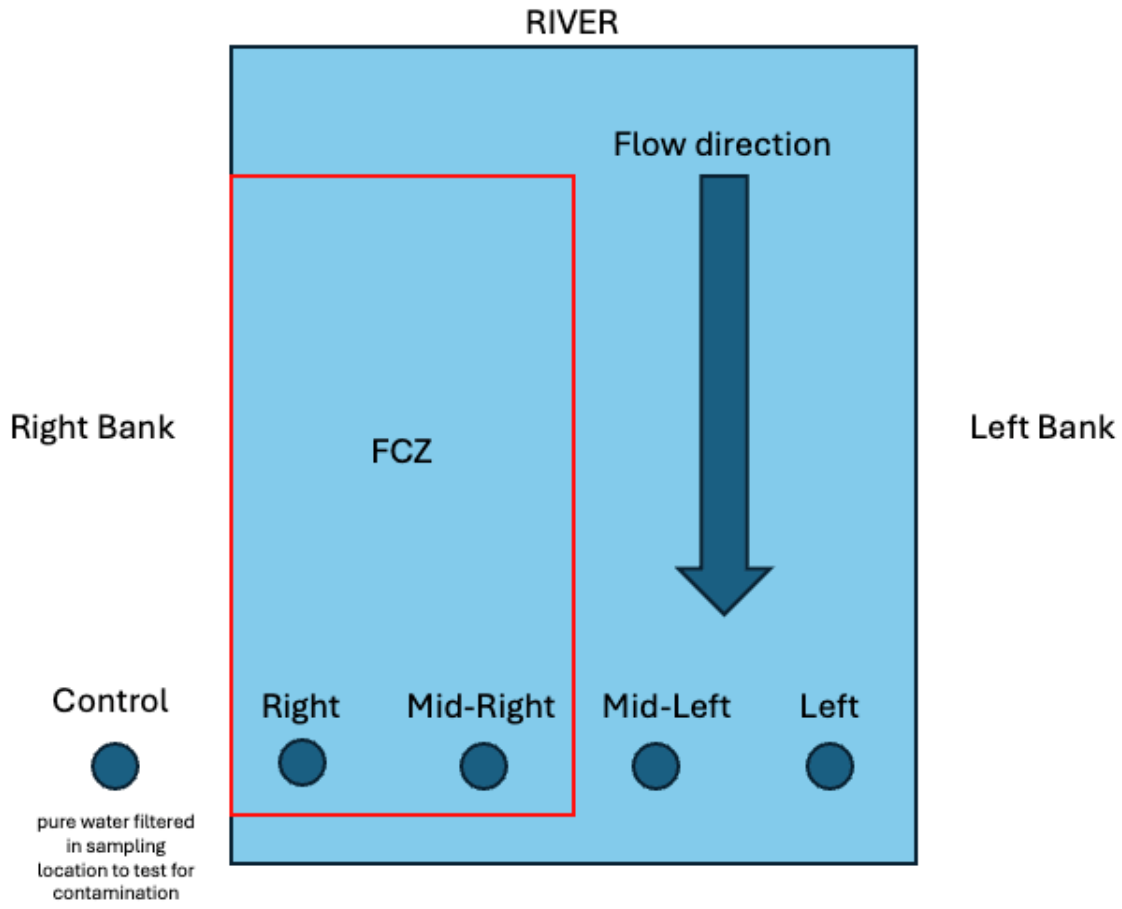


Figure 1. The sampling approach that will be used in each location.

Data Recording and Sampling Framework

Data must be carefully recorded for each sample to ensure that the resulting DNA sequences can be connected back to specific sample sites. Using the provided field data sheet (Figure 2), the field team will record the following:

1) The date and time that the sample was collected
2) The name of the site (can be a number, site code, or village name)
3) The sample kit code (this is the code that is printed on the eDNA sample collection kit)
4) The total volume of water that was pushed through the filter (in mL)
5) The GPS coordinates (latitude and longitude) where the sample was collected
6) The sample location in the river (right bank, right middle, left middle, left bank, or control)
7) Any notes on potential DNA sources in the sample area (e.g., net pens, restaurants, observed species)

A total of 70 samples will be collected across 14 sites (Table 1). These will include the four samples collected across the width of the river, and the single negative control collected in each site. Additional kits will be provided in case a mistake is made or a kit is defective. If there is any concern that a kit may be defective or that a sample was collected incorrectly, the field crew may start over and use a new kit. However, take care to follow the procedure closely to ensure that kits are not wasted.

Collected samples must be meticulously labeled and stored in a cool, dark place in order to reduce the risk of contamination, and sample data recorded on the field data sheets will be quickly entered into a shared electronic database. All samples and associated data sheets should be safely stored, and should be photographed to provide a backup record in case the sheet is lost. Samples will be transported back to Vientiane, and from there will be sent back to the US for analysis. It is critical that the sample kit codes are accurately recorded, as these will be essential for connecting the resulting DNA data to the specific sample location.

eDNA Sampling Field Datasheet **Field Crew:**_____

Date	Time	Site	Sample Kit Code *(code printed on the sample)*	Volume Filtered (mL)	Latitude	Longitude	Sample Collection Location *(circle one)*	Notes *(animals observed, potential sources of DNA like net pens or restaurants, etc.)*
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	
							Right / MidRight / MidLeft / Left / Control	

Figure 2. Proposed eDNA sample collection datasheet.

Table 2. The proposed eDNA sampling framework.

Village	District	Province	Country	Samples to be Collected
Koum	Khong Chiam	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Ta Mui	Khong Chiam	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Tha Long	Khong Chiam	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Kan Tha Kaewn	Khong Chiam	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Pha Chan	Pho Sai	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Pak Ka Lang	Pho Sai	Ubon Ratchathani	Thailand	4 in river, 1 negative control
Thaphea	Songkhone	Savannakhet	Laos	4 in river, 1 negative control
Heunhin	Xayphouthong	Savannakhet	Laos	4 in river, 1 negative control
Sivlilay	Xaybouly	Savannakhet	Laos	4 in river, 1 negative control
Kudjub	Nongbok	Khammoaun	Laos	4 in river, 1 negative control
Namuang	Thakaek	Khammoaun	Laos	4 in river, 1 negative control
Poung-Nua	Hinboun	Khammoaun	Laos	4 in river, 1 negative control
Boungkuang	Pakkading	Bolixhamxay	Laos	4 in river, 1 negative control
Kaengsadok	Paksan	Bolixhamxay	Laos	4 in river, 1 negative control
			TOTAL	**70**

Appendix 2 – Detailed Molecular Methodology

Sample Process

1.4.5

Sample barcodes were recorded and assigned a corresponding lysate tube. Sample filters, lysis buffer, and proteinase K were heated to 56 C for one hour. Under a laminar flow hood, warm lysis buffers were pushed through the filter housing, and all supernatant was collected in the corresponding lysate tube. Tubes were placed in an incubator overnight at 56 C. After incubation the lysate tubes were immediately processed.

Extraction

2.6.1

Genomic DNA from samples was extracted using the Omega Biotek Mag-Bind Blood & Tissue DNA HDQ 96 Kit (4x96 Preps) (Cat. No. / ID: M6399-01) according to the manufacturer's protocol. Whole (25mm or 47mm) filters were used for genomic DNA extraction. The extraction protocol was automated and completed using a Hamilton Microlab Starlet. Genomic DNA was eluted into 100 µl and frozen at -20 C.

PCR

3.18.3

Forward Primer: GTCGGTAAAACTCGTGCCAGC

Reverse Primer: CATAGTGGGGTATCTAATCCCAGTTTG

Primer reference: Miya et al 2015

Portions of hyper-variable regions of the mitochondrial 12S ribosomal RNA (rRNA) gene were PCR amplified from each genomic DNA sample using the MiFishUF and MiFishUR primers with spacer regions. Both forward and reverse primers also contained a 5' adaptor sequence to allow for subsequent indexing and Illumina sequencing. PCR amplification was performed in replicates of six and all six replicates were not pooled and kept separate. Each 25 µL PCR reaction was mixed according to the Promega PCR Master Mix specifications (Promega catalog # M5133, Madison, WI) which included 12.5ul Master Mix, 0.5 µM of each primer, 1.0 µl of gDNA, and 10.5 µl DNase/RNase-free H2O. DNA was PCR amplified using the following conditions: initial denaturation at 95C for 3 minutes, followed by 45 cycles of 20 seconds at 98C, 30 seconds at 60C, and 30 seconds at 72C, and a final elongation at 72C for 10 minutes. Added 11/2019.

Gel

4.1.1

To determine amplicon size and PCR efficiency, each reaction was visually inspected using a 2% agarose gel with 5µl of each sample as input.

PCR Amplicon Cleanup

5.1.1

Amplicons were then cleaned by incubating amplicons with Exo1/SAP for 30 minutes at 37C following by inactivation at 95C for 5 minutes and stored at -20C.

Barcoding PCR

6.2.1

A second round of PCR was performed to complete the sequencing library construct, appending the final Illumina sequencing adapters and integrating sample-specific, dual index sequences (2 x 10bp). The indexing PCR included Promega Master mix, 0.5 µM of each primer and 2 µl of template DNA (cleaned amplicon from the first PCR reaction) and consisted of an initial denaturation of 95 °C for 3 minutes followed by 8 cycles of 95 °C for 30 sec, 55 °C for 30 seconds and 72 °C for 30 seconds.

PCR Normal Pool

8.2.1

Final indexed amplicons from each sample were cleaned and normalized using mag-bind normalization. A 15µl aliquot of PCR amplicon was purified and normalized using Cytiva SpeedBead magnetic carboxylate modified particles (#45152105050250). Samples were then pooled together by adding 5µl of each normalized sample to the pool.

Sequencing

9.7.1

Sample library pools were sent for sequencing on an Illumina MiSeq (San Diego, CA) at the Texas A&M Agrilife Genomics and Bioinformatics Sequencing Core facility using the v2 500-cycle kit (cat# MS-102-2003). Necessary quality control measures were performed at the sequencing center prior to sequencing.

Bioinformatics

10.11.2

Raw sequence data were demultiplexed using pheniqs v2.1.0 [1], enforcing strict matching of sample barcode indices (i.e, no errors). Cutadapt v3.4 [2] was then used remove gene primers from the forward and reverse reads, discarding any read pairs where one or both primers (including a 6

bp, fully degenerate prefix) were not found at the expected location (5') with an error rate < 0.15. Read pairs were then merged using vsearch v2.15.2 [3], discarding resulting sequences with a length of < 130 bp, > 210 bp, or with a maximum expected error rate [4] > 0.5 bp. For each sample, reads were then clustered using the unoise3 denoising algorithm [5] as implemented in vsearch, using an alpha value of 5 and discarding unique raw sequences observed less than 8 times. Counts of the resulting exact sequence variants (ESVs) were then compiled and putative chimeras were removed using the uchime3 algorithm, as implemented in vsearch. For each final ESV, a consensus taxonomy was assigned using a custom best-hits algorithm and a reference database consisting of publicly available sequences (GenBank [6]) as well as Jonah Ventures voucher sequences records. Reference database searching used an exhaustive semi-global pairwise alignment with vsearch, and match quality was quantified using a custom, query-centric approach, where the % match ignores terminal gaps in the target sequence, but not the query sequence. The consensus taxonomy was then generated using either all 100% matching reference sequences or all reference sequences within 1% of the top match, accepting the reference taxonomy for any taxonomic level with > 90% agreement across the top hits.

References:

1. Galanti, Shasha and Gunsalus (2021). Pheniqs 2.0: accurate, high-performance Bayesian decoding and confidence estimation for combinatorial barcode indexing. BMC Bioinformatics 22.

2. Martin (2011), Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17.

3. Torbjørn et al. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4 e2584 (2016).

4. Edgar and Flyvbjerg (2015), Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31.

5. Edgar (2016), UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. doi: https://doi.org/10.1101/081257.

6. Benson et al. (2005), GenBank. Nucleic Acids Res 33.

Appendix 3 – 2023 Last Sighting Survey Data

Scientific Name	IUCN Status	Naphan 1	Thapho 2	Khanthachan 3	Vernkhoun 4	Laomarkhoud 5	Heunhin 6	Savang 7	Keangkabao 8	Syvilay 9	Parksebangfai 10	Nameuong 11	Houaymangli 12	Chomchaeng 13	Kudchup 14	DonkiewTai 15	Bungsantha 16	Hinboun Neua- Hinboun Tai 17	BoungNeua 18	Hinlad 19	Bungkouang 20	Thakathin 21	Xod 22	Kengsadok 23	Paksan 24	KouayOudom 25	Vernthad 26	Thabok-Phapong 27	Thaphonsan 28	Last sighting station report 2023
Pangasianodon gigas	CR				1								1									1			1	1				5
Pangasius sanitwongsei	CR			1	1			1			1	1	1			1			1				1	1		1		1		12
Probarbus jullieni	CR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Pelochelys cantorii	CR	1	1	1	1		1	1	1	1									1	1								1		11
Probarbus labeamajor	EN		1	1			1	1			1					1	1		1	1		1	1			1				12
Hemitrygon laosensis	EN		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1		25
Urogymnus polylepis	EN	1		1	1			1		1	1		1	1						1		1			1					11
Chitala lopis	EN		1		1	1	1		1							1			1	1	1	1	1	1			1			13
Wallago attu	VU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Incisilabeo behri	VU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1				24
Datnioides undecimradiatus	VU	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		27
Wallagonia leerii	LC	1			1	1	1	1	1	1	1					1		1		1	1	1		1	1		1	1		17
Hemibagrus wyckioides	LC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Bagarius lica	LC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Hemisilurus mekongensis	LC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Mekongina erythrospila	NT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1		1	1			23
Boesemania microlepis	DD				1	1	1												1		1		1							6
		11	12	13	16	12	14	14	12	12	13	10	12	10	9	13	10	10	14	13	10	14	13	12	10	12	10	10	5	
Invasive Species																														
Red Tilapia		1	1	1		1	1	1							1					1	1									9
Tilapia		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		27
Indian Carp																														0
Cirrhinus mrigal, Labeo rohita		1					1	1	1	1		1			1	1	1					1			1	1	1	1		14
Cirrhinus mrigal		1		1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1		24
Labeo rohita							1		1																					2
Chinese carp																														0
Bighead carp					1			1			1				1	1	1	1		1	1	1	1					1		12
Grass carp		1	1		1		1	1	1		1				1		1	1	1		1						1			13
Silver carp		1			1		1	1	1		1			1		1	1	1			1	1	1	1		1	1			16
African catfish….				1								1			1	1	1	1	1	1	1	1	1	1		1	1	1		15
Sucker fish								1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1		20
		6	3	4	5	3	7	8	7	4	6	5	2	3	8	6	8	7	5	6	8	7	6	5	4	6	7	6	0	