Environmental DNA Freshwater Biodiversity Assessment in the Mekong between Kratie and Stung Treng Provinces, Cambodia

August 2024

Submitted To:
WWF Cambodia

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 50 environmental DNA (eDNA) samples were collected between 4 June and 8 June, 2024 across 10 sites (six sites in Stung Treng Province and four sites in Kratie Province). 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 focused on evaluating fish diversity the Mekong as part of the World Wide Fund for Nature's effort to build a freshwater biodiversity baseline for the basin by 2025. In addition to evaluating the fish community, sampling was intended to the detect the presence of Asian Species Action Partnership (ASAP), threatened, data deficient, economically and culturally valuable, and invasive species.

Review and cleaning of the sequence data obtained from the samples resulted in the detection of a total of 117 distinct fish taxa representing 12 orders, 30 families, 72 genera, and 80 species. Note that these counts should be regarded as minimums, as taxonomic assignments of some of the detected sequences to the order, family, and genus level may represent multiple species. In particular, 21 sequences could only be assigned to the order level, and 60 could only be assigned to the family level. Additional sequences assigned only to the genus level represented 26 distinct genera. Finally, sequences assigned to the species level represented 80 different species. A total of 38.6% of the detected fish taxa belonged to the family Cyprinidae. The number of fish taxa detected per site ranged from 35 to 68, with a mean of 50. The greatest level of fish diversity was detected in Koh Khnhae (n = 68 taxa), followed by Anlung Cheuteal (66 taxa) and Kampi (55 taxa). A total of seven non-native fish taxa were detected, although three of these were species that are not capable of surviving in the Mekong (i.e., marine and coldwater species) and their DNA may have been introduced from restaurants or fish feed. However, three of these detected taxa (*Hypophthalmichthys* sp., *Cyprinus* sp., and *Labeo catla*) likely reflect true presence in the sites where they were detected.

The use of negative controls at each field site bolster confidence in the accuracy of results, as three of the 10 negative controls detected no sequences, five detected only mammalian sequences (human, pig, and horse), and only two detected sequences assigned to fish (*Channa micropeltes* and *Pangasius nasutus*). Notably *C. micropeltes* was not detected in any of the other samples at the site where the negative control detected it, and *P. nasutus* was detected in only one of four samples collected at the site where the negative control detected it, so there is no indication of broad contamination of samples. Given these findings, all fish sequences were retained for analyses.

Three of the 20 priority species of threatened freshwater fish identified as occurring in Cambodia by WWF – goonch (*Bagarius yarrelli*), Jullien's golden carp (*Probarbus jullieni*), and Laotian shad (*Tenualosa thibaudeaui*) – were definitively identified with species-level sequence assignments. DNA sequences that could only be assigned to the genus level may have included three additional target species – Mekong tiger perch (*Datnioides undecimradiatus*), *Glyptothorax fuscus*, and *Puntioplites bulu*. Though they carry even less certainty, family-level taxonomic assignments may have included sequences belonging to an additional eight of the priority species.

These include sequences assigned to family Cyprinidae that may have derived from bala sharkminnow (*Balantiocheilos melanopterus*), giant barb (*Catlocarpio siamensis*), thicklip barb (*Probarbus labeamajor*), thinlip barb (*P. labeaminor*), and tiger barb (*Puntius partipentazona*), sequences assigned to family Pangasiidae that may have derived from Mekong giant catfish (*Pangasianodon gigas*), and sequences assigned to family Sisoridae that may have derived from dwarf goonch (*Bagarius bagarius*) or crocodile catfish (*B. suchus*). Finally, order-level taxonomic assignments of sequences to the orders *Osteoglossiformes* and *Siluriformes* may have derived from the priority species Asian bonytongue (*Scleropages formosus*) or any of the six target catfish species noted above, respectively. Importantly, only 11 of these priority species have full mitochondrial reference genomes available on GenBank, whereas eight only have partial sequences available and one has no sequences available. The 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.

This project represents a valuable continuation towards establishing a baseline for molecular monitoring of fish diversity in the Cambodian Mekong. Further, modification of the sample collection and analysis protocols between this study and past projects in the region (e.g., Eschenroeder et al. 2024) 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 preparing to collect eDNA samples.

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.

Methodology

Field Methods

Study Design and Sampling Protocol

The design of this study was intended to allow for rapid sampling across a large spatial area, while capturing as much of the fish community in each site as possible. 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 required, 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.

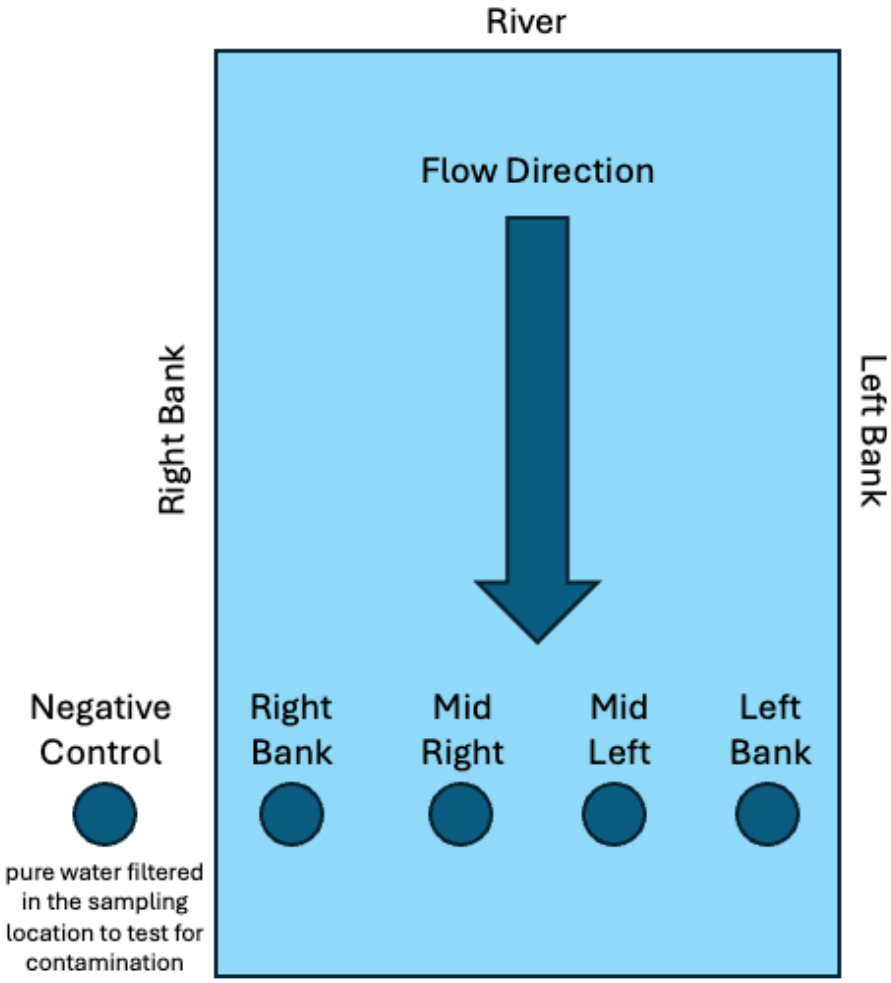

Figure 1. A diagram of the general sampling approach used in each location. Note that spacing among the four in-river points varied depending on width of the channel, but nearshore samples were generally collected ~2 m from shore and mid-channel samples were spaced evenly across the width of the river.

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 two meters from shore where possible. 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.

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 Kratie, 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., *Hypophthalmichthys* species like silver and bighead carp that are known to occur in the study region) or from contamination in the field or laboratory. Detected sequences belonging to non-native species were reviewed and interpreted based on expert opinion, and review of notes on species observations from the field sampling team.

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.

Results

Sample Collection

A total of 50 samples were collected, 30 of which were collected in sites in Stung Treng Province and 20 of which were collected in sites in Kratie Province (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.

Table 1. Locations where eDNA samples were collected. Four replicates and one negative control were collected in each location, resulting in a total of 50 samples, 10 of which were negative controls.

Date	Village	District	Province
7 June 2024	Anlung Cheuteal	Thala Barivat	Stung Treng
7 June 2024	Koh Hip – Ramsar Site	Thala Barivat	Stung Treng
8 June 2024	Stung Treng	Krong Stung Treng	Stung Treng
6 June 2024	Koh Santuk	Siem Bouk	Stung Treng
6 June 2024	Kang Konsat	Siem Bouk	Stung Treng
6 June 2024	Tbaung Khlar	Siem Bouk	Stung Treng
5 June 2024	Koh Khnhae	Sambo	Kratie
5 June 2024	Khsach Makak	Sambo	Kratie
4 June 2024	Koh Pdao	Sambo	Kratie
4 June 2024	Kampi	Chet Borey	Kratie

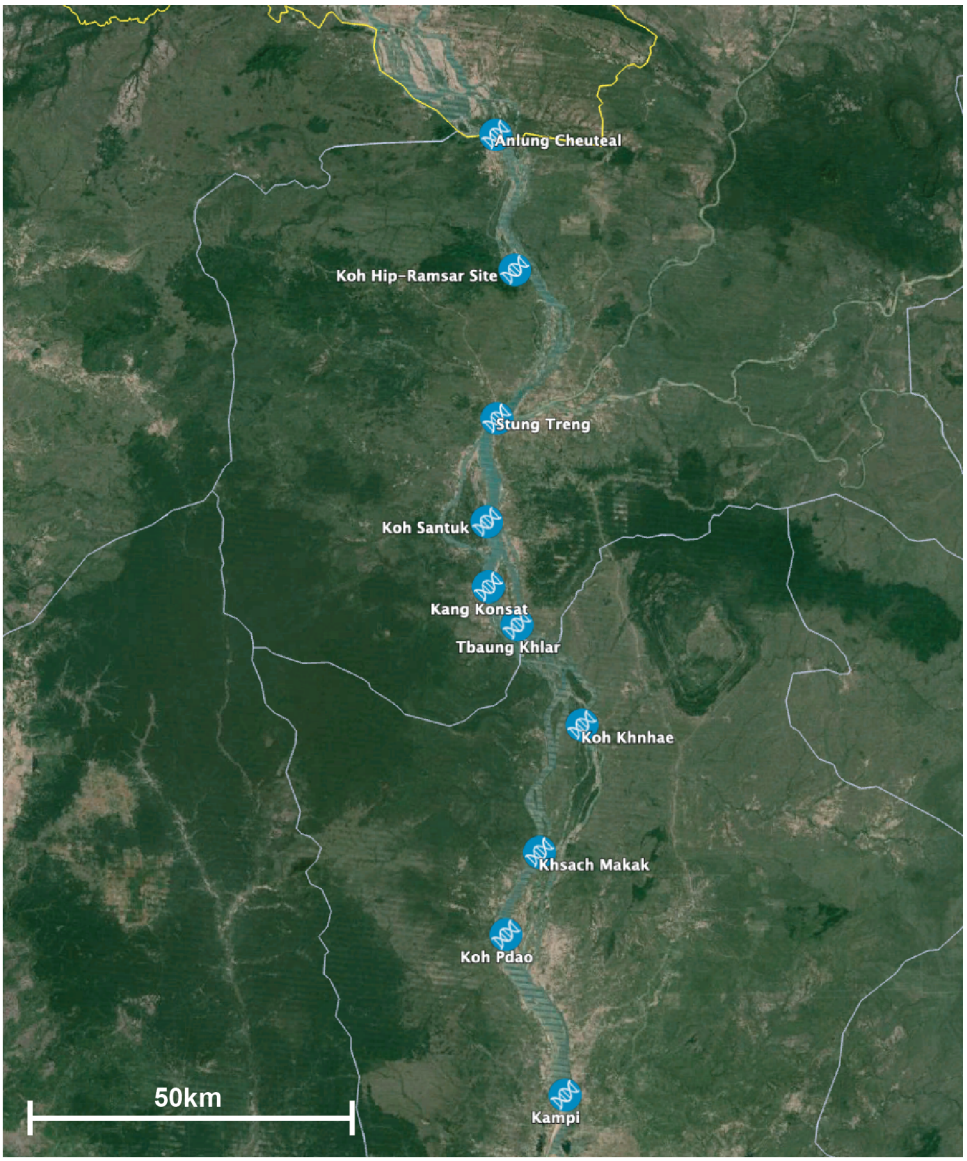

Figure 2. A map of the locations where eDNA samples were collected. See Table 1 for site details.

Analyses of Sequence Data

Sequence Detections and Data Cleaning

A total of 2,069 sequences were detected in the 50 samples collected across 10 sites. All sequence and sampling data have been uploaded to the NCBI SRA database (BioProject ID: PRJNA1140659) and are publicly available. In total, three of the samples yielded no detectable sequences, all of which were negative control samples and therefore were not expected to contain DNA. Of the 2,069 total sequences, 95 belonged to mammalian taxa, including domestic pigs (*Sus scrofa*), domestic cattle (*Bos* spp.), domestic horses (*Equus* spp.), and humans (*Homo sapiens*). No other non-fish taxa were detected, and the remaining 1,974 sequences were all assigned to fish. The filtering out of sequences with <97% match to available genetic reference sequences removed

a total of 342 sequences, leaving 1,612 that were retained for subsequent analyses. Of these, 13 were identified to the class level (Actinopteri), 21 were identified to the order level, 61 to the family level, 262 to the genus level, and 1,255 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 five sequences likely arising from contamination: one belonging to family Cottidae (sculpins), one belonging to white perch (*Morone americana*), two belonging to wahoo (*Acanthocybium solandri*), and one belonging to coho salmon (*Oncorhynchus kisutch*). The sculpin, wahoo, and coho sequences were all detected at Koh Santuk, and the white perch sequence was detected at Koh Hip. Because these species are found in marine environments (wahoo) and cold-water northern environments (sculpins, white perch, and coho salmon), It is not possible living individuals were present in the sampled locations. However, the fact that they were not detected in the negative control samples suggests that their DNA may have been present in the sampled location. This DNA may have arisen from fish feed, bait, or restaurants. Regardless of the source, these sequences were removed from the dataset and not considered in subsequent analyses.

Other sequences assigned to non-native fish taxa included one belonging to genus *Hypophthalmichthys*, six belonging to genus *Cyprinus*, and five belonging to *Labeo catla*. All of these sequences arose from non-native carp species. Presumably the sequences assigned to the genus *Cyprinus* derived from common carp (*Cyprinus carpio*), but the sequencees assigned to *Hypophthalmichthys* may have come from either bighead (*H. nobilis*), silver carp (*H. molitrix*), or both. Because these species are known to have been introduced to the region for aquaculture and have been detected by past eDNA studies, these sequences were retained for subsequent analyses.

Another species level assignment belonged to a species not known to occur in the Mekong, but likely arose from limitations in available reference libraries and/or poor resolution for certain genera. This was the Chiangmai stream goby (*Rhinogobius chiengmaiensis*), which is found only in the Chao Phraya basin in Thailand. However, but this genus is represented by a different species that does occur in the Mekong – *Rhinogobius mekongianus*. Because it could not be definitively determined whether the detections of this species represent novel introductions or simply misidentification of native species (which have been detected in previous surveys in the region, see Bezuijen et al. 2008), it was retained at the genus level for subsequent analyses (i.e., *Rhinogobius* sp.). All other taxonomic IDs were reviewed and deemed to be accurate, leaving a total of 1,604 fish sequences in the final cleaned data set.

Negative Controls

Out of the 10 negative control samples collected (one from each village), the controls from Koh Santuk, Koh Khnhae, and Stung Treng were found to contain no detectable DNA. Of the remaining seven, a total of five – those from Kampi, Koh Hip, Khsach Makak, Koh Pdao, and Kang Konsat – contained only DNA belonging to mammal species, including domestic pigs (*Sus scrofa*), domestic horses (*Equus* sp.) and humans. These mammalian 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.

Only two of the negative control samples – those collected at Anlung Cheuteal and Tbaung Khlar – were found to contain DNA belonging to fish taxa. The Anlung Cheuteal negative control contained two sequences assigned to giant snakehead (*Channa micropeltes*), and the Tbaung Khlar negative control contained one sequence assigned to long nosed pangasius (*Pangasius nasutus*). Giant snakehead was not detected in any of the in-river samples collected at Anlung Cheuteal. In fact, the only other locations where sequences assigned to this species were found were Kampi and Koh Khnhae, both of which are over 120 km away from Anlung Cheuteal. Long nosed pangasius was detected in only one of the in-river samples at Tbaung Khlar, and was also observed in in-river samples from eight other locations. Taken together, this information suggests that there was no widespread contamination of DNA from these species, and therefore their sequences in the in-river samples were retained for analysis. 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 16 to 89, with a mean of 40.1 (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 150 to 660 mL, with a mean of 311.63 mL. Comparison of the volume of water filtered with the total number of sequences detected using a linear model indicated a positive relationship ($p = 0.0026$; Figure 3). However, the model generated a multiple R-squared value of 0.2154, which indicates that sample volume explains only 21.54% of the variability in sequence detections. Therefore, volume alone is not a strong predictor of the total number of detected sequences.

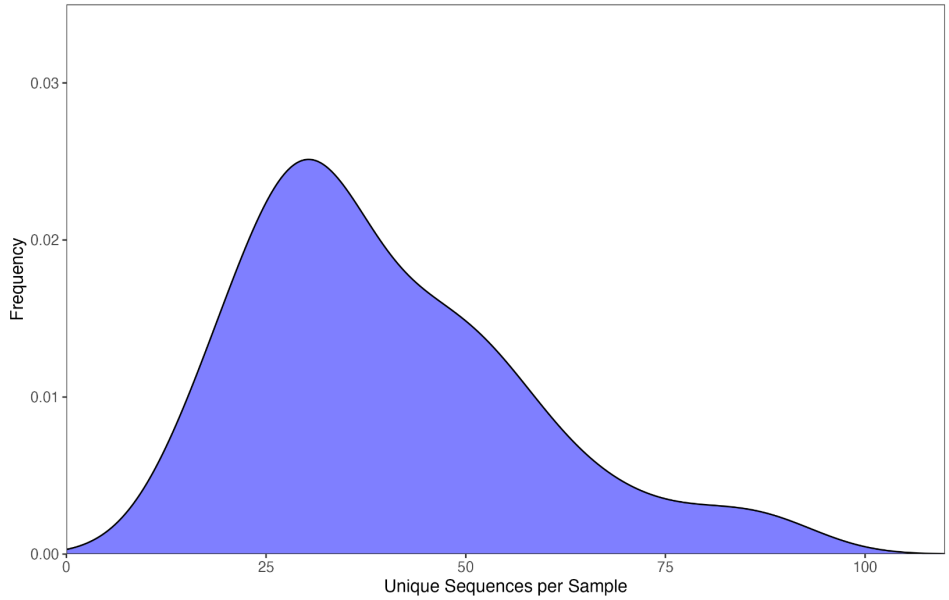

Figure 2. A density curve depicting the number of unique sequences detected per sample. Note that negative controls were excluded from this visualization, and only in-river samples were considered (n = 40).

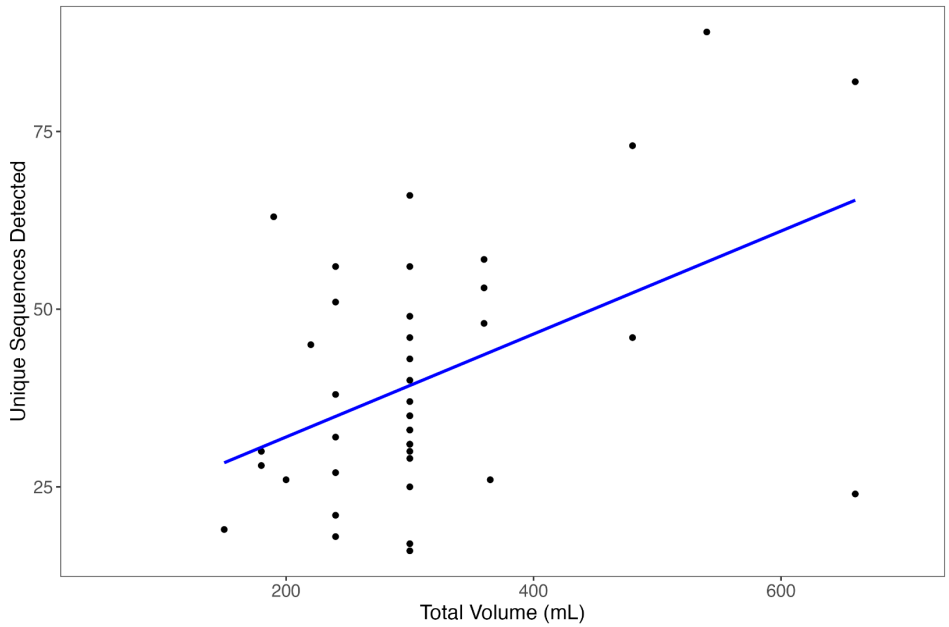

Figure 3. Total sequences detected versus volume of water filtered for each sample, and a line depicting the relationship identified by the linear model. Note that negative controls were excluded from this visualization and model, and only in-river samples were considered (n = 40).

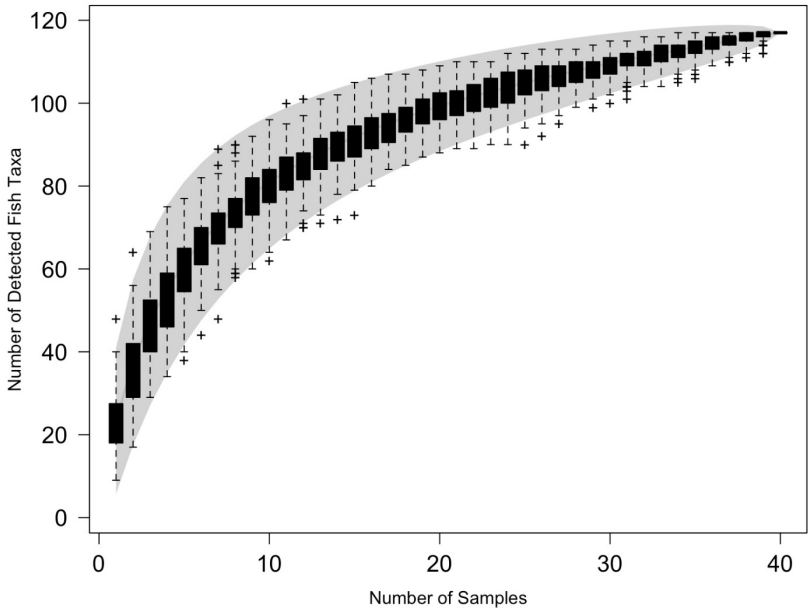

Figure 4. An accumulation curve depicting the number of additional fish taxa detections added with each additional eDNA sample collected. Note that negative controls were excluded from this visualization, and only in-river samples were considered (n = 40).

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 40 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. Notably, the inclusion of four samples at each location appears to have captured significantly more of the fish community diversity than past studies in the same region which used only three samples per site (Eschenroeder et al. 2024), as evidenced by the greater leveling off of the accumulation curve in this study.

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 (Leptobarbidae, Osphronemidae, Polynemidae, and Toxotidae) were only represented by taxa detected in nearshore samples, and one family (Ambassidae) was only represented by taxa detected in mid-channel samples (Figure 5A). In terms of total unique taxa detections, 21 taxa were detected only in nearshore samples, and 12 were detected only in mid-channel samples (Figure 5B).

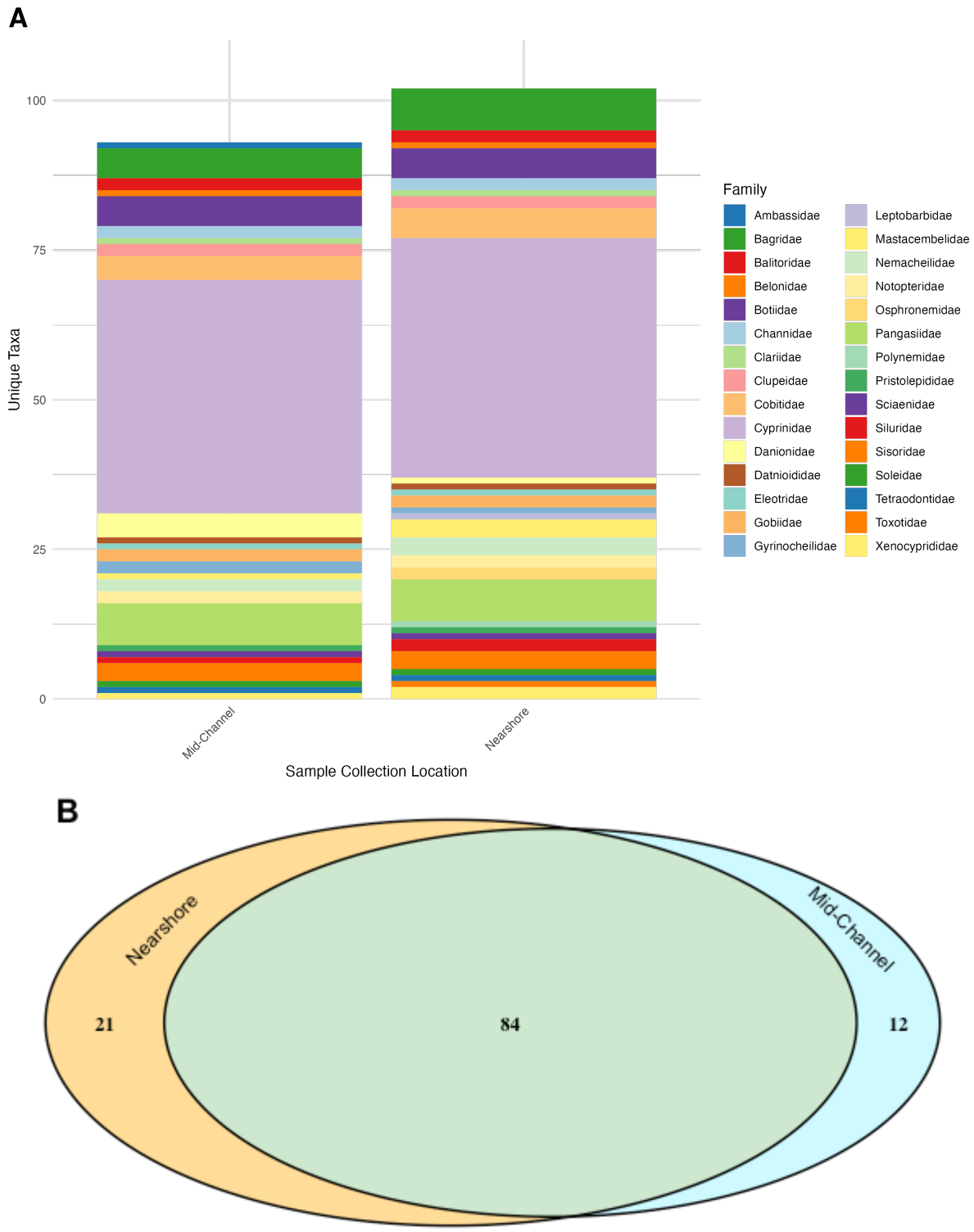

Figure 5. A) Total unique fish taxa detections by sample location (nearshore versus mid-channel), grouped by family. B) Venn diagram showing overlap of unique taxa detected in mid-channel and nearshore samples.

Diversity by Site

The 1,604 sequences in the cleaned dataset contained 724 unique sequence variants, which were assigned to 117 distinct fish taxa representing 12 orders, 30 families, 72 genera, and 80 species. Of the 21 sequences assigned to the order level, 17 belonged to Siluriformes and four belonged to Osteoglossiformes. Among family level assignments, nine belonged to Balitoridae, two to Cobitidae, 32 to Cyprinidae, 12 to Pangasiidae, and five to Sisoridae. The 262 sequences assigned to the genus level represented 26 genera, and the remaining 1,261 sequences assigned to the species level represented 80 species (Table 2; Table 3).

Notably, the total diversity detected was significantly higher than that reported in a past study that included the same region (117 taxa in this study versus 63 taxa in Eschenroeder et al. 2024), and the level of taxonomic resolution appeared to be considerably higher (80 species-level IDs in this study versus 55 species-level IDs in Eschenroeder et al. 2024). Although these studies are not directly comparable to one another given variation in the seasonal timing and locations of sample collection the detection of a larger number of taxa by this effort may reflect the value of a greater total number of samples and increased replication at the site level, a hypothesis that is further supported by the greater leveling off of the accumulation curve (Figure 4). The vast majority of the detected fish taxa belonged to the family Cyprinidae (33.3%; Figure 6). This pattern is similar to that observed in past studies in the region.

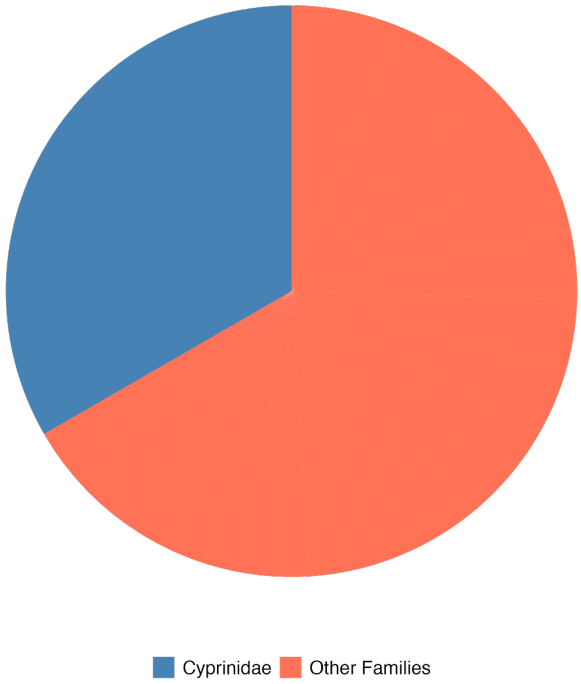

Figure 6. The percentage of taxa within the family Cyprinidae that were detected in the samples.

Table 2. Detected fish taxa by sample site. Note that all order and family level assignments fell within the groups represented by the genera and species level assignments. As such, 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 = 94 sequences) are not explicitly 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	Anlung Cheuteal	Koh Hip Ramsar Site	Stung Treng	Koh Santuk	Kang Konsat	Tbaung Khlar	Koh Khnhae	Khsach Makak	Koh Pdao	Kampi
Ambassidae	*Parambassis* sp.	–	–	–	–	–	X	–	–	–	–
Bagridae	*Hemibagrus* sp.	X	X	X	X	X	–	X	X	–	X
	Hemibagrus spilopterus	X	–	X	–	–	–	X	–	X	X
	Hemibagrus wyckii	X	X	X	–	X	X	X	–	X	X
	Mystus albolineatus	X	X	–	–	–	–	–	–	–	–
	Mystus singaringan	–	X	–	–	–	–	–	–	–	X
	Mystus sp.	–	–	–	–	–	–	–	–	–	X
	Pseudomystus siamensis	X	X	–	X	–	X	X	X	–	X
Balitoridae	*Homalopteroides smithi*	X	–	–	X	–	–	–	–	–	–
Belonidae	*Xenentodon cancila*	X	X	–	–	–	–	X	–	–	–
Botiidae	*Syncrossus helodes*	X	X	X	–	X	–	X	X	X	X
	Yasuhikotakia caudipunctata	–	–	X	–	–	–	–	–	–	X
	Yasuhikotakia eos	X	X	–	X	–	X	–	X	X	X
	Yasuhikotakia lecontei	–	–	X	X	–	–	–	–	X	X
	Yasuhikotakia sp.	X	X	–	–	X	X	X	X	X	X
Channidae	*Channa micropeltes*	–	–	–	–	–	–	X	–	–	X

	Channa sp.	X	–	–	–	–	–	–	–	–	–
	Channa striata	X	–	X	X	–	–	X	–	–	X
Clariidae	*Clarias batrachus*	–	–	–	–	X	–	X	–	–	–
Clupeidae	*Clupeichthys aesarnensis*	–	–	X	X	X	–	X	–	–	–
	Tenualosa thibaudeaui	X	X	X	X	X	–	X	–	–	–
Cobitidae	*Acantopsis ioa*	–	–	–	–	–	–	–	–	–	X
	Acantopsis sp.	–	X	X	X	X	–	X	X	X	X
	Acantopsis thiemmedhi	–	–	–	X	–	–	–	–	–	–
	Pangio anguillaris	–	–	–	X	–	–	–	–	–	–
	Pangio sp.	–	–	–	X	–	–	X	X	–	–
Cyprinidae	*Bangana* sp.	–	–	–	X	–	–	–	–	–	–
	Barbichthys laevis	–	–	–	X	–	–	–	–	–	–
	Barbodes binotatus	–	–	–	–	–	–	X	–	–	–
	Barbonymus altus	X	X	X	X	–	–	X	X	X	X
	Barbonymus gonionotus	X	X	X	X	–	X	X	–	–	X
	Cirrhinus microlepis	X	X	–	–	X	–	X	–	X	–
	Cosmochilus harmandi	X	X	X	X	X	X	X	X	X	X
	Crossocheilus reticulatus	X	–	X	X	–	–	X	X	–	X
	Cyclocheilichthys armatus	–	–	–	–	–	–	–	–	–	X
	Cyclocheilichthys enoplos	–	–	X	X	–	–	X	X	–	–

Cyclocheilichthys repasson	–	X	X	–	X	–	X	X	–	–
Cyclocheilichthys sp.	–	–	X	–	–	–	–	X	–	–
Cyprinus sp.	X	X	X	X	–	–	–	–	–	–
Epalzeorhynchos sp.	–	–	–	–	–	–	X	–	–	X
Garra fasciacauda	X	X	–	–	–	X	–	–	–	–
Hampala dispar	X	–	–	–	–	X	X	–	–	–
Hampala macrolepidota	–	X	–	–	X	–	–	–	–	–
Henicorhynchus lobatus	–	–	–	–	–	X	X	X	–	X
Henicorhynchus siamensis	–	–	–	X	–	–	X	–	X	–
Hypsibarbus malcolmi	X	X	X	X	X	X	X	X	X	X
Hypsibarbus vernayi	X	–	X	–	–	–	X	X	–	–
Labeo catla	X	–	–	X	–	–	–	–	–	–
Labeo chrysophekadion	X	X	X	X	X	X	X	X	X	X
Labiobarbus leptocheilus	X	X	X	X	X	X	X	X	X	X
Labiobarbus sp.	–	–	–	–	–	–	X	X	–	X
Lobocheilos melanotaenia	–	–	X	–	X	–	X	X	–	X
Mekongina erythrospila	X	X	–	–	–	–	–	–	–	–
Mystacoleucus ectypus	X	X	–	–	–	–	–	–	–	–
Mystacoleucus marginatus	X	X	X	X	X	X	X	X	X	X
Mystacoleucus sp.	X	–	–	–	–	–	–	–	–	–

	Onychostoma meridionale	X	–	–	–	–	–	X	X	X	X
	Osteochilus melanopleurus	X	–	–	–	–	–	–	–	–	–
	Osteochilus schlegelii	X	X	–	–	–	–	X	–	X	–
	Poropuntius normani	X	X	X	–	X	X	–	–	–	–
	Probarbus jullieni	X	X	X	X	–	X	X	X	X	X
	Puntioplites falcifer	X	X	X	X	X	X	X	X	X	X
	Puntioplites proctozystron	X	–	X	X	X	X	X	X	X	X
	Puntioplites sp.	–	X	X	–	–	X	X	X	–	X
	Scaphognathops bandanensis	X	X	X	X	X	X	X	X	–	–
	Sikukia gudgeri	X	X	X	X	–	X	X	–	X	–
	Systomus orphoides	–	–	–	X	–	–	–	–	–	X
	Systomus sp.	–	–	–	–	–	–	–	–	–	X
	Thynnichthys thynnoides	X	–	–	–	–	–	–	–	–	–
Danionidae	*Opsarius koratensis*	X	–	–	–	–	–	–	–	–	–
	Opsarius pulchellus	–	–	–	X	–	–	–	–	–	–
	Raiamas guttatus	X	X	–	X	–	–	X	X	–	–
	Rasbora sp.	–	–	–	–	–	–	X	–	–	–
Datnioididae	*Datnioides* sp.	X	–	–	–	–	–	–	–	–	X
Eleotridae	*Oxyeleotris marmorata*	–	–	–	X	–	–	X	–	–	–
Gobiidae	*Papuligobius ocellatus*	X	–	X	X	X	X	X	X	X	X

	Rhinogobius sp.	–	X	–	–	–	–	–	X	–	–
Gyrinocheilidae	*Gyrinocheilus pennocki*	–	–	–	–	–	–	–	–	X	–
	Gyrinocheilus sp.	X	–	–	–	X	–	–	X	–	X
Leptobarbidae	*Leptobarbus hoevenii*	–	–	–	–	–	–	X	–	–	–
Mastacembelidae	*Macrognathus siamensis*	X	–	–	–	–	–	–	–	–	X
	Macrognathus sp.	–	–	–	–	–	–	–	–	–	X
	Mastacembelus sp.	X	X	X	–	X	X	X	X	X	X
Nemacheilidae	*Nemacheilus masyai*	–	–	–	–	–	–	–	–	–	X
	Schistura sp.	X	X	X	X	X	X	X	X	X	X
Notopteridae	*Chitala ornata*	X	X	–	X	–	–	X	X	–	–
	Notopterus notopterus	X	X	–	–	–	–	–	–	–	–
Osphronemidae	*Trichopodus* sp.	–	X	–	–	–	–	–	–	–	–
	Trichopodus trichopterus	–	–	–	X	–	–	X	–	–	–
Pangasiidae	*Pangasius bocourti*	–	–	–	X	–	–	X	X	X	–
	Pangasius conchophilus	–	X	–	–	–	–	–	X	–	–
	Pangasius macronema	X	X	X	X	X	X	X	X	X	X
	Pangasius nasutus	X	X	X	X	X	X	–	X	X	X
	Pangasius sp.	–	X	–	X	X	X	–	X	X	–
	Pseudolais pleurotaenia	X	X	X	–	–	X	X	–	–	–
Polynemidae	*Polynemus aquilonaris*	–	–	–	–	–	–	X	–	–	–

Pristolepididae	*Pristolepis fasciata*	X	–	–	X	–	X	X	X	X	X
Sciaenidae	*Boesemania microlepis*	–	–	X	X	X	–	X	–	–	–
Siluridae	*Kryptopterus geminus*	–	–	X	–	–	–	X	–	–	–
	Ompok siluroides	–	–	–	–	–	–	–	X	–	–
Sisoridae	*Bagarius yarrelli*	X	X	X	–	X	X	–	X	–	X
	Glyptothorax sp.	X	–	X	–	–	X	X	X	–	X
Soleidae	*Achiroides leucorhynchos*	X	–	X	X	X	–	X	X	–	–
Tetraodontidae	*Pao* sp.	X	X	–	X	X	X	X	X	X	X
Toxotidae	*Toxotes chatareus*	–	–	–	–	–	–	X	–	–	–
Xenocyprididae	*Hypophthalmichthys* sp.	–	–	–	–	–	–	X	–	–	–
	Paralaubuca typus	X	X	X	X	X	X	X	X	X	X

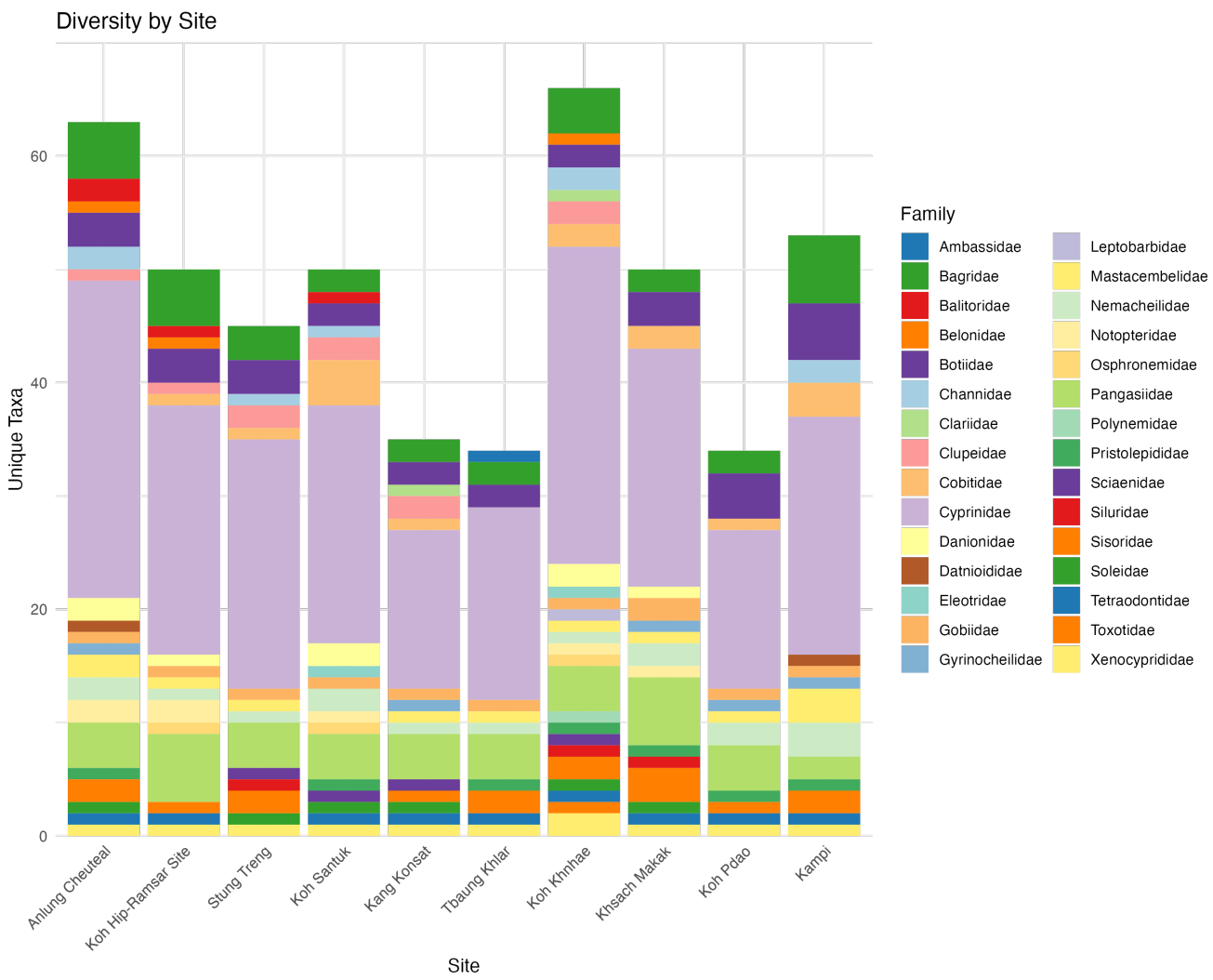

Figure 7. Total unique fish taxa detections by site, grouped by family. Sites are ordered left to right from upstream to downstream.

Table 3. The 80 species-level identities assigned from the sequence data, arranged by family. Non-native species are indicated with bolded red text, whereas priority species are indicated with bolded green text.

Family	Species-Level Detections	Total Species-Level IDs
Bagridae	*Hemibagrus spilopterus, Hemibagrus wyckii, Mystus albolineatus, Mystus singaringan, Pseudomystus siamensis*	**5**
Balitoridae	*Homalopteroides smithi*	**1**
Belonidae	*Xenentodon cancila*	**1**
Botiidae	*Syncrossus helodes, Yasuhikotakia caudipunctata, Yasuhikotakia eos, Yasuhikotakia lecontei*	**4**
Channidae	*Channa micropeltes, Channa striata*	**2**
Clariidae	*Clarias batrachus*	**1**
Clupeidae	*Clupeichthys aesarnensis,* ***Tenualosa thibaudeaui***	**2**
Cobitidae	*Acantopsis ioa, Acantopsis thiemmedhi, Pangio anguillaris*	**3**
Cyprinidae	*Barbichthys laevis, Barbodes binotatus, Barbonymus altus, Barbonymus gonionotus, Cirrhinus microlepis, Cosmochilus harmandi, Crossocheilus reticulatus, Cyclocheilichthys armatus, Cyclocheilichthys enoplos, Cyclocheilichthys repasson, Garra fasciacauda, Hampala dispar, Hampala macrolepidota, Henicorhynchus lobatus, Henicorhynchus siamensis, Hypsibarbus malcolmi, Hypsibarbus vernayi,* ***Labeo catla****, Labeo chrysophekadion, Labiobarbus leptocheilus, Lobocheilos melanotaenia, Mekongina erythrospila, Mystacoleucus ectypus, Mystacoleucus marginatus, Onychostoma meridionale, Osteochilus melanopleurus, Osteochilus schlegelii, Poropuntius normani,* ***Probarbus jullieni****, Puntioplites falcifer, Puntioplites proctozystron, Scaphognathops bandanensis, Sikukia gudgeri, Systomus orphoides, Thynnichthys thynnoides*	**35**
Danionidae	*Opsarius koratensis, Opsarius pulchellus, Raiamas guttatus*	**3**
Eleotridae	*Oxyeleotris marmorata*	**1**
Gobiidae	*Papuligobius ocellatus*	**1**
Gyrinocheilidae	*Gyrinocheilus pennocki*	**1**
Leptobarbidae	*Leptobarbus hoevenii*	**1**
Mastacembelidae	*Macrognathus siamensis*	**1**
Nemacheilidae	*Nemacheilus masyai*	**1**
Notopteridae	*Chitala ornata, Notopterus notopterus*	**2**
Osphronemidae	*Trichopodus trichopterus*	**1**
Pangasiidae	*Pangasius bocourti, Pangasius conchophilus, Pangasius macronema, Pangasius nasutus, Pseudolais pleurotaenia*	**5**
Polynemidae	*Polynemus aquilonaris*	**1**
Pristolepididae	*Pristolepis fasciata*	**1**
Sciaenidae	*Boesemania microlepis*	**1**
Siluridae	*Kryptopterus geminus, Ompok siluroides*	**2**
Sisoridae	***Bagarius yarrelli***	**1**
Soleidae	*Achiroides leucorhynchos*	**1**
Toxotidae	*Toxotes chatareus*	**1**
Xenocyprinidae	*Paralaubuca typus*	**1**
	Total	**80**

The number of taxa detected per site ranged from 35 to 68, with a mean of 50 (Figure 7). Across all the sample sites, the greatest level of fish diversity was detected in Koh Khnhae Village (68 taxa), followed by Anlung Cheuteal (66 taxa) and Kampi (55 taxa; Figure 7). The lowest number of taxa was observed in Tbaung Klar (35 taxa; Figure 7).

Some of the detected taxa were broadly distributed across the entire sampling region, and a total of nine taxa were observed in all 10 sample sites (*Cosmochilus harmandi*, *Hypsibarbus malcomi*, *Labeo chrysophekadion*, *Labiobarbus leptochilus*, *Mystacoleucus marginatus*, *Puntioplites falcifer, Paralaubuca typus*, *Pangasius macronema*, and *Schistura* sp.; Table 2). However, many other taxa were detected in only a single location. Detections restricted to a single site included 16 species-level assignments – *Acantopsis ioa*, *Acantopsis thiemmedhi*, *Barbichthys laevis*, *Barbodes binotatus*, *Cycloheilichthys armatus, Gyrinocheilus penncocki*, *Leptobarbus hoevenii*, *Nemacheilus masyai*, *Ompok siluroides*, *Opsarius koratensis*, *Opsarius pulchellus*, *Osteochilus melanopleurus*, *Pangio anguillaris, Polynemus aquilonaris*, *Thynnichthys thynnoides*, and *Toxotes chatareus*. An additional 10 genus-level assignments were also observed in only a single site each, including *Bangana* sp., *Channa* sp., *Hypophthalmichthys* sp., *Macrognathus* sp., *Mystacoleucus* sp., *Mystus* sp., *Parambassis* sp., *Rasbora* sp., *Systomus* sp., and *Trichopodus* sp. (Table 2).

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

Family	Detected Species	Sites Where Detected
Cyrpinidae	*Cyprinus* sp.	Anlung Cheuteal, Koh Hip, Koh Santuk, Stung Treng
	Labeo catla	Anlung Cheuteal, Koh Santuk
Xenocyprinidae	*Hypophthalmichthys* sp.	Koh Khnhae
	Total Detected Taxa = ≥3*	**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 three taxonomic assignments, as some may represent multiple species within the same genus.

At least three non-native fish species were also detected based on two genus-level sequence assignments and one species-level sequence assignment (Table 4). These included *Cyprinus* sp., *Hypophthalmichthys* sp., and *Labeo catla*. Of these detections, those belonging to genus *Cyprinus* were most widespread, occurring in Anlung Cheuteal, Koh Hip, Stung Treng, and Koh Santuk (Table 2; Table 4). These detections may be due to the presence of common carp (*Cyprinus carpio*), which are known to occur in the region. Sequences belonging to *Labeo catla* – a species that was introduced for aquaculture – were also detected in Anlung Cheuteal and Koh Santuk. The sole detection of a sequence belonging to *Hypophthalmichtys* – the genus containing silver and bighead carps – occurred in Koh Khnhae. All of these detections represent known invasive species, and no novel non-native invaders were detected.

Priority Species

Three of the 20 priority species of threatened freshwater fish identified as high priority by WWF Cambodia were definitively identified with species-level sequence assignments (Table 5). Sequences assigned to the Vulnerable species goonch (*Bagarius yarrelli*) were detected in Anlung Cheuteal, Koh Hip, Kang Konsat, Tbaung Khlar, Khsach Makak, and Kampi (Table 2). Sequences assigned to the Critically Endangered Jullien's golden carp (*Probarbus jullieni*) were detected in samples from nine of the 10 sampled locations (all but Kong Konsat; Table 2). Finally, sequences

assigned to the Vulnerable species Laotian shad (*Tenualosa thibaudeaui*) were detected in Anlung Cheuteal, Koh Hip, Stung Treng, Koh Santuk, Kang Konsat, and Koh Khnhae (Table 2).

Sequences that could only be assigned to the genus level may represent an additional three priority species – Mekong tiger perch (*Datnioides undecimradiatus*) in Anlung Cheuteal and Kampi; *Glyptothorax fuscus* in Anlung Cheuteal, Stung Treng, Tbaung Khlar, Koh Khnhae, Khsach Makak, and Kampi; and *Puntioplites bulu* in Koh Hip, Stung Treng, Tbaung Khlar, Koh Khnhae, Khsach Makak, and Kampi (Table 2). It is not possible to determine whether the *Datnioides* sequences derived from Mekong tiger perch (*D. undecimradiatus*) or Siamese tiger perch (*D. pulcher*), as both of these species occur in the Lower Mekong, although either would be noteworthy given that *D. pulcher* is an ASAP species. Similarly, it is not possible to determine whether the *Glyptothorax* and *Puntioplites* sequences derived from the respective priority species, as multiple members of each of these genera exist in the region. Notably, sequences belonging to two other *Puntioplites* species (*P. falcifer* and *P. proctozystron*) were identified with species-level assignments (Table 2; Table 3).

Though they carry even less certainty, family-level taxonomic assignments may have included sequences belonging to an additional eight of the priority species. These include multiple sequences assigned to family Cyprinidae that could have potentially derived from bala sharkminnow (*Balantiocheilos melanopterus*), giant barb (*Catlocarpio siamensis*), thicklip barb (*Probarbus labeamajor*), thinlip barb (*P. labeaminor*), or tiger barb (*Puntius partipentazona*). In addition, multiple sequences assigned to family Pangasiidae may have derived from Mekong giant catfish (*Pangasianodon gigas*), and multiple sequences assigned to family Sisoridae may have derived from dwarf goonch (*Bagarius bagarius*) or crocodile catfish (*B. suchus*).

Finally, order-level taxonomic assignments of sequences to the orders Osteoglossiformes and Siluriformes may have derived from the priority species Asian bonytongue (*Scleropages formosus*) or any of the six target catfish species noted above, respectively. Importantly, Asian bonytongue, also known as the Asian arowana, is the only member of the order Osteoglossiformes known to occur in the region, and therefore these detections may very well have derived from that species. The Siluriformes sequences carry far less certainty, as many catfish species within this order are found throughout the Lower Mekong.

There are several factors that may have precluded the ability to detect these 20 priority species. First of all, only 11 of these species have full mitochondrial reference genomes available on GenBank, whereas eight only have partial sequences available and one has no sequences available (Table 5). In addition, at least one of these species is known to be restricted to estuaries (sawfish; *Pristis microdon*), one is known to occur in coastal streams (tiger barb; *Puntius partipentazona*), one is restricted to foothill streams (*Glyptothorax fuscus*), and one is believed to be extirpated from Cambodia (*Puntioplites bulu*; Table 5). The gaps in existing genetic reference databases for many of these species precludes confident determination of whether they were detected or not.

Table 5. Detections and potential detections of the 20 priority freshwater fish species identified by WWF Cambodia.

Species	Common Name	Status	Reference Availability	Detected?
Osphronemus exodon	Elephant ear gourami	Vulnerable	Partial	No
Osphronemus goramy	Giant gourami	Least Concern	Yes	No
Tenualosa thibaudeaui	Laotian shad	Vulnerable	Yes	Yes
Lycothrissa crocodilus	Sabertooth thryssa	Least Concern	Yes	No
Balantiocheilos melanopterus	Bala sharkminnow	Vulnerable	Yes	Potential (family-level: Cyprinidae)
Catlocarpio siamensis	Giant barb	Critically Endangered	Yes	Potential (family-level: Cyprinidae)
Probarbus jullieni	Jullien's golden barb	Critically Endangered	Yes	Yes
Probarbus labeamajor	Thicklip barb	Endangered	No	Potential (family level: Cyprinidae)
Probarbus labeaminor	Thinlip barb	Near Threatened	Partial	Potential (family level: Cyprinidae)
Puntioplites bulu	No common name	Least Concern	Partial	No; presumed extirpated from Cambodia
Puntius partipentazona	Tiger barb	Least Concern	Yes	No; only found in coastal streams
Scleropages formosus	Asian bonytongue	Endangered	Yes	Potential (order level: Osteoglossiformes)
Pristis microdon	Sawfish	Critically Endangered	Partial	No; estuarine species
Pangasianodon gigas	Mekong giant catfish	Critically Endangered	Yes	Potential (family level: Pangasiidae)
Wallago leerii	Tapah	Endangered	Partial	Potential (order level: Siluriformes)
Bagarius bagarius	Dwarf goonch	Vulnerable	Partial	Potential (family level: Sisoridae)
Bagarius suchus	Crocodile catfish	Near Threatened	Partial	Potential (family level: Sisoridae)
Bagarius yarrelli	Goonch	Vulnerable	Yes	Yes
Glyptothorax fuscus	No common name	Least Concern	Partial	Potential (genus level: *Glyptothorax*)
Datnioides undecimradiatus	Mekong tiger perch	Vulnerable	Yes	Potential (genus level: *Datnioides*)

In addition to the challenges caused by gaps in reference databases for these 20 priority species, other important fish species in the region also lack reference library coverage. These include Mekong stingray (*Hemitrygon laosensis*), for which there no available reference sequences, and giant freshwater whipray (*Urogymnus polylepis*), for which there are only partial reference sequences. 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.

Discussion

This study represents a valuable continuation of eDNA-based monitoring in the Cambodian Mekong River, and provides both a useful update on the fish communities in various, important habitats between the Lao border and Kampi, as well as insight into variation in the resolution of data provided by varying field sampling procedures. The use of single-use sample collection kits to collect four replicates per site paired with analysis using the MiFish primers resulted in the detection of 117 fish taxa, and a total of 80 sequences could be identified to the species-level. This represents an improvement species detections compared with past studies that collected fewer replicates (Eschenroeder et al. 2024).

In addition to the detection of ASAP and priority species such as the critically endangered Jullien's golden carp and vulnerable goonch, eDNA samples were able to detect invasive species (*Cyprinus* sp., *Hypopothalmichthys* sp., and *Labeo catla*). This demonstrates the utility of eDNA monitoring mapping and monitoring the distribution of both important natives and potentially harmful non-natives, and the ability to do so is of great value for tracking the impacts of infrastructure (e.g., dams), climate change, and conservation efforts on the fish community.

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 10 discrete locations across many hundreds of river kilometers, the assessment of taxa accumulation across the 40 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 30 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. Improved resolution of species 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* in the region based on past fisher surveys (Lee et al. 2023). 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 even 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. 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.

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 by WWF Cambodia may be modified and improved based on the best available science.

As the Lower Mekong experiences increasing anthropogenic changes arising from a growing human population, land use alteration, sand mining, hydropower development, and climate change, obtaining reliable data on fish species ranges and habitat occupancy will be critical for informing adaptive management of the region's important fisheries. In particular, effective evaluations of the impacts of planned dams and other infrastructure require significant improvements in understanding of the seasonal distributions of threatened species and species important to local fisheries. Environmental DNA sampling stands as a potential means of generating this data. Therefore, continuing data collection and archiving of both extracted DNA and sequence data will be of great value for anticipating the impacts of changes to the region, as well as for monitoring changes in the fish community in response to local management practices and large-scale shifts in the climate and hydrology of the region.

Literature Cited

Bezuijen, M. R., R. Timmins and, T. Seng, editors. 2008. Biological surveys of the Mekong River between Kratie and Stung Treng Towns, northeast Cambodia, 2006-2007. WWF Greater Mekong - Cambodia Country Programme, Cambodia Fisheries Administration and Cambodia Forestry Administration, Phnom Penh.

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.

Lee, D., Eschenroeder, J. C., Baumgartner, L. J., Chan, B., Chandra, S., Chea, S., Chea, S., Chhut, C., Everest, E., Hom, R., Heng, K., Lovgren, S., Ounboundisane, S., Robinson, W., Seat, L., Soth, S., and Hogan, Z. S. 2023. World heritage, hydropower, and Earth's largest freshwater fish. Water. 15(10): 1936.

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.

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.

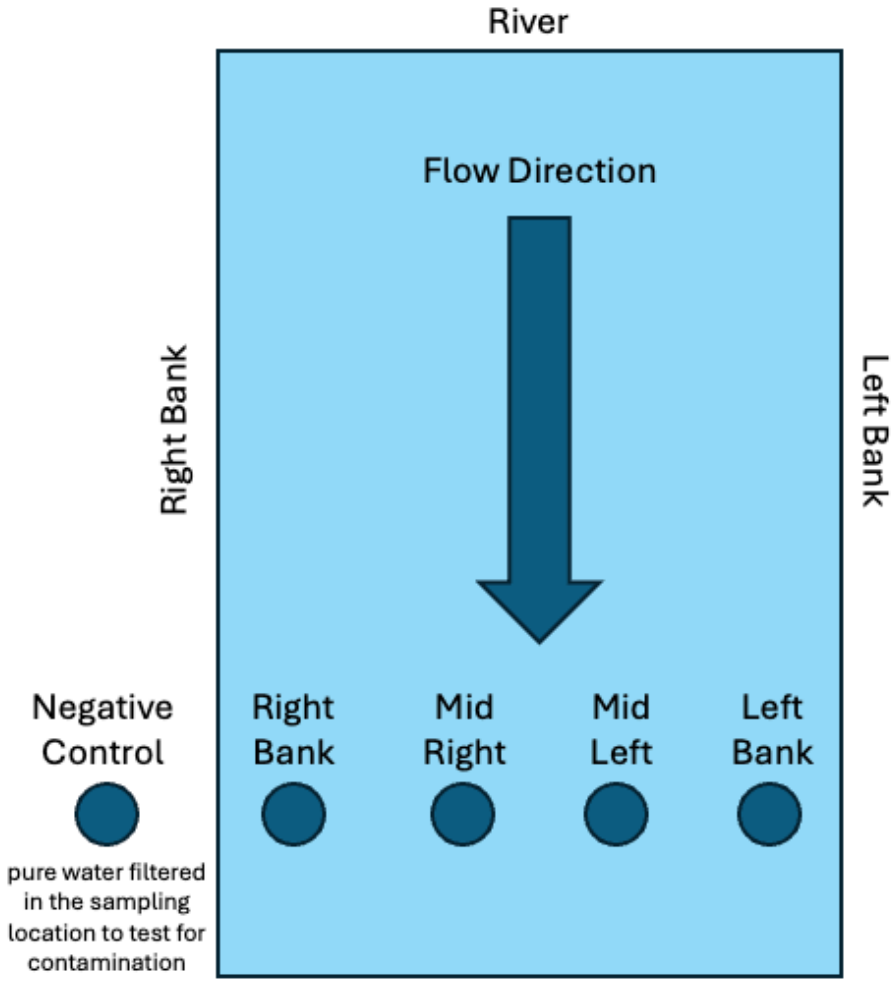

Figure 1. The sampling approach that will be used in each location.

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 50 samples will be collected across 10 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	
							Right / MidRight / MidLeft / Left / Control	

Figure 2. Proposed eDNA sample collection datasheet.

Table 2. The proposed eDNA sampling framework.

Village	District	Province	Samples to be Collected
Anlung Cheuteal	Thala Barivat	Stung Treng	4 in river, 1 negative control
Koh Hip – Ramsar Site	Thala Barivat	Stung Treng	4 in river, 1 negative control
Stung Treng	Krong Stung Treng	Stung Treng	4 in river, 1 negative control
Koh Santuk	Siem Bouk	Stung Treng	4 in river, 1 negative control
Kang Konsat	Siem Bouk	Stung Treng	4 in river, 1 negative control
Tbaung Khlar	Siem Bouk	Stung Treng	4 in river, 1 negative control
Koh Khnhae	Sambour	Kratie	4 in river, 1 negative control
Khsach Makak	Sambour	Kratie	4 in river, 1 negative control
Koh Pdao	Sambour	Kratie	4 in river, 1 negative control
Kampi	Sambour	Kratie	4 in river, 1 negative control
			50

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.