Title: Evidence for population genetic structure in two exploited Mekong River fishes across a natural riverine barrier

Biesack, Ellen E.[1], Dang, Binh T.[2], Ackiss, Amanda S.[1], Bird, Christopher E.[3], Chheng, Phen[4], Phounvisouk, Latsamy[5], Truong, Oanh T.[2], Carpenter, Kent E.[1]

[1]4402 Elkhorn Avenue, Department of Biological Sciences, Old Dominion University, Norfolk, VA, USA

[2]02 Nguyen Dinh Chieu Street, Institute for Biotechnology and Environment, Nha Trang University, Nha Trang, Vietnam

[3]6300 Ocean Drive, Department of Life Sciences, Texas A&M University – Corpus Christi, Corpus Christi, TX

[4]Inland Fisheries Research and Development Institute, Phnom Penh, Cambodia

[5]Nongteeng Sangthong Street, Living Aquatic Resources Research Centre, Nong Thang Village, Vientiane, Lao PDR

Corresponding author: E. E. Biesack - 1355 Greate Rd, Virginia Institute of Marine Science, Gloucester Point, VA, USA - eebiesack@vims.edu

This research was funded by the United States Agency for International Development supported Partnerships for Enhanced Engagement in Research Project 2-7 under USAID Cooperative Agreement AID-OAA-A-11-00012 and the National Science Foundation grant OISE-0730256 and administered by the National Academy of Sciences.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jfb.14424

Abstract

Impacts of urban development on aquatic populations are often complex and difficult to ascertain, but population genetic analysis has allowed researchers to monitor and estimate gene flow in context of existing and future hydroelectric projects. The Lower Mekong Basin is undergoing rapid hydroelectric development with around 50 completed and under construction dams and 95 planned dams. We investigated the baseline genetic diversity of two exploited migratory fishes, the Mud Carp *Henicorhynchus lobatus* (five locations) and the Rat-faced Pangasiid *Helicophagus leptorhynchus* (two locations), in the Lower Mekong Basin using the genomic ddRAD method. In both species, fish sampled upstream of Khone Falls are differentiated from those collected at other sites, and N_e estimates at the site above the falls are lower than at other sites. This was the first study to utilize thousands of RAD-generated SNPs to indicate that the Mekong's Khone Falls are a potential barrier to gene flow for these two moderately migratory species. The recent completion of the Don Sahong dam across one of the only channels for migratory fishes through Khone Falls may further exacerbate signatures of isolation and continue to disrupt the migration patterns of regionally vital food fishes. Additionally, *Hen. lobatus* populations downstream of Khone Falls, including the 3S basin and Tonle Sap system, displayed robust connectivity. Potential obstruction of migration pathways between these river systems resulting from future dam construction may limit dispersal, which has led to elevated inbreeding rates and even local extirpation in other fragmented riverine species.

Keywords: dams, fishes, fragmentation, Khone Falls, Mekong River, population genetics

Introduction

The Mekong River Basin supports the world's most productive inland fishery and is the second most biodiverse freshwater habitat, surpassed only by the Amazon River Basin (Campbell, 2009). The river system encompasses a large variety of habitats, including permanent river channels, swamps and marshes, lakes, rapids, floodplains, seasonally isolated pools, and flooded forest, all of which experience drastic annual shifts during the monsoon and dry seasons. Formative historic changes in hydrology and structure due to climatic shifts throughout the Pleistocene and Holocene epochs (Ta *et al*., 2002; Voris, 2000) have resulted in its elevated aquatic diversity, which includes over 100 species of migratory fishes (Valbo-Jørgensen *et al*.; 2009; Ziv *et al*.; 2012). This diversity, however, is under threat from the construction of manmade structures as the nations of the Mekong develop hydroelectric projects on the river, and conservation measures will be required to mitigate these impacts.

The Mekong and tributaries south of Pakse in Lao PDR are collectively known as the Lower Mekong Basin, which encompass several major hydrological features traversed by migratory fishes. Khone Falls is the largest waterfall along the mainstream Mekong River and has been hypothesized to be a major barrier to migration and a potential source of population diversification (Hurwood *et al.*, 2007; Poulsen & Valbo-Jørgensen, 2001; Chhea *et al.*, 2000). Below Khone Falls, one of the largest tributary systems is the 3S Basin, comprised of the Sekong, Sesan, and Srepok rivers (Figure 1). The 3S rivers are major tributaries that provide up to 20% of annual flow and sediment to the Lower Mekong Basin (Piman *et al*., 2016). The 3S Basin also provides habitat to 40% of Mekong fish species, including many species that migrate through the 3S rivers to reach their spawning ground (Baran *et al*., 2013). Tonle Sap River is known to support a large inland fishery throughout its tributaries and Tonle Sap Lake and is

characterized by a seasonal flood-pulse hydrology that is used by many fish species in their migrations (Arias *et al.*, 2013; Junk *et al.*, 1989; Poulsen *et al.*, 2002).

Changes to the natural hydrography of this system brought about by hydroelectric development have had measurable impacts on the composition of fish communities in this region, including decreasing overall diversity and favoring generalist species (Ngor *et al.*, 2018a). Currently, there are more than 50 existing dams in the Mekong River Basin and over 100 more planned or under construction (WLE Greater Mekong, 2016) with changes to downstream flow and sediment transportation already being observed (Fan *et al*., 2015; Li *et al*., 2017). Eleven of the planned or under construction dams are for the main river channel and the remaining are for tributaries of the Mekong, including 21 for the 3S Basin alone (Piman *et al*., 2016). A dam has just been completed in the Hou Sahong Channel, one of the more navigable passages through Khone Falls, and its construction may potentially obstruct an important migratory route for fishes navigating the main stem of the Mekong (Baird, 2011; International Rivers, 2017). The development of hydroelectric infrastructure is considered important for energy independence in the region, where water is an abundant resource. However, studies by the Mekong River Commission (1997) concluded that impacts from dam construction and urbanization were significantly greater threats to Mekong River biodiversity than ongoing intensive direct exploitation. Despite efforts to minimize biological impact on migratory species, many dams lack fish ladders and those that exist have been largely ineffective at maintaining migratory routes (Amornsakchai *et al*., 2000). A case study on the effects of dams on migratory species predicts that hydroelectric development in the Lower Mekong Basin will negatively impact fisheries development in the region (Ferguson *et al.*, 2011). The continuing impacts of dam-induced fragmentation and habitat alteration to fish communities in the Mekong River, where fisheries species compose up to 80%

of protein consumption for the human populations, will result in the need for costly compensatory action (Orr *et al*., 2012).

Here, we investigate genetic structure of two exploited fishes exhibiting varying levels of migratory behavior in the Lower Mekong Basin in order to hypothesize potential impacts from damming. The Mud Carp *Henicorhynchus lobatus* Smith 1945 is a small cyprinid (maximum size 15 cm standard length) and one of the most abundant fishes in the Mekong River system. It is the most economically important fish in the region (Roberts and Baird, 1995) and is especially important for the Tonle Sap fishery in Cambodia and in the Khone Falls area of southern Lao PDR where it accounts for up to 50% of the total wild fish catch (Baran *et al*., 2005). *Henicorhynchus lobatus* undergoes mass migrations (Baran *et al.*, 2006) and has been observed swimming through the few passable channels (Hou Sahong and Hou Sadam) in Khone Falls (Baird *et al*., 2003). This is notable since the falls are known to be a formidable barrier to riverine fish migration, and previous studies of genetic structure in the Mekong and adjacent systems have found no genetic differentiation in *Hen. lobatus* populations across the falls or along the main stem of the Lower Mekong using mitochondrial ATPase6 sequences (Hurwood *et al.*, 2007) and microsatellite markers (Iranawati, 2014).

In comparison, relatively little is known of the ability of the recently described Rat-Faced Pangasiid Catfish *Helicophagus leptorhynchus* Ng & Kottelat 2000 to traverse Khone Falls. *Helicophagus leptorhynchus* is endemic to the Mekong and Chao Phraya watersheds in Southeast Asia (maximum size 47 cm standard length). Many pangasiids are known to successfully migrate upstream through Khone Falls at the beginning of the rainy season (Baird, 2001; Baird and Flaherty, 2004; Hogan et al. 2004), and these fishes support a wing trap fishery specifically designed to target the seasonal migration through Khone Falls (Baird *et al.*, 2004). It

is unknown, however, whether *Hel. leptorhynchus* is among the pangasiids capable of traversing the falls. *Helicophagus leptorhychus* primarily inhabits deep permanent tributaries and main river channels, migrating upstream during the rainy season and avoiding seasonally flooded habitat (Ng & Kottelat, 2000; Rainboth, 1996).

To gain a better understanding of gene flow patterns in these harvested Mekong River fishes, we used double-digest restriction site associated DNA (ddRAD, Peterson et al., 2012) data to document genetic patterns in populations that were unimpacted by dam-induced fragmentation at the time of collection. Given differing migratory capabilities and exploitation between these fishes, we examined levels of genetic divergence, relatedness, and effective population size, with particular attention to overlapping collections from mainstream populations upstream of Khone Falls and upstream in the Sekong River of the 3S Basin. We also utilized the power of thousands of ddRAD-obtained SNPs to detect population structure in *Hen. lobatus*, where no genetic structure has been detected by previous studies using mitochondrial or microsatellite markers. We hypothesized that *Hen. lobatus* will exhibit limited population differentiation throughout the study area due to its documented large-scale migrations and based on previous studies detecting no genetic structure using mitochondrial and microsatellite markers. We then hypothesized that the larger and less numerous *Hel. leptorhynchus* may comprise multiple somewhat differentiated stocks, as it is not known to undertake basin-wide migrations. These results would have different implications for long-term conservation of the two species in a region that will soon be heavily dammed.

Materials and Methods

Samples of *Hen. lobatus* were field-identified and purchased at fish markets in Attapeu and Pakse in Lao PDR, and from three sites, Kratie, Stung Treng and Siem Reap, in Cambodia (Figure 1, Table 1). Samples of *Hel. leptorhynchus* were collected in the same manner from Attapeu and Pakse in Lao PDR. All fish were collected during the period of October 2016 to January 2017. Vendors reported that specimens originated from fisheries near the city of collection. Muscle tissue was taken and stored in 95% molecular-grade ethanol for transport to the molecular lab at Nha Trang University, Vietnam. DNA isolation was carried out using the Qiagen DNeasy Blood and Tissue kit (Hilden, Germany) and high-quality extracts were confirmed with gel electrophoresis, stained with ethidium bromide. Extracted DNA appearing as high-weight bands with very little degradation on a gel was shipped to the Texas A&M University – Corpus Christi (TAMUCC) Genomics Core Laboratory for library preparation. Double-digest RAD (ddRAD; Peterson *et al*., 2012) library prep was employed to reduce genome complexity, with minor modifications on the original protocol. Extracted DNA was again checked for size distribution after shipping to TAMUCC using standard gel electrophoresis. Samples with visible DNA fragments below 3000 bp were SPRI-selected (0.4x, Beckman-Coulter). Samples that were not SPRI-selected were cleaned using AmpureXP paramagnetic beads (Beckman-Coulter). In both cases, the paramagnetic beads were left in the cleaned samples to minimize DNA loss and reactivated, as required, by adding 3M NaCl 20% PEG (Fisher *et al*., 2011). The DNA was quantified using the Biotium AccuBlue HighSensitivity kit on a Spectramax M3 fluorescent plate reader. Samples were normalized and 50 ng of DNA per fish were digested with the Hin1II and EcoRI restriction enzymes. Each fish was individually barcoded and groups of 48 fish were combined and indexed via polymerase chain reaction (PCR), as described by Peterson *et al*. (2012). DNA fragments from 400-500bp were isolated

using a Pippin Blue automated electrophoresis rig and the molar concentration of DNA in each library was determined using the qPCR KAPA Library Quantification Kit for Illumina Platforms. Libraries were subsequently normalized and combined in equal proportions into a super library, which was sequenced on two lanes of an Illumina HiSeq 4000 (paired-end 150bp) at New York University's Genome Technology Center.

All bioinformatics and statistical analyses were completed by the molecular systematics laboratory at Old Dominion University. Successful sequences were demultiplexed by Illumina dual-indices and barcodes using scripts developed by the TAMUCC bioinformatics team. Quality trimming, *de novo* reference assembly, mapping, and variant calling were carried out in the dDocentHPC pipeline (https://github.com/cbirdlab/dDocentHPC) a fork of the dDocent pipeline (Puritz *et al.*, 2014). The settings used to process the FASTQ files, create the reference genomes, and call variants are included in the dDocentHPC configuration file (Supplemental S1). Briefly, for the assembly of the reference genomes, reads were subjected to a base call quality (5' phred 20, 3' phred 15, 20 nucleotide sliding window at phred 20) and adapter trimming with any read trimmed shorter than 146 nucleotides being removed. The reference genomes were assembled for each species from unique DNA sequences present in at least four individuals and with a depth of coverage of at least 5x using rainbow (Chong *et al.*, 2012) and CD-HIT (Fu *et al.*, 2012). For mapping and variant calling, reads were subjected to adapter and quality trimming as described above, but with a 3' phred quality threshold of 20, and any read trimmed shorter than 75 bp was removed. The trimmed reads were mapped to the reference genome using bwa mem (Li *et al.*, 2013; -A1 -B6 -O10 -T50 -L30,5) and the binary sequence alignment maps were filtered to remove improperly paired reads, secondary alignments, and read pairs with an unmapped read using samtools view (Li *et al.*, 2009; -q20 -F264 -f2). The variant

calling settings in freebayes (Garrison & Marth, 2012) were tuned to explicitly call SNPs rather than multinucleotide polymorphisms (MNPs) because there is no VCF filtering software to handle MNPs, which typically have more than two allelic states, and there is no software to deconstruct MNPs into SNPs that maintains or generates the statistical information associated with each SNP for downstream filtering.

SNPs underwent stringent filtering following O'Leary et al. (2018), employing vcftools (Danecek *et al.*, 2011), vcflib (https://github.com/vcflib/vcflib), and rad_haplotyper (Willis *et al.*, 2017). The filtering commands (Supplemental S2) provide all information required to replicate our filters. Briefly, genotype calls based on fewer than 3 reads were discarded initially to reduce memory required for subsequent filters. Variant positions meeting any of the following criteria were removed: mean depth of coverage less than 5x, phred-scaled probability of the presence of at least one alternate allele less than a 30, alternate allele frequency less than 0.05 or greater than 0.95, allele balance greater than 0.75 or less than 0.25, significant deviation from Hardy-Weinberg Equilibrium, and more than two haplotypes in greater than one individual. All indels and SNPs with more than two allelic states were removed. Parameter values for several other filters differed between the two datasets and can be found in the supplemental material (Supplemental S2). Following Mastretta-Yanes *et al.* (2015), of the remaining contigs with SNPs, we removed positions above 330 (*Hen. lobatus*) and 337 (*Hel. leptorhynchus*) for all contigs due to an excessively high concentration of variable positions (Supplemental S3). Finally, one SNP was selected from each contig for analysis and PGDSpider (Lischer and Excoffier, 2012) was used to convert the VCF file to the formats required by subsequent analyses. BayeScan (Foll and Gaggiotti, 2008) was used to look for SNPs showing higher differentiation (F_{ST}) than expected under neutrality in both species' datasets, with a false

discovery rate (FDR) correction of 0.05. No loci were identified as outliers from either dataset, and so all the following methods were completed on the full SNP datasets.

Indices of genetic diversity (expected heterozygosity, observed heterozygosity, and inbreeding coeffiecient [G_{IS}]) were calculated for loci, individuals, and populations in GenoDive v3.0 (Meirmans and Van Tienderen, 2004). An analysis of molecular variance (AMOVA, Excoffier *et al.*, 1992) among all sites and genetic differentiation (F_{ST}) between all pairs of sites were also calculated in GenoDive (p-values were generated with 9,999 permutations). Maximum likelihood triadic estimates of relatedness between pairs of individuals were made with the coancestry function in the R package "related" (Wang, 2007) with 500 bootstraps for confidence intervals. Estimates of effective population size (N_e) were made in the program NeEstimator (Do *et al.*, 2014) using the linkage disequilibrium method. N_e estimates were made for each individual sampling site, and with and without putatively related individuals.

Population structure analyses were carried out using the R package "adegenet" (Jombart and Ahmed, 2011) to perform principle component analyses (PCAs), and in the program STRUCTURE (Hubisz *et al.*, 2009; Pritchard *et al.*, 2000). STRUCTURE analyses utilized 100,000 MCMC iterations following 50,000 burn-in reps per *K* (*K* values 1 through 6 for *Hen. lobatus* and 1 through 4 for *Hel. leptorhynchus*). Each of these runs was repeated 10 times to maximize parameter space and location information was not used as a prior. Structure Harvester was used to determine the most likely K-value from STRUCTURE output using the Evanno method (Earl *et al.*, 2012; Evanno *et al.*, 2005), and the replicate *Q*-scores for individuals and populations for the chosen *K* were aggregated with CLUMPP (Jakobsson and Rosenberg, 2007). STRUCTURE plots were constructed using distruct (Rosenberg, 2004). *Post-hoc* clusters of genetically similar sample sites were formed based on results of population structure analyses

and tested using AMOVA (Excoffier *et al.*, 1992) in GenoDive to find the clustering scheme that explains the most variance. Mantel and partial Mantel tests were conducted in the R package "ecodist" (Goslee & Urban, 2007) to examine potential effects of isolation by distance (IBD) on patterns of structure by measuring the correlation between genetic differentiation (pairwise F_{ST} values) and geographic distance (measured as shortest over-water distance using the "measure distance" function in Google Maps). The same *post-hoc* clustering schemes created for AMOVAs were also used to test for IBD while accounting for significant structure using partial Mantel tests.

Results

Henicorhynchus lobatus

One hundred and thirty fish were successfully sequenced, demultiplexed, mapped to the *de novo* reference genome, and passed stringent filtering (Table 1). No outlier SNPs were identified by BayeScan, and 88 SNPs were removed from 10 terminal positions on all reads (Supplemental Figure S3), resulting in a panel of 1,971 putatively neutral SNPs (one per contig) to be used in the remaining analyses. Observed heterozygosity was statistically similar among locations, ranging from 0.26 (Pk) to 0.27 (ST; average 0.26). The inbreeding coefficient (G_{IS}) was lowest at Stung Treng (0.12) and highest at Pakse (0.14) and non-overlapping confidence intervals indicate a statistically significant difference in G_{IS} between these two sites.

Indices of genetic differentiation, relatedness, and inferred population size all indicated that the *Hen. lobatus* from the site above Khone Falls, Pakse, were differentiated and isolated with respect to those from the other sites. Analysis of molecular variance (AMOVA) indicated that significant genetic differentiation exists among sampling locations when unclustered (F_{ST} =

0.0019, p = 0.0001). Pairwise comparisons of neutral SNPs indicated that Pakse was differentiated from the other sites (F_{ST} = 0.0028 - 0.0050, $p < 0.0001$; Table 2), while the remaining sites were not significantly differentiated. Estimates of relatedness revealed a pair of closely related fish within Pakse with a triadic related value of 0.458 (Table 3), but not at the other locations. This level of relatedness would be seen among siblings or a parent and its offspring. The estimate of N_e was substantially lower at Pakse (N_e = 6.4E2, Table 4) than the other Lao PDR sites (N_e = 1.1E4 - 1.3E4) or the Cambodia sites downstream (N_e = 3.7E4 - 4.3E5). However, when an individual from the related pair is removed from Pakse, the N_e estimate increases to 2.4E3, though it is still at least an order of magnitude lower than estimates from other sites. Due to the lack of genetic structure among the sites below Khone Falls, N_e was estimated for all of these sites, but that resulted in an infinite estimate (Table 4) suggesting that either N_e is very large, or that more sampling is required to obtain a bounded estimate.

In a principle component analysis (PCA), Pakse was distinguishable from other sites along PC1, which explained 1.23% of the variation in the data (Figure 2). A related pair separated from the remaining individuals on PC2 (1.10% of the variation; Figure 2), and when one individual from the pair was removed from analysis, Pakse is still distinguishable from the cluster of other sites (Supplemental S4). The STRUCTURE analysis, plotted with a *k* of 2 as chosen by the Evanno method, also showed a clear distinction between Pakse and the remaining four sites, with a greater proportion of individuals assigning more strongly to one group (top color) than in the other locations with the exception of a handful of individuals below the Khone Falls that also exhibited a Pakse-like pattern (Figure 3). An AMOVA finds that the variance is maximized when Pakse is held in a separate cluster from all other sites (F_{CT} = 0.0041, p = 0.0001). All Mantel (Mantel correlation coefficient r = 0.3004, p = 0.2670) and partial Mantel (with Pakse

held in a separate cluster from all other sites; $r = 0.4675$, p = 0.2330) tests were non-significant, indicating that IBD is not a driving factor of diversification in this species.

Helicophagus leptorhynchus

Thirty-nine individuals from Attapeu and 40 from Pakse were successfully sequenced, demultiplexed, mapped to the *de novo* reference, and passed stringent filtering (Table 1). No outlier SNPs were identified by BayeScan, and 531 SNPs were removed from 12 terminal positions on all reads (Supplemental S3), resulting in a panel of 3,768 putatively neutral SNPs (one per contig) to be used in the remaining analyses. Observed heterozygosities for the sites were statistically similar (H_O = 0.29 - 0.30), as were the inbreeding coefficients (G_{IS} = 0.047 - 0.048; Table 1).

As with *Hen. lobatus*, the genetic analyses indicated that the site above Khone Falls was differentiated from the Sekong River site. An AMOVA found there was a significant amount of genetic differentiation among the two locations (F_{ST} = 0.012, $p < 0.001$). Estimates of relatedness revealed six pairs of closely related individuals within sites and no pairs of related individuals among sites (Table 3). Relatedness values for these pairs ranged from 0.21 to 0.93, which is the equivalent of half-siblings to identical. Estimates of N_e for the two locations were similar (N_e = 3.3E2 – 4.0E2; Table 4), but marginally lower at Pakse than at Attapeu. When an individual from all related pairs were removed from Attapeu, the N_e estimate increased to 2.6E3, and when the same was done for the related pair from Pakse, the N_e estimate increased to 4.3E3.

Hel. lephtorhynchus from Attapeu and Pakse were clearly differentiated along both PC1 and PC2 in a principle component analysis (PCA; Figure 2). The STRUCTURE analysis, plotted with a *k* of 2 as chosen by the Evanno method, also distinguished the two sites (Figure 3). Unlike with *Hen. lobatus*, the *Hen. leptorhynchus* from Pakse strongly assigned to a different genetic cluster

than most of those from Attapeu. There was one fish from Attapeu that more strongly assigned to the Pakse cluster than the other cluster, as well as two that assigned to the Pakse cluster in at least 20% of the replicates, suggesting limited admixture. No clustered AMOVA or IBD analyses were completed due to this dataset only having two sampling sites.

Discussion

Genetic structure and Khone Falls

For the first time for both *Henicorhynchus lobatus* and *Helicophagus leptorhynchus*, a genetic analysis in the Lower Mekong Basin has detected population structure. In particular, the findings contrast with our hypothesis and with previous studies of *Hen. lobatus* that did not detect genetic structure in the Mekong Delta (Hurwood *et al.*, 2005, 2007; Iranawati, 2014). The panmixia observed in *Hen. lobatus* among the four sites below Khone Falls, however, does remain consistent with previous surveys, including between the Stung Treng and Tonle Sap sites which were also analyzed by Hurwood *et al.* (2007). The ability to detect neutral genetic differentiation between Pakse and the other locations sampled here is likely due to the improved sampling of the genome provided by RADseq as compared to previous studies using low-throughput methods (Hurwood *et al*., 2007; Iranawati, 2014), rather than recently emergent population structure. High-throughput methods similar to those utilized in this study are increasing in use in the Mekong (Ackiss *et al.*, 2019, Dang *et al.*, 2019) and are needed to measure baseline genetic diversity, population size, and structuring for fishes in this rapidly changing riverscape. Despite our findings of statistically significant differentiation, it is clear from the PCA and STRUCTURE results from both species that some mixing of lineages through Khone Falls is evident (Figures 2, 3). In the case of *Hel. leptorhynchus*, this admixture appears to be

unidirectional, from Pakse to Attapeu which connects to the Lower Mekong, as evidenced by the absence of individuals with intermediate assignment in Pakse. *Helicophagus leptorhynchus* also exhibits two times more partitioning (F_{ST} = 0.010) than *Hen. lobatus* (F_{ST} = 0.0050) between the sites shared by both datasets, which is also evident in the PCA and STRUCTURE analyses (Figures 2, 3). Differences in the magnitude of F_{ST} and pvalues between these datasets may be attributable to varying numbers of SNPs and sample sizes between the two datasets (more SNPs and larger sample sizes in *Hel. leptorhynchus*; Table 1).

Obstruction of gene flow by Khone Falls is a plausible cause of the genetic differentiation measured between Pakse and the other sampling sites in both species (Figure 1). Based upon the pattern of genetic structure across the five sampling sites for *Hen. lobatus*, parsimony dictates that there is a gene flow restriction between Pakse and the Sekong-Mekong confluence. For *Hel. leptorhynchus*, with only two sampling sites, the gene flow restriction lies between Pakse and Attapeu. In both species, Khone Falls is the most obvious natural feature that might restrict gene flow. As the largest series of waterfalls and cascades in Southeast Asia, Khone Falls has previously been hypothesized to be a modulating feature of gene flow in the Mekong River Basin. Indeed, the falls have been shown to restrict migration and gene flow along the main stem Mekong River for relatively small potamodromous fishes such as the sister species of *Hen. lobatus*, the Siamese Mud Carp *Henicorhynchus siamensis* (Adamson *et al.*, 2009; Hurwood *et al.*, 2005) but not for larger fishes such as the Mekong Giant Catfish *Pangasianodon gigas* (Ngamsiri *et al.*, 2007) and some other pangasiids (Chhea, 2000).

Alternatively, there may be another natural impediment to gene flow that is not as obvious as Khone Falls, although it remains unclear as to what this would be. Like its sister species, *Hen. lobatus* is also relatively small, however, mass migrations of *Hen. lobatus* have been well-

documented through sections of Khone Falls via channels such as the Hou Sahong and to a lesser extent the Hou Sadam (Baird *et al*., 2003). These channels are likely to be important migration pathways in maintaining gene flow between the Upper and Lower Mekong for other species and expanding surveys to include additional taxa and sampling sites will be important in determining the extent of this feature as a barrier to gene flow. There is substantial distance between Pakse and Stung Treng or Attapeu via the direct Mekong – Sekong confluence, though there is no evidence of isolation by distance (IBD) in *Hen. lobatus* below Khone Falls. This could be further tested with additional sampling along this corridor. There is also some evidence that until recently the Mekong River maintained an anastomosing connection to the upper Sekong (Workman, 1997), but since this potential connection was diverted, the only main corridor for migration between the Upper and Lower Mekong is through Khone Falls.

At Khone Falls, the recent completion of the Don Sahong dam on the Hou Sahong channel may restrict natural levels of gene flow and have detrimental impacts on migratory fishes known to use this corridor (Fukushima *et al*., 2014). As sampling took place the same year as construction on the dam began, there is no evidence that damming has caused the genetic discontinuities observed by this study for *Hen. lobatus* or *Hel. leptorhynchus* as it is extremely unlikely that genetic drift could impact allele frequencies over such a short time span (Maruyama & Fuerst 1985; Nei *et al*., 1975; Tajima 1989). This does not mean, however, that damming has not affected population connectivity in other locations or cannot affect population connectivity in the future at Khone Falls. The genetic structure we observed along the main stem of the Mekong may become more pronounced after the construction of several planned dams. Impacts on the genetic diversity of these species should be expected both up- and downstream of planned dams

at Stung Treng and Pakse, and these impacts may extend to other basins such as the Mun River in eastern Thailand and the Xe Bang Fai tributary in Lao PDR (Figure 1; Ng & Kottelat, 2000). The effects of dams – including population fragmentation, blocked access to spawning grounds, and altered habitat both up- and downstream – can be particularly perilous for riverine species, which are confined to narrow, often relatively shallow habitat and for whom isolation can lead to local extirpation (Jager *et al*., 2001; Neraas & Spruell, 2001). Many genetic studies have investigated the actual and potential effects of damming on river fish populations with a variety of outcomes (e.g. Ferierra *et al*., 2017; Gouskov *et al*., 2016; Khedkar *et al*., 2014; Ruzich et al., 2019), however, examples from heavily dammed rivers, like the Columbia and Fraser Rivers of northwest North America, have shown that migratory fish populations are particularly vulnerable to negative effects like severe inbreeding and complete extirpation upstream of dams (Ferguson *et al*., 2011). In the 3S Basin, a population of the catfish *Hemibagrus spilopterus* sampled upstream of dams on the Srepok River was found to have diminished effective population size and elevated relatedness and inbreeding (Ackiss *et al.*, 2019).

The apparent panmixia observed in *Hen. lobatus* sampled from sites downstream of Khone Falls and in major tributary systems (3S Rivers and Tonle Sap) indicate strong connectivity despite distances of up to 650 km across varied hydrological features. *Henicorhynchus lobatus* is known to undertake migrations spanning 400 to 1000 km (Baird *et al.* 2003), and previous otolith microchemistry and genetic analyses have provided evidence for strong population connectivity, possibly via a single natal origin (Fukushima *et al.*, 2014; Hurwood *et al.*, 2007; Iranawati, 2014). However, two planned dams threaten to interrupt migration pathways between the sites sampled by this study in the Lower Mekong, the Stung Treng and the Sambor dams (Figure 1), and other dams are planned for or under construction in the region occupied by this species

(International Rivers, 2017). The Stung Treng dam will be erected at the confluence of the 3S Rivers just before they flow into the main stem of the Mekong, threatening to restrict this highly important migratory pathway from the 3S Basin. The Sambor dam planned for the Mekong main stem north of Kratie has caused concern among local advocates and scientists because of its potential to interrupt a major migratory pathway for fishes that move in and out of the Tonle Sap flood plain (International Rivers, 2017). The lake system is vitally important to local fisheries, providing over half of the fisheries production in Cambodia. Without connectivity to the greater Mekong, fish populations in the Tonle Sap region, including the highly valuable *Hen. lobatus*, may be at risk of detrimental effects from fragmentation.

Conservation status and effective population size

The effective population size (N_e) estimates for both species follow an expected pattern and can contribute to our understanding of their conservation status. The smaller, very numerous *Hen. lobatus* (Ngor *et al.*, 2018b) had a much higher N_e (Table 4) than the larger, less common *Hel. leptorhynchus* (Baird *et al.*, 2004). The IUCN Red List extinction risk assessment of *Hen. lobatus* is understandably "Least Concern" because of its present abundance estimates for the Mekong River Basin while *Hel. leptorhynchus* is listed as "Data Deficient" because it is uncommon and lacks population estimates (IUCN Red List, 2020). The 50/500 N_e rule introduced by Franklin (1980) states that N_e should be at least 50 to maintain short-term population viability, and at least 500 for long-term evolutionary viability, though more recent evaluations indicate that these benchmarks are too low (Frankham *et al.*, 2014). The very high N_e estimates for nearly all sites supports the Least Concern conservation status for *Hen. lobatus*. However, the relatively low N_e estimates for this species in general, and especially at Pakse (Table 4), may indicate the need for conservation efforts for this important fisheries species in

certain areas, especially since its population trend is considered declining (IUCN Red List, 2020). These estimates are below or near the 500 N_e rule for long term population viability and therefore indicates that this IUCN Red List Data Deficient species' conservation status warrants new consideration. Unfortunately, no census sizes or N_e/N ratios have been estimated for our two study species, nor for most Mekong fish species, indicating a critical gap in knowledge for understanding their vulnerability to exploitation and damming.

The N_e and conservation status assessments for *Hen. lobatus* and *Hel. leptorhynchus* are precautionary signals for fisheries management. The Mekong River is home to the largest inland fishery in the world, supporting the diets and livelihoods of millions of people. *Henicorhynchus lobatus* is considered the most important fisheries species in the region, composing large percentages in fishing industries, especially in Khone Falls and Tonle Sap River (Baird, 2011). A small minnow species, *Hen. lobatus* catches have remained consistently high as catches for larger fish have decreased with growing fishing pressure in Tonle Sap (Ngor *et al.*, 2018b). The share of fisheries catch from this species is likely to continue to grow as larger fish populations decline, and so informed management is vital, particularly above Khone Falls where N_e sizes may be smaller. Pangasiids are important products of the wing trap fishery in the network of channels within Khone Falls, though *Hel. leptorhynchus* is not one of the most commonly caught species (Baird *et al.*, 2004). As hydroelectric development increases, overall fisheries production is likely to decrease with damming (Ferguson *et al.*, 2011).

The relatively low N_e of *Hen. lobatus* in Pakse and the overall low N_e of *Hel. leptorhynchus* indicates these species may be vulnerable to population fragmentation due to damming. Ziv *et al.* (2012) predict that hydroelectric projects on Mekong tributaries may have an even greater impact on biodiversity than those proposed for the mainstream Mekong due to isolation of

subpopulations in the upper tributaries. Extirpation of local populations has been observed in the Columbia River system following the blocking of upstream spawning grounds by dams (Ferguson *et al*., 2011) and freshwater fish are approximately four times more likely than marine to be classified as vulnerable to extinction (Olden *et al*., 2007). Incipient extirpation due to damming has been detected in at least one species of economically important catfish in the Mekong Basin (Ackiss *et al.*, 2019).

The presence of putative related individuals in both *Hen. lobatus* (at Pakse) and *Hel. leptorhynchus* (at Pakse and Attapeu) suggests that individuals at these sites are unable to consistently disperse as far as their life histories likely allow (Chhea, 2000). Inbreeding coefficients (G_{IS}) in both species, however, are not excessively high (Table 1), so any potential bottleneck events would have been very recent and not severe enough to cause substantial inbreeding. However, the low N_e estimates for *Hen. lobatus* above Khone Falls may also be a function of restricted gene flow, thereby indicating that further decreases in migration due to damming could lead to increased inbreeding and alter the long-term viability of this species. Continued sampling is needed to confirm if populations are subdivided in the upper tributaries due to existing dams and other infrastructure, and if a planned dam on the Sekong River (Figure 1) may result in new isolation of upstream populations. Low levels of gene flow should be maintained to ensure genetic diversity in populations in the upper 3S Rivers and prevent severe inbreeding and local extinction in the tributaries (Ferguson *et al*., 2011). These precautions are particularly relevant since future effects from climate change, human population growth, and increased exploitation of aquatic resources are unpredictable and therefore planning for aquatic infrastructure requires special attention (Hoang *et al*., 2016; Pech and Sunada, 2008; Syvitski *et al*., 2009). *Hen. lobatus* and *Hel. leptorhynchus* are representatives of two of the most

economically important and species-rich fish families found in the Mekong Basin (Ngor *et al*., 2018a) and the usefulness of high throughput sequencing to inform population structure and management precautions indicated for these species, especially regarding potential for fragmentation due to damming, are likely to be relevant to many other species in the region.

Acknowledgements

This project was funded by the United States Agency for International Development supported Partnerships for Enhanced Engagement in Research Project 2-7 under USAID Cooperative Agreement AID-OAA-A-11-00012 awarded to B. T. Dang and the National Science Foundation grant OISE-0730256 awarded to K. E. Carpenter. The research was administered by the National Academy of Sciences. Base maps were created by M. Harvey of the IUCN Global Marine Species Assessment group at Old Dominion University, Norfolk, VA. We thank T. Warren for vital help in collections in Lao PDR, and Hung T. K. Le for supporting sampling in Cambodia. Analyses were conducted on the ODU Turing HPC cluster, and assistance was provided by M. Dong, J. Pratt, and T. Stillwell. Thanks also to B. Stockwell and S. Furiness-Magnuson for lab and data analysis work.

Supporting Information

Supplemental S1. The dDocentHPC configuration file containing all settings used to process the FASTQ files, create the reference genomes, and call variants.

Supplemental S2. Vcftools, vcflib, and custom SNP filtering commands for (A) *Henicorhynchus lobatus* and (B) *Helicophagus leptorhynchus*.

Supplemental S3. Scatterplots of PC2 versus PC1 for (A) *Henicorhynchus lobatus* and (B) *Helicophagus leptorhynchus* prior to removal of terminal positions. Each point represents a single fish. Colors represent sequencing pool. Ellipses are the 50% confidence intervals of the centroid. Barplots for the first 50 eigenvalues are displayed in the bottom right corner of each plot. Histograms depict the number of contigs (y-axis) with SNPs at a given position on the contig (x-axis) for (C) *Henicorhynchus lobatus* and (D) *Helicophagus leptorhynchus*. Red bars indicate positions that were removed from consideration across all contigs. Scatterplots of PC2 versus PC1 for (E) *Henicorhynchus lobatus* and (F) *Helicophagus leptorhynchus* after removal of terminal positions. Again, each point represents a single fish, colors represent sequencing pool, and ellipses are the 50% confidence intervals of the centroid. Barplots for the first 50 eigenvalues are displayed in the bottom right corner of each plot.

Supplemental S4. Scatterplots of PC2 versus PC1 for (A) *Henicorhynchus lobatus* and (B) *Helicophagus leptorhynchus* without highly related individuals. Each point represents a single fish. Colors represent sampling site. Ellipses are the 50% confidence intervals of the centroid. Barplots for the first 20 eigenvalues are displayed in the bottom right corner of each plot.

Supplemental S5. Table of log likelihood (LnP) and ΔK values as calculated by STRUCTURE Harvester for (A) *Henicorhynchus lobatus* and (B) *Helicophagus leptorhynchus*. Results are presented for each K value tested in STRUCTURE and the K values chosen by the Evanno method (Evanno *et al.*, 2005) are bolded.

Contributions

E. E. B. conducted data analyses, created tables and figures, and wrote the majority of the manuscript.

B. T. D. procured funding for the research, collected samples from the field, coordinated DNA extractions, and reviewed the manuscript.

A. S. A. contributed to data analyses, made figures, and assisted writing the manuscript.

C. E. B. coordinated the library preparation, designed the sample pools for sequencing, and assisted writing the manuscript.

P. C. organized the sample collection expedition in Cambodia and reviewed the manuscript.

L. P. collected samples from the field in Lao PDR and reviewed the manuscript.

O. T. T. collected samples from the field, conducted DNA extractions, and reviewed the manuscript.

K. E. C. designed the sampling strategy, collected samples from the field, procured funding for the research, and assisted writing the manuscript.

References

Ackiss, A. S., Dang, B. T., Bird, C. E., Biesack, E. E., Chheng, P., Phounvisouk, L., Vu, Q. H., Uy, S., & Carpenter, K. E. (2019). Cryptic lineages and a population dammed to incipient extinction? Insights into the genetic structure of a Mekong River catfish. *Journal of Heredity*, **110**, 535-547

Adamson, E. A., Hurwood, D. A., Baker, A. M., & Mather, P. B. (2009). Population subdivision in Siamese mud carp *Henicorhynchus siamensis* in the Mekong River basin: implications for management. *Journal of Fish Biology*, **75**, 1371-1392

Amornsakchai, S., Annez, P., Vongvisessomjai, S., Choowaew, S., Thailand Development Research Institute (TDRI), Kunurat, P., Nippanon, J., Schouten, R., Sripapatrprasite, P., Vaddhanaphuti, C., Vidthayanon, C., Wirojanagud, W., & Watana, E. (2000). *Pak Mun*

Dam, Mekong River Basin, Thailand. A World Commission on Dams (WCD) Case Study prepared as an input to the World Commission on Dams. World Commission on Dams, Cape Town, South Africa. Available at: http://www2.centre-cired.fr/IMG/pdf/F8_PakMunDam.pdf

Arias, M. E., Cochrane, T. A., Norton, D., Killeen, T. J., & Khon, P. (2013). The flood pulse as the underlying driver of vegetation in the largest wetland and fishery of the Mekong Basin. *Ambio*, **42**, 864-76

Baird I.G. (2001). Aquatic Biodiversity in the Siphandone Wetlands. In G. Daconto (Ed.), *Siphandone Wetlands* (pp. 61-74). Bergamo, Italy: CESVI

Baird, I. G. (2011). The Don Sahong Dam. *Critical Asian Studies*, **43**, 211-235

Baird, I. G. & Flaherty, M. S. (2004). Beyond national borders: important Mekong River medium sized migratory carps (Cyprinidae) and fisheries and Laos and Cambodia. *Asian Fisheries Science*, **17**, 279-298

Baird, I. G., Flaherty, M. S., & Phylavanh, B. (2003). Rhythms of the river: lunar phases and migrations of small carps (Cyprinidae) in the Mekong River. *Natural History Bulletin of the Siam Society*, **51**, 5-36

Baird, I. G., Flaherty, M. S., & Phylavanh, B. (2004). Mekong River Pangasiidae catfish migrations and Khone Falls wing trap fishery in southern Laos. *Natural History Bulletin of the Siam Society*. **52**, 81-109

Baran, E., So, N., Arthur, R., Leng, S. V., & Kura, Y. (2006). *Bioecology of 296 fish species of the Tonle Sap Great Lake (Cambodia)*. Report for the project "Study of the Influence of Built Structures on the Fisheries of the Tonle Sap". WorldFish Center, Phnom Penh, Cambodia. 167 pp

Baran, E., Baird, I. G., & Cans, G. (2005). *Fisheries bioecology at the Khone Falls (Mekong River, Southern Laos)*. Worldfish Center, Phnom Penh, Cambodia. 84 pp

Baran, E., Samadee, S. A., Jiau, T. S., & Cong, T. T. (2013). *Fish and fisheries in the Sekong, Sesan and Srepok Basins (3S Rivers, Mekong Watershed), with special reference to the Sesan River*. WorldFish Center, Phnom Penh, Cambodia. Available at: http://www.optimisingcascades.org/wp-content/uploads/2013/11/FE2Sample-paper.pdf

Campbell, I. C. (2009). *The Mekong. Biophysical environment of a transboundary river*. New York, NY: Elsevier Publishers.

Chhea, C. K. (2000). *Fisher's knowledge about migration patterns of three important Pangasius catfish species in the Mekong mainstream*. Mekong River Commission. Available at: http://info.mrcmekong.org/assets/midocs/0003155-biota-fisher-s-knowledge-about-migration-patterns-of-three-important-pangasius-catfish-species-in-the-mekong-mainstream.pdf

Chong, Z., Ruan, J., & Wu, C. I. (2012). Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, **28**, 2732-2737

Danecek, P., Auton. A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158

Dang, B. T., Vu, Q. H., Biesack, E. E., Doan, T. V., Truong, O. T., Tran, T. L., Ackiss, A. S., Stockwell, B. L., & Carpenter, K. E. (2019). Population genomics of the peripheral freshwater fish *Polynemus melanochir* (Perciformes, Polynemidae) in a changing Mekong Delta. *Conservation Genetics*, **20**, 961-972

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: reimplementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, **14**, 209-214

Earl, D. A. & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491

Fan, H., He, D., & Wang, H. (2015). Environmental consequences of damming the mainstream Lancang-Mekong River: a review. *Earth-Science Reviews*, **146**, 77-91

Ferguson, J. W., Healey, M., Dugan, P., & Barlow, C. (2011). Potential effects of dams on migratory fish in the Mekong River: lessons from salmon in the Fraser and Columbia Rivers. *Environmental Management*, **47**, 141-159

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., Young, G., Fennell, T. J., Allen, A., Ambrogio, L., Berlin, A. M., Blumenstiel, B.,Cibulskis, K., Friedrich, D., Johnson, R., Juhn, F., Reilly, B., Shammas, R., Stalker, J., Sykes, S. M., Thompson, J., Walsh, J., Zimmer, A., Zwirko, Z., Gabriel, S., Nicol, R., & Nusbaum, C. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology*, **12**, R1

Foll, M. & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977-993

Frankham, R., Bradshaw, C.J., & Brook, B.W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, **170**, 56-63

Franklin, I. R. (1980). Evolutionary change in small populations. In M. E. Soulé & B. A. Wilcox (Eds.), *Conservation biology: an evolutionary-ecological perspective* (pp. 135–149). Sunderland, MA: Sinauer Associates Inc.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150-3152

Fukushima, M., Jutagate, T., Grudpan, C., Phomikong, P., & Nohara, S. (2014). Potential effects of hydroelectric dam development in the Mekong River Basin on the migration of Siamese Mud Carp (*Henicorhynchus siamensis* and *H. lobatus*) elucidated by otolith microchemistry. *PloS One*, **9**, e103722

Garrison, E. & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*

Goslee, S. C. & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1-19

Gouskov, A., Reyes, M., Wirthner-Bitterlin, L., & Vorburger, C. (2016). Fish population genetic structure shaped by hydroelectric power plants in the upper Rhine catchment. *Evolutionary Applications*, **9**, 394-408

Hoang, L. P., Lauri, H., Kummu, M., Koponen, J., van Vliet, M. T. H., Supit, I., Leemas, R., Kabat, P., & Ludwig, F. (2016). Mekong River flow and hydrological extremes under climate change. *Hydrology and Earth System Sciences*, **20**, 3027-3041

Hogan, Z., Baird, I. G., Radtke, R., & Vander Zanden, M. J. (2004). Long distance migration and marine habitation in the tropical Asian catfish, *Pangasius krempfi*. *Journal of Fish Biology*, **71**, 818-832

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332

Hurwood, D. A., Adamson, E. A., & Mather, P. B. (2005). Identifying stock structure of two *Henicorhynchus* species in the Mekong River using mitochondrial DNA. *Proceedings of 7th Technical Symposium on Mekong Fisheries*, 253-263

Hurwood, D. A., Adamson, E. A., & Mather, P. B. (2007). Evidence for strong genetic structure in a regionally important, highly vagile cyprinid (*Henicorhynchus lobatus*) in the Mekong River Basin. *Ecology of Freshwater Fish*, **17**, 273-283

International Rivers (2017). Factsheet: A Dangerous Trajectory for the Mekong River. Available at: https://www.internationalrivers.org/sites/default/files/attached-files/mekongmainstreamdamsupdatejune2017_english.pdf

Iranawati, F. (2014). *An assessment of the geographical scale of recurrent gene flow in wild populations of two species of Mekong River carps (*Henicorhynchus *spp.)* (Doctoral dissertation, Queensland University of Technology).

Jager, H. I., Chandler, J. A., Lepla, K. B., & Van Winkle, W. (2001). A theoretical study of river fragmentation by dams and its effects on white sturgeon populations. *Environmental Biology of Fishes*, **60**, 347-361

Jakobsson, M. & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics,* **23**, 1801-1806

Jombart, T. & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070-3071

Junk, W. J., Bayley, P. B., & Sparks, R. E. (1989). The flood pulse concept in river-floodplain systems. *Canadian Special Publication of Fisheries and Aquatic Sciences,* **106**, 110-27

Khedkar, G. D., Jamdade, R., Kalyankar, A., Tiknaik, A., Ron, T. B., & Haymer, D. (2014). Genetic fragmentation in India's third longest river system, the Narmada. *SpringerPlus*, **3**, 385

Li, D., Long, D., Zhao, J., Lu, H., & Hong, Y. (2017). Observed changes in flow regimes in the Mekong River basin. *Journal of Hydrology*, **551**, 217-232

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*:*1303.3997*

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079

Lischer, H. E. L. & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298-299

Maruyama, T. & Fuerst, P. A. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, **111**, 675-689

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28-41

Meirmans, P. G. & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792-794

Mekong River Commission. (1997). Mekong River basin diagnostic study, final report. *Mekong River Commission*. Bangkok, Thailand.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10

Neraas, L. P. & Spruell, P. (2001). Fragmentation of riverine systems: the genetic effects of dams on bull trout (*Salvelinus confluentus*) in the Clark Fork River system. *Molecular Ecology*, **10**, 1153-1164

Ng, H. H. & Kottelat, M. (2000). *Helicophagus leptorhynchus*, a new species of molluscivorous catfish from Indochina (Teleostei: Pangasiidae). *Raffles Bulletin of Zoology*, **48**, 55-58

Ngamsiri, T., Nakajima, M., Sukmanomon, S., Sukumasavin, N., Kamonrat, W., NANAKORN University, & Taniguchi, N. (2007). Genetic diversity of wild Mekong giant catfish *Pangasianodon gigas* collected from Thailand and Cambodia. *Fisheries Science*, **73**, 792-799

Ngor, P. B., Legendre, P., Oberdorff, T. & Lek, S. (2018a). Flow alterations by dams shaped fish assemblage dynamics in the complex Mekong-3S river system. *Ecological Indicators* **88,** 103-114

Ngor, P. B., McCann, K., Grenouillet, G., So, N, McMeans,. B. C., Fraser, E, & .Lek, S. (2018b) Evidence of indiscriminate fishing effects in one of the world's largest inland fisheries. *Scientific Reports*, **8**, 8947

Olden, J.D., Hogan, Z.S., & Zanden, M.J.V. (2007). Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. *Global Ecology and Biogeography*, **16**, 694-701

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, **27**, 3193-3206

Orr, S., Pittock, J., Chapagain, A., & Dumaresq, D. (2012). Dams on the Mekong River: Lost fish protein and the implications for land and water resources. *Global Environmental Change*, **22**, 925-932

Pech, S. & Sunada, K. (2008). Population growth and natural-resources pressures in the Mekong River Basin. *AMBIO: A Journal of the Human Environment*, **37**, 219-224

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, **7**, e37135

Piman, T., Cochrane, T.A., & Arias, M.E. (2016). Effect of proposed large dams on water flows and hydropower production in the Sekong, Sesan and Srepok Rivers of the Mekong Basin. *River Research and Applications*, **32**, 2095-2108

Poulsen, A., & Valbo-Jørgensen, J. (Eds.). (2001). *Fish migrations and spawning habits in the Mekong mainstream: a survey using local knowledge.* DeKalb, IL: Cultivate Understanding Multimedia

Poulsen, A. F., Poeu, O., Viravong, S., Suntornratana, U., & Tung, N. T. (2002). Fish migrations of the lower Mekong River Basin: implications for development, planning and environmental management. MRC Technical Paper No. 8, Mekong River Commission, Phnom Penh. 62 pp. ISSN: 1683-1489

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959

Puritz, J., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e314v1

Rainboth, W. J. (1996). *Fishes of the Cambodian Mekong*. FAO Species Identification Field Guide for Fishery Purposes. FAO, Rome.

Roberts, T. R. & Baird, I. G. (1995). Traditional fisheries and fish ecology on the Mekong River in Khone Waterfalls in southern Laos. *Natural History Bulletin of Siam Society*, **43**, 219-262

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Resources*, **4**, 137-138

Ruzich, J., Turnquist, K., Nye, N., Rowe, D., & Larson, W.A. (2019). Isolation by a hydroelectric dam induces minimal impacts on genetic diversity and population structure in six fish species. *Conservation Genetics*, **20**, 1421-1436

Syvitski, J. P., Kettner, A. J., Overeem, I., Hutton, E. W. H., Hannon, M. T., Brakenridge, G. R., Day, J., Vörösmarty, C., Saito, Y., Gosan, L., & Nicholls, R. J. (2009). Sinking deltas due to human activities. *Nature Geoscience*, **2**, 681

Ta, T. K., Nguyen, V. L., Tateishi, M., Kobayashi, I., Tanabe, S., & Saito, Y. (2002). Holocene delta evolution and sediment discharge of the Mekong River, southern Vietnam. *Quaternary Science Reviews*, **21**, 1807-1819

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597-601

Valbo-Jørgensen, J., Coates, D., & Hortle, K. G. (2009). Fish diversity in the Mekong River basin. In I. Campbel (Ed.), *The Mekong: Biophysical Environment of an International River Basin* (pp. 161-196). Amsterdam: Elsevier Publishers

Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153-1167

Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, **89**, 135-153

Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Gold, J. R. & Portnoy, D. S. (2017). Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage. *Molecular Ecology Resources*, **17,** 955–965

WLE Greater Mekong. (2016). *Dataset on the Dams of the Irrawaddy, Mekong, Red and Salween River Basins. Vientiane, Lao PDR.* CGIAR Research Program on Water, Land and Ecosystems (Greater Mekong). Available at: https://wle-mekong.cgiar.org/maps/. (last accessed August 2017).

Workman, D. R. (1997). *Laos*. In: Encyclopedia of European and Asian Regional Geology. Dordrecht, Netherlands: Springer.

Ziv, G., Baran, E., Nam, S., Rodríguez-Iturbe, I., & Levin, S. A. (2012). Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. *Proceedings of the National Academy of Sciences*, **109**, 5609-5614

Figure 1. Map of the Mekong River Basin. Circles and triangles indicate *Henicorhynchus lobatus* and *Helicophagus leptorhynchus* sampling locations, respectively. Dams, both planned (marked by **) or under construction at the time of sampling, are labeled in smaller black font with the megawatt hours (International Rivers, 2017) which are indicators of the amount of water restricted by the dams. The Mekong River is indicated by black, while the other rivers and tributaries are grey.

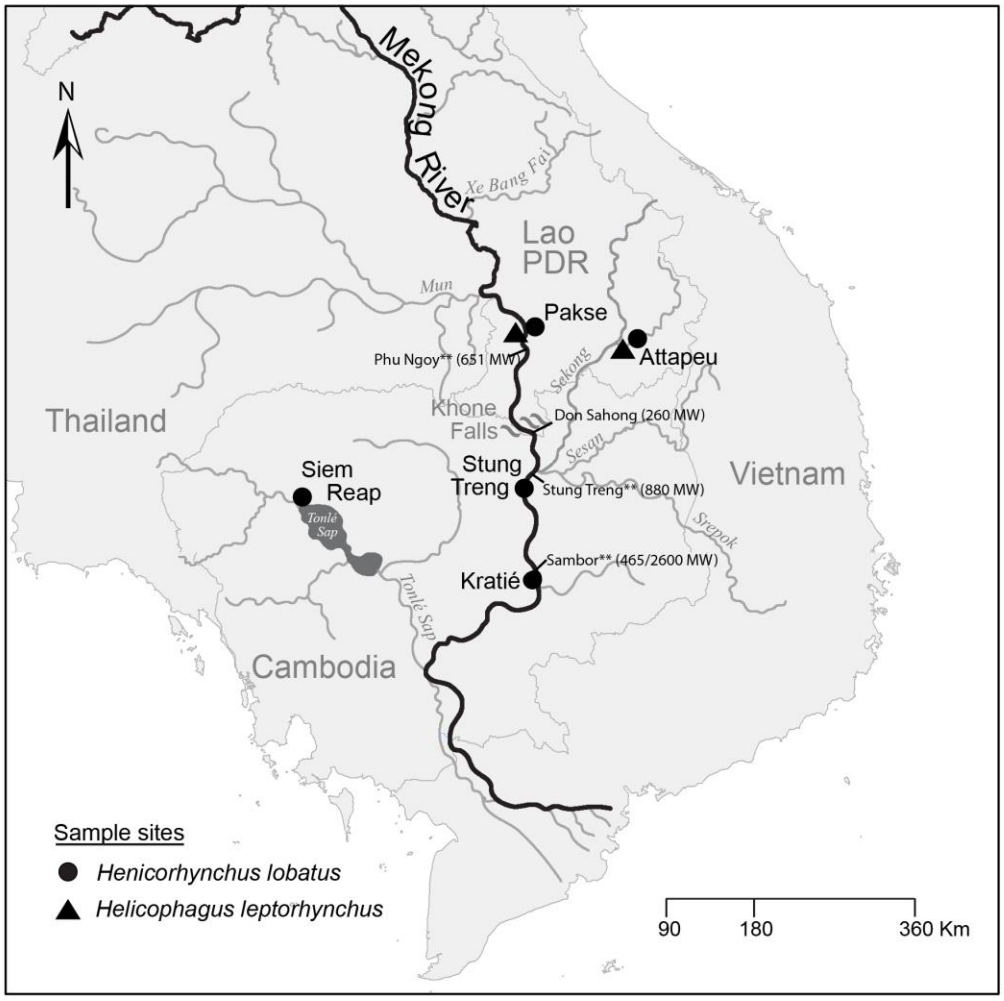

Figure 2. Scatterplots of PC2 versus PC1 for (B) *Henicorhynchus lobatus* and (A) *Helicophagus leptorhynchus*. Each point represents a single fish. Colors represent sampling site. Ellipses are the 50% confidence intervals of the centroid. Barplots for the first 20 eigenvalues are displayed in the bottom right corner of each plot.

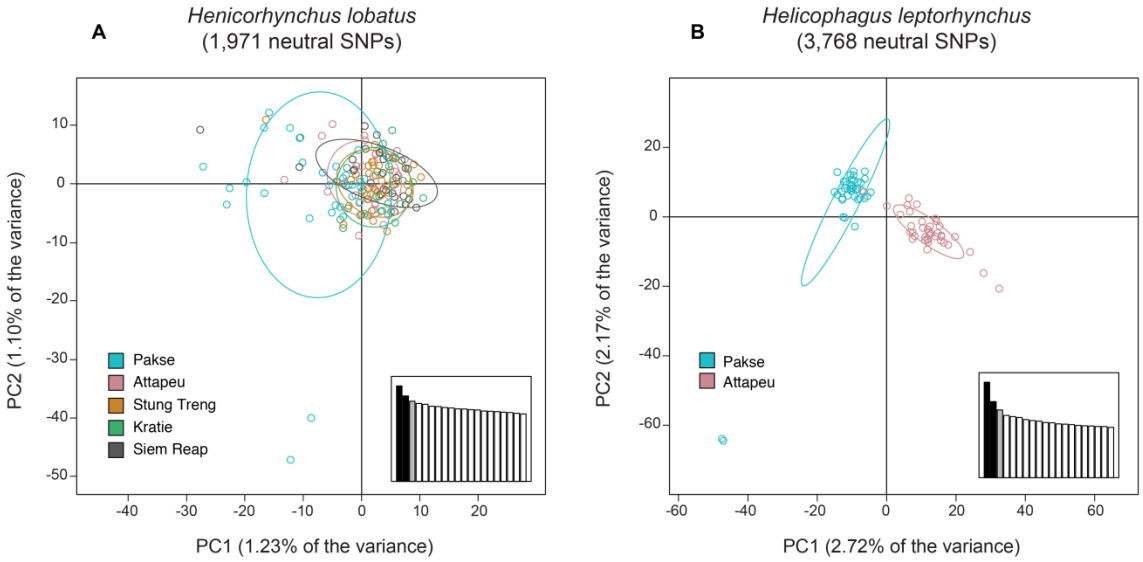

Figure 3. STRUCTURE barplots for (A) *Henicorhynchus lobatus* and (B) *Helicophagus leptorhynchus*. Bars depict the probability of assignment (y axis) of each individual fish (x axis) to each genetic cluster (color). *K* is the number of genetic clusters, chosen as two by the Evanno method (Evanno *et al.*, 2005).

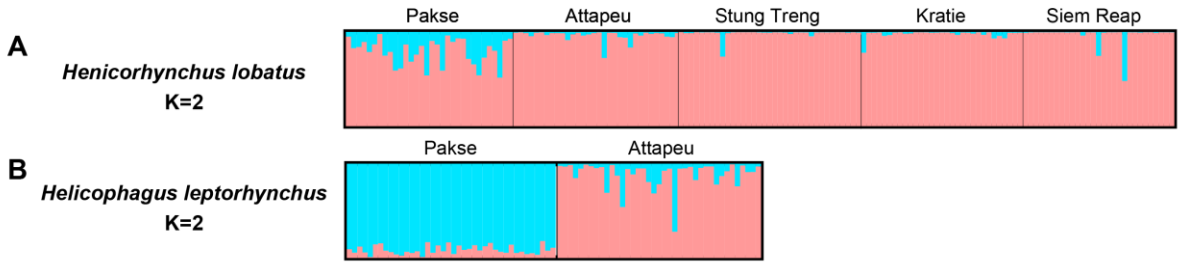

Table 1. Summary statistics by site, including numbers of individuals collected, sequenced, and analyzed, for *Henicorhynchus lobatus* and *Helicophagus leptorhynchus*. H_O is observed heterozygosity. H_e is expected heterozygosity. G_{IS} is the inbreeding coefficient.

Henicorhynchus lobatus

Site	Collected	Sequenced	Analyzed	H_O	H_S	G_{IS}
Lao PDR						
Attapeu	58	45	32	0.2630	0.3041	0.1353
Pakse	54	41	32	0.2645	0.3051	0.1332
Cambodia						
Kratie	54	42	31	0.2648	0.3016	0.1221
Siem Reap	60	40	29	0.2691	0.3047	0.1168
Stung Treng	60	46	35	0.2628	0.3055	0.1398

Helicophagus leptorhynchus

Site	Collected	Sequenced	Analyzed	H_O	H_S	G_{IS}
Lao PDR						
Attapeu	64	46	39	0.3023	0.3173	0.0471
Pakse	56	48	40	0.2940	0.3088	0.0478

Table 2. Pairwise comparisons for genetic differentiation among sites for *Henicorhynchus lobatus*. Above the diagonals are F_{ST} values and below the diagonals are *p* values. Significant genetic differentiation is indicated by bold font.

Henicorhynchus lobatus	Attapeu	Kratie	**Pakse**	Siem Reap	Stung Treng
Attapeu	--	0.0002	**0.0045**	0.0004	0.0001
Kratie	0.3667	--	**0.0028**	-0.0003	0.0007
Pakse	**0.0001**	**0.0001**	--	**0.0045**	**0.0050**
Siem Reap	0.2699	0.6695	**0.0001**	--	0.0004
Stung Treng	0.4669	0.1235	**0.0001**	0.2680	--

Table 3. Pairs of individuals with a triadic relatedness with confidence intervals above or including 0.25 for *Henicorhynchus lobatus* and *Helicophagus leptorhynchus*. Columns contain the identities of individuals, their source location (Pk = Pakse; At = Attapeu), and the relatedness estimates with 95% confidence intervals.

Henicorhynchus lobatus

Individual 1	Individual 2	Group	trioml r (95% CIs)
Pk_Hlo018	Pk_Hlo040	Pk-Pk	0.4580 (0.4067 - 0.5135)

Helicophagus leptorhynchus

Individual 1	Individual 2	Group	trioml r (95% CIs)
Pk_Hle210	Pk_Hle222	Pk-Pk	0.9296 (0.9047-0.9514)
At_Hle314	At_Hle316	At-At	0.8495 (0.8061-0.8916)
At_Hle276	At_Hle309	At-At	0.2568 (0.2212-0.2944)
At_Hle299	At_Hle319	At-At	0.2532 (0.224-0.2918)
At_Hle295	At_Hle298	At-At	0.2526 (0.2183-0.291)
At_Hle274	At_Hle289	At-At	0.2111 (0.1752-0.2523)

Table 4. Estimates of effective population size (N_e) with 95% confidence intervals for *Henicorhynchus lobatus* and *Helicophagus leptorhynchus*. Below Khone Falls is composed of all locations except Pakse because it was genetically differentiated. Also included are analyses in which one individual from each sibling pair was removed, indicated by "w/o related pair(s)".

Henicorhynchus lobatus

Population		**N_e (95% CI)**
Lao PDR		
	Attapeu	10687.8 (3087.3 - Infinite)
	Pakse	642.4 (558.7 - 755.2)
	w/o related pair	2423.2 (1493.9 - 6369.4)
Cambodia		
	Stung Treng	13287.1 (3569.6 - Infinite)
	Kratie	432369.8 (4213.8 - Infinite)
	Siem Reap	37347.8 (3390.2 - Infinite)
Below Khone Falls		Infinite (25688.0 - Infinite)

Helicophagus leptorhynchus

Population		**N_e (95% CI)**
Lao PDR		
	Attapeu	396.7 (382.3 - 412.3)
	w/o related pairs	2551.3 (1995.9 - 3532.0)
	Pakse	331.9 (321.5 - 342.9)
	w/o related pair	4251.7 (3023.8 - 7148.5)