3-19-2019

Cryptic Lineages and a Population Damned to Incipient Extinction? Insights into the Genetic Structure of a Mekong River Catfish

Amanda S. Ackiss
Old Dominion University, aacki002@odu.edu

Binh T. Dang

Christopher E. Bird

Ellen E. Biesack
Old Dominion University, ebies001@odu.edu

Phen Chheng

See next page for additional authors

Original Publication Citation

Ackiss, A. S., Dang, B. T., Bird, C. E., Biesack, E. E., Chheng, P., Phounvisouk, L., ... 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*. doi:10.1093/jhered/esz016

Cryptic lineages and a population dammed to incipient extinction? Insights into the genetic structure of a Mekong River catfish

Amanda S. Ackiss[1], Binh T. Dang[2], Christopher E. Bird[3], Ellen E. Biesack[1], Phen Chheng [4], Latsamy Phounvisouk [5], Quyen H.D. Vu[2], Sophorn Uy[4], Kent E. Carpenter[1]

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

[2]Department of Biology, Institute of Biotechnology and Environment, Nha Trang University, 02 Nguyen Dinh Chieu, Nha Trang, Vietnam

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

[4]Inland Fisheries Research and Development Institute (IFReDI), Fisheries Administration, Phnom Penh, Cambodia

[5]Living Aquatic Resources Research Center, Nong Thang Village, Vientiane, Lao PDR

Corresponding authors: Kent E. Carpenter & Amanda S. Ackiss

Emails: kcarpent@odu.edu, aackiss@uwsp.edu

Abstract

An understanding of the genetic composition of populations across management boundaries is vital to developing successful strategies for sustaining biodiversity and food resources. This is especially important in ecosystems where habitat fragmentation has altered baseline patterns of gene flow, dividing natural populations into smaller sub-populations and increasing potential loss of genetic variation through genetic drift. River systems can be highly fragmented by dams built for flow regulation and hydropower. We used reduced-representation sequencing to examine genomic patterns in an exploited catfish, *Hemibagrus spilopterus*, in a hotspot of biodiversity and hydropower development- the Mekong River basin. Our results revealed the presence of two highly-divergent coexisting genetic lineages which may be cryptic species. Within the lineage with the greatest sample sizes, pairwise F_{ST} values, principal components analysis, and a STRUCTURE analysis all suggest that long-distance migration is not common across the Lower Mekong Basin, even in areas where flood-pulse hydrology has limited genetic divergence. In tributaries, effective population size estimates were at least an order of magnitude lower than in the Mekong mainstream indicating these populations may be more vulnerable to perturbations such as human-induced fragmentation. Fish isolated upstream of several dams in one tributary exhibited particularly low genetic diversity, high amounts of relatedness, and a level of inbreeding (G_{IS} = 0.51) that has been associated with inbreeding depression in other outcrossing species. Our results highlight the importance of assessing genetic structure and diversity in riverine fisheries populations across proposed dam development sites for the preservation of these critically-important resources.

Keywords: conservation genetics, genomics, spatial genetic structure, effective population size, inbreeding, RAD sequencing

Introduction

Preserving the evolutionary potential of species in the face of a rapidly changing world is a major goal for resource managers and conservation geneticists (Hedrick & Miller, 1992; Lankau, Jørgensen, Harris, & Sih, 2011; Pertoldi, Bijlsma, & Loeschcke, 2007; Petit, Mousadik, & Pons, 1998; Ryman, 1991). The ability to adapt and succeed in new environmental conditions is rooted in a species' underlying genetic variation. Arising from random mutation and persisting through the processes of selection and drift, variation in the genetic architecture of a species provides the building blocks of evolution as individuals with beneficial alleles survive to pass their genes on to offspring.

One of the biggest threats to genetic variation is human-induced habitat fragmentation (DiBattista, 2008; Gibbs, 2001; Young, Boyle, & Brown, 1996). The division of natural populations into subpopulations can result in partial or complete isolation between subpopulations and greatly reduced subpopulation sizes. If a subsequent subpopulation is small enough, drift will remove variation faster than it can be introduced by random mutation, resulting in declining genetic diversity, increasing homozygosity, and inbreeding depression (Charlesworth & Willis, 2009; Ralls, Brugger, & Ballou, 1979; Reed, Nicholas, & Stratten, 2007).

Freshwater and diadromous fishes of lotic ecosystems are particularly vulnerable to fragmentation since the causes - including weirs and hydroelectric dams - often result in complete isolation between populations above and below these barriers. Flow modification is considered a major threat to global freshwater biodiversity (Dudgeon et al., 2006), and the installation of tens of thousands of dams in river systems over the past century resulted in large-scale disruption of migratory fish movement and subdivision of resident fish populations. Damming a river can remove hundreds of miles of fluvial habitat from a migratory species' range and often leads to the localized extinction of diadromous species upstream of an impassable barrier (Gerke, Gilligan, & Barwickm 2002; Morita & Yamamoto, 2002). Both expectations and observations of the detrimental impacts this fragmentation has on exploited fish populations contributed to the recent rise in dam removal in many countries (Bellmore, Duda, Craig, Greene, Torgersen, Collin & Vittum, 2017; Doyle, Stanley, Harbor, & Grant, 2003; O'Connor, Duda, & Grant, 2015). However, whether a dam will actually have a negative effect on subdivided resources is difficult to predict. Many studies examining the influence of a dam on the genetic composition of populations on either side have found no major genetic differentiation or only weak differentiation and no effect on genetic diversity (Diener, Garza, Coey,

& Girman, 2007; Reid, Wilson, Mandrak, & Carl, 2008), making the evolutionary impacts of these barriers on fishery populations difficult to predict.

The Mekong River is one of the largest and most biodiverse freshwater ecosystems in the world and is currently experiencing a period of rapid hydropower development. The portion of this river system that runs through Myanmar, Lao PDR, Thailand, Cambodia, and Viet Nam is known as the Lower Mekong Basin (LMB), and estimates over the last few decades put this drainage as supporting anywhere from 1000-1700 species of estuarine and freshwater fish (Rainboth, 1996; Sverdrup-Jensen, Bishop, Clayton, Barlow, & Mekong River Commission, 2002). These fishes are a vital protein source for more than 60 million people living in the watershed (Mekong River Commission, 2010), and like other large river systems, habitat fragmentation has become a leading threat to the fisheries of the LMB as the demand for hydroelectric energy has increased. The past three decades have seen the construction of dozens of dams across the Mekong mainstream and its tributaries, and over 100 additional proposals are under consideration (Barlow, Baran, Halls, & Kshatriya, 2008; Dugan et al., 2010; Li & He, 2008). During the onset of hydropower development in this region, consultants assessing the impacts of 10 proposed dams on Mekong fishes concluded that biological data were insufficient to evaluate or prioritize among sites (Hill & Hill, 1994), and, currently, assessing the impacts of hydropower on aquatic habitats and species in the LMB is considered one of the highest research priorities for the region (Beilfuss & Triet, 2014). Generating these data is critical to the proper protection of biodiversity and food resources as evidence indicates potentially catastrophic impacts as a result of hydropower development (Ziv, Baran, Nam, Rodriguez-Iturbe, & Levin, 2012).

The LMB is characterized by complex hydrological patterns with great potential to influence the natural levels of gene flow between populations. The largest waterfall in Southeast Asia, the Khone Falls (Fig. 1), is a 40 km wide series of falls and cascades on the Mekong mainstream that is only passable by certain species such as migratory pangasiid catfishes (Baird, Flaherty, & Phylavanh, 2004) and is likely a strong barrier to upstream movement for many species (Sverdrup-Jensen et al., 2002). South of the Khone Falls, the Mekong River experiences strong influences from two systems that may enhance gene flow in this region. The three tributaries of the '3-S' basin - the Sekong, Sesan, and Srepok rivers - make up the largest hydrological input into the LMB and provide over 25% of the mean annual flow to Kratié (Mekong River Commission 2005). South of Kratié is the Mekong's connection to the largest freshwater lake in Southeast Asia, the Tonlé Sap, which is characterized by a flood-pulse

system (Junk, Bayley, & Sparks, 1989). During the wet season, the majority of the inflow into the lake (57%) comes from the Mekong mainstream. When the Mekong recedes in the dry season, this flow reverses and approximately 88% of the water volume of the Tonlé Sap returns to the Mekong (Arias, Cochrane, Norton, Killeen, & Khon, 2013; MRCS/WUP-FIN 2006). Understanding the baselines of genetic structure and variation in species inhabiting this hydrologically diverse riverscape is essential to correctly predicting and interpreting their responses to human-induced fragmentation.

Despite the scale of biodiversity and fisheries within the LMB, only a small number of mitochondrial- and microsatellite-based genetic analyses have been published that examine the genetic composition of populations across this complex landscape. Results indicate a range of connectivity in the LMB from highly structured populations (Adamson, Hurwood, Baker, & Mather, 2009; Adamson, Hurwood, & Mather, 2012; Hurwood, Adamson, & Mather, 2007; Tagaki et al., 2006) to panmictic (Ngamsiri et al., 2007; So, Van Houdt, & Volckaert, 2006), highlighting a critical need for more genetic studies across a broad taxonomic range to aid our understanding of how both natural and man-made hydrological features influence populations of fisheries species in the LMB.

In terms of value and volume of wild-caught fishes, the catfish *Hemibagrus spilopterus* (Ng & Rainboth, 1999) consistently ranks as one of the most important fisheries resources in the LMB, but next to nothing is known about its population structure throughout this region. A survey in Lao PDR indicated that *H. spilopterus* was one of 15 species of fish that comprised 90% of the total fish abundance surveyed (Hortle et al., 2015), and a year-long survey in Cambodia placed *H. spilopterus* within the top ten species and groups accounting for the largest volume and value of fisheries in the LMB (Mille, Hap, & Loeng, 2016). Currently, the only published genetic study that includes *H. spilopterus* is a phylogenetic evaluation of the genus (Dodson & Lecomte, 2015). The authors sequenced the mitochondrial cytochrome *b* gene in 18 *H. spilopterus* from Southeast Asia and found three subclades – two subclades in the LMB separated by 2% sequence divergence and a third subclade in a separate, Cambodian river system showing 5% sequence divergence from Mekong River *H. spilopterus*. A more thorough understanding of the patterns and processes impacting populations of *H. spilopterus* is vital for the proper management of this resource in the face of growing hydropower development.

To address this need, we used double-digest restriction site-associated DNA sequencing (ddRAD, Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) to sample loci across the genome of *H. spilopterus* from Lao PDR,

Cambodia, and Viet Nam in order to evaluate levels of genetic structure, diversity, relatedness, and effective population size found in populations across the LMB. Our goals for this study were two-fold: (1) to establish an understanding of baseline levels of gene flow among *H. spilopterus* populations across areas of predicted high and low gene flow in the LMB before dams that are planned or under construction are built and (2) to examine the adaptive potential of isolated populations upstream of existing dams.

Methods

Sample collection

We collected tissue samples from *H. spilopterus* at regional markets or landings from six locations on the Mekong River system (Fig. 1, Table 1): two sites on the mainstream - one above the Khone Falls (Pakse) and one below (Kratié), one site on the Tonlé Sap (Siem Reap), one site near the downstream convergence of the 3-S rivers (Sesan), and one upstream site from two of the three main tributaries of the 3-S Basin (Sekong and Dak Lak) the latter of which is isolated from the rest of the LMB by four dams. We attempted tissue collection above established dams in the headwaters of the third major 3-S tributary (the Sesan River) but were unsuccessful at finding specimens despite reports of its historical occurrence in the region. Great care was taken in choice of vendors and the condition of fish to ensure those sampled were fresh and locally caught since the distribution and marketing of fishes within the LMB is highly complex. Vendors and fishermen were asked where specimens had been caught by local scientists familiar with the fishery (Cambodia: S.Uy; Laos: L Phounvisouk; Viet Nam: B.T. Dang, Q.H.D. Vu). If fish were reliably determined by the regional scientist(s) to have been captured locally, dissected muscle tissue was taken and preserved in 95% ethanol in separate vials. DNA from muscle tissue was extracted using a QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Inc., USA). Extract quality was examined using gel electrophoresis, and elutions with visible high-weight DNA and little degradation were used for preparation of restriction site-associated DNA (RAD) libraries.

Library preparation

Double-digest RAD (ddRAD, Peterson et al., 2012) library preparation was employed to reduce genome complexity, with minor modifications from the original protocol detailed here. Visible DNA fragments below 3000 bp were removed using SPRIselect beads (0.4x, Beckman Coulter, USA). High quality samples that were not SPRIselected were cleaned using AMPureXP paramagnetic beads (Beckman-Coulter). In both cases, the paramagnetic beads were left in the cleaned samples and reactivated, as necessary, by adding 3M NaCl 20% PEG (Fisher et al., 2011). DNA extracts were quantified using the AccuBlue High-Sensitivity kit (Biotium, USA) on a Spectramax M3 fluorescent plate reader (Molecular Devices, USA). Samples were normalized, and 50 ng of DNA per sample were digested with the *HindIII* and *EcoRI* restriction enzymes. Each sample was individually barcoded and groups of 48 samples were combined and indexed via polymerase chain reaction (PCR), as described by Peterson et al. (2012). Half of the volume of each individually barcoded sample was saved for renormalization following a test sequence run which is detailed below.

To identify and eliminate index mis-assignment (Hussmann, 2015; Kircher, Sawyer, & Meyer, 2011; Wright & Vetsigian, 2016), a dual-indexing system was used where each unique 11 bp index (Quail et al., 2014) was only paired with one other unique index, forming matched index pairs. Employing this design, index mis-assignment can be identified during sample demultiplexing by targeting and removing samples from the dataset with unmatched index pairs.

Following PCR, DNA fragments from 400-500bp were isolated using a BluePippin automated electrophoresis rig. The molar concentration of DNA in each library was determined using the KAPA Library Quantification Kit for Illumina Platforms. Libraries were subsequently normalized and combined in equal proportions into a super library. In order to confirm a similar concentration of DNA from each sample, we combined 1 part super library with 99 parts of another super library from another project. The super library cocktail, with a DNA concentration of at least 2 nM, was sequenced on one lane of an Illumina HiSeq 4000 (PE 150) at New York University's Genome Technology Center. The resulting read depth was used to re- normalize the individually-barcoded samples, which were subsequently processed as described above. The renormalized super library was sequenced on two lanes of an Illumina HiSeq 4000 (PE 150).

Read processing

Sequenced reads were demultiplexed by index and barcode, filtered for the expected forward read cut site, and truncated to 90 bp in the Stacks subprogram process_radtags (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). The dDocent pipeline was used to generate a *de novo* assembly, map reads, and call variants (Puritz, Hollenbeck, & Gold, 2014). Default parameters were modified as required, including changing the CD-Hit mapping parameter (c) to 0.92 and increasing the base quality score (q) to 20 for a variant to be called with Freebayes (Garrison & Marth 2012).

SNP Filtering

Preliminary filtering of variants was performed with vcftools (Danecek et al. 2011) and vcflib (Garrison 2012). Filtering parameters included removing individuals with low representation and keeping only bi-allelic loci present in 95% or more of the remaining individuals with a minor allele frequency of 0.01 or greater. Additional filtering removed loci missing in more than 10% of the individuals from a single site, loci covered by both forward and reverse reads, heterozygous loci with an allele balance less than 0.25 or greater than 0.75, and loci that deviated by population from Hardy-Weinberg Equilibrium (HWE).

After preliminary filtering, the dataset was checked for potential contamination and paralogous loci using two methods. First, heterozygosity for all samples and loci was examined and any outliers removed. In addition, individuals were haplotyped (https://github.com/chollenbeck/rad_haplotyper/), and those with more than the expected number of haplotypes were removed. RAD tags with more than two individuals containing more observed haplotypes than expected were removed. Finally, one random SNP per paired-end tag was selected for the final panel to remove any effect of linkage among SNPs on the same RAD tag. File format conversions were executed in the Java conversion tool PGDSpider v 2.0.8.3 (Lischer & Excoffier, 2012).

Identifying loci under selection

Our final filtered panel of SNPs was run in both LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008; Beaumont & Nichols, 1996) and BayeScan v2.1 (Foll & Gaggiotti, 2008) to identify loci under divergent or balancing selection. LOSITAN was run with the recommended parameters of forced and "neutral" mean F_{ST} for 100,000 simulations with 95% confidence intervals. BayeScan was run under default parameter settings. A false discovery rate (FDR) of 5% was applied in both programs. All loci putatively identified as outliers after FDR correction by either program were removed from the dataset to generate a panel of neutral SNPs.

Genetic distance, diversity, and differentiation

Measures of genetic distance among individuals, genetic diversity and inbreeding within sampled sites, and pairwise F_{ST} values between sampled sites were generated in GenoDive v. 2.0b27 (Miermans & Van Tienderen, 2004). Genetic similarity of sampled individuals was examined with agglomerative hierarchical clustering (hclust) using a matrix of Euclidean genetic distances from GenoDive and the average linkage method (UPGMA, Sokal & Michener, 1958) in R. The significance of F_{ST} values was tested with 10,000 permutations of the dataset and Benjamini and Hochberg's (1995) method for correcting the FDR was applied to p-values.

Two maximum likelihood estimates of relatedness (dyadic and triadic; Milligan, 2003; Wang, 2007) were generated with the coancestry function in the R package "related" (Pew, Muir, Wang, & Frasier, 2015; Wang, 2011) with parameters set to account for inbreeding, an error rate of 0.001, and the generation of 95% confidence intervals with 500 bootstrapped replicates. Estimates of contemporary effective population (N_e) size were calculated using a bias-corrected linkage disequilibrium (LD) method in the program NeEstimator v2.1 (Do et al., 2014).

Gene flow between sampled sites in the LMB

Population structure among the six sampled sites was examined using principal component analysis (PCA) and Bayesian clustering. PCAs were generated for both neutral and divergent SNP panels in the R package "adegenet" (Jombart & Ahmed, 2011). The program STRUCTURE v2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000) was used to identify distinct genetic clusters and assign individuals to

populations using the neutral SNP panel under the assumption of population admixture and correlated allele frequency. Runs consisted of a burn-in period of 50,000 MCMC iterations followed by an additional 100,000 steps for inferred clusters (*K*) 1 to 10, with 10 replicates each. Results were collated in STRUCTURE HARVESTER (Earl & vonHoldt, 2012), and the Δ*K* statistic (Evanno, Regnaut, & Goudet, 2005) was used to determine the most likely number of clusters. Replicate Q-matrices for the optimal *K* were processed with the Greedy algorithm in CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007).

Results

Genetic distance, diversity, and differentiation

A total of 248 libraries were sequenced and 203 individuals were successfully genotyped at 756 bi-allelic loci. An initial analysis of genetic similarity among individuals with a UPGMA cluster dendrogram revealed the presence of two highly divergent genetic clusters co-existing in at least three of the six locations we sampled (Fig. 2, Cluster A and B). Cluster A contained two divergent subgroups, one comprised of an individual sampled in Pakse and 13 individuals sampled in Sekong, and one comprised of 19 individuals sampled in Siem Reap. Individuals assigned to Cluster B were found across all six sampled sites. A Bayesian clustering analysis run in STRUCTURE with no location priors produced concordant results with an optimal *K* of 2 and all individuals being assigned to Cluster A or B with 92.2% certainty or greater (Fig. 2). An analysis of genetic differentiation between the two clusters generated an F_{ST} value of 0.463 (p=0.001). We concluded that there were multiple morphologically-cryptic lineages present in our samples, and it would violate the assumptions of the analyses to analyze fish collected from each location as samples from the same randomly-mating population.

Due to the low representation across our sample sites of individuals belonging to Cluster A and the confounding factor of two strongly divergent genetic clusters, subsequent analyses were performed using only the individuals assigned to Cluster B. Individuals from Cluster A were removed from the original variant file, filtering steps were repeated, and a new panel of 916 bi-allelic SNPs was generated for the 170 individuals assigned to Cluster B. LOSITAN and BayeScan identified a total of 128 loci as candidates for selection after FDR correction, 50 loci putatively under balancing selection and 78 loci putatively under divergent selection. All potential outliers were

removed to produce a panel of 788 neutral SNPs and a panel of 78 putatively divergent SNPs identified by LOSITAN and BayeScan for further analysis.

Summary statistics by site are reported in Table 1. Observed and expected heterozygosity (H_o / H_e) were similar across all sites except Dak Lak (Table 1), with site-specific H_o measuring 0.0447 in Dak Lak and ranging from 0.1140-0.1487 in all other sampled sites. Dak Lak also exhibited the lowest percentage of observed alleles (67.7%; other sites 73.5-94.6%) and had the highest inbreeding coefficient (G_{IS} = 0.5070). Upon further examination of individuals from Dak Lak, we observed that 11 individuals from Dak Lak exhibited extremely low levels of heterozygosity (< 2% of loci were heterozygous). The percentage of loci genotyped for these 11 individuals was high, ranging from 86.4-99.7% (average 94.8%). A high level of successfully genotyped loci suggests that the extremely low level of heterozygosity measured in these 11 individuals is likely due to a high level of fixation and not allelic dropout. The unique genetic signature resulting from a high proportion of fixation can be observed by the placement of these 11 individuals in a divergent group within Cluster B in the UPGMA cluster dendrogram (*, Fig. 2).

Pairwise F_{ST} values indicate the presence of high levels of regional genetic variance (Table 2), which supports the findings of multiple groups within Cluster B in the UPGMA cluster dendrogram. All comparisons between sites were significant ($p \leq 0.0001$), with F_{ST} values ranging from 0.0149 (Kratié/Siem Reap) to 0.4403 (Dak Lak/Siem Reap).

Both maximum likelihood estimates of relatedness produced concordant results and indicate that Dak Lak contained the most genetically related fish with a mean coefficient of relatedness (r) of 0.5555, the level that would be expected from full-sibling or parent-offspring relationships ($r > 0.5$, Fig. 3). Pairwise coefficients of relatedness across all individuals in Cluster B ranged from 0.000-0.9719. Most pairwise comparisons were deemed unrelated (86.6%) in large part due to insignificant levels of estimated relatedness ($r \leq 0.1250$) between individuals from different sites. The vast majority of between-site related pairs were between Kratié and Siem Reap (409 pairs, r=0.1251-0.402). Additional related pairs were found between Pakse and Sekong (34 pairs, r=0.1252-0.2032), Kratié and Sesan (20 pairs, r=0.1291-0.1976), Siem Reap and Sesan (13 pairs, r=0.1304-0.1942), Dak Lak and Sekong (11 pairs, r=0.1269-0.1446), Pakse and Sesan (6 pairs, r=0.1266-0.1587), Sesan and Sekong (6 pairs, r=0.1262-0.1602), Dak Lak and Pakse (2 pairs, r=0.1272 and r=0.1303). Most between-site related pairs fall within

a value associated with that of first-cousin relatedness ($r \cong 0.1250$-0.250), so these relationships should be viewed very conservatively. While empirical evidence has shown that panels of ~500 SNPs are needed to estimate a relationship at the parent-offspring/full-sibling/half-sibling levels with high confidence, more than twice the number in our panel (788 SNPs) is needed to determine a first-cousin relationship with confidence of 90% or greater (1,858-3,200 SNPs; Kopps, Kang, Sherwin, & Palsbøll, 2015; Mo et al., 2016). Values for mean coefficients of relatedness within site for both estimators with standard errors are reported in Table A in supporting information.

Estimates of effective population size (N_e) for the six sites ranged from less than one in Dak Lak to 3,353 in Kratié (Table 3). Harmonic mean sample size (S) ranged from 14.6-38.8. Kratié and Siem Reap (S=26.2 and 14.6, respectively) had upper bound confidence intervals of 'infinite.'. 'Infinite' estimates are often an indication that the harmonic mean sample size is too small to generate an estimate of N_e using the LD N_e method (Do et al., 2014; Waples & Do, 2010).

Gene flow among sampled sites in the LMB

A principal component analysis (PCA) with our neutral SNP panel indicated regional clustering of individuals. Along the first principal component, these individuals were grouped by sampling sites and spread across the axis from north to south in the LMB, accounting for 7.4% of the total variance (Fig. 4a & 4b). Graphing against the second principal component resulted in a cluster of closely associated individuals from Pakse and Sekong, a cluster of individuals from Sesan, another cluster of closely associated individuals from Kratié and Siem Reap, and two clusters of individuals within Dak Lak – a distant cluster of the 11 individuals with extremely low heterozygosity (denoted with *, Fig. 4a) and a cluster of the remaining 5 individuals sampled at this site. Graphing against the third principal component removed all but two of the Sekong individuals to a cluster separated from Pakse (Fig. 4b). A PCA with the 78 divergent SNPs produced almost identical regional clustering with the first and second principal components, which accounted for 21.5% and 14.5% of the total variance within this panel, respectively (Fig. 4c). Graphing against the third principal component produced the same split between Pakse and Sekong individuals but removed the split between the two Dak Lak clusters (Fig. 4d). Overall, the similarity between the PCAs with putative outliers and the PCAs with neutral SNPs may indicate that the strong spatial

structure present among sites in the LMB resulted in false positives in our outlier analyses. Highly structured riverscapes have been shown to bias F_{ST}-based outlier detection with the number of false positives increasing with the number of sampled demes (Fourcade, Chaput-Bardy, Secondi, Fleurant, & Lamaire, 2013).

STRUCTURE results were consistent over ten replicates and similar to those from the PCAs and groupings found within the Cluster B lineage in the UPGMA dendrogram, with ΔK indicating an optimal K of 5 (Fig. 5). All individuals from Pakse and two individuals from Sekong were assigned to a distinct lineage with high certainty. The remaining individuals from Sekong were assigned with high certainty to another lineage. All individuals from the flood-pulse zone in Kratié and Siem Reap were assigned to a third lineage. Five individuals from Dak Lak were assigned to a fourth lineage along with six individuals from Sesan. The remaining individuals from Dak Lak were the eleven exhibiting extremely low heterozygosity and were assigned to a fifth lineage only present in this sampling site. Sesan appears to be a site of admixture as the remaining 34 individuals from this location could not be assigned with a high level of certainty to any lineage sampled within the other locations.

Discussion

Baseline patterns of genetic composition and diversity

The discovery of two highly-divergent genetic clusters of *H. spilopterus* within the Mekong River system is consistent with previous phylogenetic research (Dodson & Lecomte, 2015). However, the authors of that research based their evaluation on the relationships of only 13 individuals sampled from the LMB. They hypothesized the split between the two LMB subclades occurred due to vicariance during the Pleistocene epoch (see Voris, 2001) followed by secondary contact in the region north of the Khone Falls and that the level of mitochondrial divergence (2%) was consistent with intraspecific divergence. Our sampling indicates that the sympatry between these two lineages is much more extensive than previously measured with representatives of both linages coexisting as far south as Siem Reap. In addition, the STRUCTURE assignment of all individuals to either Cluster A or B with greater than 92% certainty using a panel of SNPs sampled across the genome indicates that hybridization between members of the two lineages may not be successful. Secondary contact with hybridization over thousands of years should result in lower assignment scores (greater admixture) as F1 hybrids with 50% representation of both lineages

successfully reproduce (Barilani, Sfougaris, Giannakopoulos, Mucci, Tabarroni, & Randi, 2007; Rodriguez, Cedeño- Vázquez, Forstner, & Densmore III, 2008). More detailed morphological analysis and assessment of reproductive behavior and timing may be required to determine the exact mechanism preventing successful hybridization between these lineages.

Within Cluster B, genetic patterns appear to indicate that *H. spilopterus* in different tributaries and regions of the LMB maintain distinct populations despite potential for admixture. It has been observed that members of the genus *Hemibagrus* move into flooded forests to spawn during the wet season with young first documented in August and adults returning to rivers in November and December (Rainboth, 1996). If these spawning migrations are occurring in *H. spilopterus*, significant pairwise differences measured between all sites do not appear to indicate that they are migrating very great distances. Across sites that were unimpacted by dam-driven fragmentation at the time of sampling, pairwise F_{ST} values increased the further the sites were oriented north and south from one another. This trend was also evident in all PCAs.

Connectivity among *H. spilopterus* across the Khone Falls is uni-directional from up- to downstream. Two individuals sampled in Sekong, below Khone Falls, consistently aligned with samples from Pakse, above Khone Falls, in PCAs of both neutral and divergent SNPs. In the STRUCTURE analysis, these two individuals were assigned with 93% and 98% certainty to the genetic lineage representative of all individuals sampled in Pakse. Vendors for locally caught versus regionally caught fishes are found on opposite sides of the Sekong market, and as *H. spilopterus* is neither an aquaculture nor aquarium species, the probability that these two were sampled in Sekong due to human-mediated translocation is low. Assuming these samples represent natural movement, two possible routes exist for this migration. One possibility is that during the summer wet season when adults migrate to flooded forest to spawn, a handful of adults or offspring may have been directed through narrow, ephemeral tributaries that connect the mainstream of the Mekong near Pakse to the Sekong river. A second possibility is that these fish migrated downstream through the Khone Falls and unable to pass back upstream through the falls, diverted upstream to Sekong. The apparent admixture of multiple lineages near the convergence of the 3-S rivers in Sesan as well as downstream in Kratié indicates that some migration is occurring. The individuals assigned to Pakse that were collected in Sekong in October could represent rare migrants that, if reproductively successful the following

summer, contribute to the low levels of genetic representation of this Pakse lineage found in a handful of other individuals sampled in Sekong.

Levels of genetic differentiation and diversity along the three rivers of the 3-S basin also appear to be influenced by up- or downstream position, thereby supporting evidence from a variety of taxa that indicate gene flow in river systems is largely driven by the direction of water flow (Hänfling & Weetman, 2006; Alp, Keller, Westram, & Robinson, 2012). The most northern of the 3-S rivers - the Sekong - was sampled more than 300 km upstream from the Mekong mainstream, and STRUCTURE assigned all but the two individuals with similarities to Pakse to a unique genetic cluster with high certainty. In contrast to the sampling locations on the Sekong and Srepok, the middle of the three rivers - the Sesan - was sampled downstream near the convergence of the 3-S rivers, and the STRUCTURE plot and centralized position of this location in PCAs indicates it experiences a large amount of genetic admixture. Studies of genetic patterns associated with river architecture indicate that diversity is highest at the confluence of tributaries (Crispo, Bentzen, Reznik, Kinnison, & Hendry, 2006; Thomas, Christie, & Knowles, 2016), and both allelic richness and observed heterozygosity was highest in Sesan.

Regionalism among sampled *H. spilopterus* even appears to influence the genetic composition of sites with strong flood-pulse connections. Siem Reap (on the western end of the Tonlé Sap) and Kratié (on the Mekong mainstream) were both assigned to the same genetic cluster by STRUCTURE, were closely associated in all PCAs, and had the lowest pairwise genetic distance, which are all likely due to the influence of the flood-pulse system between the Tonlé Sap and the Mekong mainstream near Kratié. However, the F_{ST} value measured between these two sites was still significant, indicating that despite the seasonal hydrological mixing in this region of the LMB, migration and gene flow between these regions is not strong enough to induce complete panmixia.

Potential for adaptation to a fragmented landscape

The strongest genetic pattern observed across all analyses of Cluster B was the differentiation and relatedness of the individuals sampled in Dak Lak. The UPGMA dendrogram, STRUCTURE analysis, and PCAs clearly illustrate the differentiation driven by low levels of heterozygosity. The eleven individuals exhibiting less than 2% heterozygosity across loci in Dak Lak were even assigned to a distinct STRUCTURE cluster, which is most

likely an artefact of drift and fixation rather than a historic genetic lineage since no other sites had any percentage of representation of this cluster. Though all sampled sites indicated some level of relatedness and inbreeding among sampled individuals, a mean average relatedness of the level expected from full siblings in Dak Lak points to significant levels of inbreeding in this population and is further supported by the large inbreeding coefficient (G_{IS}) measured at this site. Both the value of G_{IS} in Dak Lak as well as the magnitude of difference between G_{IS} in Dak Lak versus G_{IS} in our other sites are at levels that have been associated with decreased fitness and inbreeding depression in a variety of other outcrossing species (Charlesworth & Willis, 2009; Hoffman *et al*., 2014; Naish, Seamons, Dauer, Hauser, & Quinn, 2013; Pekkala, Knott, Kotiaho, Nissinen, & Puurtinen, 2014).

One of the most likely causes of such high levels of inbreeding in Dak Lak compared to the upstream population sampled in the Sekong River is the isolation of this population from the downstream portion of Srepok river with the completion of the 12MW Dray H'linh 1 dam in 1990 (Middleton, 2007), followed by the Dray H'linh 2, Srepok 3 and Srepok 4 dams (Fig. 1). The catastrophic impacts of dams on upstream fish populations have been well-established elsewhere, with documented extinctions of fish populations upstream of dams in rivers in Australia, Canada, and the United States (Beamish & Northcote, 2011; Ferguson, Healey, Dugan, & Barlow, 2011; Gehrke, Gilligan, & Barwick, 2002). The Dak Lak region of the Srepok was further isolated when construction began upstream on the 280MW Buon Koup dam at the end of 2003 (SWECO Grøner, 2006a). Just as downstream dams can isolate upstream populations, upstream dams have the potential to both isolate downstream populations and significantly alter downstream habitat and water quality for tens of kilometers (Brandt, 2000; De Jalon, Sanchez, & Camargo, 1994; Lessard & Hayes, 2003; Power, Dietrich, & Finlay, 1996; Wyatt & Baird, 2007). Without a baseline assessment of genetic variation and effective population size in Dak Lak before the installation of dams, it is impossible to know for sure what an unimpacted population in this region should look like. Our results from this site underscore the importance of establishing these baselines before substantial changes to the environment are made.

Anecdotal evidence from upstream of several dams in the headwaters of another 3-S river – the Sesan - suggests local extinction driven by fragmentation may have already occurred in this species. During the collection of fishes from upstream sites in the 3-S rivers, it was noted that *H. spilopterus* and other species present in the Srepok (Dak Lak) and Sekong (Sekong) were not found in local catches in the headwaters of the Sesan in Viet Nam (B.T.D.

& K.E.C). Fishermen and fish mongers in this area commented that species such as *H. spilopterus* used to be present but were no longer caught there. Isolation of the Sesan headwaters by hydrological development began in November 1993 with the construction of the 720MW Yali Falls dam in Viet Nam (SWECO Grøner, 2006b; Wyatt & Baird, 2007), followed by the Se San 3, Se San 3A, Se San 4, and Se San 4A dams (Fig.1). Subjective information on the absence of *H. spilopterus* above the Yali Falls dam is not conclusive evidence for local extinction but, coupled with our findings in Dak Lak and growing research indicating that headwater populations are particularly vulnerable to fragmentation (Junker, Peter, Wagner, Mwaiko, Germann, Seehausen, & Keller, 2012), warrants further investigation into the previous records and current composition of species found above established dams in the tributaries of the LMB.

Hemibagrus spilopterus found anywhere in tributaries of the LMB may be at inherently greater risk than mainstream populations to inbreeding depression due to fragmentation. Estimates of effective population size (N_e) indicated much larger regional populations on the mainstream of the Mekong River than tributaries. The two mainstream sites - Pakse and Kratié – had N_e estimates at least an order of magnitude greater than any other sampled site (526 and 3,353, respectively). In comparison, unimpacted sites sampled in tributaries - the Sekong (upstream) and Sesan (downstream) - produced estimates of N_e ranging from 58-97. Conservation geneticists have set guidelines for minimum viable population size at N_e >50 over short-term periods and N_e >500 in the long-term in order to avoid inbreeding depression and extinction (Franklin, 1980; Jamieson & Allendorf, 2012). Given that estimates of N_e within the Sekong and Sesan tributaries fell closer to 50 than 500, any perturbations that may further decrease N_e such as damming or overfishing could put these populations at risk. This finding, coupled with the depressed genetic diversity found in Dak Lak and anecdotal evidence of a possible local extinction above several dams in the headwaters of the Sesan, lends further support to the indication that tributary populations of *H. spilopterus* are especially vulnerable to fragmentation.

The results of this study emphasize the value of using a genomics approach to assess the evolutionary potential of an important fishery species in a rapidly changing landscape. In addition to uncovering putative cryptic species in *H. spilopterus*, tributary populations within the most well-sampled of the two species clusters exhibited several risk factors that indicate their ability to adapt to stressors such as human-induced fragmentation is limited, and this must be considered for the proper management of this fishery. The genetic patterns observed in Dak Lak do

not bode well for the future of this population. If swift action is not taken to re-introduce genetic variability into this population, it is likely the combined pressures of isolation, inbreeding, and fishing will drive this population to extinction resulting in the loss of this population as a source of harvestable protein for the people living in this region of Viet Nam. In order to avoid similar circumstances in other tributaries of the LMB, development of hydrological power should be pursued with careful consideration for the genetic impacts on upstream food resources.

Funding

This research was supported by a Partnerships for Enhanced Engagement in Research (PEER) 2-7 grant to B.T. Dang under a USAID Cooperative Agreement AID-OAA-A-11-00012 with NSF grant OISE-0730256 to K.E. Carpenter.

Acknowledgements

This research was supported by the Turing High Performance Computing cluster at Old Dominion University, and we are grateful to M. Dong, T. Stilwell, W. Purwanto, and J. Pratt for their HPC support. S. F. Magnuson at the Genomics Core Laboratory at TAMUCC preformed library preparation. B. Stockwell provided many useful comments on analyses and manuscript content.

Data Availability

Unmodified sequence data in fastq format and corresponding metadata were archived with the publicly accessible Genomic Observatories Metadatabase (GeOMe, GUID: https://n2t.net/ark:/21547/Chw2, NCBI BioProject ID: PRJNA5271333).

References

Abidin, M.Z., Hashim, R., & Chien, A.C.S. (2006). Influence of dietary protein levels on growth and egg quality in broodstock female bagrid catfish (*Mystus nemurus* Cuv. & Val.). *Aquaculture Research*, 37, 416–418.

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.

Adamson, E.A., Hurwood, D.A., & Mather, P.B. (2012). Insights into historical drainage

evolution based on the phylogeography of the chevron snakehead fish (*Channa striata*) in the Mekong Basin. *Freshwater Biology*, 57, 2211-2229.

Alp, M., Keller, I., Westram, A. M., & Robinson, C. T. (2012). How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology*, 57, 969-981.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: a workbench to detect molecular adaptation based on a F ST-outlier method. *BMC Bioinformatics*, *9*(1), 323.

Arias, M.E., Cochrane, T., 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, 864–876.

Baird, I.G., Flaherty, M.S., & Phylavanh, B. (2004). Mekong River Pangasiidae catfish

migrations and the Khone Falls wing trap fishery in southern Laos. *Natural History Bulletin of the Siam Society*. 52, 81-109.

Barilani, M., Sfougaris, A., Giannakopoulos, A., Mucci, N., Tabarroni, C., & Randi, E. (2007). Detecting introgressive hybridisation in rock partridge populations (Alectoris graeca) in Greece through Bayesian admixture analyses of multilocus genotypes. *Conservation Genetics*, 8, 343-354.

Barlow, C., Baran, E., Halls, A., & Kshatriya, M. (2008). How much of the Mekong fish catch is at risk from mainstem dam development? *Catch and Culture*, 14, 16–21.

Beaumont, M.A., & Nichols, R.A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, 263, 1619-1626.

Beilfuss, R., & Triet, T. (2014). Climate change and hydropower in the Mekong River Basin: A synthesis of research. *Deutsche Gesellschaft fürInternationale Zusammenarbeit (GIZ), Vientiane*.

Brandt, S. A. (2000). Classification of geomorphological effects downstream of dams. *Catena*, *40*, 375-401.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140.

Charlesworth, D., & Willis, J.H. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics*, 10:783.

Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., & Hendry, A. P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular ecology*, 15, 49-62.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., … 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156-2158.

De Jalon, D. G., Sanchez, P., & Camargo, J. A. (1994). Downstream effects of a new hydropower impoundment on macrophyte, macroinvertebrate and fish communities. *Regulated Rivers: Research & Management*, *9*, 253-261.

Deiner, K., Garza, J.C., Coey, R., & Girman, D.J. (2007) Population structure and genetic diversity of trout (*Oncorhynchus mykiss*) above and below natural and man-made barriers in the Russian River, California. *Conservation Genetics*, 8, 437-454.

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 (N_e) from genetic data. *Molecular Ecology Resources*,14, 209-214.

Dodson, J. J., & Lecomte, F. (2015). A DNA barcode-based evaluation of the Southeast Asian catfish genus Hemibagrus Bleeker, 1862 (Teleostei: Siluriformes; Bagridae). *Advances in Evolutionary Biology*, 2015. DOI: 10.1155/2015/490158

Doyle, M.W., Stanley, E.H., Harbor, J.M., & Grant, G.S. (2003) Dam removal in the United States: emerging needs for science and policy. *Eos, Transactions American Geophysical Union*, 84, 29-33.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., … Sullivan, C.A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 8, 163-182.

Dugan, P.J., Barlow, C., Agostinho, A.A., Baran, E., Cada, G.F., Chen, D., Cowx, I.G., … Marmulla, G. (2010). Fish migration, dams, and loss of ecosystem services in the Mekong basin. *Ambio*, 39, 344-348.

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.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

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., … 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.

Fourcade, Y., Chaput-Bardy, A., Secondi, J., Fleurant, C., & Lemaire, C. (2013) Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology*, 22, 2065-2073.

Franklin, I.R. (1980). Evolutionary change in small populations. In Conservation Biology: an Evolutionary–Ecological Perspective (Soule´, M.E. and Wilcox, B.A., eds), pp. 135–150, Sinauer Associates

Garrison, E. (2012). Vcflib: A C++ library for parsing and manipulating VCF files. *GitHub https://github.com/vcflib/vcflib (accessed January 21, 2018)*.

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

Gehrke, P.C., Gilligan, D.M., & Barwick, M. (2002). Changes in fish communities of the Shoalhaven River 20 years after construction of Tallowa Dam, Australia. *River Research and Applications*, 18, 265-286.

Gibbs, J.P. (2001) Demography versus habitat fragmentation as determinants of genetic variation in wild populations. *Biological Conservation*, 100, 15-20.

Hänfling, B., & Weetman, D. (2006). Concordant genetic estimators of migration reveal anthropogenically-enhanced source-sink population structure in the river sculpin, Cottus gobio. Genetics, 173, 1487-1501.

Hedrick, P.W., & Miller, P.S. (1992) Conservation genetics: techniques and fundamentals. *Ecological Applications*, 2, 30-46.

Hill, M.T., & Hill, S.A. (1994). *Fisheries ecology and hydropower in the Mekong River: an evaluation of run-of-the-river projects*. Don Chapman Consultants, Boise, Idaho

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A., ... & Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, *111*(10), 3775-3780.

Hortle, K.G., Khounsavanh, O., Chanthasone, P., Phommachanh, P., Singhanouvong, D.,

& Viravong, S. (2015). Larval and juvenile fish in the Mekong River in Northern Lao PDR.MRC Technical Paper No. 46. Mekong River Commission, Phnom Penh, Cambodia. 87 pp.

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.S., & Mather, P.B. (2007). Evidence for strong genetic structure in a regionally important, highy vagile cyprinid (*Henicorhynchus lobatus*) in the Mekong River Basin. *Ecology of Freshwater Fish*, 27, 1153-1167.

Hussmann, J.A. (2015). *Expanding the applications of high-throughput DNA sequencing* (Doctoral dissertation)

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.

Jamieson, I.G., & Allendorf, F.W. (2012). How does the 50/500 rule apply to MVPs? *Trends in Ecology and Evolution*, 27, 578-584.

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., Bayley, P.B., & Sparks, R.E. (1989). The flood pulse concept in river floodplain systems. Pages 110-127 in D.P. Dodge, ed. Proceedings of the International Large River Symposium (LARS). Canadian Special Publication of Fisheries and Aquatic Sciences 106

Junker, J., Peter, A., Wagner, C. E., Mwaiko, S., Germann, B., Seehausen, O., & Keller, I. (2012). River fragmentation increases localized population genetic structure and enhances asymmetry of dispersal in bullhead (*Cottus gobio*). *Conservation Genetics*, 13, 545-556.

Kircher, M., Sawyer, S., & Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40, e3-e3.

Kopps, A.M., Kang, J., Sherwin, W.B., & Palsbøll, P.J. (2015). How well do molecular and pedigree relatedness correspond, in populations with variable mating systems, and types and quantities of molecular and demographic data? *G3: Genes, Genomes, Genetics*, g3-115.

Lankau, R., Jørgensen, P.S., Harris, D.J., & Sih, A. (2011) Incorporating evolutionary principles into environmental management and policy. *Evolutionary Applications*, 4, 315-325.

Lessard, J. L., & Hayes, D. B. (2003). Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. *River research and applications*, *19*(7), 721-732.

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, S., & He, D. (2008). Water level response to hydropower development in the upper

Mekong river. *Ambio*, 37, 170–177.

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

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 (2005) Overview of the Hydrology of the Mekong Basin.
Mekong River Commission, Vientiane, Lao PDR, November 2005, 73pp.

Mekong River Commission (2010) State of the Basin Report 2010, Mekong River

Commission, Vientiane, Lao PDR

Middleton, C. (2007). The ADB/WB/MRC 'Mekong Water Resources Assistance Strategy': Justifying Large Water Infrastructure with Transboundary Impacts. In *International Conference: Critical Transitions in the Mekong Region", 29-31 January*.

Mille, G., Hap, N., & Loeng, N. (2016). Economic value of fish in Cambodia and value added along the trade chain. Inland Fisheries Research and Development Institute (Fisheries Administration) and WorldFish. Phnom Penh, Cambodia. 62 pp.

Milligan, B.G. (2003) Maximum-likelihood estimation of relatedness. *Genetics*, 163, 1153-1167.

Mo, S. K., Liu, Y. C., Wang, S. Q., Bo, X. C., Li, Z., Chen, Y., & Ni, M. (2016). Exploring the efficacy of paternity and kinship testing based on single nucleotide polymorphisms. *Forensic Science International: Genetics*, *22*, 161-168.

Mo, T. (1991). Anatomy, Relationships and Systematics of the Bagridae (Teleostei:

Siluroidei) with a Hypothesis of Siluroid Phylogeny, vol. 17 of Theses Zoologicae, Koeltz Scientific Book

Morita, K., & Yamamoto, S. (2002) Effects of habitat fragmentation by damming on the persistence of stream-dwelling charr populations. *Conservation Biology*, 16, 1318-1323.

MRCS/WUP-FIN (2006) Technical Paper No. 5- Tonle Sap Lake Water Balance

Calculations, WUP-FIN Phase II-Hydrological, Environmental and Socio-Economic Modelling [sic] Tools for the Lower Mekong Basin Assessment. Mekong River Commission and Finnish Environment Institute Consultancy Consortium, Vientiane, 32pp.

Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., & Quinn, T. P. (2013). Relationship between effective population size, inbreeding and adult fitness- related traits in a steelhead (Oncorhynchus mykiss) population released in the wild. *Molecular Ecology*, *22*(5), 1295-1309.

Ng, H.H., & Rainboth, W.J. (1999). The bagrid catfish genus *Hemibagrus* (Teleostei:

Siluriformes) in central Indochina with a new species from the Mekong River. *Raffles Bulletin of Zoology*, 47, 555-576.

Ng, W.K., Soon, S-C., & Hashim, R. (2001). The dietary protein requirement of a bagrid

catfish, *Mystus nemurus* (Cuvier & Valenciennes), determined using semipurified diets of varying protein level. *Aquaculture Nutrition*, 7, 45–51.

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

O'Connor, J.E., Duda, J.J., & Grant, G.E. (2015) 1000 dams down and counting. *Science*, 348, 496–497.

Pekkala, N., Knott, K. E., Kotiaho, J. S., Nissinen, K., & Puurtinen, M. (2014). The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients. *Evolutionary Applications*, *7*(9), 1107-19.

Pertoldi, C., Bijlsma, R., & Loeschcke, V. (2007) Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodiversity and Conservation*, 16, 4147-4163.

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.

Petit, R.J., El Mousadik, A., & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, 12, 844-855.

Pew, J., Muir, P., Wang, J., & Frasier, T.R. (2015). related: an R package for analyzing pairwise relatedness data based on codominant molecular markers. *Molecular Ecology Resources*, 15, 557-561.

Piman, T., Cochrance, 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. DOI: 10.1002/rra

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

Power, M. E., Dietrich, W. E., & Finlay, J. C. (1996). Dams and downstream aquatic biodiversity: potential food web consequences of hydrologic and geomorphic change. *Environmental management*, *20*, 887-895.

Puritz, J.B., Hollenbeck, C.M., & Gold, J.R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431.

Quail, M.A., Smith, M., Jackson, D., Leonard, S., Skelly, T., Swerdlow, H.P., … Ellis, P.

(2014). SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics*, 15,110.

Ralls, K., Brugger, K., & Ballou, J. (1979) Inbreeding and juvenile mortality in small populations of ungulates. *Science*, 206:1101-1103.

Rainboth, W.J. (1996). Fishes of the Cambodian Mekong. FAO Species Identification Field Guide for Fishery Purposes. FAO, Rome. Xi+265, 27pls.

Reed, D.H., Nicholas, A.C., & Stratton, G.E. (2007) Genetic quality of individuals impacts population dynamics. *Animal Conservation*, 10, 275-283.

Reid, S.M., Wilson, C.C., Mandrak, N.E., & Carl, L.M. (2008) Population structure and genetic diversity of black redhorse (*Moxostoma duquesnei*) in a highly fragmented watershed. *Conservation Genetics*, 9, 531.

Rodriguez, D., Cedeño- Vázquez, J. R., Forstner, M. R., & Densmore III, L. D. (2008). Hybridization between Crocodylus acutus and Crocodylus moreletii in the Yucatan Peninsula: II. Evidence from microsatellites. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 309, 674-686.

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

Ryman, N. (1991) Conservation genetics considerations in fishery management. *Journal of Fish Biology*, 39, 211-224.

So, N., Van Houdt, J.K., & Volckaert, F.A. (2006). Genetic diversity and population history of the migratory catfishes *Pangasianodon hypophthalmus* and *Pangasius bocourti* in the Cambodian Mekong River. *Fisheries Science*, 72, 469-476.

Sverdrup-Jensen, S., Bishop, A., Clayton, T., Barlow, C., & Mekong River Commission (2002). Fisheries in the Lower Mekong Basin: status and perspectives.

SWECO Grøner (2006a). FINAL REPORT: Environmental Impact Assessment on the Cambodian part of the Srepok River due to Hydropower Development in Vietnam. Available at: https://www.internationalrivers.org/sites/default/files/attached-files/final_report_srepok_eia_sweco.pdf. Retrieved February 2019.

SWECO Grøner (2006b). FINAL REPORT: Environmental Impact Assessment on the Cambodian part of the Se San River due to Hydropower Development in Vietnam. Available at: https://www.internationalrivers.org/sites/default/files/attached-files/sesaneiadec2006.pdf. Retrieved February 2019.

Takagi, A.P., Ishikawa, S., Nao, T., Hort, S., Nakatani, M., Nishida, M., & Kurokura, H. (2006). Genetic differentiation of the bronze featherback *Notopterus notopterus* between Mekong River and Tonle Sap Lake populations by mitochondrial DNA analysis. *Fisheries Science*, 72, 750-754.

Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution*, *70*, 731-739.

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.

Wang, J. (2011). COANCESTRY: a program for simulating, estimating and analyzing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 1, 141-145.

Wild, T.B., & Loucks, D.P. (2015). Mitigating Dam Conflicts in the Mekong River Basin. In Conflict Resolution in Water Resources and Environmental Management 2015 (pp. 25-48). Springer International Publishing.

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/. Retrieved March 2018.

Wright, E.S., & Vetsigian, K.H. (2016). Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics*, 17, 876.

Wyatt, A.B., & Baird, I.G. (2007) Transboundary Impact Assessment in the Sesan River Basin: The Case of the Yali Falls Dam. *International Journal of Water Resources Development*, 23, 427-442.

Yang, S. L., Milliman, J. D., Xu, K. H., Deng, B., Zhang, X. Y., & Luo, X. X. (2014). Downstream sedimentary and geomorphic impacts of the Three Gorges Dam on the Yangtze River. *Earth-Science Reviews*, *138*, 469-486.

Young, A., Boyle, T., & Brown, T. (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, 1, 413-418.

Zakaria, Z.A., Chong, A.S.C., & Ali, A.B. (2004). Influence of stocking density on

various growth parameters and water quality of the bagriid [sic] catfish *Hemibagrus nemurus* (Cuvier & Valenciennes) in canvas tanks. *Malaysian Applied Biology*, 29, 35–42.

Ziv, G., Baran, E., Nam, S., Rodriguez-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.

Tables

Table 1. Summary statistics by location of *H. spilopterus* from Cluster B. Number of individuals successfully genotyped and used in analyses of Cluster B (N), percentage of observed alleles represented within site (A_o), observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (G_{IS}, an analogue to F_{IS} based on G_{ST}: Nei 1973). The location sampled upstream of existing dams is denoted with *.

Site	Country	LMB Location	Collection Date	N	A_o (%)	H_o	H_e	G_{IS}
Pakse	Lao PDR	Mekong Mainstream	Nov 2016	41	85.7	0.1276	0.1509	0.1544
Sekong	Lao PDR	Sekong River (3-S)	Oct 2016	30	85.0	0.1471	0.1593	0.0765
Sesan	Cambodia	Sesan River (3-S)	Dec 2016	40	94.6	0.1487	0.1684	0.1170
Dak Lak*	Viet Nam	Srepok River (3-S)	Nov 2016	16	67.7	0.0447	0.0907	0.5070
Kratié	Cambodia	Mekong Mainstream	Dec 2016	28	80.2	0.1140	0.1399	0.1851
Siem Reap	Cambodia	Tonlé Sap Lake	Dec 2016	15	73.5	0.1167	0.1335	0.1258

Table 2. Pairwise genetic differentiation between sampled sites using individuals assigned to Cluster B. All pairwise F_{ST} values were significant (p<0.0001). Values from our isolated site, Dak Lak, are in bold.

Site	**Pakse**	**Sekong**	**Sesan**	**Dak Lak**	**Kratié**	**Siem Reap**
Pakse	-					
Sekong	0.0852	-				
Sesan	0.0648	0.0604	-			
Dak Lak	**0.3257**	**0.3112**	**0.2948**	-		
Kratié	0.2027	0.1866	0.0804	**0.3964**	-	
Siem Reap	0.2340	0.2154	0.1051	**0.4403**	0.0149	-

Table 3. N_e for *H. spilopterus* in Cluster B. *S* is the harmonic mean sample size for each population. Effective population size estimates (N_e) and 95% jackknifed confidence intervals (CI) were generated with from the linkage disequilibrium method in NeEstimator with a minor allele frequency cutoff of 0.01. The location sampled upstream of existing dams is denoted with *.

Site	***S***	**LMB Location Type**	**N_e (CI)**
Pakse	38.8	Mainstream	526 (217-Inf)
Sekong	28.8	Tributary	58 (34-143)
Sesan	38.4	Tributary	97 (52-355)
Dak Lak*	14.6	Tributary	0.6 (0.4-1.3)
Kratié	26.2	Mainstream	3,353 (121-Inf)
Siem Reap	14.6	Lake tributary	796 (92-Inf)

Figure Captions

Figure 1. Sample sites for *H. spilopterus* along the Mekong River system. INSET: Sampling region (black box) within Southeast Asia. Only existing and planned dams affecting connectivity between the six sampled sites are labeled in this figure (Li, Long, Zhao, Lu, & Hong, 2017; Piman, Cochrane, & Arias, 2016; WLE Greater Mekong, 2017).

Figure 2. UPGMA dendrogram and STRUCTURE plot of all sampled *H. spilopterus* indicating two highly divergent genetic clusters (Cluster A and Cluster B). Membership to each approximate subgrouping (black bars) is denoted with a label of: location (# of individuals). The most divergent group within Cluster B (*) consists of 11 individuals from Dak Lak with extremely low levels of heterozygosity. See online version for full colors.

Figure 3. Relatedness among Cluster B *H. spilopterus* within sampled sites. Mean coefficients of relatedness (*r*) were generated with two maximum likelihood estimators in the R package 'related.' Standard error of the mean is denoted by bars (values are less than 0.005 for all but our isolated (*) site, Dak Lak).

Figure 4. Principal component analyses of neutral and divergent SNP panels for *H. spilopterus* in Cluster B. For each panel, the first principal component (PC 1) is graphed on the x-axis against the second principal components (PC 2, a & c) and the third principal components (PC 3, b & d). The * denotes the location of the 11 individuals from Dak Lak with extremely low heterozygosity in graphs A-C. Insets on the bottom right contain barplots of the first 30 eigenvalues. See online version for full colors.

Figure 5. STRUCTURE results for *H. spilopterus* from Cluster B across the six locations sampled in the LMB. K=5 generated the largest ΔK (graph). Plot indicates probability of lineage membership for each individual by location, pie charts indicate the proportion of representation of the five lineages at each site. See online version for full colors.

Figure 1

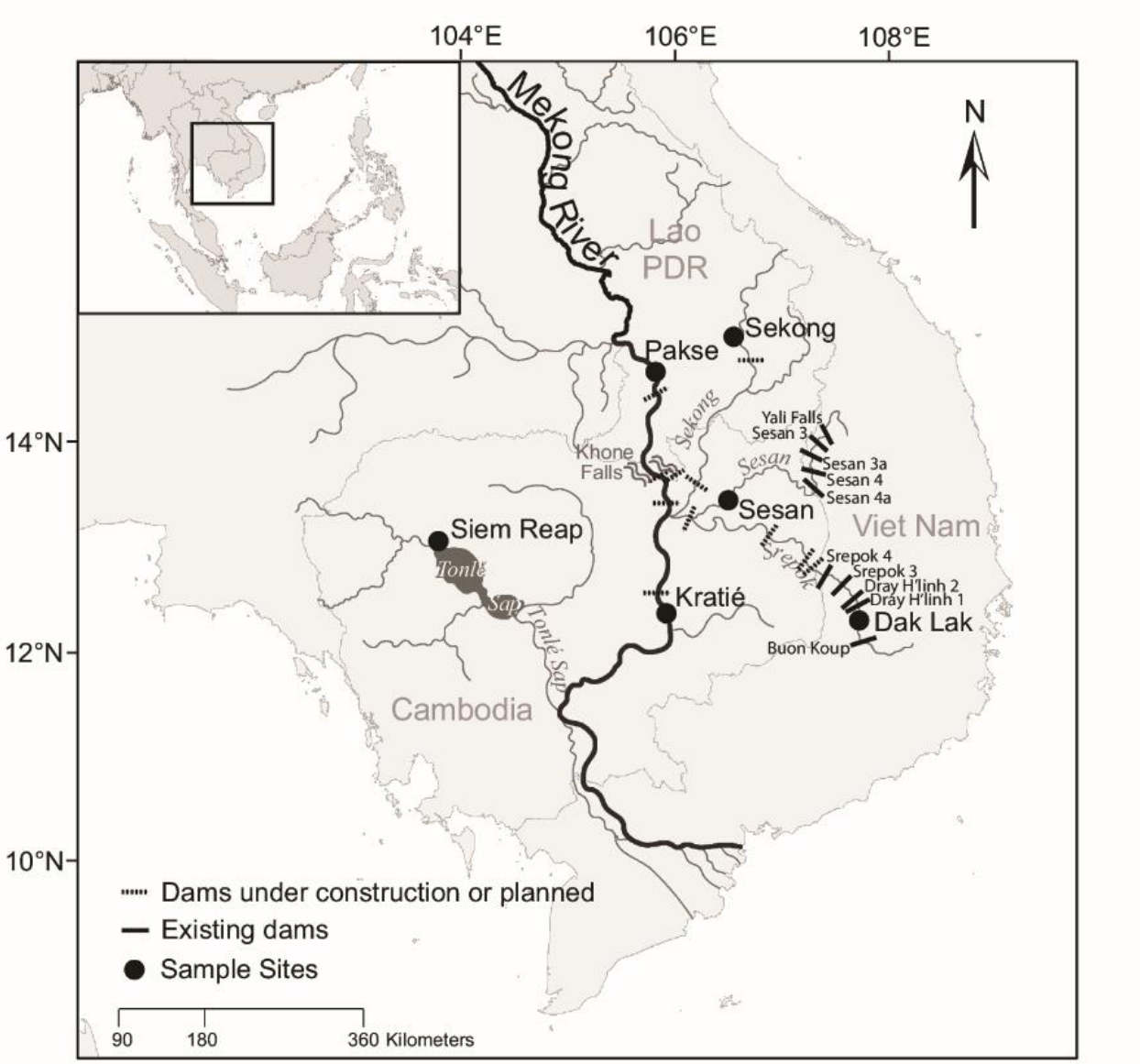

Figure 2

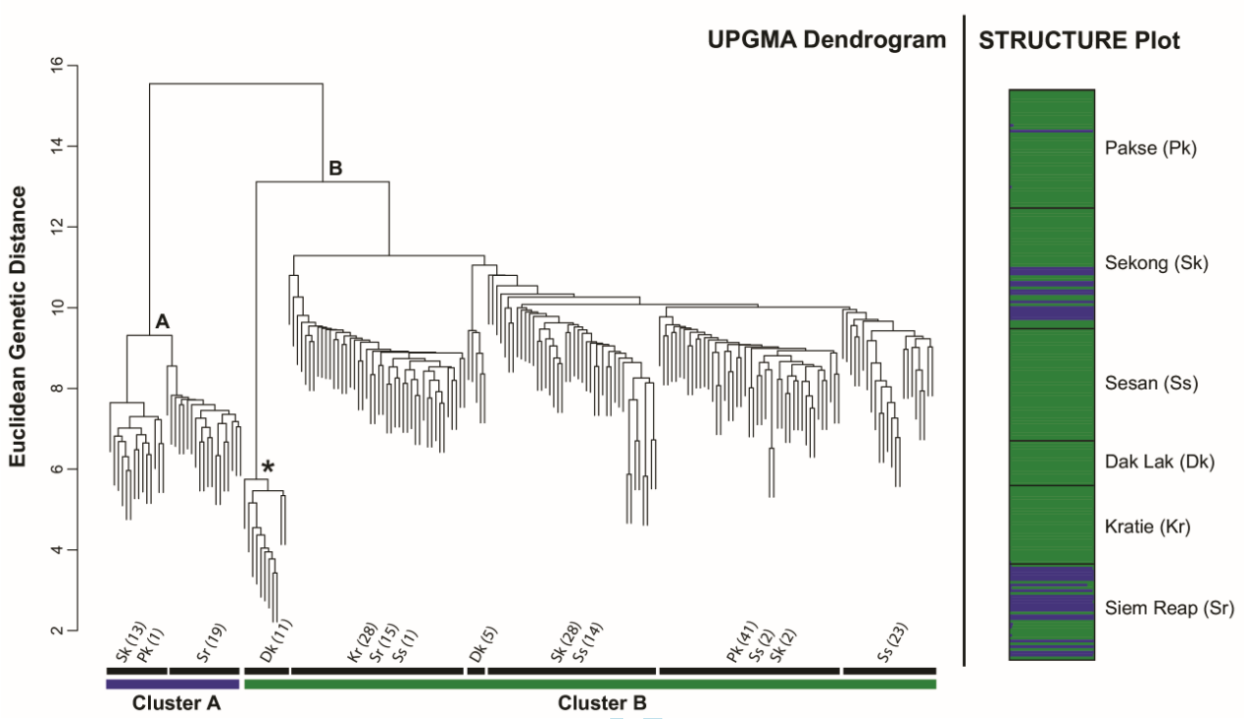

Figure 3

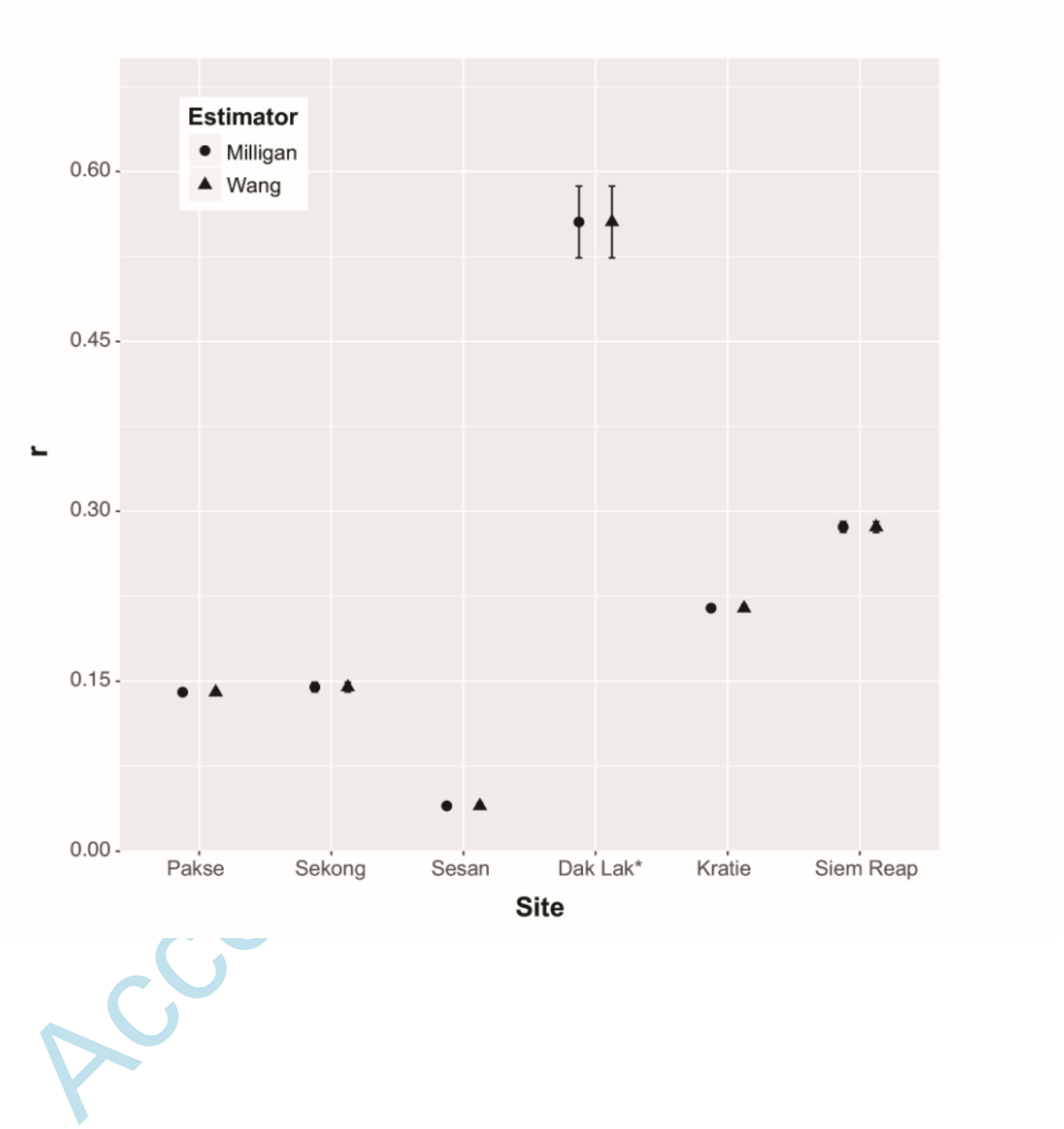

Figure 4

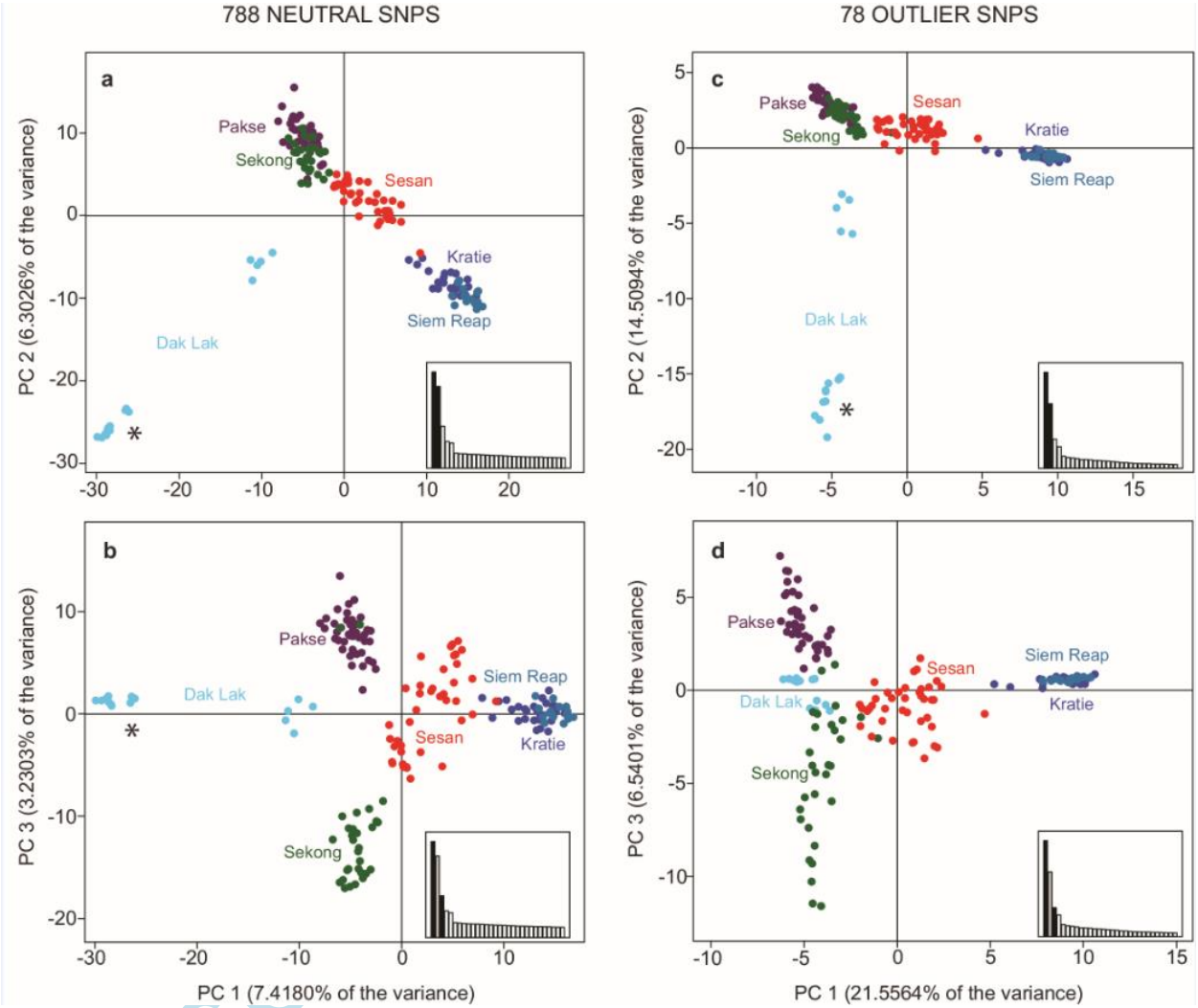

Figure 5