Zoological Studies **64:**18 (2025)

Genome-wide SNP Analyses Reveal High Gene Flow of Endemic Smallscale Croaker (*Boesemania microlepis*) in the Lower Mekong Basin

Quyen Dang Ha Vu[1], Christopher E. Bird[2], Oanh Thi Truong[1], Sang Quang Tran[1], Ellen E. Biesack[3], Amanda S. Ackiss[3], Kent E. Carpenter[3,*], and Binh Thuy Dang[1,*]

[1]*Department of Biotechnology, Institute for Biotechnology and Environment, Nha Trang University, 02 Nguyen Dinh Chieu, Nha Trang, Vietnam. *Correspondence: E-mail: binhdt@ntu.edu.vn (Dang)*
E-mail: quyenvdh@ntu.edu.vn (Vu); oanhtt@ntu.edu.vn (Truong); sangtq@ntu.edu.vn (Tran)

[2]*Department of Life Sciences, Texas A&M University– Corpus Christi, Corpus Christi, TX 78412, USA. E-mail: chris.bird@tamucc.edu (*Bird*)*

[3]*Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529 -0266, USA. *Correspondence: E-mail: kcarpent@odu.edu (*Carpenter*)*
E-mail: ebies001@odu.edu (Biesack); aacki002@odu.edu (Ackiss)

(Received 30 July 2024 / Accepted 6 April 2025 / Published -- 2025)
Communicated by Jen-Pan Huang

ORCiD
Quyen Dang Ha Vu: https://orcid.org/0000-0003-4073-7114
Oanh Thi Truong: https://orcid.org/0000-0002-3933-2948
Sang Quang Tran: https://orcid.org/0000-0002-9586-2433
Christopher E. Bird: https://orcid.org/0000-0003-0228-3318
Kent E. Carpenter: https://orcid.org/0000-0003-3618-1811
Ellen E. Biesack: https://orcid.org/0000-0001-7932-9471
Amanda S. Ackiss: https://orcid.org/0000-0002-8726-7423

The Smallscale Croaker, *Boesemania microlepis*, is a valuable fishery resource in the Mekong River basin that lacks clear biological data to understand its basic ecology and conservation management priorities. This species is common in the southernmost floodplain of the Mekong which extends from Tonlé Sap in Cambodia to the Mekong Delta in Vietnam. This floodplain is of particular biodiversity conservation concern because of the many upstream dams that restrict water flow in an ecology that relies heavily on a flood-pulse cycle. The literature regarding the biology *B. microlepis* in this region contains conflicting ideas that populations are both highly localized and exhibit extensive migratory behavior. We used restriction-site associated DNA to test the hypothesis that localized populations exist in the southernmost floodplain of the Mekong basin. Our data indicates high connectivity among eight sites sampled in this region supporting overall

panmixia. Our results suggest a potential upstream source of propagules for this floodplain and the need for further research to clarify mechanisms driving gene flow.

Keywords: *Boesemania microlepis*, EzRAD, Lower Mekong Basin, Gene flow, Population genetics

Citation: Vu QDH, Bird CE, Truong OT, Tran SQ, Biesack EE, Ackiss AS, Carpenter KE, Dang BT. 2025. Genome-wide SNP analyses reveal high gene flow of endemic Smallscale Croaker (*Boesemania microlepis*) in the Lower Mekong Basin. Zool Stud **64:**18.

BACKGROUND

The Mekong River Basin and its biota face many environmental and conservation challenges (Hughes 2024), and these are amplified in its southernmost floodplain region. The Lower Mekong Basin (LMB) includes the part of the Mekong River and its tributaries from Myanmar south to Vietnam, and the largest continuous extent of floodplain in the LMB is from Tonlé Sap to the Mekong Delta (Grill et al. 2014). This southernmost floodplain of Cambodia and Vietnam supports one of the largest and most biodiverse inland fisheries globally, but it is threatened by changing hydrological conditions imposed from dam construction and unsustainable fishing pressure (Campbell and Barlow 2020; Chan et al. 2020; Vu et al. 2021; Baird and Hogan 2023; Chevalier et al. 2023). Dramatic cyclical flood-pulse water level changes in the LMB floodplain are dampened by the restriction in water flow caused by upstream dams reducing the nutrient availability that supports high productivity and the fisheries this sustains (Pokhrel et al. 2018; Chua et al. 2022; Dang et al. 2022).

Changes in water depth and nutrient availability due to hydrological disruptions, and unsustainable fisheries can particularly impact large, high trophic level fishes such as the Smallscale Croaker (*Boesemania microlepis*) in the LMB floodplain (Chan et al. 2020; Thapa et al. 2024). This species is widespread in Southeast Asia, throughout most of the LMB, Chao Phraya river basin, and in Sumatra island (west Malaysia and Indonesia), can reach up to 100 cm total length, and feeds mostly on crustaceans, insects, gastropods, and small fishes (Rainboth 1996; Poulsen et al. 2004; Baird 2021). It is important in fisheries in the LMB as a highly valued food fish and also prized for its swimbladder (Baird 2001; Vann et al. 2005; Suvarnaraksha 2011). *B. microlepis* is regularly present in catch lists, as well as indicator species for annual yields in important fishing grounds in the LMB, such as Tonlé Sap (Ou 2012), Laos and Cambodia (Baird et al. 2004). While local

ecological knowledge provides substantial information (Poulsen et al. 2004; Baird and Flaherty 2005), the population status of *B. microlepis* remains largely unknown (Baird 2021). Although still common in parts of its reported range, *B. microlepis* is sensitive to pollution and overfishing, leading to notable declines and local extirpations (Baird 2021), including declines of up to 45% in the Mekong Delta (Mai et al. 2018).

Unlike many of the important food fishes in the LMB which are highly migratory such as Mekong giant catfish (*Pangasianodon gigas*) (Mitamura et al. 2006) and *Pangasius krempfi* (Duong et al. 2023), *B. microlepis* appears to mostly migrate short distances according to fisheries surveys and local ecological knowledge (Baird 2001, 2021; Poulsen et al. 2004; Baird and Flaherty 2005). It is thought to migrate during the late dry to early flood season upstream above the Khone Falls and downstream below the Khone Falls and opposite directions in these segments during the late flood to early dry seasons (Poulsen et al. 2004). These short migrations do not involve shoaling or schooling behavior and may be for feeding habits rather than reproduction and may also be influenced by lunar cycles (Poulsen and Valbo-Jørgensen 1999). *B. microlepis* is also presumed to migrate from the Tonlé Sap Lake to the Mekong River at the beginning of the dry season according to fishery data (Baird et al. 2001). Probably because of its large size it is sometimes included among the highly migratory "white fishes" within the Tonlé Sap and Mekong Delta region (Heng et al. 2018; Nguyen et al. 2022). Within the Tonlé Sap basin, it is known to move into shallow grassland and marsh areas during the flood-pulse cycle (Lamberts and Koponen 2008). This species also prefers deep pools, not only during the dry season but also to spawn at the height of the dry season, and it resides permanently in deep pools in southern Laos and northeastern Cambodia (Poulsen et al. 2004) where these deep pools are common (Halls et al. 2013). This relatively sedentary behavior led to speculation that its population structure consists of many localized populations (Poulsen et al. 2004). Deep pools are also common in the Tonlé Sap, Hau and Tien rivers, and *B. microlepis* is reported as common in the deep pools of the Vam Nao in the Delta (Vu et al. 2009). Peak spawning occurs around May, and the larval duration of this species is reported to be 30 days (Sittajaruwat 1995). Currently, the only confirmed spawning locations of this species are seven locations above the Khone Falls in Laos (Baird 2021), and migration or dispersal patterns from these areas remain unknown.

Genetic data is increasingly used to understand the influence of migration or larval dispersal within the confines of riverine topography on the population structure of fishes (Davis et al. 2017; Shelley et al. 2022). This includes the use of Single Nucleotide Polymorphisms (SNPs) derived from Restriction site-associated DNA sequencing (RAD-seq) to investigate population differentiation and genetic diversity of Mekong fishes (Ackiss et al. 2019; Dang et al. 2019; Biesack et al. 2020). Among RAD-seq techniques (GBS, ddRAD, 2bRAD (Peterson et al. 2012)), ezRAD

provides a strong combination of simplicity (ready-made kit), cost-effectiveness, and flexibility, as an appealing choice for working with non-model organisms (Toonen et al. 2013).

Understanding the population structure of *B. microlepis* in the southernmost floodplain region of the LMB is important for conservation management given the many threats to this species. The Tonlé Sap river and the Mekong Delta contains many deep pools (Halls et al. 2013) ideal for *B. microlepis* residency, and its river pathways are both highly dendritic and anastomosing with highly variable river flow in different branches (MRC 2004). This, combined with the hypothesized numerous isolated populations of this species (Poulsen et al. 2004), could require a complex conservation management scheme to avoid local extirpations of unique gene pools. The purpose of this study was to leverage genetic data to test the hypothesis that isolated populations exist in the complex riverscape of the Tonlé Sap-Mekong Delta floodplain.

MATERIALS AND METHODS

Sampling sites and tissue collection

Boesemania microlepis were collected from markets located along local riversides or directly from fishing boats docked at markets from the Bassac (Hau) and Mekong (Tien) Rivers in the Vietnamese Mekong Delta (MD) between 2013 and 2016. Local markets supplied by small fishing boats were targeted, and all vendors were interviewed to confirm that fish were caught locally. A total of 187 individuals were sampled from Hau River (An Giang, Can Tho, and Soc Trang) and Tien River (Dong Thap, Vinh Long, and Ben Tre-Tra Vinh) in the MD. An additional 19 individuals were sampled at a local fish landing in Tonlé Sap (Siem Reap, Cambodia) (Table 2, Fig. 1). Tissue samples were taken from fresh fish and preserved in 95% ethanol immediately after sampling.

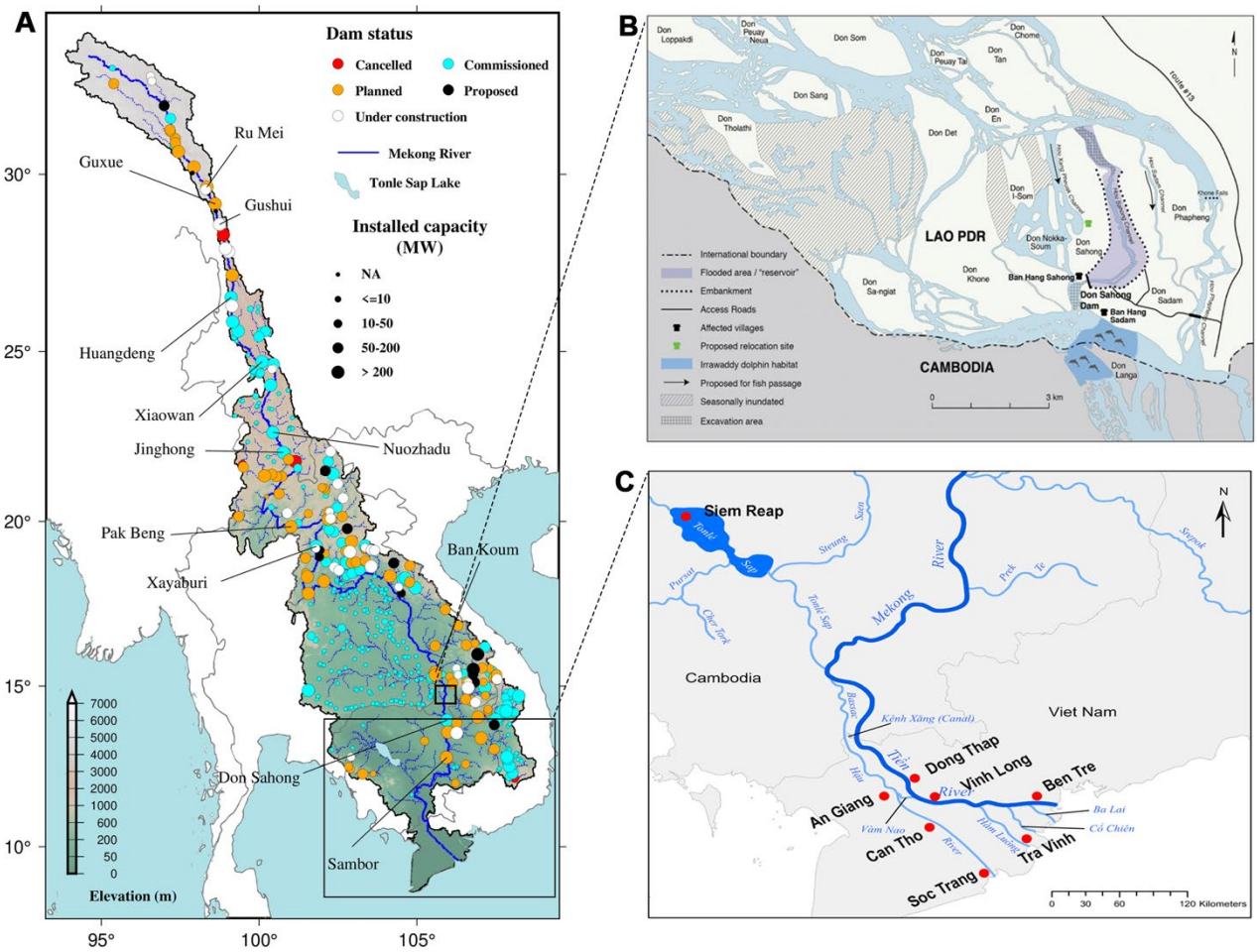

Fig. 1. Sampling map of *Boesemania microlepis* in the Lower Mekong Basin. (A) Dam status on Mekong River; (B) Zoon in "Khone Falls"; (C) Sampling sites (red dots): An Giang (AG), Can Tho (CT), Soc Trang (ST) at Bassac (Hau) river, Dong Thap (DT), Vinh Long (VL), Ben Tre - Tra Vinh (BV) at Mekong (Tien) river, and Siem Reap (SR) at Tonlé Sap, Cambodia.

RAD library preparation

Genomic DNA was extracted from preserved tissue samples using the Qiagen DNeasy Blood and Tissue kit (Hilden, Germany) following the manufacturer's instructions and treated with RNase (100 mg/mL) to remove residual RNA. Extracted DNA was eluted off the filter in four separate aliquots (100 µl elution/time) to aid in the removal of degraded DNA fragments prior to Restriction-site Associated DNA (RAD) library preparation. All four elutions for each sample were inspected by gel electrophoresis (1.5% agarose), and the elution that contained high-quality genomic DNA with the least amount of smaller, degraded DNA fragments was selected. The concentration of this selected elution for each sample was then measured using a Qubit® 2.0 Fluorometer with dsDNA High Sensitivity Assay kit (Invitrogen, Thermo Fisher Scientific, Inc.).

Selected DNA (100 ng, ≥ 3ng/µl) was used for EzRAD library preparation (Toonen et al. 2013; Dang et al. 2019). The library preparation process followed the same protocols detailed in Dang et al. (2019), which included randomly fragmenting the genomic DNA with the

isoschizomeric restriction enzymes *MboI* and *Sau3AI* (New England Biolabs, USA), followed by end-repair, fragment size selection, poly-A tailing, ligation of dual-indexed Illumina adapters, and PCR amplification using the Illumina TruSeq Nano DNA Library Prep kit (Illumina, USA). The libraries were then sent to the Genomics Core Laboratory (Texas A&M University, Corpus Christi, USA) for paired-end sequencing on the Illumina HiSeq 2500/4000 system.

SNP discovery and filtering

Data processing, including sequence quality trimming, *de novo* reference assembly, mapping, and variant calling, was performed using dDocent v2.24 (Puritz et al. 2014) following the protocols outlined in Dang et al. (2019). Raw SNP calls were filtered with vcftools v0.1.11 (Danecek et al. 2011) and vcffilter in vcflib v1.0.0 (Garrison et al. 2021). The SNP filtering process was carried out in several steps. Initially, raw SNPs were filtered to retain only biallelic loci, followed by the removal of insertion/deletion polymorphisms (INDELs). SNPs with a minimum quality score (MinQ > 30) and a mean depth (mean $5 \leq DP \leq 10$) were retained. A maximum missingness threshold (Max-missing of 95%) was applied to include loci with genotype data present in at least 95% of individuals. Further filtering included the application of an allelic balance filter (AB < 0.3) and a mapping quality filter ($0.1 < MQM < 1.1$), as well as the exclusion of SNPs based on paired-end read consistency (PAIR < 0.05). A minor allele frequency threshold ($MAF \geq 0.05$) was applied to ensure the inclusion of common variants. Subsequently, individuals with excessive missing data were excluded. SNPs deviating from Hardy-Weinberg equilibrium (HWE, $p < 0.001$) were removed, and loci were analyzed using the RAD haplotyper tool to eliminate possible paralogous loci. To minimize linkage disequilibrium, one SNP per contig was retained. Finally, a heterozygosity filter (Ho > 0.35) was applied to exclude individuals with excessively excess heterozygosity, resulting in the final dataset used for analyses. The pipeline steps and script descriptions are detailed in Supplementary Index 1.

Outlier loci detection and linkage-disequilibrium (LD) analysis

Outlier loci potentially under selection were identified using LOSITAN (Antao et al. 2008) and BayeScan v2.1 (Foll and Gaggiotti 2008). LOSITAN was run using the infinite alleles mutation model with parameter settings of 'neutral', 500,000 simulations, and a confidence interval of 0.95. BayeScan was run with default parameters, and a false discovery rate (FDR) correction of 0.05.

Linkage disequilibrium (LD) was measured as the squared pairwise correlation coefficient between loci (r^2) calculated using the 'LD' function in the R package 'genetics' (Gregory et al.

2019). Selected Outlier Clusters (SOC) and Compound Outlier Clusters (COC) were identified by LD network analysis using the R package 'LDna' (Kemppainen et al. 2015), and an optimal value of φ and |E|min parameter and LD threshold was set up for SOC. LD networks were constructed using the R package 'igraph' (Csardi and Nepusz 2006).

All loci putatively under selection or exhibiting linkage were removed from the dataset to generate a panel of neutral, unlinked SNPs. A dataset of outlier SNPs identified by all two analyses was also generated to examine samples for selection-associated differentiation.

Relatedness

High levels of relatedness can impact analyses of population structure. We performed simulations using neutral loci dataset of identical twins, parent-offspring, full-siblings, half-siblings, cousins, second cousins, and unrelated individuals (1,000 simulated dyads per category) to evaluate the performance of seven relatedness estimators, and selected the best estimator using COANCESTRY (Wang 2011). The results were visualized using ggplot2 in the tidyverse package (Wickham et al. 2019) showing dyadic is best approach (Fig. S1). Based on the simulations, we delineated upper and lower relatedness value ranges that were consistent with the relatedness categories (identical twins, full-siblings, half-siblings, and unrelated). Any putatively related pairs at the half-sibling level or higher were broken by removing one individual from analyses.

Population genetic analyses

Measures of genetic diversity were generated with the neutral SNP dataset, including numbers of alleles (N_A), effective numbers of alleles (N_E), expected (H_e) and observed (H_o) heterozygosity, and inbreeding coefficients (G_{IS}), were computed in GenoDive v3.04 (Meirmans and Van Tienderen 2004). Genetic differentiation (F_{ST}) between all pairs of *B. microlepis* populations was calculated using ARLEQUIN v3.5 (Excoffier and Lischer 2010). A FDR correction was applied to all *p*-values to mitigate the risk of false positives due to multiple comparisons (Benjamini and Hochberg 1995). An analysis of molecular variance (AMOVA) testing for significant structure within and among individuals and populations was carried out using ARLEQUIN v3.5 with 9,999 permutations.

To investigate population structure, we performed two clustering approaches with different methodologies using both neutral and divergent SNP panels. First, the program Structure v2.3.4 (Pritchard et al. 2000) was used to identify distinct genetic clusters and assign individuals to populations under the assumption of population admixture and correlated allele frequency. Runs consisted of a burn-in period of 5,000 Markov Chain Monte Carlo iterations followed by an

additional 50,000 steps for inferred clusters (K) 1 - 7, with 10 replicates each. Results were collated in STRUCTURE HARVESTER (Earl and Vonholdt 2012), and the ΔK statistic (Evanno et al. 2005) was used to determine the most likely number of clusters. Second, principal components analysis (PCA) was run using the R package "adegenet" (Jombart and Ahmed 2011). These analyses provide a graphic description of the allelic divergence among populations in multivariate space.

Estimates of effective population size (N_e) were generated with NeEstimator v2.1 (Do et al. 2014) using the bias-corrected linkage disequilibrium (LD) method with a minor allele frequency cutoff of 0.05. Effective population size was calculated for the Hau (AG, CT, ST), and Tien (DT, VL, BV) Rivers, the Mekong Delta populations combined, and for all sampled individuals combined into a single population.

Migration patterns

We estimated the direction and magnitude of the gene flow among sampling sites using three methods.

Firstly, gene flow between populations was estimated using Bayesian inference implemented in MIGRATE-n v3.6.11 (Beerli and Felsenstein 2001). Sample sizes were reduced for each population to obtain 250 loci genotyped in 100% of individuals used for the analysis. The run was performed using 500,000 recorded genealogies sampled every 100 steps, preceded by a burn-in of 20,000. Four hot chains were used with temperatures: T1 = 1.0, T2 = 1.5, T3 = 3.0 and T4 = 1.0×10^6. After optimization, the maximum mutation-scaled effective populations size (θ) prior was set at 0.1 while the maximum mutation-scaled migration (M) prior was set at 20,000. Ten hypotheses of migration among populations were tested: (1) symmetric migration rates between all sites (Panmixia Model), (2) non-symmetric migration rates between all sites (Full Model), (3) migration between all sites within each of the rivers (Hau and Tien), but no migration between rivers (Rivers Separate), (4) migration occurring only between neighboring, downstream sites and between rivers (Downstream Open), (5) migration occurring only between neighboring, downstream sites but no migration between rivers (Downstream Closed), (6) migration occurring only between neighboring, upstream sites and between rivers (Upstream Open), (7) migration occurring only between neighboring, upstream sites but no migration between rivers (Upstream Closed), (8) migration occurring among all sites found in each river, however migration only occurs from the Tien to Hau River sites (Tien Source), (9) migration occurring among all sites found in each river, however migration only occurs from the Hau to Tien River sites (Hau Source), (10) migration occurring among all sites, but no migration to Siem Reap. The most likely model was

chosen using the Bezier approximation score produced by Migrate-n and migrants per generation for the chosen model were calculated according to Beerli (2009).

Second, the relative migration levels and migration patterns were estimated based on neutral SNPs using the divMigrate function (Sundqvist et al. 2016) in the R package "Diversity" (Keenan et al. 2013). Gene flow patterns were visualized using network graphics produced using the R package "qgraph" (Epskamp et al. 2012). All simulations of network graphics were run following the instructions on GitHub (https://github.com/quangsang52sh/pyMigSim), with 10,000 repetitions employing cosine distance calculations in a Python script. The best simulation from these migrations was selected based on statistical analysis using the Ordinary Least Squares (p-value ≤ 0.05) method.

Thirdly, TreeMix (Pickrell and Pritchard 2012) – an algorithm simultaneously infers a tree of relationships and migration events – was applied to infer the split and admixture among sampling sites. Based on the allele frequency data, an unrooted maximum likelihood population tree was approximated. The stepwise likelihood method was subsequently used to examine the effect of migration events (m = 0–7) on the residual covariance matrix. The significance of migration events was tested using a jackknife approach. In order to determine the optimal number of migration events (m-model), *ad hoc* statistic (Δm value) (Fitak 2021) was estimated from the maximum composite likelihood.

RESULTS

SNP filtering, outlier detection, and tests for linkage

A total of 206 ezRAD libraries (across 7 sample sites: Siem Reap (SR), An Giang (AG), Can Tho (CT), Soc Trang (ST), Dong Thap (DT), Vinh Long (VL), Ben Tre - Tra Vinh (BV)) of *B. microlepis* were sequenced. After filtering, 113 individuals were successfully genotyped at 638 validated SNPs. Nine individuals exhibited excess heterozygosity (Ho > 0.35) and probable contamination were dropped from the dataset. Information on number of SNPs and individuals removed at each step of filtering and data analysis is presented in table S1 and table S2.

LOSITAN identified 47 putative outlier loci, 38 loci putatively under balancing selection and 9 loci putatively under divergent selection. BayeScan identified 12 SNPs as outliers (FDR ≤ 0.05). LD network presented 01 SOC including 47 loci (φ = 5 and |E|min = 0, λmin = 0.822) (Fig. S2). In total, 97 loci were removed, and the panel of 541 remaining loci was assumed to be comprised of

neutral, unlinked SNPs. A dataset of eight loci identified as divergent outliers by all three approaches was used for additional analyses.

Relatedness

Relatedness analysis detected 3 pairs of putative full siblings and 2 pairs of putative half siblings. One individual from each putative full sibling and half sibling were removed to obtain a final dataset of 99 genotyped individuals used in population genetic analyses (Table 1).

Table 1. Related pairs identified by the Dyadml likelihood estimator with 95% confidence intervals. The most likely relationship for each pair is also shown

Specimen Pairs	Removed	Dyadml (95% CI)	Relationship
SR311-SR318	SR311	0.3971 (0.3012–0.4963)	Full sibling
ST309-ST327	ST309	0.335 (0.2414–0.4332)	Full sibling
VL202-VL204	VL202	0.3311 (0.2313–0.4354)	Full sibling
AG117-CT113	AG117	0.1872 (0.0881–0.2835)	Half sibling
ST307-ST310	ST307	0.1841 (0.0851–0.2469)	Half sibling

Abbreviations: SR, Siem Reap; ST, Soc Trang; VL, Vinh Long; AG, An Giang; and CT, Can Tho

Population genetic analysis

Estimates of genetic diversity in *B. microlepis* are presented in table 2. The mean observed and expected heterozygosity of the populations were 0.262 and 0.272, respectively. Observed heterozygosity within sites ranged from 0.251 (DT) to 0.268 (ST) and expected heterozygosity from 0.258 (AG) to 0.288 (BV). The average number of alleles (N_A) was 1.875; and effective number of alleles (N_E) was 1.431. Inbreeding coefficients ranged from 0.008 (AG) to 0.089 (DT), with an overall G_{IS} for all individuals at 0.038.

Table 2. *Boesemania microlepis* sample site information and genetic diversity indices

River location		Sampling sites (population code)	Geographic coordinates		N	Nse	N_A	N_E	*Ho*	*He*	G_{IS}
Tonlé Sap - Cambodia		Siem Reap (SR)	13°0'N	104°3'E	19	15	1.904	1.438	0.266	0.275	0.030
Mekong Delta - Vietnam	Hau River	An Giang (AG)	10°28'N	105°12'E	30	12	1.865	1.421	0.253	0.258	0.008
		Can Tho (CT)	10°02'N	105°45'E	38	15	1.898	1.437	0.265	0.276	0.033
		Soc Trang (ST)	9°32'N	105°56'E	29	24	1.959	1.437	0.268	0.276	0.027
	Tien River	Dong Thap (DT)	10°35'N	105°36'E	29	6	1.756	1.417	0.251	0.273	0.089
		Vinh Long (VL)	10°06'N	106°01'E	26	15	1.867	1.427	0.258	0.267	0.034
		Ben Tre – Tra Vinh (BV)	10°08'N 9°47'N	106°29'E 106°20'E	35	12	1.878	1.438	0.266	0.288	0.070
Overall					206	99	1.875	1.431	0.262	0.272	0.038

Abbreviations: N, Number of samples collected; Nse, number of individuals successfully genotyped and relatedness removed used in analyses; N_A, number of alleles; N_E, effective number of alleles; *Ho*, observed heterozygosity; *He*, expected heterozygosity; and G_{IS}, inbreeding coefficient.

Pairwise genetic differentiation (F_{ST}) among sites ranged from < 0.001 to 0.003, with all pairwise comparisons showing non-significance ($p > 0.05$) (Table 3). Hierarchical AMOVA results (Table 4) showed the majority of the variation in *B. microlepis* was found among individuals within populations (16%), and within individuals (84%), and highly significant in both cases ($p = 0.001$), suggesting high connectivity among geographic defined populations.

Table 3. Pairwise genetic differentiation (below the diagonal) and respective p-values (above the diagonal) between sampled sites in the Lower Mekong Basin using neutral SNP panels
Sample site codes are given in table 2

Population	SR	AG	DT	CT	VL	ST	BV
SR	--	0.371	0.371	0.262	0.241	0.293	0.369
AG	< 0.001	--	0.418	0.410	0.400	0.310	0.418
DT	< 0.001	< 0.001	--	0.393	0.347	0.371	0.392
CT	0.003	< 0.001	< 0.001	--	0.369	0.267	0.398
VL	0.003	< 0.001	< 0.001	< 0.001	--	0.284	0.393
ST	0.001	0.001	< 0.001	0.002	0.002	--	0.312
BV	< 0.001	< 0.001	< 0.001	< 0.001	<0.001	0.002	--

Table 4. Hierarchical analysis of molecular variance (AMOVA) in *Boesemania microlepis* using the neutral SNP panels

Source of variation	Degrees of freedom	Sum of square	Variant components	% of variation	Fixation index	*p* value
Among populations	6	571.415	0	0	F_{ST} = -0.001	0.621
Among individuals within populations	92	9040.838	13.261	16	F_{IS} = 0.156	0.001
Within individuals	99	7103.000	71.747	84	F_{IT} = 0.155	0.001
Total	197	16715.253	85.009	100		

Two population clustering methods using the neutral SNP panel of *B. microlepis* produced concordant results indicating highly connectivity across the sampling locations. STRUCTURE showed high admixture (Fig. S3) and PCA revealed no detectable allelic divergence among the southernmost populations in the LMB (Mekong Delta – Tonlé Sap Lake) (Fig. 2A). When repeated with the dataset of eight putative outlier loci, PCA showed a similar pattern to the neutral loci (Fig. 2B).

Estimates of effective population size (*N*e) for proposed groups ranged from 1,732 (Tien River) to 12,268 (Hau River) and 4,764 for all sites combined, indicating a relatively large population of *B. microlepis* in the sampled region of the LMB (Table 5).

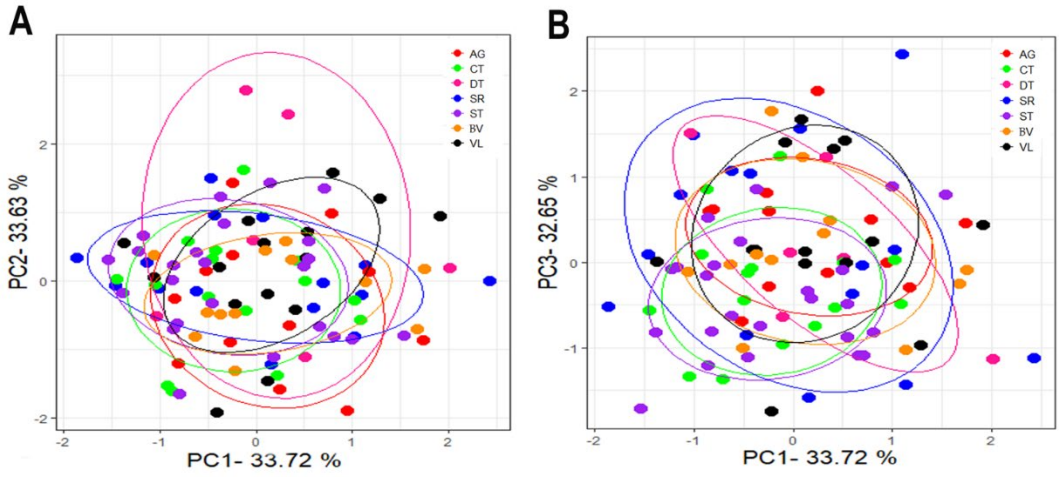

Fig. 2. Principal Component Analysis (PCA) of *B. microlepis* population structure in the Lower Mekong Basin using neutral SNP panels (A) and outlier loci dataset (B). Dots represent individuals and colored by sampling sites (see Fig. 1 for site codes). Ellipses contain 95% of the variation. Percentage of variance explained by each principal component is shown on the axes.

Table 5. Estimates of *Boesemania microlepis* effective population size (N_E) calculated from 541 neutral SNPs. Sample sizes (N) are presented in parentheses with the name of each analysis and 95% confidence intervals (CI) are presented in parentheses with estimates of *N*e

Dataset (N)	N_E (95% CIs)
Hau River (51)	Infinite (1,875 - infinite)
Tien River (33)	4,876 (846 - infinite)
Mekong Delta (84)	14,928 (2,373 - infinite)
All sites (99)	17,741 (2,850 - infinite)

Migratory patterns

The results of MIGRATE-n analysis strongly supported the Panmixia model out of 10 hypothesized models we tested based on the highest Bezier approximation score (ln = -60587.71), in which migration was maintained among all sites with random mating between individuals (Table 6). Significant relative rate (a–b) from divMigrate (Fig. S4) was also supported the highly migratory pattern between geographic defined populations. Using Treemix, two migration events were supported based on Δm value (Fig. S5). Both downstream migrations were observed from an intermediate site between VL and SR toward CT and ST (Fig. 3).

Table 6. Migrate-n results for ten models testing historical migration of *Boesemania microlepis* between sites. Ranks were assigned using marginal log-likelihoods based on the Bezier approximation score

Model #	Model name	Bezier ln	Rank
1	Full	-62200.68	10
2	**Panmixia**	**-60587.71**	**1**
3	River Separate	-61148.70	2
4	Downstream Open	-62895.64	9
5	Downstream Closed	-61285.79	4
6	Upstream Open	-61160.99	3
7	Upstream Closed	-61691.55	5
8	Tien Source	-61704.65	6
9	Hau Source	-62396.69	8
10	Siem Reap Separate	-61849.69	7

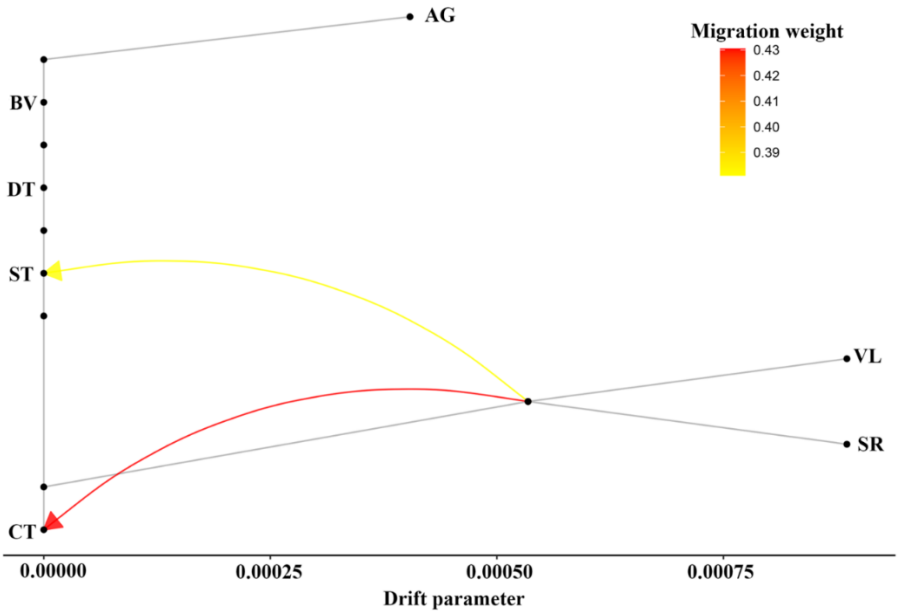

Fig. 3. Maximum likelihood tree of the genetic relationships among the *Boesemania microlepis* populations with two migration events (m = 2) inferred from the TreeMix analysis. The graph shows the topology and branch lengths according to the drift parameter. Migration arrows depict potential gene flow events that increase the likelihood of the tree, color coded according to their weight. Intermediate nodes are hypothetical ancestral populations that are inferred but were not sampled. Abbreviations at nodes are sample sites as depicted in figure 1.

DISCUSSION

The hypothesis that there are many isolated populations of *B. microlepis* sheltering in the numerous deep pools of the LMB (Poulsen et al. 2004; Baird 2021) is not supported by our data. This is despite the observation that there are many deep pools in the southernmost floodplains of the LMB including the Tonlé Sap River and in the Mekong Delta (Halls et al. 2013). The observed panmixia in our data indicates that either adult migration or passive dispersal of larvae is taking place within this population. However, after many years of observations on the biology of this

species active adult migration in this species is limited to only short distances (Poulsen et al. 2004, Baird 2021), making long range migration an unlikely explanation for the observed panmixia. This aligns with findings of panmixia in *Polynemus melanochir* (Dang et al. 2019) and *Anabas testudineus* (Pham and Duong 2014) within the Mekong Delta, suggesting high gene flow likely facilitated by the delta's river/canal system.

Passive early life stage dispersal originating upstream from the region we sampled is an alternative explanation for the observed panmixia since active widespread adult migration is not considered likely. The reproductive biology of *B. microlepis* is largely unstudied except above the Khone Falls where there are seven confirmed spawning areas (Baird et al. 2001). Passive dispersal from reproductive sites above Khone Falls is feasible but it is also possible that the numerous deep pools below Khone Falls and north of Kratie (Halls et al. 2013) could also serve as breeding grounds for this species. The Treemix results (Fig. 3) support the notion that migration to both Siem Reap and the Mekong Delta originates at some unsampled point between these two areas and a potential explanation is that this unknown point is at the confluence of the Tonlé Sap River and the Mekong River. This confluence could receive dispersal propagules originating upstream that ultimately flow into the Tonlé Sap during the wet season and into the Mekong Delta year-round. Peak spawning for this species is reported in May (Poulsen and Valbo-Jørgensen 1999) at the beginning of the wet season when flow into Tonlé Sap from the Mekong tributaries. We suggest that future studies expand our population genetic sample to points north of the Tonlé Sap-Mekong rivers confluence. This would allow for a direct test of the hypothesis that panmixia in the southernmost floodplains extends to the concentrated areas of deep pools in the Mekong River between Kratie and Khone Falls (Halls et al. 2013).

The contribution of reported seasonal migrations (upstream during late flood to early dry season, and downstream during late dry to early flood season, Poulsen et al. (2004)) to the observed panmixia in the southern LMB remains unclear. While *B. microlepis* is generally considered to exhibit limited, localized feeding movements (Poulsen and Valbo-Jørgensen 1999), existing reproductive biology data originates from studies above the Khone Falls (Baird et al. 2001; Baird 2021), potentially not representative of the southern floodplains. Our knowledge of migration in *B. microlepis* depends mostly on local ecological knowledge – LEK (Poulsen et al. 2004; Baird and Flaherty 2005) and fragmentary fishery landing data. Although LEK data is abundant in the Mekong and considered reasonably reliable (Valbo-Jørgensen and Poulsen 2000; Baird 2007; Silvano and Valbo-Jørgensen 2008), it is possible that *B. microlepis* may rely on deep channels or parts of the river that are not easily observed and are capable of evading fishing methods used by local fishermen during potential migrations, especially in the southernmost floodplain region. Landings by local fishermen are a prominent source of LEK data. In addition, juveniles of this

species have been observed aggregating near shore in some areas (Baird et al. 2001) and movement of juveniles enhanced by river flow may also facilitate mixing of individuals from different areas. Additional sampling of this species throughout Laos and Cambodia for genetic analyses may help elucidate the extent of panmixia within the LMB. Tagging and reproductive biology data may also help to understand the limits of migration in this species. These additional data may be particularly useful in the Tonlé Sap and Mekong Delta floodplain where these species have been reported as being highly migratory based on limited data (Heng et al. 2018; Nguyen et al. 2022).

In addition to increased spatial sampling and corroborating migratory studies, our dataset could be improved by more comprehensive genome-sampling to increase the relatively small number of loci in our study. The number of neutral loci sampled in our study (541) should be able to detect discrete differences among populations that have been reproductively isolated for many generations (Morin et al. 2009), but they may be insufficient to detect evolutionarily recent patterns of population structure or populations experiencing higher levels of gene flow, both of which may be relevant in this species. For example, studies examining a post-Pleistocene radiation of salmonids in the Laurentian Great Lakes were first able to detect clear differentiation among species with RAD-seq panels comprised of ~30,000 SNPs where for decades mitochondrial and microsatellite loci could not (Ackiss et al. 2020; Lachance et al. 2021). While the hundreds of SNPs genotyped here is larger than a typical microsatellite panel comprised of tens of loci, simulations have shown SNP panels up to 10,000 loci are unable to fully resolve populations at shorter divergence times, particularly when there was a high proportion (~20%) of low minor allele frequency SNPs (Haasl and Payseur 2011). Future research on *B. microlepis* that samples a larger portion of the genome could confirm if the patterns we found in our study are true representations of gene flow in this species or an artifact of the number of loci genotyped.

The conservation and management implications of our findings are complex. Our rejection of the hypothesis that numerous isolated populations may exist for this species (Poulsen et al. 2004) in the southernmost floodplains of the LMB simplifies the management strategy. Conservation of many discrete populations would require more complicated management schemes to ensure that unique gene pools are not threatened. In addition, our data indicates that the panmictic population in these floodplains has healthy levels of genetic diversity, and effective population sizes appear to be large at present (Tables 2 and 5). It is notable, however, potential consequence of population isolation, driven by anthropogenic pressures including dam construction, intensive agriculture, urbanization, and climate change impacts like sea level rise and altered river flow regimes, is the increased likelihood of encountering full and half siblings. This suggests that recruitment is occurring from increasingly restricted gene pools, leading to higher levels of inbreeding. This reinforces the necessity of proactive conservation measures for reproductive sources (*e.g.*, Ozerov

et al. 2015; Cossu et al. 2021). If long range migration of adults around the southernmost floodplain does occur and is the cause of panmixia despite what is reported about their migratory habits, then the stock could be managed in its entirety. However, if recruitment within the southernmost floodplain depends on successful recruitment of larvae or juveniles from upstream reproductive events then this southernmost region could be considered a population sink, and management of upstream populations would be most important. In this instance, additional measures may be needed to ensure that flow of water in the Mekong River is not impeded as this flow facilitates recruitment to the southernmost floodplains. This precautionary note is particularly relevant since many recent studies warn of increasing consequences of continuing damming of the Mekong River and the LMB (Campbell and Barlow 2020; Chan et al. 2020; Vu et al. 2021; Baird and Hogan 2023; Chevalier et al. 2023).

Although valuable new population information has been gained for *B. microlepis*, particularly in the LMB, additional population and reproductive biology studies are still needed. Despite the wealth of LEK data available for this species, it appears as if our understanding of the population biology of *B. microlepis* remains somewhat disjunction between the regions of its range that have been more thoroughly studied. This has impacted extinction risk assessments for this species as well as a previous assessment considered it Near Threatened but a more recent assessment pronounced it Data Deficient (Baird 2021). The Smallscale Croaker is a valued fishery resource in the LMB, and support should be allocated to address the population and reproductive biology questions that remain.

Acknowledgments: We thank the many people from the communities of the Mekong River Delta who helped us collect tissues from fish markets and team members of Biodiversity and Conservation, Institute for Biotechnology and Environment, Nha Trang University for project support. This project was funded by the United States Agency for International Development supported Partnerships for Enhanced Research Project 2–7 under USAID Cooperative Agreement AID-OAA-A-11-00012.

Authors' contributions: QVDH collected specimens, performed analyses and wrote the manuscript. CEB, EEB, and ASA participated in data analysis, and edited the manuscript. OTT collected specimens, participated in data analysis and edited the manuscript. STQ performed analyses and edited the manuscript. KEC collected specimens, wrote and edited the manuscript. BTD project PI, collected specimens, participated in data analysis, wrote and edited the manuscript.

Competing interests: All authors declare that they have no competing interests.

Availability of data and materials: Requests for access to raw data should be directed to the corresponding author.

Consent for publication: All authors consent to the publication of this manuscript.

Ethics approval consent to participate: All of the fish samples utilized in this study were collected from local fishermen or market vendors. As food fishes, they were dead at the time of sampling; therefore, no ethical approval was required for sampling.

REFERENCES

Ackiss AS, Dang BT, Bird CE, Biesack EE, Chheng P et al. 2019. Cryptic lineages and a population dammed to incipient extinction? Insights into the genetic structure of a Mekong River catfish. J Hered **110:**535–547. doi:10.1093/jhered/esz016.

Ackiss AS, Larson WA, Stott W. 2020. Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes **13:**1037–1054. doi:10.1111/eva.12919.

Antao T, Lopes A, Lopes RJ, Beja-pereira A, Luikart G. 2008. LOSITAN: A workbench to detect molecular adaptation based on a Fst - outlier method. BMC Bioinformatics **5:**1–5. doi:10.1186/1471-2105-9-323.

Baird I. 2021. *Boesemania microlepis*. IUCN Red List Threat Species e.T181232A1711758. doi:10.2305/IUCN.UK.2021-1.RLTS.T181232A1711758.en.

Baird IG. 2001. The ecology and conservation of the smallscale croaker *Boesemania microlepis* (Bleeker 1858-59) in the mainstream Mekong River, Southern Laos. Nat Hist Bull Siam Soc **49:**161–176.

Baird IG. 2007. Local Ecological Knowledge and Small-Scale Freshwater Fisheries Management in the Mekong River in Southern Laos. *In*: Putting fishers' knowledge to work. pp 87–99.

Baird IG, Flaherty MS. 2005. Mekong River Fish Conservation Zones in southern Laos: Assessing effectiveness using local ecological knowledge. Environ Manage **36:**439–454. doi:10.1007/s00267-005-3093-7.

Baird IG, Flaherty MS, Phylavanh B. 2004. Mekong River Pangasiidae catfish migrations and the Khone Falls wing trap fishery in southern Laos. Nat Hist Bull Siam Soc **52:**81–109.

Baird IG, Hogan ZS. 2023. Hydropower dam development and fish biodiversity in the Mekong

River basin: A review. Water **15:**1352. doi:10.3390/w15071352.

Beerli P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? *In*: Bertorelle G, Bruford MW, Hauffe HC et al. (ed) Population Genetics for Animal Conservation. Cambridge University Press, Cambridge, pp. 42–79.

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proc Natl Acad Sci U S A **98:**4563–4568. doi:10.1073/pnas.081068098.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B **57:**289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Biesack EE, Dang BT, Ackiss AS, Bird CE, Chheng P et al. 2020. Evidence for population genetic structure in two exploited Mekong River fishes across a natural riverine barrier. J Fish Biol **97:**696–707. doi:10.1111/jfb.14424.

Campbell I, Barlow C. 2020. Hydropower development and the loss of fisheries in the Mekong River basin. Front Environ Sci **8:**1–10. doi:10.3389/fenvs.2020.566509.

Chan B, Brosse S, Hogan ZS, Ngor PB, Lek S. 2020. Influence of local habitat and climatic factors on the distribution of fish species in the Tonle Sap Lake. Water **12:**786. doi:10.3390/w12030786.

Chevalier M, Ngor PB, Pin K, Touch B, Lek S et al. 2023. Long-term data show alarming decline of majority of fish species in a Lower Mekong basin fishery. Sci Total Environ **891:**164624. doi:10.1016/j.scitotenv.2023.164624.

Chua SDX, Lu XX, Oeurng C, Sok T, Grundy-Warr C. 2022. Drastic decline of flood pulse in the Cambodian floodplains (Mekong River and Tonle Sap system). Hydrol Earth Syst Sci **26:**609–625. doi:10.5194/hess-26-609-2022.

Cossu P, Mura L, Scarpa F, Lai T, Sanna D et al. 2021. Genetic patterns in *Mugil cephalus* and implications for fisheries and aquaculture management. Sci Rep **11:**1–10. doi:10.1038/s41598-021-82515-7.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal Complex Syst **1695:**1–9.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E et al. 2011. The variant call format and VCFtools. Bioinformatics **27:**2156–2158. doi:10.1093/bioinformatics/btr330.

Dang BT, Vu QHD, Biesack EE, Doan TV, Truong OT et al. 2019. Population genomics of the peripheral freshwater fish *Polynemus melanochir* (Perciformes, Polynemidae) in a changing Mekong Delta. Conserv Genet **20:**961–972. doi:10.1007/s10592-019-01189-x.

Dang H, Pokhrel Y, Shin S, Stelly J, Ahlquist D et al. 2022. Hydrologic balance and inundation

dynamics of Southeast Asia's largest inland lake altered by hydropower dams in the Mekong River basin. Sci Total Environ **831:**154833. doi:10.1016/j.scitotenv.2022.154833.

Davis CD, Epps CW, Flitcroft RL, Banks MA. 2017. Refining and defining riverscape genetics: How rivers influence population genetic structure. Wiley Interdiscip Rev Water **5:**1–20. doi:10.1002/WAT2.1269.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ et al. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour **14:**209–214. doi:10.1111/1755-0998.12157.

Duong TY, Nguyen NTT, Tran DD, Le TH, Nor SAM. 2023. Multiple genetic lineages of anadromous migratory Mekong catfish *Pangasius krempfi* revealed by mtDNA control region and cytochrome *b*. Ecol Evol **13:**1–14. doi:10.1002/ece3.9845.

Earl D, Vonholdt B. 2012. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour **4:**1–3. doi:10.1007/s12686-011-9548-7.

Epskamp S, Cramer A, Waldorp L, Schmittmann, VD Borsboom D. 2012. qgraph: Network Visualizations of Relationships in Psychometric Data. J Stat Softw **48:**1–18.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol **14:**2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour **10:**564–567. doi:10.1111/j.1755-0998.2010.02847.x.

Fitak RR. 2021. OptM: Estimating the optimal number of migration edges on population trees using Treemix. Biol Methods Protoc **6:**1–6. doi:10.1093/biomethods/bpab017.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics **180:**977–993. doi:10.1534/genetics.108.092221.

Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. 2021. Vcflib and tools for processing the VCF variant call format. bioRxiv 2021–05. doi:10.1101/2021.05.21.445151.

Gregory W, Gorjanc G, Leisch F, Man M. 2019. Genetics: Population Genetics. R package version 1.3.8.1.

Grill G, Ouellet Dallaire C, Fluet Chouinard E, Sindorf N, Lehner B. 2014. Development of new indicators to evaluate river fragmentation and flow regulation at large scales: A case study for the Mekong River basin. Ecol Indic **45:**148–159. doi:10.1016/j.ecolind.2014.03.026.

Haasl RJ, Payseur BA. 2011. Multi-locus inference of population structure: a comparison between

single nucleotide polymorphisms and microsatellites. Heredity **106:**158–171. doi:10.1038/hdy.2010.21.

Halls AS, Conlan I, Wisesjindawat W, Phouthavongs K et al. 2013. Atlas of deep pools in the Lower Mekong River and some of its tributaries. Mekong River Commission, Phnom Penh, Cambodia, 69 pp. ISSN: 1683-1489.

Heng K, Chevalier M, Lek S, Laffaille P. 2018. Seasonal variations in diet composition, diet breadth and dietary overlap between three commercially important fish species within a flood-pulse system: The Tonle Sap Lake (Cambodia). PLoS ONE **13:**1–16. doi:10.1371/journal.pone.0198848.

Hughes K. 2024. The Mekong's forgotten fishes and the emergency recovery plan to save them. WWF, Gland, Switzerland.

Jombart T, Ahmed I. 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics **27:**3070–3071. doi:10.1093/bioinformatics/btr521.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol **4:**782–788. doi:10.1111/2041-210X.12067.

Kemppainen P, Hlaing T, Walton C, Somboon P, Knight CG et al. 2015. Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. Mol Ecol Resour **15:**1031–1045. doi:10.1111/1755-0998.12369.

Lachance H, Ackiss AS, Larson WA, Vinson MR, Stockwell JD. 2021. Genomics reveals identity, phenology and population demographics of larval ciscoes (*Coregonus artedi*, *C. hoyi*, and *C. kiyi*) in the Apostle Islands, Lake Superior. J Great Lakes Res **47:**1849–1857. doi:10.1016/j.jglr.2021.09.012.

Lamberts D, Koponen J. 2008. Flood pulse alterations and productivity of the Tonle Sap ecosystem: A model for impact assessment. Ambio **37:**178–184. doi:10.1579/0044-7447(2008)37[178:FPAAPO]2.0.CO;2.

Mai V Van, Nguyen TV, Nguyen HL, Nguyen NH, Dang TP. 2018. The smallscale croaker *Boesemania microlepis* (Bleeker 1858) exploitation status on the Bassac river, Mekong Delta. Can Tho Univ J Sci **54:**108–114. doi:10.22144/ctu.jvn.2018.102.

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes **4:**792–794. doi:10.1111/j.1471-8286.2004.00770.x.

Mitamura H, Mitsunaga Y, Arai N, Viputhanumas T. 2006. Comparison of two methods of attaching telemetry transmitters to the Mekong giant catfish, *Pangasianodon gigas*. Zoolog Sci **23:**235–238. doi:10.2108/zsj.23.235.

Morin PA, Martien KK, Taylor BL. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour **9**:66–73. doi:10.1111/j.1755-0998.2008.02392.x.

MRC. 2004. Tonle Sap Lake and Vicinities Project. Consolidation of Hydrometeorological Data and Multi-Functional Hydrologic Roles of Tonle Sap Lake and Vicinities, Phase III

Nguyen XD, Veettil BK, Ngo XQ, Nguyen MT. 2022. Fish species composition, diversity, and migration in the Mekong Delta: a study in the Cua Tieu River, Vietnam. Environ Monit Assess **194:**769. doi:10.1007/s10661-022-10174-w.

Ou R. 2012. Fish Trade on Fishing Products in Tonle Sap Great Lake. Int J Environ Rural Dev **3:**49–54

Ozerov M, Jürgenstein T, Aykanat T, Vasemägi A. 2015. Use of sibling relationship reconstruction to complement traditional monitoring in fisheries management and conservation of brown trout. Conserv Biol **29:**1164–1175. doi:10.1111/cobi.12480.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE **7:**e37135. doi:10.1371/journal.pone.0037135.

Pham TTN, Duong TY. 2014. Genetic diversity of climbing perch (*Anabas testudineus*, Bloch 1792) populations based on RAPD and ISSR markers. CTU J Sci **Special issue:**101–108

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet **8:**e1002967. doi:10.1371/journal.pgen.1002967.

Pokhrel Y, Shin S, Lin Z, Yamazaki D, Qi J. 2018. Potential Disruption of Flood Dynamics in the Lower Mekong River Basin Due to Upstream Flow Regulation. Sci Rep **8:**1–13. doi:10.1038/s41598-018-35823-4.

Poulsen AF, Hortle KG, Valbo-Jørgensen J, Chan S, Chhuon CK et al. 2004. Distribution and ecology of some important riverine fish species of the Mekong River Basin. MRC Technical Paper No. 10. ISSN: 1683-1489.

Poulsen AF, Valbo-Jørgensen J. 1999. Fishermen's knowledge as background information in tropical fish ecology: A quantitative comparison with fish sampling results. Surv. manual-AMFP Rep. 1/99:23.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155:**945–959. doi:10.1093/genetics/155.2.945.

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ **2:**e431. doi:10.7717/peerj.431.

Rainboth WJ. 1996. Fishes of the Cambodian Mekong. FAO Species Identification Field Guide for Fishery Purposes, Rome. 265 p., 27 colour plates.

Shelley JJ, Holland OJ, Swearer SE, Dempster T, Le Feuvre MC, Sherman CDH, Miller AD. 2022. Landscape context and dispersal ability as determinants of population genetic structure in freshwater fishes. Freshw Biol **67:**338–352. doi:10.1111/fwb.13844.

Silvano RAM, Valbo-Jørgensen J. 2008. Beyond fishermen's tales: contributions of fishers' local ecological knowledge to fish ecology and fisheries management. Environ Dev Sustain **10:**657–675. doi:10.1007/s10668-008-9149-0.

Sittajaruwat D. 1995. Study on the development of stomach and first feeding of drumfish, *Boesemania microlepis* (Bleeker). Kasetsart University, Bangkok (Thailand).

Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D. 2016. Directional genetic differentiation and relative migration. Ecol Evol **6**:3461–3475. doi:10.1002/ece3.2096.

Suvarnaraksha A. 2011. Biology of two keystone fish species and fish assemblage patterns and modeling approaches in tropical river basin: case study of Ping river basin, Thailand. Evolution & Diversité Biologique (Toulouse) & Ubon Ratchathani University (Ubon Ratchathani).

Thapa J, English M, Ou C, Lanoue J, Walker RH. 2024. Assessment of the floodplain fishery of Tonle Sap Lake in Cambodia. Fish Manag Ecol **31:**e12683. doi:10.1111/fme.12683.

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-silva I, Andrews KR, Bird CE. 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ **1:**e203. doi:10.7717/peerj.203.

Valbo-Jørgensen J, Poulsen AF. 2000. Using Local Knowledge as a Research Tool in the Study of River Fish Biology: Experiences from the Mekong. Environ Dev Sustain **2:**253–376. doi:10.1023/A:1011418225338.

Vann LS, Baran E, Phen C, Thang TB. 2005. Biological reviews of important Cambodian fish species, based on Fishbase 2004. WorldFish Center Contribution.

Vu A V., Hortle KG, Nguyen DN. 2021. Factors driving long term declines in inland fishery yields in the Mekong delta. Water **13:**1005. doi:10.3390/w13081005.

Vu VA, Nguyen ND, Hidas E, Nguyen MN. 2009. Vam Nao deep pools: a critical habitat for *Pangasius krempfi* and other valuable species in the Mekong Delta, Vietnam. Asian Fish Sci **22**:631–639.

Wang J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol Ecol Resour **11:**141–145. doi:10.1111/j.1755-0998.2010.02885.x.

Wickham H, Averick M, Bryan J, Chang W, McGowan L et al. 2019. Welcome to the Tidyverse. J Open Source Softw **4:**1686. doi:10.21105/joss.01686.

Supplementary materials

Fig. S1. Coancestry simulation analysis showing stimulated relatedness values following the redefined relatedness categories (A) and difference in averages between the estimators (B). (download)

Fig. S2. Graphical representation of selection test results for *Boesemania microlepis*. (A) Values are plotted against expected heterozygosity (H_e) in Lositan with different colors showing balancing, neutral, or positive patterns of selection (yellow, gray, and red, respectively); (B) F_{ST} values are plotted against the log 10 of posterior odds (PO) for all loci in BayeScan, where values of log 10 PO > 1.25 indicate the threshold value for identifying loci under selection; (C) All λ values in increasing order with values above λmin corresponding to outlier clusters in LD analysis, parameter values for φ and |E|min are shown above plots. (download)

Fig. S3. STRUCTURE results from different K values (K=1-7, presented here K=2, 4 and 6) across the 7 sampling locations in the Lower Mekong Basin using neutral SNP panels. Bar plots showing membership probabilities of lineage membership of individuals to assigned genetic clusters. (download)

Fig. S4. Directional relative migration calculated by the divMigrate function (left), and the illustration of the significant directional asymmetric migration (right) for neutral SNP panels. Arrows indicate the direction of gene flow, and numbers show relative migration coefficient. (download)

Fig. S5. The output with m = 2 migration edges based on the second-order rate of change (Δm) across values of m. (download)

Table S1. Numbers of *Boesemania microlepis* individuals and SNPs following the filtering steps. (download)

Table S2. Number of individuals of *Boesemania microlepis* per population retained at each processing and analysis step. (download)