Hierarchical population genetic structure and signatures of adaptation in *Lates calcarifer*

Matthew A. Campbell[1*] and Joy A. Becker[1]

[1] The University of Sydney, School of Life and Environmental Sciences, 380 Werombi Road, Camden New South Wales 2570, Australia

[*] Corresponding Author
Matthew A. Campbell, m.campbell@sydney.edu.au; https://orcid.org/0000-0002-5826-0329

Joy A. Becker, https://orcid.org/0000-0003-4749-4004

Abstract

Context

Lates calcarifer known as barramundi perch, is widespread in the tropical Indo-Pacific region and is an important taxon in aquaculture, recreational and commercial fisheries. With mitochondrial barcoding study of barramundi perch, three clear mitochondrial clusters have emerged.

Aims

To test for congruent signal between mitochondrial and nuclear genomes and to test for population genetic structure. We further estimated the evolutionary significance of genetic clusters in terms of signatures of adaptation in the nuclear genome.

Methods

We conducted analyses of mitochondrial sequence data and whole-genome sequencing data from across the range of barramundi perch (n=60). We applied phylogenetic, phylogenomic and population genetic methods.

Key Results

We find congruent signal for three lineages of barramundi perch in both the mitochondrial and nuclear genomes. Substantial genetic differentiation is present between and within all three lineages of barramundi perch. Calculation of divergence times indicates a separation of ~ 7 million years ago, and another separation dating to ~1 million years ago.

Conclusions

Three lineages of barramundi perch are apparent, with candidate adaptive loci identified between and within each lineage, indicating both discreteness and evolutionary significance.

Implications

This study highlights the need for comprehensive sampling and integrative study of genotypes and phenotypes across the range of barramundi perch.

Additional Key Words

Adaptive Genetic Variation; Aquaculture; Barramundi; Conservation Genetics; Latidae; Phylogenetics; Population Genetics; Recreational Fisheries

1.0 Introduction

Lates calcarifer (Bloch 1790) is a valuable commercial taxon with numerous common names given its wide distribution and importance in historic and contemporary subsistence activities. Having been depicted in rock art and Dreamtime stories for millennia, 'barramundi' (Australian Aboriginal word meaning 'large, scaled silver fish') and also known as 'namamkorl' in the Kunwinjku dialect from the peoples in Kakadu and Western Arnhem Land (Northern Territory), hold deep cultural significance in Australian Aboriginal heritage https://pad.katalyst.com.au/plant_or_animals/barramundi (Cluney, 2004; Edwards-Vandenhoek, 2015; Shamsi et al., 2020). The common name 'barramundi' is widely used in Australian English while 'Asian sea bass' is also used in English and scientific works. We use the common name is barramundi perch (BP) in this manuscript.

Barramundi perch is distributed from the Persian Gulf eastwards to Taiwan and Australia, and into Oceania (**Figure 1**). Aquaculture beginning in Thailand in the 1970's has increased greatly and BP is widely propagated across its range with some aquaculture outside its native range. Barramundi perch is also an important taxon for conservation aquaculture with annual restocking in lakes and impoundments across northern Australia (e.g. Manton Dam, Northern Territory; Lake Kununurra, Western Australia) to support recreational fisheries and tourism in remote communities. The BP is a protandrous hemaphrodite, catadromous, and tolerant of a wide range of salinities (Moore, 1979; Moore and Reynold, 1982). The sequential hermaphrodism of the taxon requires the importation of young males into aquaculture facilities as older fish become females, a potential source of hybridization and mixing of stocks of different origin. Other qualities BP such as high fecundity, ease of culture and rapid growth (i.e. harvestable size of 3 kg in less than two years) encourage its production (Rimmer, 2009).

The widespread culture of this species has led to genetic studies to understand population genetic structure of BP. Mitochondrial barcoding revealed a deep divergence between Australian and Myanmar fish (K2P distance ~ 9.50%) (Ward et al., 2008). Broader sampling geographically of mitochondrial lineages has substantiated this divergence, showing a distinctive clade of Indian subcontinent fish including the Bay of Bengal (IND, additional locations the Andaman Islands and Myanmar) separate from Southeast Asia (SEA including Singapore, Malaysia, Thailand and

73 Indonesia) and Northern Australia (Vij et al., 2014). Vij et al. (2014) did include the sequences
74 from Ward et al. (2008) but did not include New Guinea, a region known to have biogeographic
75 similarities to Northern Australia (Wallace, 1859). Studies utilizing more complete genomic and
76 more complete geographic sampling are very limited. Restriction site-Associated DNA
77 sequencing (RADseq) was applied to population-level samples of BP (n = 131) from six
78 localities (Thailand, Malaysia, Indonesia, Western Australia, Eastern Australia and Papua New
79 Guinea) found a high degree of differentiation of SEA fish from Australia + New Guinea
80 (AUS+NG). Pairwise F_{st} within SEA was 0.005-0.009 but 0.36-0.443 between SEA and
81 AUS+NG without evidence of contemporary gene flow (Wang et al., 2016). Some investigation
82 of population genomic structure was applied to a range-wide data set at a high level as part of a
83 genome assembly paper but did not investigate concordance between mtDNA and genomic DNA
84 or population genetic metrics (Vij et al., 2016). No study to date, has analyzed both
85 mitochondrial data and genomic data from the same individuals to evaluate concordance between
86 data sources and resolve possible discordance between mitochondrial and genomic signals.
87
88 While there is evidence of genetic differentiation of BP, the evolutionary significance of these
89 differences are not substantiated and there remains a lack of recognition of distinctive species
90 that has serious implications in management of BP. For example, there is potential range overlap
91 of the IND lineage with the SEA lineage that has not been evaluated, and the forced
92 hybridization of these lineages during propagation is possible. They may exist as separate stocks
93 in a mixed stock fishery or form a natural hybrid zone. There is also the movement of broodstock
94 that may mix fish of different genetic lineages throughout the range of BP. While the recognition
95 of species is indeed important, so is the recognition of units below the species level that
96 contribute to diversity such as ecotypes or life history variants. The ecological effects of
97 intraspecific variation can meet our exceed species effect e.g. (Des Roches et al., 2018).
98 Conservation of both species-level and intraspecific variation can serve to promote ecological,
99 cultural and economic functions of organisms, resilience and genetic variation for future adaptive
100 potential e.g. (Layton et al., 2021; Lehnert et al., 2023; Schindler et al., 2010). Identification of
101 both specific and intraspecific variation can be considered both through isolation and adaptation
102 (Waples, 1991). We follow with an approach to identify units of BP through both isolation, that

is discreteness of higher-level groups, and testing for evolutionary significant differences between genetic groupings (i.e. adaptive genetic loci) (Lehnert et al., 2023).

2.0 Methods

2.1 Whole-Genome Sequence Data Set

Genome-wide sequence data from BP was downloaded from the NCBI Sequence Read Archive under the BioProject accession PRJNA311498. The 61 samples in PRJNA311498 covered a broad distribution, from the Indian subcontinent, parts of Southeast Asia, Northern Australia and Papua New Guinea. Two additional sequences of BP were obtained from Bangladesh (PRJNA1021005) and for phylogenetic analyses an outgroup obtained (*L. japonicus,* n =1, PRJDB13763). The paired-end data was aligned to the BP reference genome GCF001640805.2 (TLL_Latcal_v3) with the Burrows-Wheeler aligner (BWA) with the mem algorithm (Li and Durbin, 2010, 2009). Alignments were sorted, properly paired reads kept and PCR duplicates removed with SAMTools (Danecek et al., 2021; Li et al., 2009). Total read counts, filtered read counts and depth of coverage were calculated from alignments with SAMTools. Those samples with coverage ≤ 3X were excluded from further analyses.

2.2 Mitochondrial Phylogeny

A mitochondrial alignment from the WGS data set was created by creating a Binary Alignment Map (BAM) file for each individual with more than 3X coverage only including reads aligned to the reference mitochondrial genome (NC_007439.1, samtools view). These BAM files were converted to consensus FASTQ files with the bam2cns subprogram of proovread (Hackl et al., 2014) then to FASTA files turning all bases with quality scores below 20 to N with seqtk (seq -a -q20 -n N) (https://github.com/lh3/seqtk). The mitochondrial Cytochrome Oxidase I (mtCOI) gene was extracted from the complete mitochondrial FASTA-formatted sequences with SAMtools (faidx NC_007439.1:6329-7879). For rooting, the corresponding region of the *L. japonicus* mitochondrial genome was extracted from the NCBI reference sequence (NC_034339.1) and sequences were aligned with MAFFT (Katoh et al., 2002; Katoh and Toh, 2008). A Maximum-Likelihood (ML) phylogeny of sequences was generated with IQTREE specifying a Hasegawa-Kishino-Yano model of nucleotide substitution with gamma-distributed rate variation (-m HKY+G) and support for nodes evaluated with rapid bootstrapping (-bb

134 10000) (Hoang et al., 2017; Nguyen et al., 2014). The tree was imported into R and visualized
135 with the ggtree package (Yu et al., 2017) collapsing nodes to polytomies receiving less than 75%
136 Bootstrap Support (BS) and rooting by the longest branch.
137
138 *2.3 Genome-Wide Phylogenetic and Network Analyses*
139 We generated a genome-wide set of called SNPs with Analysis of Next Generation Sequencing
140 Data (ANGSD) from the 24 main linkage groups in the BP assembly (Korneliussen et al., 2014).
141 Quality of called SNPs was enforced by ensuring that a SNP was present in at least 90% of
142 individuals, had a minimum Minor Allele Frequency (MAF) of 0.05, a minimum mapping
143 quality of 20, minimum base quality of 20, a SNP p – val of 1.00×10^{-6} and a posterior cutoff of
144 0.90. A SAMTools model was specified and a PLINK-formatted file created as the output. We
145 converted the PLINK-formatted file to a VCF-formatted file with PLINK (Purcell et al., 2007).
146 Linked SNPS were removed with BCFTools +prune (-l 0.20 -w 10000).
147
148 The multispecies coalescent was applied to model gene-tree and species-tree discordance
149 resulting from Incomplete Lineage Sorting (ILS) e.g. (Maddison, 1997) on a set of called SNPs
150 that included the outgroup *L. japonicus*. A species-tree was generated using SVDQuartets in
151 PAUP* version 4.0a (Chifman and Kubatko, 2015, 2014; Swofford, 2003). Samples were
152 pooled by sampling region for the species tree. A random set of quartets (evalQuartets=random),
153 partially ambiguous SNPs were distributed over all compatible site patterns (ambigs=distribute),
154 and 1,000 bootstrap replicates (bootstrap=standard nreps=1000) were options specified.
155
156 For ML phylogenetic analysis and network analysis we called SNPS with BP samples and a
157 phylogeny of SNPs was inferred with IQTREE by identifying a model of nucleotide evolution
158 with ModelFinder Plus and indicating the ascertainment bias correction should be applied (-m
159 MFP+ASC) (Kalyaanamoorthy et al., 2017; Nguyen et al., 2011). IQTREE was used to further
160 filter SNPs for suitability for ascertainment bias correction. Support for nodes was evaluated
161 with 10,000 rapid boostraps (Hoang et al., 2017). As concatenated phylogenetic analyses do not
162 explicitly model hybridization or ILS, we created an implicit network to identify potential
163 hybridization or ILS in our data set with SplitsTree (Huson and Bryant, 2005). The same
164 alignment provided to IQTREE was analyzed with SplitsTree.

165

166 *2.4 Population Genetic Structure and Statistics*

167 Total signal in the data set was assessed by the generation of covariance matrix with ANGSD
168 and a PC analysis. The covariance matrix was generated by with a single read sampling approach
169 due to a wide range of individual coverage (minimum 4.73, maximum 34.81, mean 9.38 and
170 median 6.41) with the options -doIBs 1 and -doCov 1 in ANGSD. The 24 main linkage groups of
171 the reference assembly were sampled for this analysis. A variable site was required to be present
172 in >90% of individuals with a minimum MAF of 0.05. A minimum mapping quality of 10 was
173 specified with a minimum base quality of 20 and a GATK genotype likelihood model was used
174 (-GL 2). The resulting covariance matrix was imported into R for PC analysis.

175

176 To calculate genetic differentiation in terms of F_{st} was calculated by generating Site Frequency
177 Spectra (SFS) for each sampling region with ANGSD and the realSFS subprogram, splitting
178 Indonesia into two locations based on metadata available with the sequences ("K" n=5 and and
179 "SJ" n=5). Vietnam (n=2), Indonesia "SU" (n=1) and Bangladesh (n=2) were excluded due to
180 small sample sizes. Two-dimensional SFS were calculated between each possible pair of
181 sampling regions and F_{st} calculated (see code repository accompanying this paper). Nucleotide
182 diversity (π) and Tajima's D (D) estimates were calculated for each chromosome with realSFS
183 using the saf2theta option followed by the thetaStat subprogram of ANGSD and the do_stat
184 option.

185

186 Admixture analysis was conducted with NGSadmix (Skotte et al., 2013) on genotype likelihoods
187 generated with ANGSD as for the PC analysis using a single read sampling approach. The
188 resulting BEAGLE-formatted file was used with NGSadmix to produce an admixture plot of K =
189 2-5 genetic clusters.

190

191 *2.5 Tests for Adaptative Genetic Variation*

192 Genotype calls produced for phylogenetic network analysis after pruning were imported into R
193 for analysis with PCAdapt (Luu et al., 2017; Privé et al., 2020). The *pcadapt* function was then
194 used to generate a scree plot and identify an objective number of PCs with the first 20 PCs
195 examined. From the resulting object from the *pcadapt* function, outlier loci were identified by

196 filtering for a Benjamini and Hochberg (1995) and Bonferroni corrected *p* – values with
197 significance values less than 0.10. We repeated this analysis for individuals in each main genetic
198 lineage by filtering for variants with a minimum minor allele frequency of 0.05 and examining
199 the first 5 PCs.
200
201 *2.6 Divergence Time Estimation*
202 Existing efforts have covered a broad geographic and lineage sampling of *L. calcarifer* with
203 barcoding (mitochondrial *cox1*) (Vij et al., 2014; Ward et al., 2008). The effort of this study was
204 then to calculate divergence times of the three lineages (IND, SEA and AUS+NG). We
205 downloaded barcode sequences from the Barcode of Life Database and Genbank for Latidae.
206 With BLAST (blastn) we confirmed the major lineages of *Lates calcarifer* in barcode data
207 (Altschul et al., 1997) then selected representative sequences from these three lineages to
208 conduct a divergence time analysis (**Table 1**). We also searched for representation across Latidae
209 for broad taxonomic representation of this family including sequences from *L. japonicus*, *L.*
210 *microlepis* and *L. niloticus*. We included representatives of the two main mitochondrial lineages
211 present in *Psammoperca* corresponding to *P. waigiensis* and *P. datnioides*. Centropomidae has
212 been confirmed to be the sister lineage of Latidae in molecular phylogenetic study e.g.
213 (Campbell et al., 2013; Chen et al., 2007; Li et al., 2011) and inclusion allows dating the Time to
214 Most Recent Common Ancestor (TMRCA) of Centropomidae and Latidae to be 48.6 million
215 years ago (mya) (Otero, 2004). We also included sequences from the outgroup lineages
216 Channidae and Nandidae as an additional calibration point dating the TMRCA of those taxa to
217 be 48 mya (Santini et al., 2009).
218
219 Sequences were aligned with MAFFT (Katoh et al., 2002; Katoh and Toh, 2008), in frame
220 coding was confirmed by translation in the NCBI ORF finder and a NEXUS file made for a
221 time-calibrated phylogenetic analysis in MrBayes (Ronquist et al., 2012). The third codon
222 position was excluded due to the age of the data set and expected saturation occurring in
223 mitochondrial sequences from carangimorph taxa at these timescales e.g. (Campbell et al., 2014).
224 A nucleotide substitution model of General Time Reversible (GTR) with a proportion of
225 invariant sites (I) and gamma-distributed rate variation (Γ) was specified. Constraints were
226 applied to Centropomidae + Latidae to date with an exponential distribution used,

227 offsetexp(48.60, 63.18), and the outgroup, offsetxp(48.00,62.40). Initially, three runs and three
228 chains were specified with a run length of 1 million generations sampled ever 1,000 generations
229 and a 25% burnin applied. Evaluation of parameters and convergence with Tracer (Rambaut et
230 al., 2018) was assessed, and run length increased until sufficient ESS for all parameters (>200)
231 was reached. Tree summarization was done in MrBayes with the sumt command and an all
232 compatible consensus tree made.
233

234 3.0 Results

235 *3.1 Whole-Genome Sequence Data Set*

236 From the original 63 *L. calcarifer* sequences, 60 had coverage over 3X and were included in
237 additional analyses (**Table 1**). Sample sizes for regions were 13 for putative members of the
238 IND lineage, 30 for putative members of the SEA lineage, and 17 for putative members of the
239 AUS+NG lineage from ten sampling regions (**Figure 1**). The *L. japonicus* sequence aligned to
240 the *L. calcarifer* reference genome with high success with 110X coverage and was down-
241 sampled to 20X coverage with SAMtools to be more consistent with genome coverage measures
242 of *L. calcarifer* individuals.
243

244 *3.2 Mitochondrial Phylogeny*

245 The 61 samples (60 *L. calcarifer* and one *L. japonicus*) produced an alignment of 1,576 base
246 pairs in length with 71 distinct patterns and 128 parsimony-informative sites. In the alignment,
247 1,441 sites were invariant. The ML phylogeny indicates a clear division of *L. calcarifer* IND
248 individuals with two Cambodian-origin fish from all other fish, BS=85% (**Figure 2**). Support
249 for monophyly of *L. calcarifer* SEA + *L. calcarifer* AUS+NG is maximal, and support for a
250 clade of *L. calcarifer* AUS+NG is moderate (BS=80%). Individuals sourced from the putative *L.*
251 *calcarifer* SEA clade do not form clear geographic clusters. The Assemble Species by Automatic
252 Partition algorithm indicated the strongest support for two species, with a threshold distance of
253 4.10% (results provided in the Data Supplement). This division separates the *L. calcarifer* IND
254 mitochondrial lineage from *L. calarifer* SEA + *L. calcarifer* AUS+NG.
255

256 *3.3 Genome-Wide Phylogenetic and Network Analyses*

257 For analysis with SVDQuartets to generate a rooted species-tree, 8,529 variants that all have
258 unique alignment patterns were analyzed and is presented as **Figure 3**. The SNP alignment
259 consisted of 6,456 parsimony informative and 2,073 singleton sites. Three lineages are apparent
260 in the species-tree with maximal support for monophyly corresponding to the *L. calcarifer* IND,
261 *L. calcarifer* SEA and *L. calcarifer* AUS+NG. The SEA and AUS+NG lineages are highly
262 supported (BS=100%) as sister lineages. Within the IND lineage, the India Eastern Coast and
263 Bangladesh sampling locations are indicated to be sister lineages (BS=100%). The AUS+NG
264 lineage individuals from the Northern Territory and Papua New Guinea are indicated to be most
265 closely related (BS=100%). The first division within the SEA lineage is between Indonesia-SU +
266 Indonesia-K and receives maximal support for monophyly and branching patterns. The
267 Philippines sampling location separates with the next successive branching within the SEA
268 lineage (BS=100%). Relationships among Cambodia, Indonesia-SJ, Thailand and Vietnam
269 sampling locations are not clearly resolved.
270
271 For ML phylogenetic analysis and phylogenetic network analysis an alignment of 60 BP
272 individuals with 7,661 SNPs having 7,661 distinct unique alignment patterns and 5,571
273 parsimony informative sites was created. The alignment contains 2,090 singleton sites. Support
274 for monophyly of *L. calcarifer* IND is maximal, as well as geographic division with *L. calcarifer*
275 IND into Eastern and Western Indian subcontinent (**Figure 4A**). *L. calcarifer* SEA as a grouping
276 receives low support (BS = 76%) with concordance between sampling locations and placement
277 not complete. Indonesia-K and the Philippines sampling locations are highly supported to be
278 monophyletic, but Cambodia, Thailand and Indonesia-SJ are not. Support for a *L. calcarifer*
279 AUS+NG is maximal, with maximal support for division between Queensland and Northern
280 Territory + Papua New Guinea fish. Monophyly of samples from the Northern Territory is
281 poorly supported (BS = 68%).
282
283 The phylogenetic network resolves an IND lineage containing two sub-lineages (India Eastern
284 Coast + Bangladesh, India Western Coast) with separation of all three sampling locations
285 (**Figure 4B**). The SEA lineage is present, with reticulations with the IND lineage occurring with
286 a subgroup of 10 individuals, disrupting monophyly of some sampling locations. The Indonesia-
287 K and Philippines locations are monophyletic, but the Cambodia, Thailand and Indonesia-SJ

288 sampling locations are not. The AUS+NG grouping is composed of three subclades Queensland,
289 Northern Territory and Papua New Guinea. In the network, Northern Territory and Papua New
290 Guinea fish are most-closely related to each other.
291
292 *3.4 Population Genetic Structure and Statistics*
293 The PC analysis included 7,321,852 variable sites in the genome and indicated a separation into
294 three larger genetic clusters corresponding to IND, SEA and AUS+NG individuals (**Figure 5**).
295 The first PC separated AUS+NG, SEA and IND individuals (PC 1 12.57% of variation) with the
296 second PC showing variation within SEA and IND clusters (PC2 4.71%). Clear separation of
297 Indian East Coast + Bangladesh and Indian West Coast individuals is present on both PC 2 and
298 PC 3 (PC 3 3.51%).
299
300 Genetic differentiation (F_{st}) was lowest within main genetic lineages, such as Thailand and
301 Cambodia (F_{st} = 0.06) within the *L. calcarifer* SEA lineage and between the Indian East Coast
302 and Indian West Coast sampling locations (F_{st} = 0.11) within the *L. calcarifer* IND lineage
303 (**Figure 6A**). The highest levels of genetic differentiation are found between main genetic
304 lineages, between AUS+NG and SEA F_{st} = 0.15-0.35, between AUS+NG and IND F_{st} = 0.22-
305 0.33, between SEA and IND F_{st} = 0.12-0.26). Nucleotide diversity is the lowest in the AUS+NG
306 lineage (π = 20,067-22,775) and highest in the IND lineage (π = 207,299-214,485) (**Figure**
307 **6B**). Tajima's D was strongly negative (< -1) in the Indonesia-SJ, Thailand, India Eastern Coast
308 and India Western Coast sampling locations (**Figure 6C**). The AUS+NG sampling locations had
309 *D* values between -0.58 and -0.77.
310
311 Admixture analysis of 7,321,852 genotype likelihoods from 60 BP individuals for K = 2-5 is
312 presented as **Figure 7**. At K = 2, the Bangladesh, Indian Eastern Coast and Indian Western
313 Coast have high ancestry coefficients (Q) > 0.99 for a single genetic cluster and the AUS+NG
314 lineage and Indonesia-K sampling location have high ancestry coefficients for the other. The
315 remaining SEA sampling locations have a mixture of contributions from both genetic clusters.
316 At K = 3 the AUS+NG individuals form a clearly genetic cluster with Q > 0.99 for ancestry
317 coefficients, the SEA sampling locations are heterogenous mixture proportions from sampling
318 locations largely composed of the third genetic cluster, with the Indonesia-K and Indonesia-SU

319 sampling locations exhibiting a near uniform ~0.27% admixture proportion with the genetic
320 cluster typical of AUS+NG sampling locations. Higher values of *K* lead to separating of genetic
321 units within the SEA lineage that may or may not correspond to sampling locations.
322
323 *3.5 Tests for Adaptation*
324 The variation within the pruned genotype calls of 22,088 SNPs from the n=60 WGS sequences
325 with >3X coverage supported an optimal *K* = 3 PCs (**Supplemental Figure S1**). These PCs
326 separated the three main lineages of *L. calcarifer*, with the third PC separating the Indonesia-K
327 sampling location and to a lesser degree the Indonesia-SU individual from other SEA. Benjamini
328 and Hochberg (1995) adjusted *p* – values (BH) identified 134 candidate adaptive SNPs and
329 Bonferroni corrected *p* – values (BF) indicates 34 candidate adaptive SNPs across the three PC
330 axes.
331
332 Filtering to IND individuals (n=13) resulted in 12,075 SNPs with a minimum MAF of 0.05, with
333 the first PC separating the Bangladesh, India Eastern Coast and India Western Coast with 738
334 BH and 379 BF candidate adaptive SNPs identified on this axis (**Supplemental Figure S2**). The
335 second PC axis separated India Eastern Coast with 285 BH and 59 BF candidate adaptive SNPS.
336 From the SEA lineage, 30 individuals with 17,556 SNPs with a minimum MAF of 0.05 were
337 analyzed. The first three PCs were considered, separating Indonesia-K and Indonesia-SU
338 sampling locations on PC 1, the Philippines along PC 2 and some Thailand and Cambodia
339 individuals along PC 3 with 190 BH and 29 BF candidate adaptive SNPs identified in total
340 (**Supplemental Figure S3**). The data set of AUS+NG individuals (n=17) for loci with a
341 minimum MAF of 0.05 resulted in 14,642 SNPs. We found strong support for *K* = 2 PCs, the
342 first PC separating Queensland from the Northern Territory and Papua New Guinea
343 (**Supplemental Figure S4**). The second PC separated Papua New Guinea from the other
344 locations with 234 BH candidate adaptive SNPs and 57 BF candidate adaptive SNPs across the
345 two PC axes.
346
347 *3.6 Divergence Time Estimation*
348 We compiled a barcode alignment of 13 sequences with 660 characters resulting in 440
349 characters (399 invariant) when including the first and second codon positions only. Sufficient

350 effective sample sizes (ESS >200, range of 712-3,002) was reached with 20 million generations
351 and two runs of three chains each as evaluated after a 25% burnin. The resulting consensus tree
352 presented in **Figure 8.** The median divergence time of *L. calcarifer* IND from the other two
353 lineages is estimated to be 6.76 mya with a 95% Highest Posterior Density (HPD) of 1.31-14.24
354 mya. The divergence of *L. calcarifer* AUS+PNG from *L. calcarifer* SEA is estimated to be 1.02
355 mya with a 95% HPD range of 0.00-5.03 mya. The relationships of *L. calcarifer* lineages are
356 highly supported, with the placement of *L. calcarifer* having a Posterior Probability (PP) of 0.99,
357 and the sister relationship between *L. calcarifer* AUS+PNG and *L. calcarifer* SEA having a PP
358 of 0.97. The NEXUS file for use with MrBayes and the consensus tree are provided in Data
359 Supplement of this paper.
360

361 # 4.0 Discussion

362 We demonstrate concordance of nuclear and mitochondrial genomes of three main genetic
363 lineages of BP with substantial differences quantified in terms of genome-wide estimates of F_{st},
364 candidate adaptive loci and deep divergence times. We observed F_{st} values up to 0.36 and in
365 terms of divergence times, we estimated 6.76 million years for the divergence of IND from other
366 BP and a divergence time estimate of 1.02 million years for AUS+NG from SEA. The previously
367 published estimate based on RADseq data for divergence times by Wang et al. (2016) is
368 restricted to AUS+NG from the SEA lineage and is an approximate 644,100-878,900 years
369 before present. Both of these estimates, the estimate presented here from fossil-calibrated
370 phylogeny and Wang et al. (2016), support a divergence during the glacial cycling and changes
371 in sea level and conditions during the Quaternary. The three main lineages of BP exhibit
372 discreteness (genetic isolation) and evolutionary significance (i.e. adaptive differentiation).
373

374 Within each of the three main lineages, there was additional evidence for discreteness and
375 evolutionary significance. The IND sampling locations were divided phylogenetically (**Figures 3**
376 **& 4**), and in PC analysis (**Supplemental Figure S2**) with candidate adaptive SNPs identified.
377 Previously, with the IND lineage, substructure was not clear with mitochondrial data and we can
378 find a clear division within IND (Bangladesh, India Eastern Coast and India Western Coast) with
379 genomic data. The SEA lineage exhibited notable separation of the Indonesia-K and Indonesia-
380 SU sampling locations, the Philippines sampling location and some Thailand fish (**Figure 4**,

381 **Supplemental Figure S3**). Clear differentiation of all Philippines and Thailand fish may be
382 impacted by hybridization (**Figure 7**). The RADseq data set of Wang et al. (2016) did not show
383 clear evidence of genetic structure with three sampling locations within SEA, but broader
384 sampling with microsatellites did appear to support genetic structuring (Yue et al., 2009). The
385 AUS+NG sampling exhibited a clear division between Queensland and other sampling locations
386 (**Figures 3 & 4**) and all three sampling locations could be clearly differentiated on two PC axes
387 with candidate adaptive loci apparent (**Supplemental Figure S4**). Within the AUS+NG lineage
388 there are several studies demonstrating genetic differentiation within that lineage e.g.
389 (Chenoweth et al., 1998; Loughnan et al., 2019; Shaklee and Salini, 1985; Wang et al., 2016)
390 with 16 discrete genetic populations identified in Australia only (Keenan, 1994).
391
392 *4.2 What is* Lates calcarifer*?*
393 Our fundamental understanding of nature is based at the species level and the lack of recognition
394 of species is a major knowledge shortfall in conservation and management (Brown and
395 Lomolino, 1998). Molecular genetic techniques have continued to identify candidate species of
396 fishes which may or may not have diagnostic morphological characters typical of species
397 descriptions e.g. (Campbell et al., 2023; Moyle and Campbell, 2022). Instead of being a
398 historical aspect of biology, the need for continued taxonomic work is very important e.g. (Delić
399 et al., 2017; Godfray et al., 2004). Taxonomic work often involves the detailed anatomical
400 investigation of organisms that can reveal phenotypic differences related not only to speciation,
401 but also life history variation. What remains outstanding is if BP can be accurately represented
402 within Linnean taxonomy or if an alternative approach is better suited to the biological realities
403 of BP.
404
405 The most recent taxonomic effort relevant to BP led to the description of *L. lakdiva* Pethiyagoda
406 & Gill 2012 from western Sri Lanka and *L. usiwara* Pethiyagoda & Gill 2012 from eastern
407 Myanmar (Pethiyagoda and Gill, 2012). *Lates lakdiva* from Sri Lanka apparently cannot be
408 propagated in aquaculture using methods used for BP elsewhere indicative of divergences from
409 those species. The specimens of *L. usiwara* were the same as featured in the barcoding study of
410 Ward et al. (2008) highlighting deep mitochondrial divergences, but in a wider context the barcode
411 sequences of *L. usiwara* are identical to samples from the Indian subcontinent (Vij et al., 2014).

412 However, in comparison to *L. calcarifer*, a differential diagnosis was possible and the distribution
413 of *L. usiwara* is perhaps indicative of a recently derived species or ecotype. The anatomical study
414 undertaken by Vij et al. (2014) along with barcoding did not clearly separate all these newly
415 described species of *Lates* from *L. calcarifer* noting small sample sizes used in the description of
416 *L. usiwara* and *L. lakdiva* and conclude that there is substantial mismatch between anatomical
417 descriptions of new species of *Lates* and candidate species of *Lates* apparent genetically. Barcode
418 sequences, it should be noted, will not reflect proportion of ancestry as do genomic data and
419 ascertaining the ancestry proportions of fish used in taxonomic studies is critical going forward
420 due human translocations of *Lates*. Furthermore, it is possible for a high degree of genome-wide
421 differentiation or substantial reproductive isolation be present that would not be apparent in
422 barcode data. Pethiyagoda and Gill (2012) themselves observe that due to the large body size and
423 sequential hermaphodism of *L. calcarifer sensu lato*, they were restricted to examining many
424 juveniles for their study.
425
426 The three genetic lineages of BP remain unclear in distribution and have unclear anatomical
427 differentiation (Vij et al., 2014). Portions of the range, such as Taiwan, remain unsampled for
428 genomic data and genetic structuring is apparent across all three lineages (**Figures 3, 4, 5 and 6,**
429 **Supplemental Figures S1-S4**). Differentiation between and within the main lineages can be
430 clarified with additional sampling and representation of nominal taxa in genomic data sets (i.e. *L.*
431 *usiwara* and *L. lakdiva*). Given the strong genetic differentiation and deep inferred time scales
432 (**Figure 8**), one possible resolution is to recognize at least two species-level taxa that correspond
433 to the IND lineage and a combined SEA and AUS+NG lineage. At a potential species-level, the
434 taxonomy of *L. calcarifer* is complex and the recognition of types and their origins is also
435 challenging, reviewed by Pethiyagoda and Gill (2012). The type locality of *L. calcarifer* (Bloch
436 1790) may be Tamil Nadu, India (10.4°N, 79.3°E) (Fricke et al., 2023). Thus, the *L. calcarifer*
437 IND lineage may be represented by this name. There are numerous available names that may be
438 applicable to *L. calcarifer* SEA, but this requires substantial time and expertise to resolve. There
439 is, however, an available name, *L. cavifrons* (Alleyne & Macleay 1877) that could be applied to
440 *L. calcarifer* AUS+NG. The collection locality is either the Torres Strait or the coast of New
441 Guinea with a lost holotype with the drawing providing the current synonymy (Pethiyagoda and
442 Gill, 2012).

443

444 Without formal taxonomic recognition, the three main genetic lineages of BP can be recognized
445 for management purposes such as mixed stock fisheries. Vij et al. (2014) note that the
446 assumption of single species of BP has undoubtedly influenced previous studies and
447 microsatellite data from Thailand showing the highest allelic richness may reflect genotyping of
448 two divergent genetic lineages along the west coast of Thailand (Yue et al., 2009). If different
449 species of BP are recognized, then questions of biology can be researched such as the causes of
450 speciation and what the mechanisms for reproductive isolation of IND may be. Development of
451 an anatomical key usable in the field that can separate these species e.g. (Wang et al., In
452 Revision) would be of great practical utility and could be used to resolve the distributions and
453 overlap of IND and SEA lineages. A high-throughput sequencing assay that can discriminate the
454 main lineages could provide a rapid assessment of genetic lineage, hybridization levels and
455 possible provenance in studies of food adulteration.
456

457 *4.2 Admixture*

458 Overall, analyses that can quantify hybridization in this study are challenging because of the high
459 innate diversity of the SEA lineage and limited sample sizes, notably that of the IND lineage (n =
460 13). The sequences were also obtained opportunistically from the NCBI Sequence Read Archive
461 and not collected by the authors. With these cautions in mind, we find evidence of admixture
462 between major genetic lineages within our data set as evidenced by the placement of Cambodia-
463 origin samples, with a largely SEA genomic background within the mitochondrial IND lineage
464 (**Figures 2, 4 & 5**). When examining all 60 individuals at $K = 3$, the two Cambodia individuals
465 with IND mtDNA have ancestry coefficients (Q) of 0.29 and 0.43 with the IND genetic cluster,
466 however, there is such a degree of variation in the SEA group with respect to admixture
467 proportions and inconsistency across sampling locations that these measurements should not be
468 accepted at face value. The Cambodia group has two individuals with Q = 0.00 for the IND
469 genetic cluster and one with 0.17 for example when examining all individuals. Separate analysis
470 of Cambodia and *L. calcarifer* IND individuals only show $Q = 0.23$ and 0.33 for the two
471 Cambodia individuals with IND mtDNA and no evidence for nuclear genomic admixture of
472 other Cambodian individuals with IND (**Figure 7**).
473

474 Nucleotide diversity in the SEA Indonesia-SJ location is very high (**Figure 6B)** and no doubt
475 influences the reduced F_{st} of this lineage to other sampling regions (**Figure 6A**). Combined with
476 the large negative *D* value which arises from admixture e.g. (Hahn et al., 2013) this sampling
477 location may contain admixed individuals. A similar pattern of high nucleotide diversity, low-F_{st},
478 and strongly negative *D* is apparent with the Thailand sampling location (**Figure 6**). Both the
479 Indonesia-SJ and Thailand sampling locations show substantial contributions of the IND genetic
480 cluster in their ancestry (**Figure 7**), but it is not uniformly distributed across sampling locations.
481 If the distribution of ancestry is uniform across sampling locations, it would suggest an older age
482 of hybridization. The variable admixture proportions in sampling locations suggest recent
483 hybridization events, e.g. Cambodia at $K = 4$. Furthermore, the admixture analyses suggest that
484 Thailand at $K = 4$ is represented by more than one distinct genetic unit from within the SEA
485 lineage with additional contributions from the IND lineage.
486
487 Clearly, samples of known provenance, the use of informative loci that are diagnostic to
488 lineages, and boosting of sampling locations and sample sizes focused across the possible
489 intergrade zone IND and SEA would help resolve these ambiguities. Testing for levels of
490 admixture with higher sample sizes e.g. NewHybrids (Anderson and Thompson, 2002) could
491 resolve if the observed admixture is recent (F1s, F2s) or more advanced backcrosses, providing
492 evidence for disentangling artificial and natural hybridization events. Furthermore, the use of
493 archive and museum samples to provide genetic baselines prior to use of BP as an aquaculture
494 species can be used to create a genetic baseline to compare modern genetics against to identify
495 potential translocations and adaptation to changing environments e.g. (Turbek et al., 2023)
496
497 *4.3 Conclusion*
498 We corroborate the findings of mitochondrial barcoding identifying three divergent lineages of
499 BP through the analysis of genomic data. The analysis of genome-wide nuclear variants also
500 provides evidence of hierarchical divergences, that is within the IND, SEA and AUS+NG
501 lineages there is genetic structuring. The SEA lineage demonstrates a complex history with a
502 combination of natural genetic differentiation and what is likely human mediated admixture
503 events. Phenotypic and life history variation within BP remain poorly characterized and are other
504 aspects of variation that can substantially contribute to ecosystem productivity and commercial

applications. Research indicates that within Australia for instance, there is strong local adaptation related to temperature (Newton et al., 2013) that could be applied to various goals. The wide and varied use of BP in commercial, recreation and subsistence fisheries as well as aquaculture creates differing priorities across user groups. Further exploration of BP variation can support conservation aquaculture and food production as well as the preservation of native BP stocks.

Data Availability

Genomic sequence data are available from the NCBI Sequence Read Archive under BioProject accessions PRJNA311498, PRJNA1021005 and PRJDB13763. Alignments for phylogenetic analyses include the control block for SVDQuartets are included in the Data Supplement of this paper. Sequence data examined in this study for time-calibrated phylogenetic analysis are publicly available with mitochondrial sequence accessions reported in Table 1.

Declaration of Funding

Support for MC was received from the School of Life and Environmental Sciences at the University of Sydney and the Charles Gilbert Heydon Travelling Fellowship.

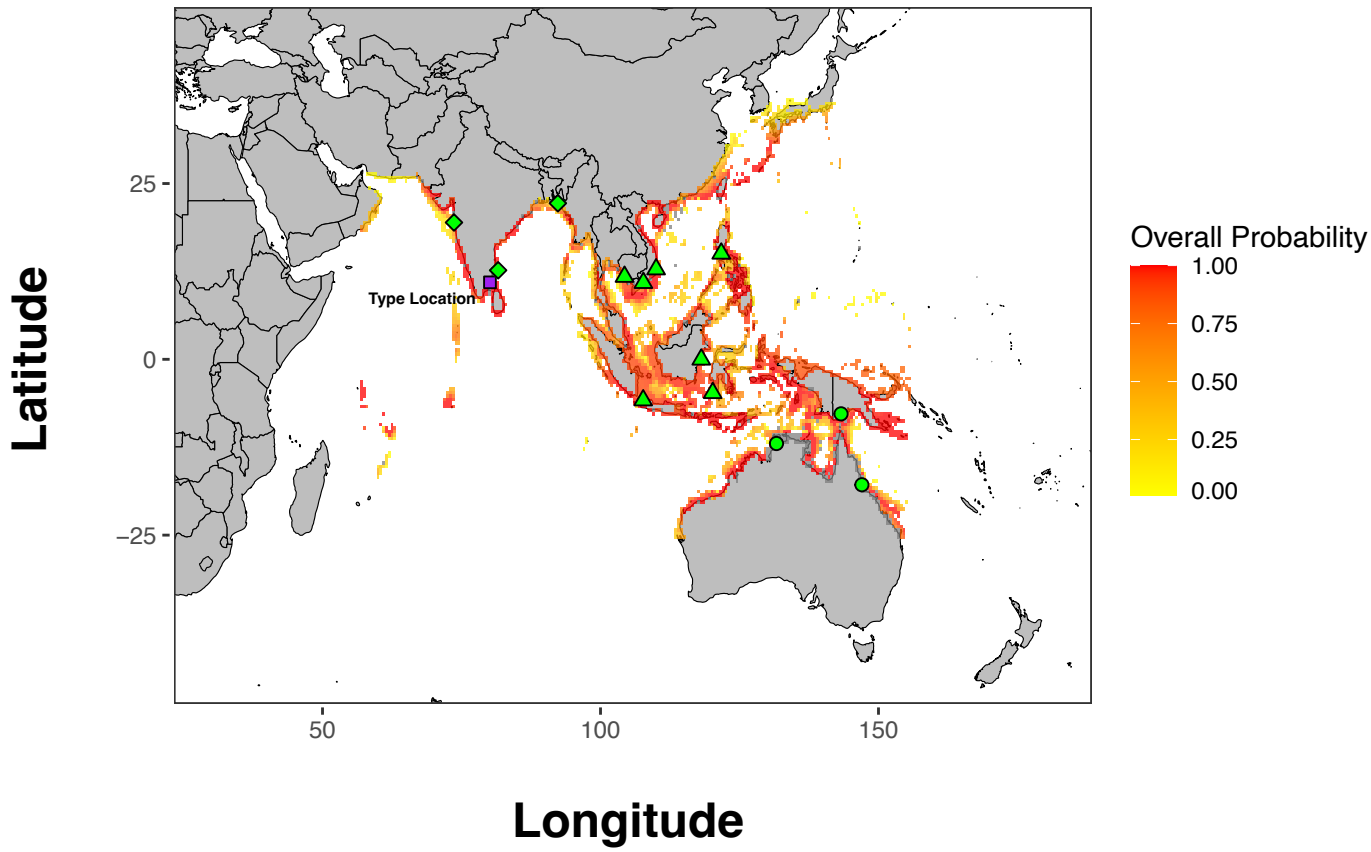
Conflicts of Interest

The authors declare no conflicts of interest.

Lay Summary

Barramundi Perch, aka Barramundi, are a widespread fish widely propagated and highly sought after for food and recreation. We find that the Barramundi contains three main genetic lineages with numerous subdivisions, and that these divisions are linked to adaptive changes and may be considered evolutionarily significant. These findings indicate diversity within Barramundi that may be useful for both conservation and exploitation.

532 **Figures**



533

534 **Figure 1** Distribution of *Lates calcarifer sensu lato* from AquaMaps (Kaschner et al., 2015) with probability of occurrence (Overall
535 Probability) colored from yellow (low) to red (high). The type locality of *L. calcarifer* is indicated by a purple square. Sampling of
536 main genetic lineages included in this paper are indicated by shape, AUS+NG circles, SEA triangles and IND diamonds. Locations of
537 WGS are approximate as specific coordinates are not available. Cambodia is represented by four separate collection locations (shown
538 as one).

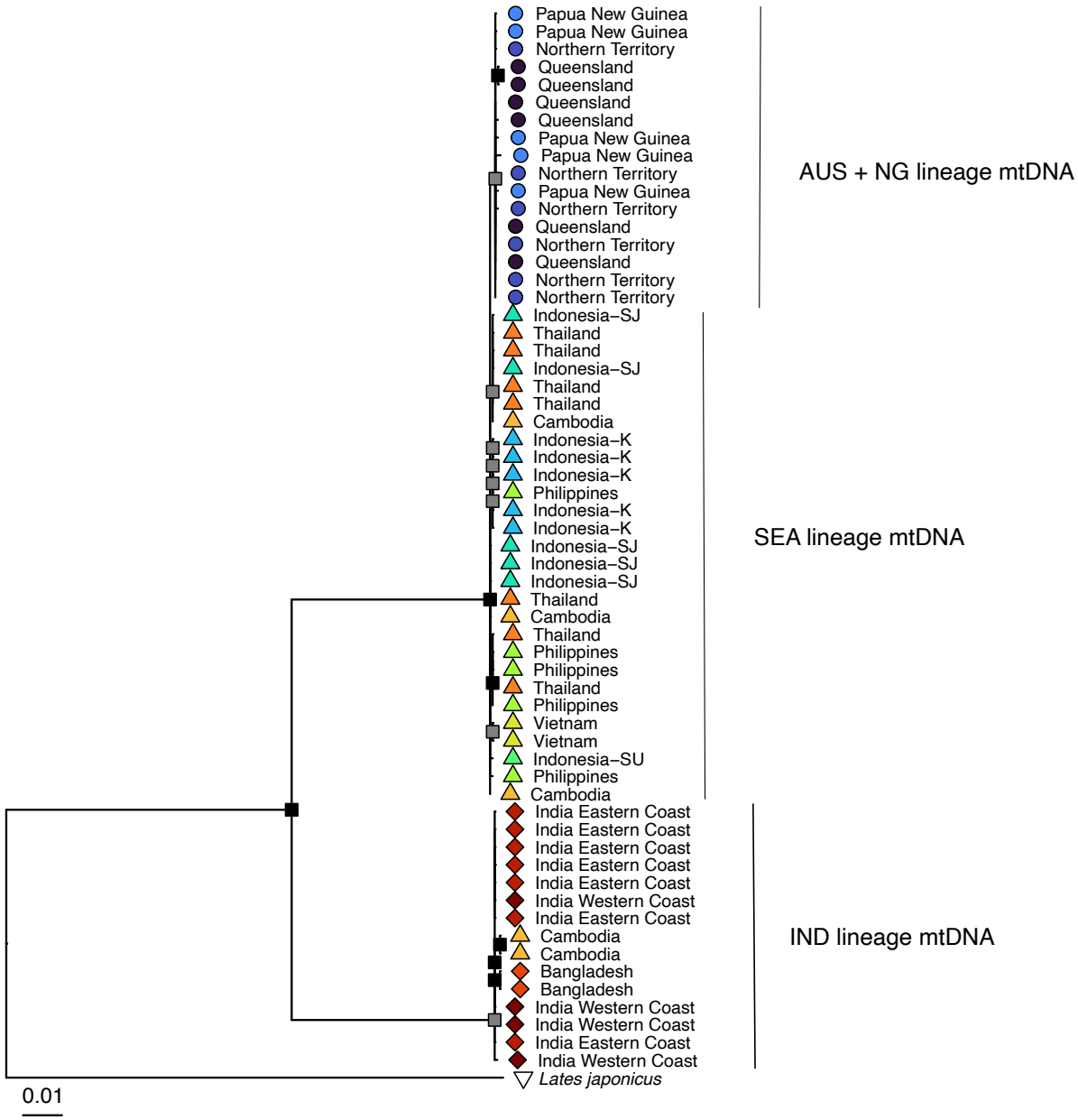


Figure 2 Maximum-Likelihood phylogeny of mitochondrial Cytochrome Oxidase I from 60 barramundi perch individuals derived from whole-genome sequence data. The tree is rooted by *L. japonicus*. Node shape indicates the putative lineage of origin and collection information is provided as tip labels. Nodes with < 75% bootstrap support are collapsed to polytomies with nodes ≥ 75% bootstrap support but < 90% indicated as grey squares and nodes ≥ 90% bootstrap support as black rectangles.

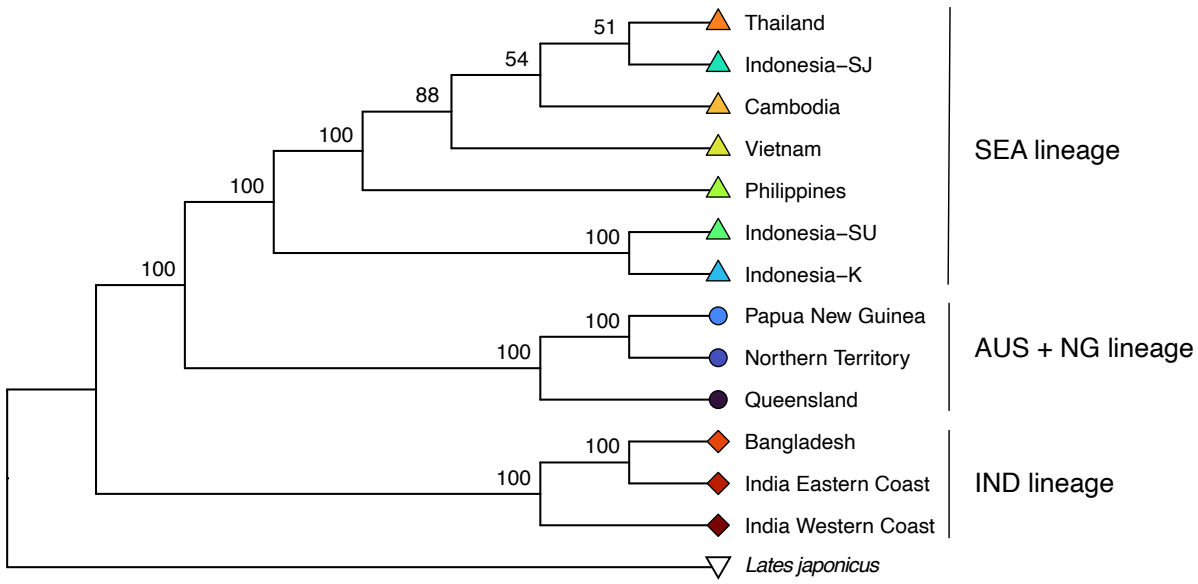


Figure 3 Species-tree from SVDQuartets of Barramundi Perch sampling locations in this study based on 8,529 variants rooted by *Lates japonicus*. Values at nodes are bootstrap support values .

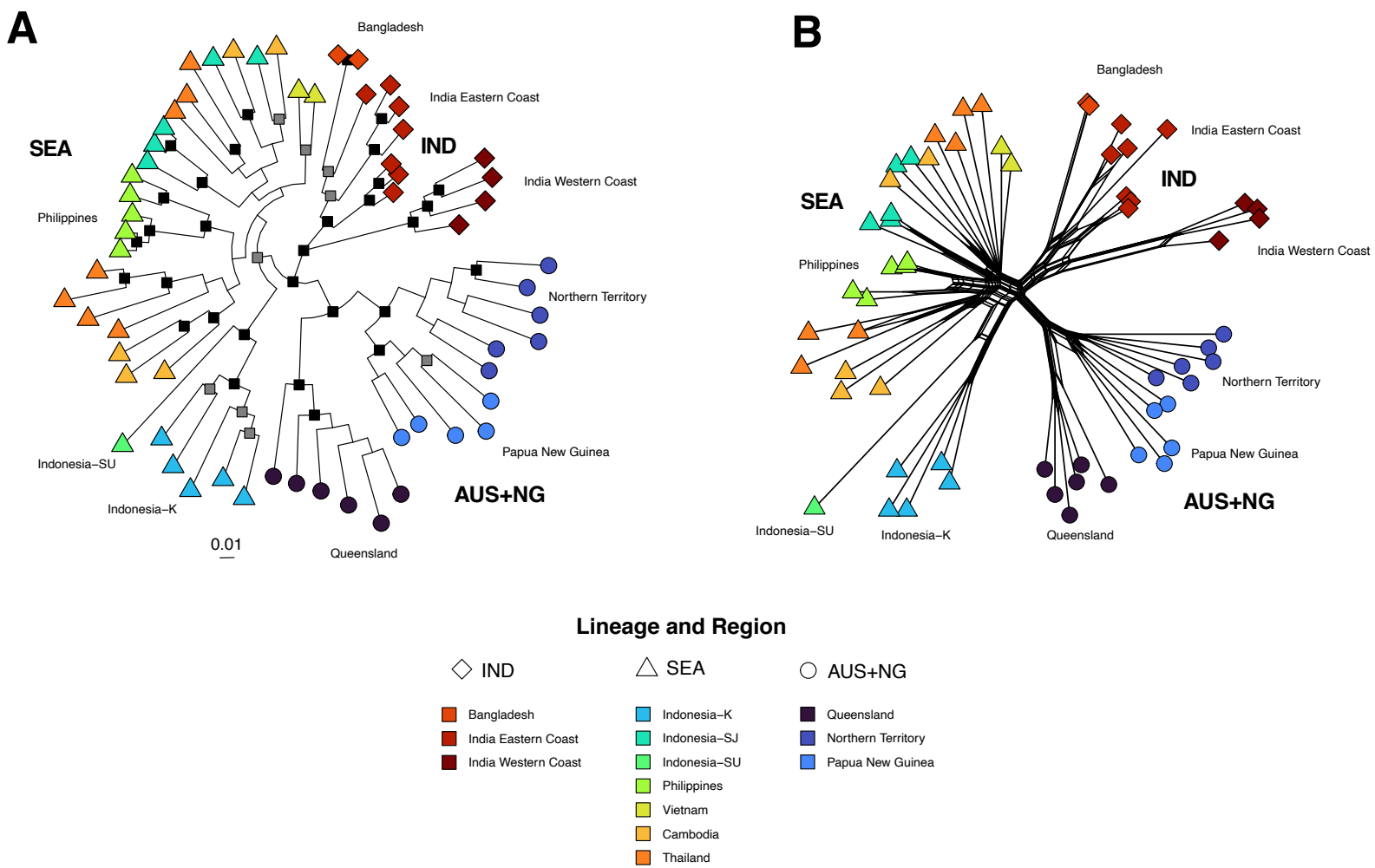


Figure 4 (A) Maximum-Likelihood phylogeny of genome-wide SNPs of 60 barramundi perch individuals. Nodes with < 75% bootstrap support are collapsed to polytomies with nodes ≥ 75% bootstrap support but < 90% indicated as grey squares and nodes ≥ 90% bootstrap support as black rectangles. **(B)** Implicit phylogenetic network of genome-wide SNPs from 60 barramundi perch individuals. Node shape indicates the putative lineage of origin and color indicates collection location.

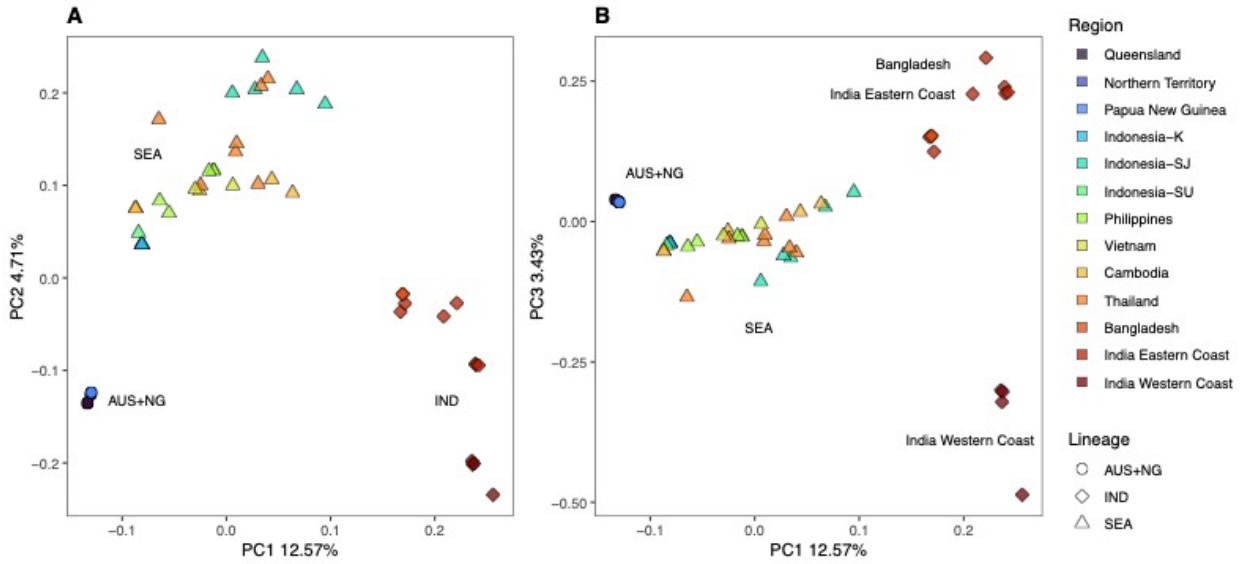


Figure 5 Principal Component (PC) analysis of the whole-genome sequence data set of 60 barramundi perch individuals utilizing 7,321,852 genotype likelihoods. The first PC is on the *x* – axis in both figures with the second PC on *y* – axis in (A) and third PC in (B).

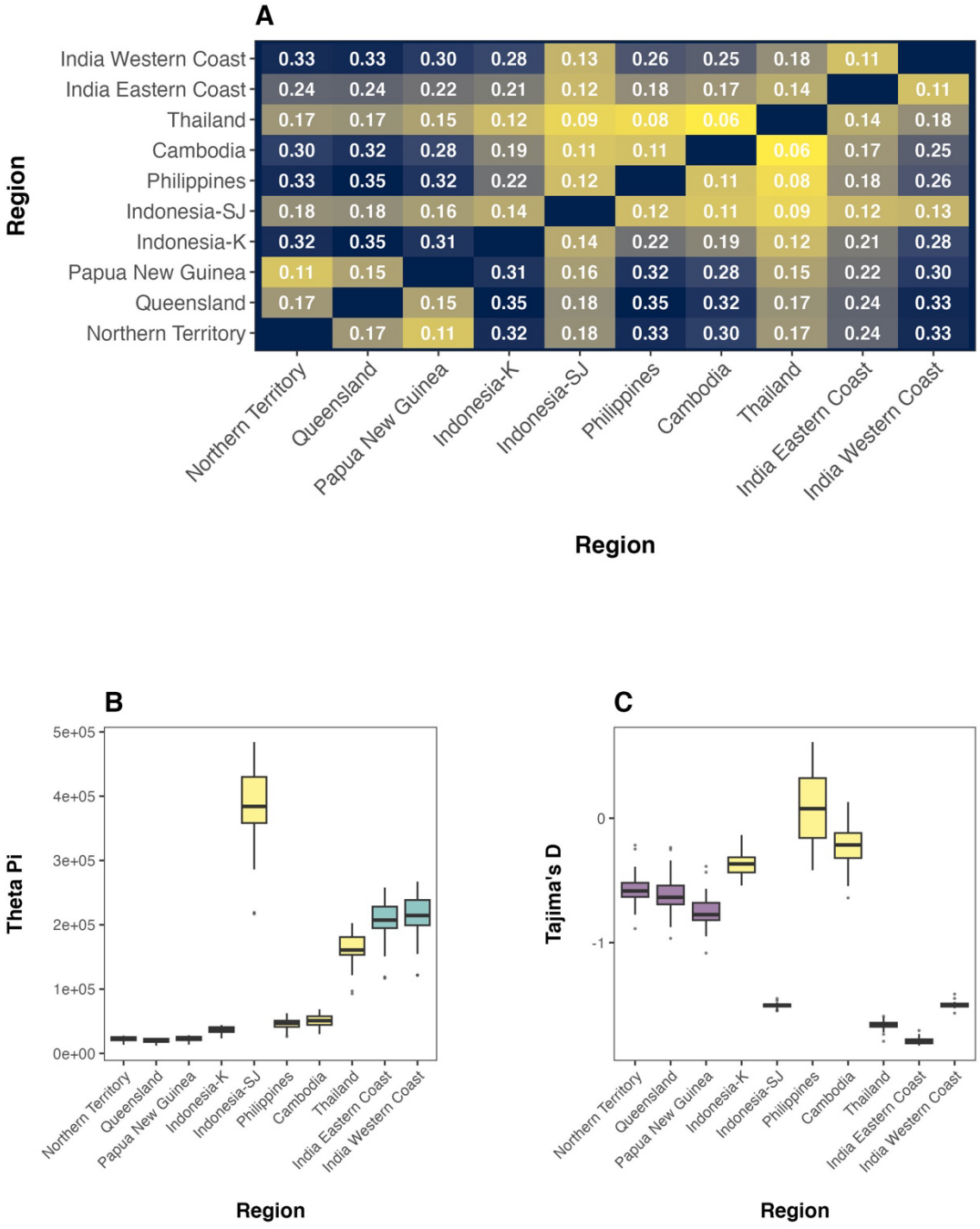


Figure 6 Population genetic statistics of barramundi perch in the WGS data set. **(A)** Pairwise F_{st} values between regions. **(B)** Nucleotide diversity of sampling regions in the WGS data set. **(C)** Tajima's D of sampling regions in the WGS data set. Indonesia is represented by two sampling regions. Indonesia-SU, Vietnam and Bangladesh were excluded due to low sample sizes ($n \leq 2$).

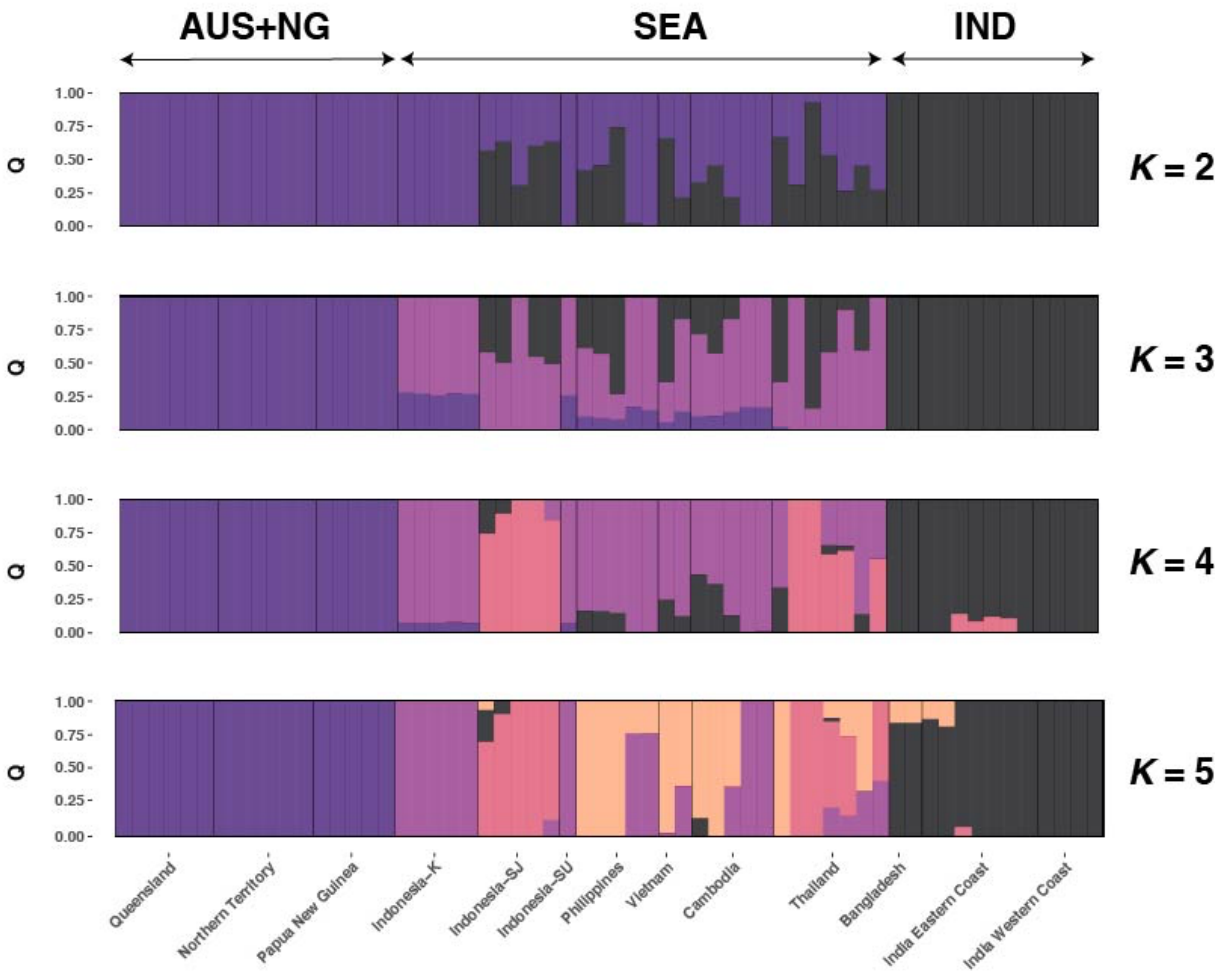


Figure 7 Genome-wide admixture analysis of the WGS data set (n=60) from 7,321,852 genotype likelihoods for K = 2 - 5 genetic clusters. Individuals are plotted on the x – axis and ancestry coefficient (Q) on the y – axis.

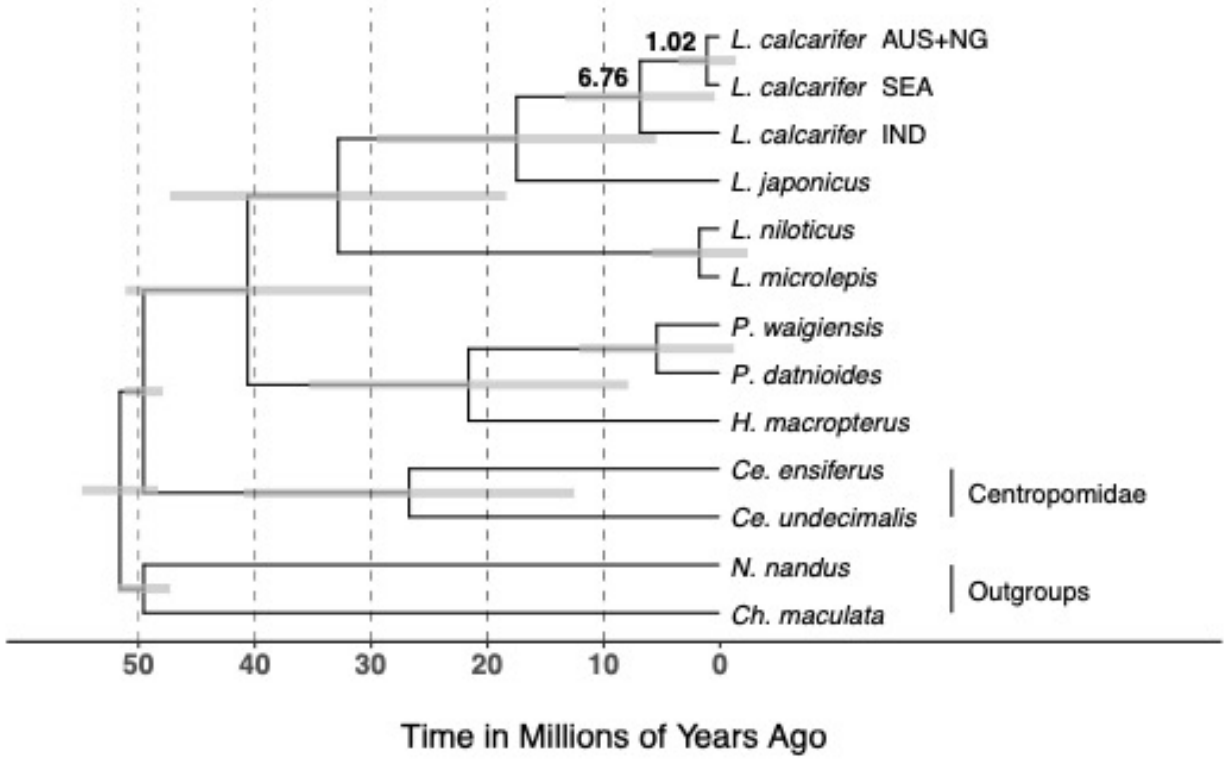
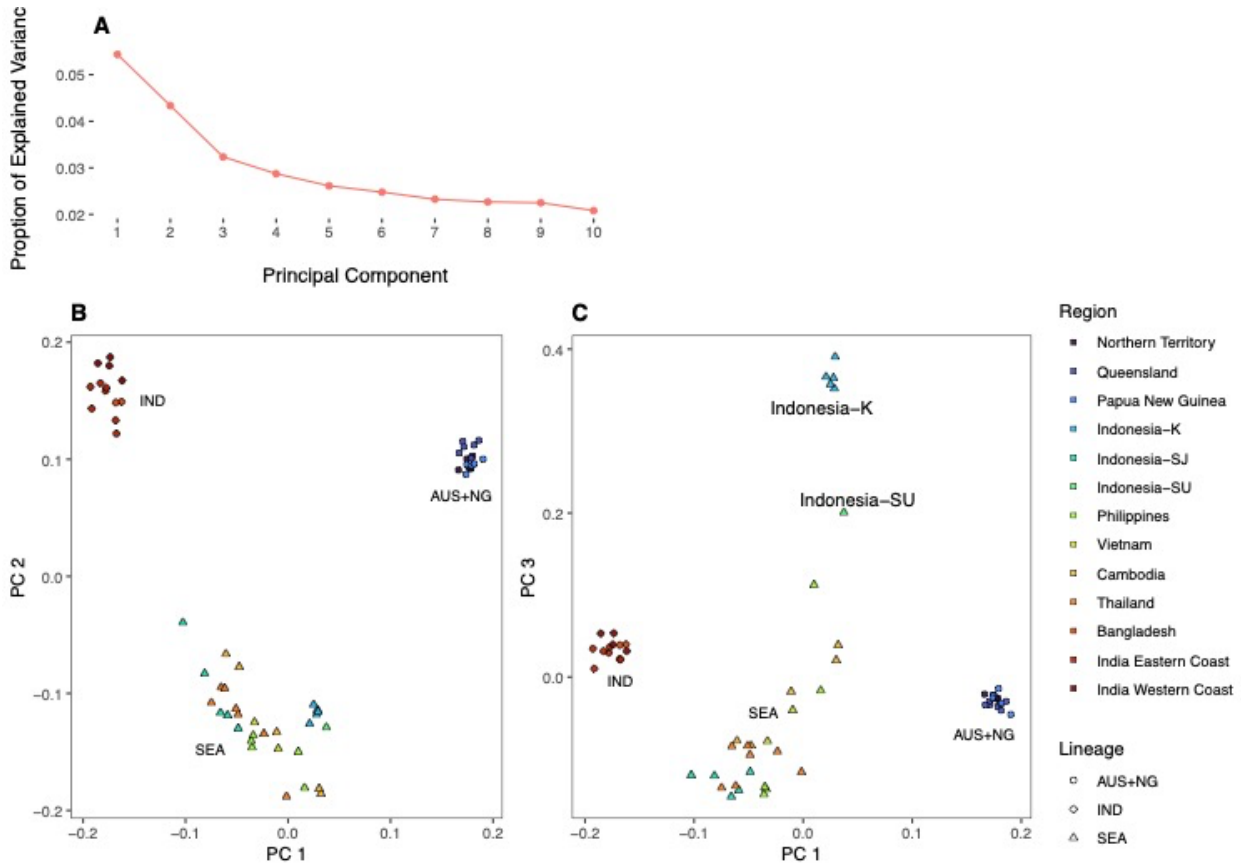
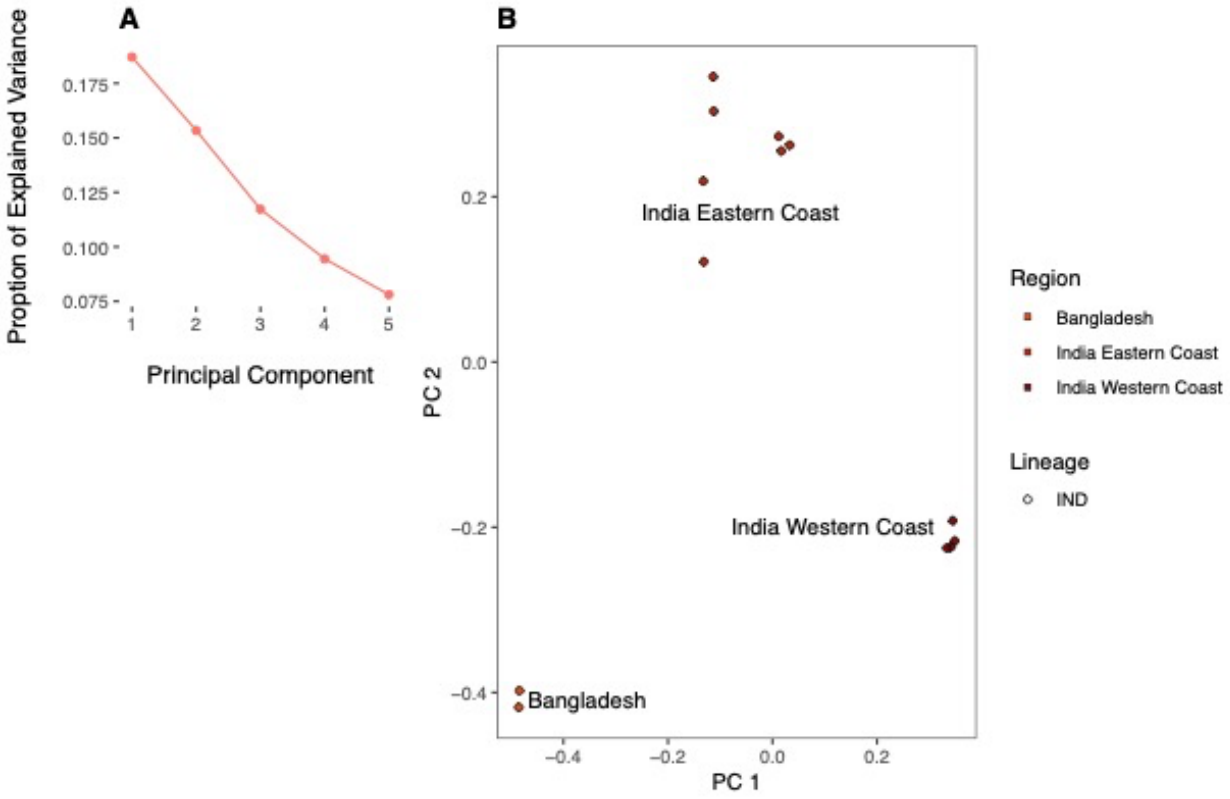


Figure 8 Time-calibrated phylogeny of main mitochondrial lineages of barramundi perch. Node age is from the posterior medians with 95% Highest Posterior Density (HPD) indicated by horizontal bars.



Supplemental Figure S1 PCAdapt analyses for identification of adaptive genetic loci. **(A)** Proportion of explained variance for each Principal Component (PC), supporting *K* = 3 PCs. **(B)** Plot of PC 1 & PC 2, IND, SEA and AUS+NG. **(C)** Plot of PC 1 & PC 3, PC 3 separates Indonesia-K and Indonesia-SU sampling locations from all other samples.

583



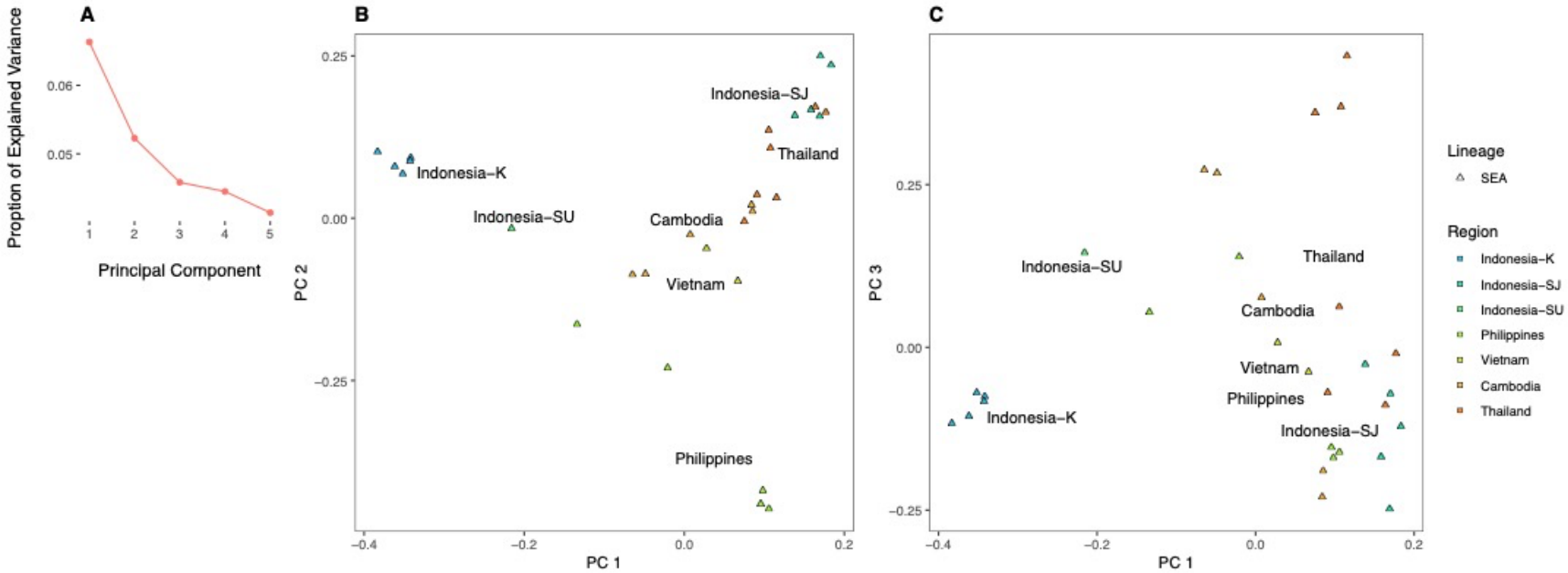
584

585

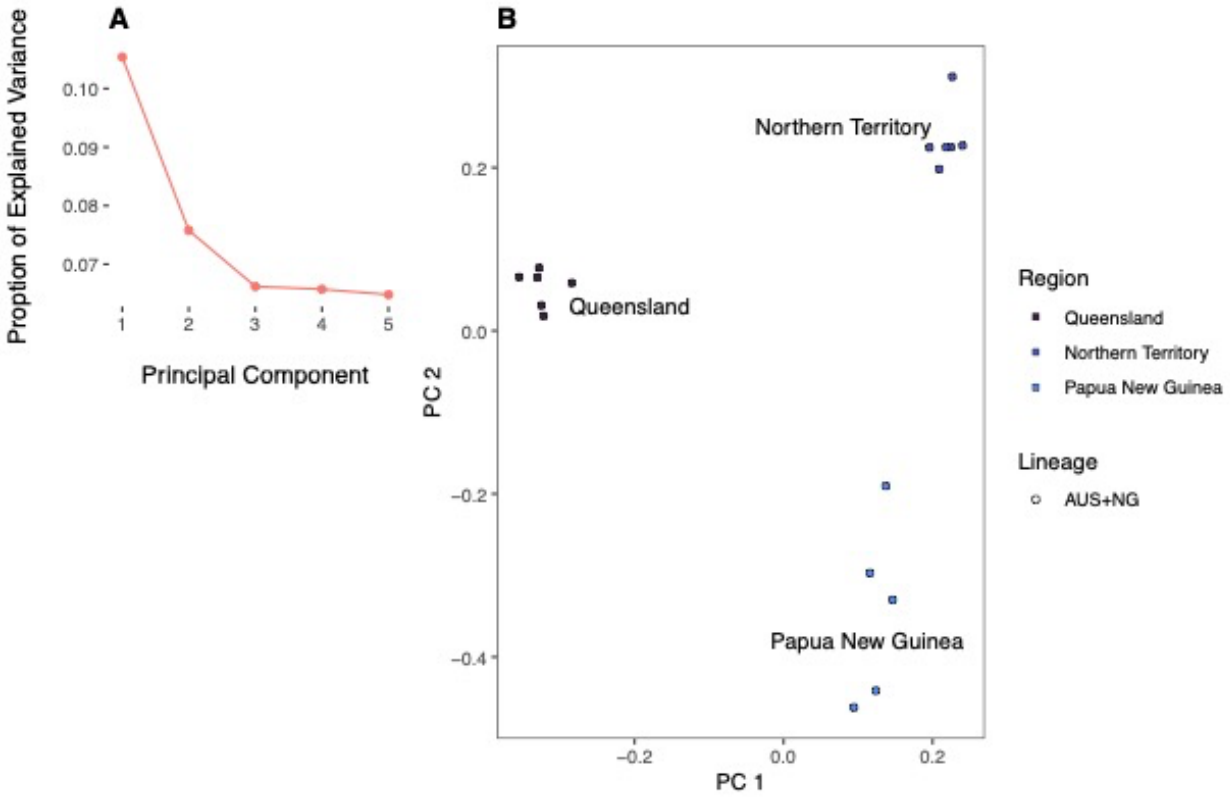
586 **Supplemental Figure S2** PCAdapt analyses for identification of adaptive genetic loci from the

587 IND genetic lineage. **(A)** Proportion of explained variance for each Principal Component (PC).

588 **(B)** Plot of PC 1 & PC 2.



Supplemental Figure S3 PCAdapt analyses for identification of adaptive genetic loci within the SEA lineage. **(A)** Proportion of explained variance for each Principal Component (PC), supporting at least *K* = 3 PCs. **(B)** Plot of PC 1 & PC 2, **(C)** Plot of PC 1 & PC 3.



Supplemental Figure S4 PCAdapt analyses for identification of adaptive genetic loci within the AUS+NG lineage. **(A)** Proportion of explained variance for each Principal Component (PC). **(B)** Plot of PC 1 & PC 2, separating Queensland, Northern Territory and Papua New Guinea sampling locations.

602 **Tables**

603 **Table 1** Sample origins by region of whole-genome sequence data set for *L. calcarifer sensu*
604 *lato*. The sample size and average coverage is presented for each sampling region.

Region	**Additional Geographic Information**	**Source**	**Sample Size**	**Average Coverage**
Northern Territory, Australia	Darwin Harbor	Wild-Caught	6	5.80
Queensland, Australia	Hinchinbrook Channel	Wild-Caught	6	5.34
Papua New Guinea	Fly River	Wild-Caught	5	5.83
Indonesia	K - Kalimantan	Unknown	5	5.73
Indonesia	SJ – South of Jakarta	Wild-Caught	5	17.8
Indonesia	SU - Sulawesi	Unknown	1	6.13
Philippines		Hatchery Broodstock	5	6.32
Vietnam		Hatchery Broodstock	2	6.21
Cambodia	Three locations	Wild-Caught	5	5.90
Thailand	East Coast	Wild-Caught	7	15.60
Bangladesh	Bakkhali River	Unknown	2	10.04
India Eastern Coast	Tamil Nadu	Wild-Caught	7	8.50
India Western Coast	Mumbai	Wild-Caught	4	20.22

605

606

607 **Table 2** Taxa and sequences included in time-calibrated phylogeny of *Lates calcarifer sensu*
608 *lato*. *Lates calcarifer* is represented by three lineages, Australia + New Guinea (AUS+NG),
609 Southeast Asia (SEA) and Indian subcontinent (IND) as described in the text. The family is
610 given, with the taxon, corresponding BOLD or NCBI accession and the geographic origin of the
611 sample if known.
612

Family	**Taxon**	**Accession**	**Origin**
Latidae	*Lates calcarifer* AUS+NG	EU189379.1	Western Australia
	Lates calcarifer SEA	KJ573921.1	Indonesia
	Lates calcarifer IND	KJ573895.1	Andaman and Nicobar Islands
	Lates japonicus	LC269832.1	Japan
	Lates microlepis	ISZA082-21	Democratic Republic of the Congo
	Lates niloticus	LT853908.1	Not Specified
	Hypopterus macropterus	LC269834.1	Australia
	Psammoperca waigiensis	OQ387793.1	Philippines, Luzon
	Psammoperca datnioides	NC035737.1	Australia, Darwin
Centropomidae	*Centropomus ensiferus*	MW183492.1	Brazil, Para, Braganca
	Centropomus undecimalis	JQ841102.1	Belize, Belize City
Channidae	*Channa maculata*	KP112195.1	China, Hubei, Wuhan
Nandidae	*Nanus nandus*	KX245097.1	India

613
614

Bibliography

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anderson, E.C., Thompson, E.A., 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160, 1217–1229. https://doi.org/10.1093/genetics/160.3.1217

Brown, J.H., Lomolino, M.V., 1998. Biogeography. Sinauer Press, Sunderland, MA.

Campbell, M.A., Badger, M.E., Buckmaster, N., Starostka, A.B., Hawks, T., Finger, A.J., 2023. Molecular phylogenetic and population genetic relationships of a putative species of sucker (*Catostomus* sp.) from Surprise Valley in the Great Basin, USA. Transactions of the American Fisheries Society 152, 273–286. https://doi.org/10.1002/tafs.10407

Campbell, M.A., Chen, W.-J., López, J.A., 2013. Are flatfishes (Pleuronectiformes) monophyletic? Mol. Phylogenet. Evol. 69, 664–673. https://doi.org/10.1016/j.ympev.2013.07.011

Campbell, M.A., López, J.A., Satoh, T.P., Chen, W.-J., Miya, M., 2014. Mitochondrial genomic investigation of flatfish monophyly. Gene 551, 176–182. https://doi.org/10.1016/j.gene.2014.08.053

Chen, W.-J., Ruiz-Carus, R., Ortí, G., 2007. Relationships among four genera of mojarras (Teleostei: Perciformes: Gerreidae) from the western Atlantic and their tentative placement among percomorph fishes. J. Fish. Biol. 70, 202–218. https://doi.org/10.1111/j.1095-8649.2007.01395.x

Chenoweth, S.F., Hughes, J.M., Keenan, C.P., Lavery, S., 1998. Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, Lates calcarifer (Australian barramundi). Heredity 80, 187–197. https://doi.org/10.1046/j.1365-2540.1998.00292.x

Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. Journal of Theoretical Biology 374, 35–47. https://doi.org/10.1016/j.jtbi.2015.03.006

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Cluney, B.J., 2004. The Northern Territory wild-catch barramundi *Lates calcarifer* fishery: a comparative investigation of the social, cultural, economic and environmental values of key stakeholders. (Master's). CDU.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

Delić, T., Trontelj, P., Rendoš, M., Fišer, C., 2017. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. Scientific Reports 7, 3391. https://doi.org/10.1038/s41598-017-02938-z

Des Roches, S., Post, D.M., Turley, N.E., Bailey, J.K., Hendry, A.P., Kinnison, M.T., Schweitzer, J.A., Palkovacs, E.P., 2018. The ecological importance of intraspecific variation. Nature Ecology & Evolution 2, 57–64. https://doi.org/10.1038/s41559-017-0402-5

Edwards-Vandenhoek, S., 2015. This my country, I'm painting here: the role of social practice art and participatory design in building resilience and place-making in the Warmun Aboriginal Community, Kimberley, Western Australia., in: Australia Council of University Art and Design Schools Conference Proceedings. Adelaide.

Fricke, R., Eschmeyer, W.N., van der Laan, R. (Eds.), 2023. Eschmeyer's Catalog of Fishes: Genera, Species, References.

Godfray, H.C.J., Knapp, S., Mace, G.M., 2004. The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359, 711–719. https://doi.org/10.1098/rstb.2003.1454

Hackl, T., Hedrich, R., Schultz, J., Förster, F., 2014. proovread : large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 30, 3004–3011. https://doi.org/10.1093/bioinformatics/btu392

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research. https://doi.org/10.1093/nar/gkt371

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518–522. https://doi.org/10.1093/molbev/msx281

Huson, D.H., Bryant, D., 2005. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254–267. https://doi.org/10.1093/molbev/msj030

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kaschner, K., Kesner-Reyes, K., Garilao, C., Rius-Barile, J., Rees, T., Froes, R., 2015. AquaMaps: Predicted range maps for aquatic species. [WWW Document]. AquaMaps. URL www.aquamaps.org (accessed 8.8.16).

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298. https://doi.org/10.1093/bib/bbn013

Keenan, C., 1994. Recent evolution of population structure in Australian barramundi, Lates calcarifer (Bloch): An example of isolation by distance in one dimension. Mar. Freshwater Res. 45, 1123–1148.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinform 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Layton, K.K.S., Snelgrove, P.V.R., Dempson, J.B., Kess, T., Lehnert, S.J., Bentzen, P., Duffy, S.J., Messmer, A.M., Stanley, R.R.E., DiBacco, C., Salisbury, S.J., Ruzzante, D.E., Nugent, C.M., Ferguson, M.M., Leong, J.S., Koop, B.F., Bradbury, I.R., 2021. Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. Nature Climate Change 11, 158–165. https://doi.org/10.1038/s41558-020-00959-7

Lehnert, S.J., Bradbury, I.R., Wringe, B.F., Van Wyngaarden, M., Bentzen, P., 2023. Multifaceted framework for defining conservation units: An example from Atlantic salmon (Salmo salar) in Canada. Evolutionary Applications n/a. https://doi.org/10.1111/eva.13587

Li, C., Ricardo, B.R., Smith, L.W., Ortí, G., 2011. Monophyly and interrelationships of Snook and Barramundi (Centropomidae sensu Greenwood) and five new markers for fish phylogenetics. Mol. Phylogenet. Evol. 60, 463–471. https://doi.org/10.1016/j.ympev.2011.05.004

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Loughnan, S.R., Smith-Keune, C., Beheregaray, L.B., Robinson, N.A., Jerry, D.R., 2019. Population genetic structure of barramundi (*Lates calcarifer*) across the natural distribution range in Australia informs fishery management and aquaculture practices. Mar. Freshwater Res. 70, 1533–1542.

Luu, K., Bazin, E., Blum, M.G.B., 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources 17, 67–77. https://doi.org/10.1111/1755-0998.12592

Maddison, W.P., 1997. Gene trees in species trees. Systematic Biology 46, 523–536. https://doi.org/10.1093/sysbio/46.3.523

Moore, R., 1979. Natural sex inversion in the Giant Perch (*Lates calcarifer*). Mar. Freshwater Res. 30, 803–813.

Moore, R., Reynold, L., 1982. Migration patterns of barramundi, *Lates calcarifer* (Bloch), in Papua New Guinea. Mar. Freshwater Res. 33, 671–682.

Moyle, P.B., Campbell, M.A., 2022. Cryptic species of freshwater sculpin (Cottidae: *Cottus*) in California. Zootaxa 5154, 501–527. https://doi.org/10.11646/zootaxa.5154.5.1

Newton, J.R., Zenger, K.R., Jerry, D.R., 2013. Next-generation transcriptome profiling reveals insights into genetic factors contributing to growth differences and temperature adaptation in Australian populations of barramundi (Lates calcarifer). Marine Genomics 11, 45–52. https://doi.org/10.1016/j.margen.2013.07.002

Nguyen, H.-V., Legras, J.-L., Neuvéglise, C., Gaillardin, C., 2011. Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380T. PLoS ONE 6, e25821. https://doi.org/10.1371/journal.pone.0025821

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268–274. https://doi.org/10.1093/molbev/msu300

Otero, O., 2004. Anatomy, systematics and phylogeny of both Recent and fossil latid fishes (Teleostei, Perciformes, Latidae). Zool. J. Linn. Soc-Lond. 141, 81–133. https://doi.org/10.1111/j.1096-3642.2004.00111.x

Pethiyagoda, R., Gill, A.C., 2012. Description of two new species of sea bass (Teleostei: Latidae: *Lates*) from Myanmar and Sri Lanka. Zootaxa 3314. https://doi.org/10.11646/zootaxa.3314.1.1

Privé, F., Luu, K., Vilhjálmsson, B.J., Blum, M.G.B., 2020. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. Molecular Biology and Evolution 37, 2153–2154. https://doi.org/10.1093/molbev/msaa053

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559–575. https://doi.org/10.1086/519795

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67, 901–904. https://doi.org/10.1093/sysbio/syy032

Rimmer, M.A., 2009. *Lates calcarifer*, in: Crespi, V., New, M. (Eds.), Cultured Aquatic Species Fact Sheets. FAO.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Santini, F., Harmon, L., Carnevale, G., Alfaro, M., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Bio. 9, 194. https://doi.org/10.1186/1471-2148-9-194

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., Webster, M.S., 2010. Population diversity and the portfolio effect in an exploited species. Nature 465, 609–612. https://doi.org/10.1038/nature09060

Shaklee, J., Salini, J., 1985. Genetic variation and population subdivision in Australian barramundi, Lates calcarifer (Bloch). Mar. Freshwater Res. 36, 203–218.

Shamsi, S.M., Williams, M., Mansourian, Y., 2020. An introduction to Aboriginal fishing cultures and legacies in seafood sustainability. Sustainability 12. https://doi.org/10.3390/su12229724.

Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions from Next Generation Sequencing data. Genetics 195, 693. https://doi.org/10.1534/genetics.113.154138

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).

Turbek, S.P., Bossu, C., Rayne, C., Gruppi, C., Kus, B.E., Whitfield, M., Smith, T.B., Paxton, E.H., Bay, R.A., Ruegg, K.C., 2023. Historical DNA reveals climate adaptation in an endangered songbird. Nature Climate Change 13, 735–741. https://doi.org/10.1038/s41558-023-01696-3

Vij, S., Kuhl, H., Kuznetsova, I.S., Komissarov, A., Yurchenko, A.A., Van Heusden, P., Singh, S., Thevasagayam, N.M., Prakki, S.R.S., Purushothaman, K., Saju, J.M., Jiang, J., Mbandi, S.K., Jonas, M., Hin Yan Tong, A., Mwangi, S., Lau, D., Ngoh, S.Y., Liew, W.C., Shen, X., Hon, L.S., Drake, J.P., Boitano, M., Hall, R., Chin, C.-S., Lachumanan, R., Korlach, J., Trifonov, V., Kabilov, M., Tupikin, A., Green, D., Moxon, S., Garvin, T., Sedlazeck, F.J., Vurture, G.W., Gopalapillai, G., Kumar Katneni, V., Noble, T.H., Scaria, V., Sivasubbu, S., Jerry, D.R., O'Brien, S.J., Schatz, M.C., Dalmay, T., Turner, S.W., Lok, S., Christoffels, A., Orbán, L., 2016. Chromosomal-level assembly of the Asian Seabass genome using long sequence reads and multi-layered scaffolding. PLOS Genetics 12, e1005954. https://doi.org/10.1371/journal.pgen.1005954

Vij, S., Purushothaman, K., Gopikrishna, G., Lau, D., Saju, J.M., Shamsudheen, K.V., Kumar, K.V., Basheer, V.S., Gopalakrishnan, A., Hossain, M.S., Sivasubbu, S., Scaria, V., Jena, J.K., Ponniah, A.G., Orbán, L., 2014. Barcoding of Asian seabass across its geographic range provides evidence for its bifurcation into two distinct species. Frontiers in Marine Science 1.

Wallace, A.R., 1859. The geographical distribution of birds. Ibis 1, 449–454.

Wang, D., Wong, M.-K., Lee, M.-Y., Campbell, M.A., Chen, W.-J., In Revision. Systematics of the superfamily Trichiuroidea (Teleostei: Scombriformes): undescribed species and a new family of commercially important demersal fishes in tropical and temperate seas. Zoological Journal of the Linnean Society.

Wang, L., Wan, Z.Y., Lim, H.S., Yue, G.H., 2016. Genetic variability, local selection and demographic history: genomic evidence of evolving towards allopatric speciation in Asian seabass. Molecular Ecology 25, 3605–3621. https://doi.org/10.1111/mec.13714

Waples, R.S., 1991. Definition of "species" under the Endangered Species Act: application to Pacific salmon. NOAA technical memorandum NMFS F/NWC ; 194.

Ward, R.D., Holmes, B.H., Yearsley, G.K., 2008. DNA barcoding reveals a likely second species of Asian sea bass (barramundi) (*Lates calcarifer*). J. Fish Biol 72, 458–463. https://doi.org/10.1111/j.1095-8649.2007.01703.x

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8, 28–36. https://doi.org/10.1111/2041-210X.12628

Yue, G.H., Zhu, Z.Y., Lo, L.C., Wang, C.M., Lin, G., Feng, F., Pang, H.Y., Li, J., Gong, P., Liu, H.M., Tan, J., Chou, R., Lim, H., Orban, L., 2009. Genetic variation and population structure of Asian seabass (Lates calcarifer) in the Asia-Pacific region. Aquaculture 293, 22–28. https://doi.org/10.1016/j.aquaculture.2009.03.053