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 is a widespread Indo-Pacific fish that is important in aquaculture, recreational and commercial fisheries. Genetic divergences from different data sources and sampling schemes have been reported.

Aims

To conduct phylogenetic and population genetic analyses from a geographically and phylogenetically representative data set to identify hierarchical divisions within *L. calcarifer*. We further test the evolutionary significance of genetic units in terms of signatures of adaptation.

Methods

Using a whole-genome sequence data set of 61 fish including an outgroup, we conducted phylogenetic and population genetic analyes. We also generated measures of F_{st}, nucleotide diversity (π) and Tajima's D (D).

Key Results

We identify three main lineages of *L. calcarifer* corresponding to the Indian subcontinent, Southeast Asia and Australasia. Subdivision within each of the three main lineages is also identified and characterized. Adaptively significant differences are indicated within and between the three main lineages.

Conclusions

L. calcarifer exhibits genetic divergences at different levels that originate before and during the Pleistocene. These divergences are associated with adaptive divergence but unclear phenotypic changes.

Implications

This study highlights the need for comprehensive sampling and integrative study of genotypes and phenotypes across the range of *L. calcarifer.*

Key Words

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

Introduction

Essential processes necessary for life on Earth are sustained by biodiversity, and biodiversity provides for human needs in various ways. Regardless of its limits and challenges, the basic unit for understanding biodiversity is the concept of species (e.g. Agapow et al., 2004). While the recognition of species is indeed important, so is the recognition of units below the species level that contribute to biodiversity. The ecological effects of intraspecific variation can meet our exceed species effect and therefore are important to document and recognize (Des Roches et al., 2018). Conservation of both species-level and intraspecific variation serves to promote ecological, cultural and economic functions of organisms, resilience and genetic variation for future adaptive potential (e.g. Layton et al., 2021; Lehnert et al., 2023; Schindler et al., 2010). Specific and intraspecific variation can be considered through both isolation and adaptation without the reliance on formal species taxonomy (Waples, 1991).

We apply these basic ideas of isolation and adaptation to identify variation to a Indo-Pacific *Lates*. Most Indo-Pacific *Lates* are classified under *Lates calcarifer* (Bloch 1790), and it is a valuable commercial taxon of cultural and recreational significance e.g. (Cluney, 2004; Edwards-Vandenhoek, 2015; Shamsi et al., 2020). The common names 'barramundi' and 'Asian sea bass' are widely used. A broad-ranging taxon, *L. calcarifer* is often considered to be distributed from the Persian Gulf eastwards to Taiwan and Australia, and into Oceania (**Figure 1**). Abed et al. (2023) note that there are no reliable reports from the Persian Gulf of native *Lates* and indicate a natural distribution through western India. Within this broader range, two regional endemics may be recognized, in Sri Lanka *L. lakdiva* Pethiyagoda & Gill 2012 and Myanmar *L. uwisara* Pethiyagoda & Gill 2012 (Pethiyagoda and Gill, 2012). We refer to all these fishes together as Indo-Pacific *Lates*.

Aquaculture, beginning in Thailand in the 1970's, has increased greatly and Indo-Pacific *Lates* are now widely propagated across their range with some aquaculture outside the native range. Indo-Pacific *Lates* are now grown in diverse locations such as the United States of America (Arizona, Hawaii and Massachusetts), Israel, Taiwan and the Persian Gulf (Iraq) (Abed et al., 2023; Chen et al., 2024; Mathew, 2009). Propagation is also undertaken for stocking programs in lakes across northern Australia (e.g. Manton Dam, Northern Territory; Lake Kununurra, Western

Australia) to support recreational fisheries and tourism in remote communities. Generally, Indo-Pacific *Lates* are considered to be protandrous hemaphrodites, 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. Movement for aquaculture, e.g. from Thailand to Taiwan also can result in the mixing of fish of different backgrounds and opportunities for release or escape (Chen et al., 2024). Other qualities of Indo-Pacific *Lates* such as high fecundity, ease of culture and rapid growth (i.e. harvestable size of 3 kg in less than two years) encourage their production (Rimmer, 2009).

The widespread culture and commercial fishing for Indo-Pacific *Lates* has led to genetic studies to understand population genetic structure. Initial mitochondrial barcoding revealed a deep divergence between Australian and Myanmar fish classified under *L. calcarifer* (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 and Bay of Bengal fish (IND, including the Andaman Islands and Myanmar) separate from Southeast Asia (SEA, e.g. Singapore, Malaysia, Thailand and Indonesia) and Northern Australia (Vij et al., 2014). Vij et al. (2014) did include the sequences from Ward et al. (2008) and demonstrated that they are similar to a more broadly distributed IND mitochondrial lineage. Restriction site-Associated DNA sequencing (RADseq) applied to population-level samples of Indo-Pacific *Lates* (n = 131) from six localities (Thailand, Malaysia, Indonesia, Western Australia, Eastern Australia and Papua New Guinea) revealed a high degree of differentiation of SEA fish from Australia + New Guinea (AUS+NG). Pairwise F_{st} within SEA was 0.005-0.009 but 0.360-0.443 between SEA and AUS+NG without evidence of contemporary gene flow (Wang et al., 2016). Some investigation of population genomic structure was applied to a range-wide data set at a high level as part of a genome assembly study but did not investigate concordance between mtDNA and genomic DNA or population genetic metrics (Vij et al., 2016). Broadly, Vij et al. (2016) identified a separation of Indian Ocean fish, Southeast Asian fish and Australasian fish into genetic clusters. Population genetic structure range-wide examined with microsatellites (K = 4 genetic clusters presented) produced rather unclear distinctions of Indian Ocean fish from parts

of Southeast Asia, and indicates a central Indonesian phylogroup and Australasian phylogroup (Jerry and Smith-Keune, 2014). No study to date, has analyzed both mitochondrial data and genomic data explicitly from the same individuals to evaluate concordance between data sources and resolve possible discordance between mitochondrial and genomic signals.

While there is evidence of genetic differentiation within Indo-Pacific *Lates* at a high phylogenetic level (Vij et al. 2014), this has only been evaluated thoroughly with molecular phylogenetic approaches using mtDNA. Within major genetic lineages, there is also genetic differentiation evident (e.g. Chenoweth et al., 1998; Shaklee and Salini, 1985; Yue et al., 2009), and the evolutionary significance of these differences are not substantiated. We apply a hierarchical approach to identify isolation (discreteness) of Indo-Pacific *Lates* genetic lineages with molecular phylogenetic and population genetic approaches, and test for adaptive significance between these lineages. Within lineages, we test for additional isolation and adaptive significance of these differences with population genetic approaches.(Lehnert et al., 2023). We further evaluate evidence of hybridization between main genetic lineages to determine if the hybridization is recent (F1, F2) and presumably a result of human activities or represents older admixture events.

Methods

We approached the question of isolation within Indo-Pacific *Lates* by first constructing phylogenies using mitochondrial DNA and genome-wide SNPs. Subsequently, population genetic techniques were used, and divisions were examined at a high level (genetic lineages) and then within those lineages. We tested for evolutionarily significance through identifying possible adaptive loci between and within main lineages of Indo-Pacific *Lates.* An outstanding question is if hybridization is recent as a result of human mediated translocations or a result of natural admixture between lineages. We applied statistical testing to this question to determine if evidence of recent introgression in mtDNA was corroborated at the genetic level. Finally, we developed a timeline for diversification of Indo-Pacific *Lates* using a time-calibrated phylogenetic approach.

134 *Whole-Genome Sequence Data Set*

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

146

147 *Mitochondrial Phylogeny*

148 A mitochondrial alignment from the WGS data was created by creating a Binary Alignment Map
149 (BAM) file for each individual with more than 3X coverage only including reads aligned to the
150 reference mitochondrial genome (NC_007439.1, samtools view). These BAM files were
151 converted to consensus FASTQ files with the bam2cns subprogram of proovread (Hackl et al.,
152 2014) then to FASTA files turning all bases with quality scores below 20 to N with seqtk (seq -a
153 -q20 -n N) (https://github.com/lh3/seqtk). The mitochondrial Cytochrome Oxidase I (mtCOI)
154 gene was extracted from the complete mitochondrial FASTA-formatted sequences with
155 SAMtools (faidx NC_007439.1:6329-7879). For rooting, the corresponding region of the *L.*
156 *japonicus* mitochondrial genome was extracted from the NCBI reference sequence
157 (NC_034339.1) and sequences were aligned with MAFFT (Katoh et al., 2002; Katoh and Toh,
158 2008). A Maximum-Likelihood (ML) phylogeny of sequences was generated with IQTREE
159 specifying a Hasegawa-Kishino-Yano model of nucleotide substitution with gamma-distributed
160 rate variation (-m HKY+G) and support for nodes evaluated with rapid bootstrapping (-bb
161 10000) (Hoang et al., 2018; Nguyen et al., 2014). The tree was imported into R and visualized
162 with the ggtree package (Yu et al., 2017) collapsing nodes to polytomies receiving less than 50%
163 Bootstrap Support (BS) and rooting by the longest branch.

164

165 *Genome-Wide Phylogenetic and Network Analyses*

166 We generated a genome-wide set of called SNPs with Analysis of Next Generation Sequencing
167 Data (ANGSD) from the 24 main linkage groups in the BP assembly (Korneliussen et al., 2014).
168 Quality of called SNPs was enforced by ensuring that a SNP was present in at least 90% of
169 individuals, had a minimum Minor Allele Frequency (MAF) of 0.05, a minimum mapping
170 quality of 20, minimum base quality of 20, a SNP p – val of 1.00×10^{-6} and a posterior cutoff of
171 0.90. A SAMTools model was specified and a PLINK-formatted file created as the output. We
172 converted the PLINK-formatted file to a VCF-formatted file with PLINK (Purcell et al., 2007).
173 Linked SNPS were removed with BCFTools +prune (-l 0.20 -w 10000).
174

175 The multispecies coalescent was applied to model gene-tree and species-tree discordance
176 resulting from Incomplete Lineage Sorting, ILS (Maddison, 1997) on a set of called SNPs that
177 included the outgroup *L. japonicus*. A species-tree was generated using SVDQuartets in PAUP*
178 version 4.0a (Chifman and Kubatko, 2015, 2014; Swofford, 2003). Samples were pooled by
179 sampling region for the species tree. A random set of quartets (evalQuartets=random), partially
180 ambiguous SNPs were distributed over all compatible site patterns (ambigs=distribute), and
181 1,000 bootstrap replicates (bootstrap=standard nreps=1000) were options specified.
182

183 For ML phylogenetic analysis and network analysis we called SNPS with BP samples and a
184 phylogeny of SNPs was inferred with IQTREE by identifying a model of nucleotide evolution
185 with ModelFinder Plus and indicating the ascertainment bias correction should be applied (-m
186 MFP+ASC) (Kalyaanamoorthy et al., 2017; Nguyen et al., 2011). IQTREE was used to further
187 filter SNPs for suitability for ascertainment bias correction. Support for nodes was evaluated
188 with 10,000 rapid boostraps (Hoang et al., 2018). As concatenated phylogenetic analyses do not
189 explicitly model hybridization or ILS, we created an implicit network to identify potential
190 hybridization or ILS in our data set with SplitsTree (Huson and Bryant, 2005). The same
191 alignment provided to IQTREE was analyzed with SplitsTree.
192

193 *Population Genetic Structure and Statistics*

194 Total signal in the data set was assessed by the generation of covariance matrix with ANGSD
195 and a PC analysis. The covariance matrix was generated by with a single read sampling approach

196 due to a wide range of individual coverage (minimum 4.73, maximum 34.81, mean 9.38 and
197 median 6.41) with the options -doIBs 1 and -doCov 1 in ANGSD. The 24 main linkage groups of
198 the reference assembly were sampled for this analysis. A variable site was required to be present
199 in >90% of individuals with a minimum MAF of 0.05. A minimum mapping quality of 10 was
200 specified with a minimum base quality of 20 and a GATK genotype likelihood model was used
201 (-GL 2). The resulting covariance matrix was imported into R for PC analysis.
202
203 To calculate genetic differentiation in terms of F_{st} was calculated by generating Site Frequency
204 Spectra (SFS) for each sampling region with ANGSD and the realSFS subprogram, splitting
205 Indonesia into two locations based on metadata available with the sequences ("K" n=5 and and
206 "SJ" n=5). Vietnam (n=2), Indonesia "SU" (n=1) and Bangladesh (n=2) were excluded due to
207 small sample sizes. Two-dimensional SFS were calculated between each possible pair of
208 sampling regions and F_{st} calculated (see code repository accompanying this paper). Nucleotide
209 diversity (π) and Tajima's D (D) estimates were calculated for each chromosome with realSFS
210 using the saf2theta option followed by the thetaStat subprogram of ANGSD and the do_stat
211 option.
212
213 Admixture analysis was conducted with NGSadmix (Skotte et al., 2013) on genotype likelihoods
214 generated with ANGSD as for the PC analysis using a single read sampling approach. The
215 resulting BEAGLE-formatted file was used with NGSadmix to produce an admixture plot of K =
216 2-5 genetic clusters.
217
218 *Tests for Adaptative Genetic Variation*
219 Genotype calls produced for phylogenetic network analysis after pruning were imported into R
220 for analysis with PCAdapt (Luu et al., 2017; Privé et al., 2020). The *pcadapt* function was then
221 used to generate a scree plot and identify an objective number of PCs with the first 20 PCs
222 examined. From the resulting object from the *pcadapt* function, outlier loci were identified by
223 filtering for a Benjamini and Hochberg (1995) and Bonferroni corrected p – values with
224 significance values less than 0.10. We repeated this analysis for individuals in each main genetic
225 lineage by filtering for variants with a minimum minor allele frequency of 0.05 and examining
226 the first 5 PCs.

227

228 *Evaluation of Hybridization Levels*

229 It is possible that Southeast Asian *Lates* exhibit a mixture of hybridization levels between IND
230 and SEA lineages, with possible F1s, F2s and advanced backcrosses. Two wild-caught samples
231 in the data set from Cambodia (sequence read accessions SRR3165998, SRR3165999) had IND
232 lineage mtDNA sequences (see results, Figure 2). We applied NewHybrids (Anderson and
233 Thompson, 2002) to test if these two Cambodian fish were recent (F1, F2) or older hybrids
234 (advanced backcrosses). Using the called and pruned SNP data sets, we excluded recently
235 reported potential chromosomal inversions (Campbell and Hale, 2024) and imported the data
236 into R and analyzed with functions of the dartR package (Gruber et al., 2018). We restricted the
237 analysis to fish from the SEA and IND lineages as NewHybrids functions with two populations
238 and mixtures of those populations then filtered out low MAF SNPS (threshold of 0.25). The
239 *gl.nhybrids* function of dartR was used to drive a NewHybrids analysis with India Western Coast
240 and Bangladesh sampling locations designated as one parental lineage (P0) and Indonesia-K and
241 Indonesia-SU as another (P1). We allowed SNPs to not be entirely fixed between parental
242 populations (threshold=0.01) and chose a subset of 200 SNPs for analysis with the 'AvgPIC'
243 criterion. Analysis included a 100,000 step burnin and 50,000 sweeps.

244

245 *Divergence Time Estimation*

246 Existing efforts have covered a broad geographic and lineage sampling of Indo-Pacific *Lates*
247 with barcoding (mitochondrial *cox1*) (Vij et al., 2014; Ward et al., 2008). The effort of this study
248 was then to calculate divergence times of the three lineages (IND, SEA and AUS+NG). We
249 downloaded barcode sequences from the Barcode of Life Database and Genbank for Latidae.
250 With BLAST (blastn) we confirmed the major lineages of Indo-Pacific *Lates* in barcode data
251 (Altschul et al., 1997) then selected representative sequences from these three lineages to
252 conduct a divergence time analysis (**Table 1**). We also searched for representation across Latidae
253 for broad taxonomic representation of this family including sequences from *L. japonicus*, *L.*
254 *microlepis* and *L. niloticus*. We included representatives of the two main mitochondrial lineages
255 present in *Psammoperca* corresponding to *P. waigiensis* and *P. datnioides*. Centropomidae has
256 been confirmed to be the sister lineage of Latidae in molecular phylogenetic study e.g.
257 (Campbell et al., 2013; Chen et al., 2007; Li et al., 2011) and inclusion of Centropomidae allows

258 dating the Time to Most Recent Common Ancestor (TMRCA) of Centropomidae and Latidae to
259 48.6 million years ago (mya) (Campbell et al. 2013; Otero, 2004). We also included sequences
260 from the outgroup lineages Channidae and Nandidae as an additional calibration point dating the
261 TMRCA of those taxa to be 48 mya (Santini et al., 2009).
262
263 Sequences were aligned with MAFFT (Katoh et al., 2002; Katoh and Toh, 2008), in frame
264 coding was confirmed by translation in the NCBI ORF finder and a NEXUS file made for a
265 time-calibrated phylogenetic analysis in MrBayes (Ronquist et al., 2012). The third codon
266 position was excluded due to the age of the data set and expected saturation occurring in
267 mitochondrial sequences from carangimorph taxa at these timescales e.g. (Campbell et al., 2014).
268 A nucleotide substitution model of General Time Reversible (GTR) with a proportion of
269 invariant sites (I) and gamma-distributed rate variation (Γ) was specified. Constraints were
270 applied to Centropomidae + Latidae to date with an exponential distribution used,
271 offsetexp(48.60, 63.18), and the outgroup, offsetxp(48.00,62.40). Initially, three runs and three
272 chains were specified with a run length of 1 million generations sampled ever 1,000 generations
273 and a 25% burnin applied. Evaluation of parameters and convergence with Tracer (Rambaut et
274 al., 2018) was assessed, and run length increased until sufficient ESS for all parameters (>200)
275 was reached. Tree summarization was done in MrBayes with the sumt command and an all
276 compatible consensus tree made.
277

278 # Results

279 *Whole-Genome Sequence Data Set*
280 From the original 63 Indo-Pacific *Lates* sequences, 60 had coverage over 3X and were included
281 in additional analyses (**Table 1**). Sample sizes for regions were 13 for putative members of the
282 IND lineage, 30 for putative members of the SEA lineage, and 17 for putative members of the
283 AUS+NG lineage from ten sampling regions (**Figure 1**). The *L. japonicus* sequence aligned to
284 the *L. calcarifer* reference genome with high success with 110X coverage and was down-
285 sampled to 20X coverage with SAMtools to be more consistent with genome coverage measures
286 of Indo-Pacific *Lates*.
287
288 *Mitochondrial Phylogeny*

289 The 61 samples (60 Indo-Pacific *Lates* and one *L. japonicus*) produced an alignment of 1,576
290 base pairs in length with 71 distinct patterns and 128 parsimony-informative sites. In the
291 alignment, 1,441 sites were invariant. The ML phylogeny indicates a clear division of *L.*
292 *calcarifer* IND individuals with two Cambodian-origin fish from all other fish (**Figure 2**).
293 Support for monophyly of this IND mtDNA lineage is maximal. Support for monophyly of SEA
294 + AUS+NG is maximal, and support for a clade of *L. calcarifer* AUS+NG is moderate
295 (BS=80%). Individuals sourced from the putative *L. calcarifer* SEA clade do not form clear
296 geographic clusters, with the exception of a Central Indonesia mtDNA lineage that receives low
297 support for monophyly (BS=57%) and includes a single sequence from the Philippines. Support
298 for the arrangement of major mtDNA lineages SEA and AUS+NG lineages is low (BS=55%).
299 The Assemble Species by Automatic Partition algorithm indicated the strongest support for two
300 species, with a threshold distance of 4.10% (results provided in the Data Supplement). This
301 division separates the IND mitochondrial lineage from the combined SEA and AUS+NG lineage.
302
303 *Genome-Wide Phylogenetic and Network Analyses*
304 For analysis with SVDQuartets to generate a rooted species-tree, 8,529 variants that all have
305 unique alignment patterns were analyzed and is presented as **Figure 3**. The SNP alignment
306 consisted of 6,456 parsimony informative and 2,073 singleton sites. Three lineages were
307 apparent in the species-tree with maximal support for monophyly corresponding to the *L.*
308 *calcarifer* IND, *L. calcarifer* SEA and *L. calcarifer* AUS+NG. The SEA and AUS+NG lineages
309 were highly supported (BS=100%) as sister lineages. Within the IND lineage, the India Eastern
310 Coast and Bangladesh sampling locations were indicated to be sister lineages (BS=100%). The
311 AUS+NG lineage individuals from the Northern Territory and Papua New Guinea were indicated
312 to be most closely related (BS=100%). The first division within the SEA lineage is between
313 Central Indonesia sampling locations (Indonesia-SU + Indonesia-K) and receives maximal
314 support for monophyly and branching patterns. The Philippines sampling location separates with
315 the next successive branching within the SEA lineage (BS=100%). Relationships among
316 Cambodia, Indonesia-SJ, Thailand and Vietnam sampling locations were not clearly resolved.
317
318 For ML phylogenetic analysis and phylogenetic network analysis an alignment of 60 BP
319 individuals with 7,661 SNPs having 7,661 distinct unique alignment patterns and 5,571

320 parsimony informative sites was created. The alignment contains 2,090 singleton sites. Support
321 for monophyly of IND lineage fish is maximal, as well as geographic divisions into Eastern and
322 Western Indian subcontinent (**Figure 4A**). Indo-Pacific *Lates* from SEA as a grouping receives
323 low support (BS = 76%) with concordance between sampling locations and placement not
324 complete. Indonesia-K and the Philippines sampling locations were highly supported to be
325 monophyletic, but Cambodia, Thailand and Indonesia-SJ were not. Support for a group of
326 AUS+NG sampling locations is maximal, with maximal support for division between
327 Queensland and Northern Territory + Papua New Guinea fish. Monophyly of samples from the
328 Northern Territory is poorly supported (BS = 68%).
329
330 The phylogenetic network resolves an IND lineage containing two sub-lineages (India Eastern
331 Coast + Bangladesh, India Western Coast) with separation of all three sampling locations
332 (**Figure 4B**). The SEA lineage is present separated between Central Indonesia and all other
333 locations as two main sub lineages, with reticulations with the IND lineage occurring with a
334 subgroup of 10 individuals, disrupting monophyly of some sampling locations. The Indonesia-K
335 and Philippines locations were monophyletic, but the Cambodia, Thailand and Indonesia-SJ
336 sampling locations were not. The AUS+NG grouping is composed of three subclades
337 Queensland (Australia), Northern Territory (Australia) and Papua New Guinea. In the network,
338 Northern Territory and Papua New Guinea fish were most-closely related to each other.
339
340 *Population Genetic Structure and Statistics*
341 The PC analysis included 7,321,852 variable sites in the genome and indicated a separation into
342 three larger genetic clusters corresponding to IND, SEA and AUS+NG individuals
343 (**Supplemental Figure S1**). The first PC separated AUS+NG, SEA and IND individuals (PC 1
344 12.57% of variation) with the second PC showing variation within SEA and IND clusters (PC2
345 4.71%). Clear separation of Indian East Coast + Bangladesh and Indian West Coast individuals is
346 present on both PC 2 and PC 3 (PC 3 3.51%).
347
348 Genetic differentiation (F_{st}) was lowest within main genetic lineages, such as Thailand and
349 Cambodia (F_{st} = 0.06) within the *L. calcarifer* SEA lineage and between the Indian East Coast
350 and Indian West Coast sampling locations (F_{st} = 0.11) within the *L. calcarifer* IND lineage

351 (**Figure 5A**). The highest levels of genetic differentiation were found between main genetic
352 lineages, between AUS+NG and SEA F_{st} = 0.15-0.35, between AUS+NG and IND F_{st} = 0.22-
353 0.33, between SEA and IND F_{st} = 0.12-0.26). Nucleotide diversity is the lowest in the AUS+NG
354 lineage (π = 20,067-22,775) and highest in the IND lineage (π = 207,299-214,485) (**Figure**
355 **5B**). Tajima's D was strongly negative (< -1) in the Indonesia-SJ, Thailand, India Eastern Coast
356 and India Western Coast sampling locations (**Figure 5C**). The AUS+NG sampling locations had
357 *D* values between -0.58 and -0.77.
358
359 Admixture analysis of 7,321,852 genotype likelihoods from 60 BP individuals for *K* = 2-5 is
360 presented as **Supplemental Figure S2**. At *K* = 2, the Bangladesh, Indian Eastern Coast and
361 Indian Western Coast have high ancestry coefficients (*Q*) > 0.99 for a single genetic cluster and
362 the AUS+NG lineage and Indonesia-K sampling location have high ancestry coefficients for the
363 other. The remaining SEA sampling locations have a mixture of contributions from both genetic
364 clusters. At *K* = 3 the AUS+NG individuals form a clearly genetic cluster with *Q* > 0.99 for
365 ancestry coefficients, the SEA sampling locations were heterogenous mixture proportions from
366 sampling locations largely composed of the third genetic cluster, with the Indonesia-K and
367 Indonesia-SU sampling locations exhibiting a near uniform ~0.27% admixture proportion with
368 the genetic cluster typical of AUS+NG sampling locations. Higher values of *K* lead to separating
369 of genetic units within the SEA lineage that may or may not correspond to sampling locations.
370
371 *Tests for Adaptation*
372 The variation within the pruned genotype calls of 22,088 SNPs from the n=60 WGS sequences
373 with >3X coverage supported an optimal *K* = 3 PCs (**Supplemental Figure S3**). These PCs
374 separated the three main lineages of *L. calcarifer*, with the third PC separating the Indonesia-K
375 sampling location and to a lesser degree the Indonesia-SU individual from other SEA fish.
376 Benjamini and Hochberg (1995) adjusted *p* – values (BH) identified 134 candidate adaptive
377 SNPs and Bonferroni corrected *p* – values (BF) indicates 34 candidate adaptive SNPs across the
378 three PC axes.
379
380 Filtering to IND individuals (n=13) resulted in 12,075 SNPs with a minimum MAF of 0.05, with
381 the first PC separating the Bangladesh, India Eastern Coast and India Western Coast with 738

382 BH and 379 BF candidate adaptive SNPs identified on this axis (**Supplemental Figure S4**). The
383 second PC axis separated India Eastern Coast with 285 BH and 59 BF candidate adaptive SNPS.
384 From the SEA lineage, 30 individuals with 17,556 SNPs with a minimum MAF of 0.05 were
385 analyzed. The first three PCs were considered, separating Indonesia-K and Indonesia-SU
386 sampling locations on PC 1, the Philippines along PC 2 and some Thailand and Cambodia
387 individuals along PC 3 with 190 BH and 29 BF candidate adaptive SNPs identified in total
388 (**Supplemental Figure S5**). The data set of AUS+NG individuals (n=17) for loci with a
389 minimum MAF of 0.05 resulted in 14,642 SNPs. We found strong support for $K = 2$ PCs, the
390 first PC separating Queensland from the Northern Territory and Papua New Guinea
391 (**Supplemental Figure S6**). The second PC separated Papua New Guinea from the other
392 locations with 234 BH candidate adaptive SNPs and 57 BF candidate adaptive SNPs across the
393 two PC axes.
394
395 *Evaluation of Hybridization Levels*
396 The NewHybrid analysis of 43 individuals from the SEA and IND lineages indicated all IND
397 lineage fish were most likely a parent population (P0, **Supplemental Figure S7**). This includes
398 India Eastern Coast, which was not designated *a priori* as a potential parent population.
399 Indonesia-K and Indonesia-SU were resolved as P1 fish with high probability. The two wild-
400 caught fish from Cambodia with IND lineage mtDNA were indicated to be F2s, with other fish
401 from Cambodia P1 (n=2) and a backcrossed P1 fish.
402
403 *Divergence Time Estimation*
404 We compiled a barcode alignment of 13 sequences with 660 characters resulting in 440
405 characters (399 invariant) when including the first and second codon positions only. Sufficient
406 effective sample sizes (ESS >200, range of 712-3,002) was reached with 20 million generations
407 and two runs of three chains each as evaluated after a 25% burnin. The resulting consensus tree
408 presented in **Figure 6.** The median divergence time of the IND lineage from the other two
409 lineages is estimated to be 6.76 mya with a 95% Highest Posterior Density (HPD) of 1.31-14.24
410 mya. The divergence of AUS+NG from SEA is estimated to be 1.02 mya with a 95% HPD range
411 of 0.00-5.03 mya. The relationships of Indo-Pacific *Lates* mitochondrial lineages were highly
412 supported, with the placement having a Posterior Probability (PP) of 0.99, and the sister

413 relationship between AUS+NG and SEA having a PP of 0.97. The NEXUS file for use with
414 MrBayes and the consensus tree are provided in Data Supplement of this paper.
415

416 Discussion

417 We identify from mitochondrial and nuclear genomic data three main genetic lineages of Indo-
418 Pacific *Lates*. These are an Indian Ocean lineage (IND), and a Pacific Ocean lineage split
419 between Southeast Asia (SEA) and Australasia (AUS+NG). A Central Indonesian genetic
420 cluster, representing Borneo and Sulawesi appears as distinctive group with placement poorly
421 supported in our mtDNA analysis (Figure 2). Phylogenetic and population genetic analyses using
422 data from the nuclear genome, place the Central Indonesian samples as more closely related to
423 Southeast Asian fish (SEA), and we treat them as part of that (SEA) lineage.
424
425 The three main genetic lineages of Indo-Pacific *Lates* are present with substantial differences
426 quantified in terms of genome-wide estimates of F_{st}, candidate adaptive loci and deep divergence
427 times. We observed F_{st} values up to 0.36 and in terms of divergence times, we estimated 6.76
428 million years for the divergence of IND from other Indo-Pacific *Lates* and a divergence time
429 estimate of 1.02 million years for AUS+NG from SEA. The previously published estimate based
430 on RADseq data for divergence times by Wang et al. (2016) is restricted to AUS+NG from the
431 SEA lineage and is an approximate 644,100-878,900 years before present. Both of these
432 estimates, the estimate presented here from fossil-calibrated phylogeny and Wang et al. (2016),
433 support a divergence during the glacial cycling and changes in sea level and conditions during
434 the Quaternary. The three main lineages of Indo-Pacific *Lates* exhibit discreteness (genetic
435 isolation) and evolutionary significance (i.e. adaptive differentiation).
436
437 Within each of the three main lineages, there was additional evidence for discreteness and
438 evolutionary significance. The IND sampling locations were divided phylogenetically (**Figures 3**
439 **& 4**), and in PC analysis (**Supplemental Figure S4**) with candidate adaptive SNPs identified.
440 We find that all three sampling locations within IND (Bangladesh, India Eastern Coast and India
441 Western Coast) can be separated clearly with genomic data but exhibit some mixing of mtDNA
442 haplotypes. The SEA lineage exhibited notable separation of the Indonesia-K and Indonesia-SU
443 sampling locations, the Philippines sampling location and some Thailand fish (**Figure 4**,

444 **Supplemental Figure S5**). Clear differentiation of all Philippines and Thailand fish may be
445 impacted by hybridization (**Supplemental Figure S2**). The RADseq data set of Wang et al.
446 (2016) did not show clear evidence of genetic structure with three sampling locations within
447 SEA, but broader sampling with microsatellites did appear to support genetic structuring (Yue et
448 al., 2009). The AUS+NG sampling exhibited a clear division between Queensland and other
449 sampling locations (**Figures 3 & 4**) and all three sampling locations could be clearly
450 differentiated on two PC axes with candidate adaptive loci apparent (**Supplemental Figure S6**).
451 Within the AUS+NG lineage there are several studies demonstrating genetic differentiation
452 within that lineage (e.g. Chenoweth et al., 1998; Loughnan et al., 2019; Shaklee and Salini, 1985;
453 Wang et al., 2016) with 16 discrete genetic populations identified in Australia only (Keenan,
454 1994).
455
456 *What is* Lates calcarifer*?*
457 Our fundamental understanding of nature is based at the species level and the lack of recognition
458 of species is a major knowledge shortfall in conservation and management (Brown and
459 Lomolino, 1998). Molecular genetic techniques have continued to identify candidate species of
460 fishes which may or may not have diagnostic morphological characters typical of species
461 descriptions (e.g. Campbell et al., 2023; Moyle and Campbell, 2022). Instead of being a
462 historical aspect of biology, the need for continued taxonomic work is very important (e.g. Delić
463 et al., 2017; Godfray et al., 2004). Taxonomic work often involves the detailed anatomical
464 investigation of organisms that can reveal phenotypic differences related not only to speciation,
465 but also life history variation. What remains outstanding is if Indo-Pacific *Lates* can be
466 accurately represented within Linnean taxonomy or if an alternative approach is better suited to
467 the biological realities of Indo-Pacific *Lates*.
468
469 The most recent taxonomic effort relevant to BP led to the description of *L. lakdiva* Pethiyagoda
470 & Gill 2012 from western Sri Lanka and *L. usiwara* Pethiyagoda & Gill 2012 from eastern
471 Myanmar (Pethiyagoda and Gill, 2012). The specimens of *L. usiwara* were the same as featured
472 in the barcoding study of Ward et al. (2008) highlighting deep mitochondrial divergences, but in a
473 wider context the barcode sequences of *L. usiwara* are identical to samples from the Indian
474 subcontinent (Vij et al., 2014). However, in comparison to *L. calcarifer*, a differential diagnosis

475 was possible and the distribution of *L. usiwara* is perhaps indicative of a recently derived species
476 or ecotype. The anatomical study undertaken by Vij et al. (2014) along with barcoding did not
477 clearly separate all these newly described species of *Lates* from *L. calcarifer* noting small sample
478 sizes used in the description of *L. usiwara* and *L. lakdiva* and conclude that there is substantial
479 mismatch between anatomical descriptions of new species of *Lates* (within the Indian Ocean) and
480 candidate species of *Lates* apparent genetically (between the Indian and Pacific oceans at the
481 broadest scale). Barcode sequences, it should be noted, will not reflect proportion of ancestry as
482 do genomic data and ascertaining the ancestry proportions of fish used in taxonomic studies is
483 critical going forward due human translocations of *Lates*. Furthermore, it is possible for a high
484 degree of genome-wide differentiation or substantial reproductive isolation be present that would
485 not be apparent in barcode data. Pethiyagoda and Gill (2012) themselves observe that due to the
486 large body size and sequential hermaphodism of *L. calcarifer sensu lato*, they were restricted to
487 examining many juveniles for their study.
488
489 The three genetic lineages of Indo-Pacific *Lates* remain unclear in distribution and have unclear
490 anatomical differentiation (Vij et al., 2014). Portions of the range, such as Taiwan, remain
491 unsampled for genomic data and genetic structuring is apparent across all three lineages (**Figures**
492 **3 & 4, Supplemental Figures S1-S6**). Differentiation between and within the main lineages can
493 be clarified with additional sampling and representation of nominal taxa in genomic data sets
494 (i.e. *L. usiwara* and *L. lakdiva*). Given the strong genetic differentiation and deep inferred time
495 scales (**Figure 6**), one possible resolution is to recognize at least two species-level taxa that
496 correspond to the IND lineage and a combined SEA and AUS+NG lineage. At a potential
497 species-level, the taxonomy of Indo-Pacific *Lates* is complex and the recognition of types and
498 their origins is also challenging, reviewed by Pethiyagoda and Gill (2012). The type locality of *L.*
499 *calcarifer* (Bloch 1790) may be Tamil Nadu, India (10.4°N, 79.3°E) (Pethiyagoda and Gill,
500 2012). Thus, the IND lineage may be represented by this name. There are numerous available
501 names that may be applicable to the SEA lineage, but this requires substantial time and expertise
502 to resolve. There is, however, an available name, *L. cavifrons* (Alleyne & Macleay 1877) that
503 could be applied to AUS+NG fish. The collection locality is either the Torres Strait or the coast
504 of New Guinea with a lost holotype with the drawing providing the current synonymy
505 (Pethiyagoda and Gill, 2012).

506

507 Without formal taxonomic recognition, the three main genetic lineages of Indo-Pacific *Lates* can
508 be recognized for management purposes such as mixed stock fisheries. Vij et al. (2014) note that
509 the assumption of single species of Indo-Pacific *Lates* has undoubtedly influenced previous
510 studies and microsatellite data from Thailand showing the highest allelic richness may reflect
511 genotyping of two divergent genetic lineages along the west coast of Thailand (Yue et al., 2009).
512 If different species of Indo-Pacific *Lates* are recognized, then questions of biology can be
513 researched such as the causes of speciation and what the mechanisms for reproductive isolation
514 of IND may be. Development of an anatomical key usable in the field that can separate these
515 species e.g. (Wang et al., In Revision) would be of great practical utility and could be used to
516 resolve the distributions and overlap of IND and SEA lineages. A high-throughput sequencing
517 assay that can discriminate the main lineages could provide a rapid assessment of genetic
518 lineage, hybridization levels and possible provenance in studies of food adulteration.
519

520 *Admixture*

521 Overall, analyses that can quantify hybridization in this study are challenging because of the high
522 innate diversity of the SEA lineage and limited sample sizes, notably that of the IND lineage (n =
523 13). The sequences were also obtained opportunistically from the NCBI Sequence Read Archive
524 and not collected by the authors. With these cautions in mind, we find evidence of admixture
525 between major genetic lineages within our data set as evidenced by the placement of Cambodia-
526 origin samples, with a largely SEA genomic background within the mitochondrial IND lineage
527 (**Figures 2, 3 & 4**). When examining all 60 individuals at $K = 3$, the two Cambodia individuals
528 with IND mtDNA have ancestry coefficients (Q) of 0.29 and 0.43 with the IND genetic cluster,
529 however, there is such a degree of variation in the SEA group with respect to admixture
530 proportions and inconsistency across sampling locations that these measurements should not be
531 accepted at face value. The Cambodia group has two individuals with Q = 0.00 for the IND
532 genetic cluster and one with 0.17 for example when examining all individuals. Separate analysis
533 of Cambodia and *L. calcarifer* IND individuals only show $Q = 0.23$ and 0.33 for the two
534 Cambodia individuals with IND mtDNA and no evidence for nuclear genomic admixture of
535 other Cambodian individuals with IND. This finding is congruent with the NewHybrids analysis

536 indicating the presence of two F2 hybrids in the Cambodian fish, in line with the ancestry
537 proportions of these fish analyzed with IND fish separately (**Supplemental Figure S7**).
538
539 Nucleotide diversity in the SEA Indonesia-SJ location is very high (**Figure 5B)** and no doubt
540 influences the reduced F_{st} of this lineage to other sampling regions (**Figure 5A**). Combined with
541 the large negative *D* value which arises from admixture (Hahn et al., 2013), this sampling
542 location may contain admixed individuals. A similar pattern of high nucleotide diversity, low-F_{st},
543 and strongly negative *D* is apparent with the Thailand sampling location (**Figure 5**). Both the
544 Indonesia-SJ and Thailand sampling locations show substantial contributions of the IND genetic
545 cluster in their ancestry (**Supplemental Figure S2**), but it is not uniformly distributed across
546 sampling locations. If the distribution of ancestry is uniform across sampling locations, it would
547 suggest an older age of hybridization. In NewHybrid analyses, Indonesia-SJ and Thailand
548 contain numerous recent (F2) hybrids (Supplemental Figure S7). Furthermore, the admixture
549 analyses suggest that Thailand at $K = 4$ is represented by more than one distinct genetic unit from
550 within the SEA lineage with additional contributions from the IND lineage. The dearth of F1
551 hybrids found by NewHybrids was surprising but may have to do with a high degree of diversity
552 within the SEA lineage as a result of (1) retained ancestral variation and (2) historical admixture.
553 Furthermore, the selection of a random set of 200 SNPs due to computational limitations may not
554 capture informative loci for hybridization assessment.
555

556 # Conclusions

557 We provide several advances in the study of Indo-Pacific *Lates*, presenting molecular
558 phylogenetic analyses of a geographically and genomically representative data set and testing for
559 adaptive differentiation and recent or more advanced hybridization. The analysis of genome-
560 wide nuclear variants also provides evidence of hierarchical divergences, that is within the IND,
561 SEA and AUS+NG lineages there is genetic structuring. The SEA lineage demonstrates a
562 complex history with a combination of natural genetic differentiation and what is likely human
563 mediated admixture events. Clearly, samples of known provenance, the use of informative loci
564 that are diagnostic to lineages, and boosting of sampling locations and sample sizes focused
565 across the possible intergrade zone IND and SEA would help resolve these ambiguities. Future
566 studies making use of archive and museum samples to provide genetic baselines prior to use of

BP as an aquaculture species can be used to create a genetic baseline to compare modern genetics against to identify potential translocations and adaptation to changing environments (Turbek et al., 2023).

Phenotypic and life history variation within Indo-Pacific *Lates* remain poorly characterized and are other 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 Indo-Pacific *Lates* in commercial, recreation and subsistence fisheries as well as aquaculture creates differing priorities across user groups. Further exploration of genotypic and phenotypic variation can support conservation aquaculture and food production as well as the preservation of productivity and adaptive potential of Indo-Pacific *Lates* 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.

Acknowledgements

The authors would like to recognize two anonymous reviewers that provided valuable feedback for improving the manuscript.

Figures

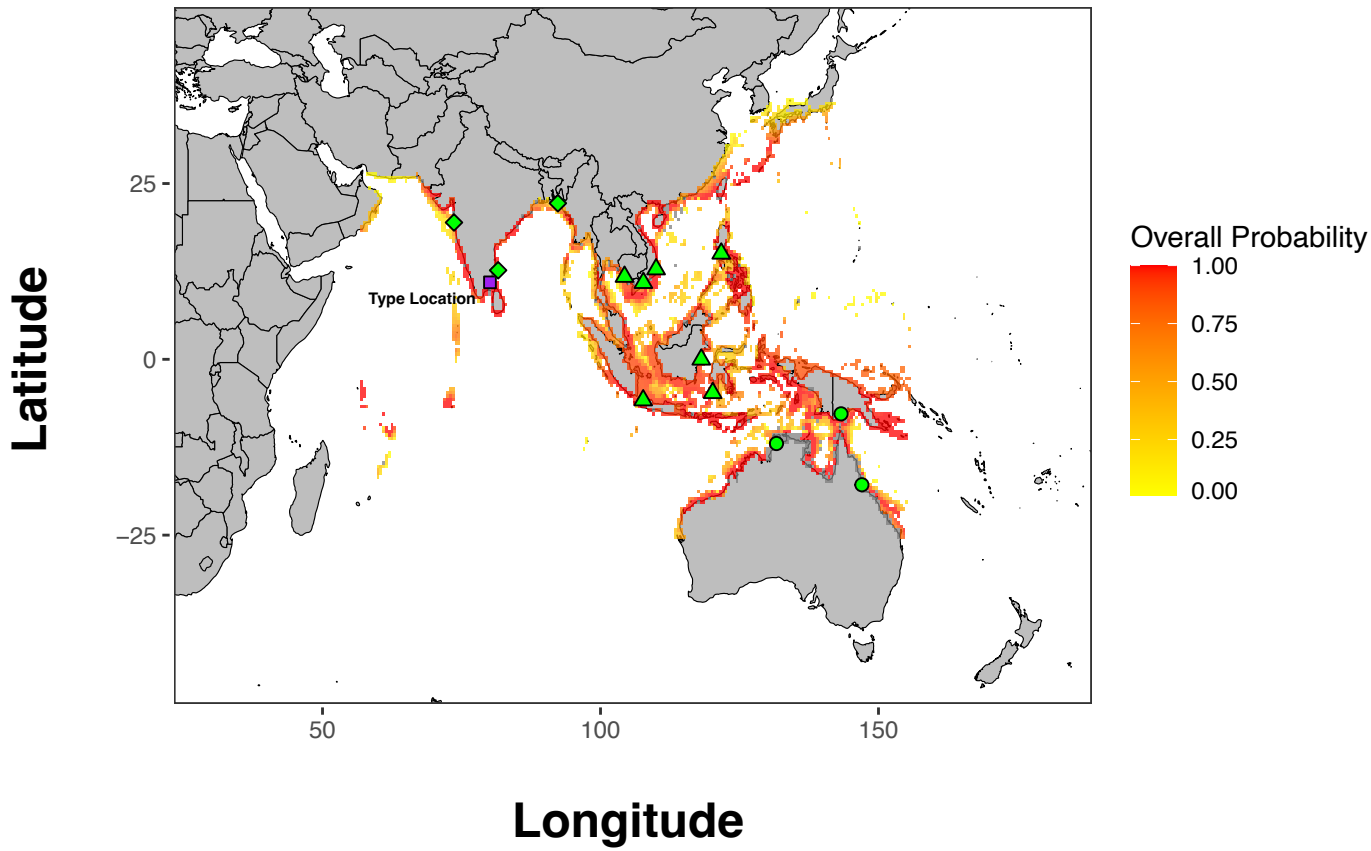


Figure 1 Distribution of *Lates calcarifer sensu lato* from AquaMaps (Kaschner et al., 2015) with probability of occurrence (Overall Probability) colored from yellow (low) to red (high). The type locality of *L. calcarifer* is indicated by a purple square. Sampling of main genetic lineages included in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

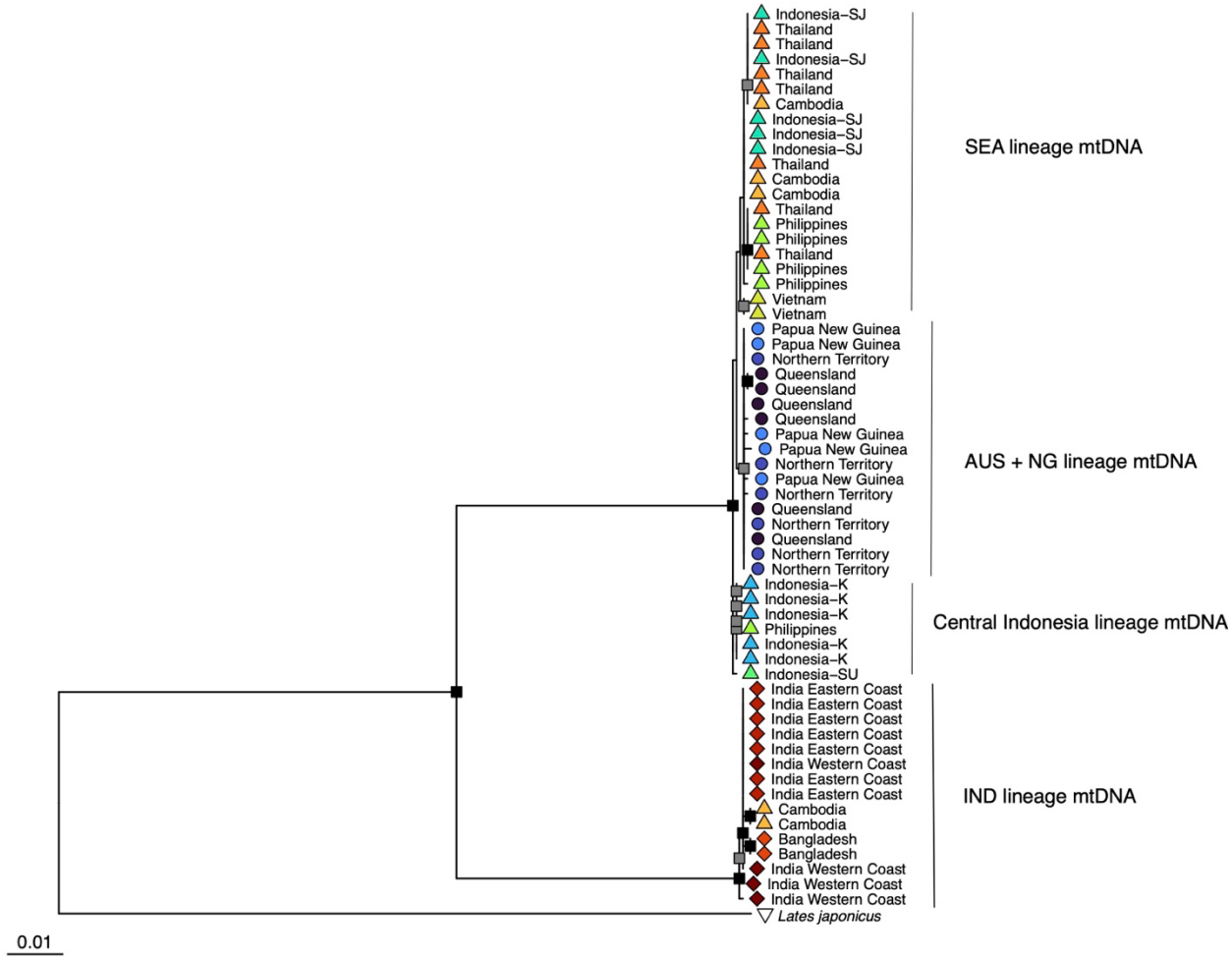


Figure 2 Maximum-Likelihood phylogeny of mitochondrial Cytochrome Oxidase I from 60 Indo-Pacific *Lates* individuals derived from whole-genome sequence data. The tree is rooted by *L. japonicus*. Node shape indicates the putative lineage of origin and collection information is provided as tip labels. Nodes with < 50% bootstrap support are collapsed to polytomies. Nodes with ≥ 75% bootstrap support but < 90% bootstrap support are indicated as grey squares and nodes ≥ 90% bootstrap support as black squares.

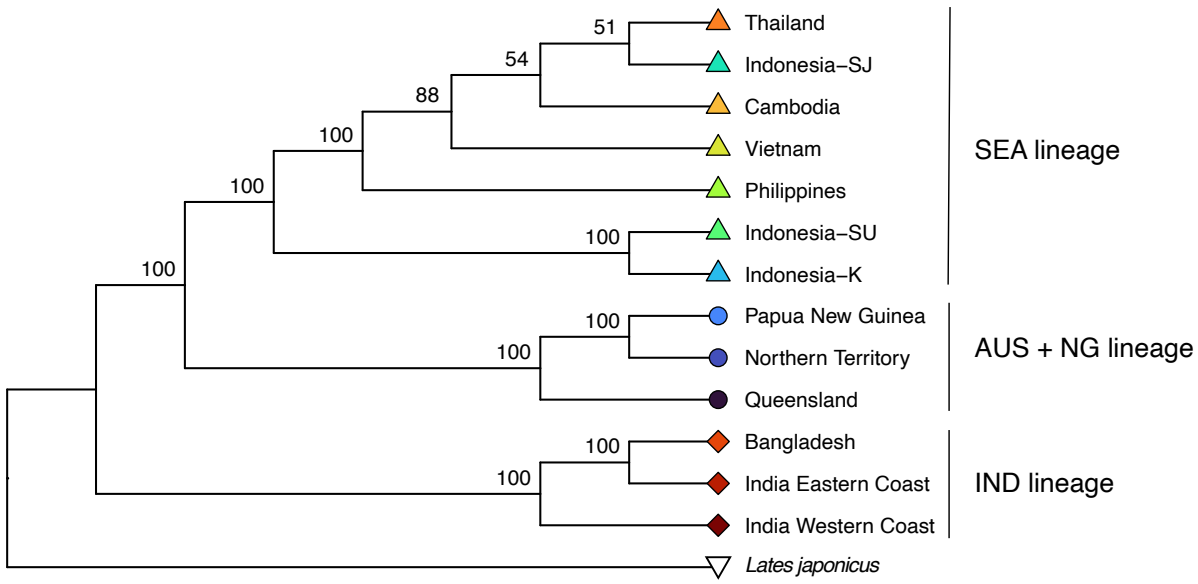


Figure 3 Species-tree from SVDQuartets of Indo-Pacific *Lates* sampling locations in this study based on 8,529 variants rooted by *Lates japonicus*. Values at nodes are bootstrap support values.

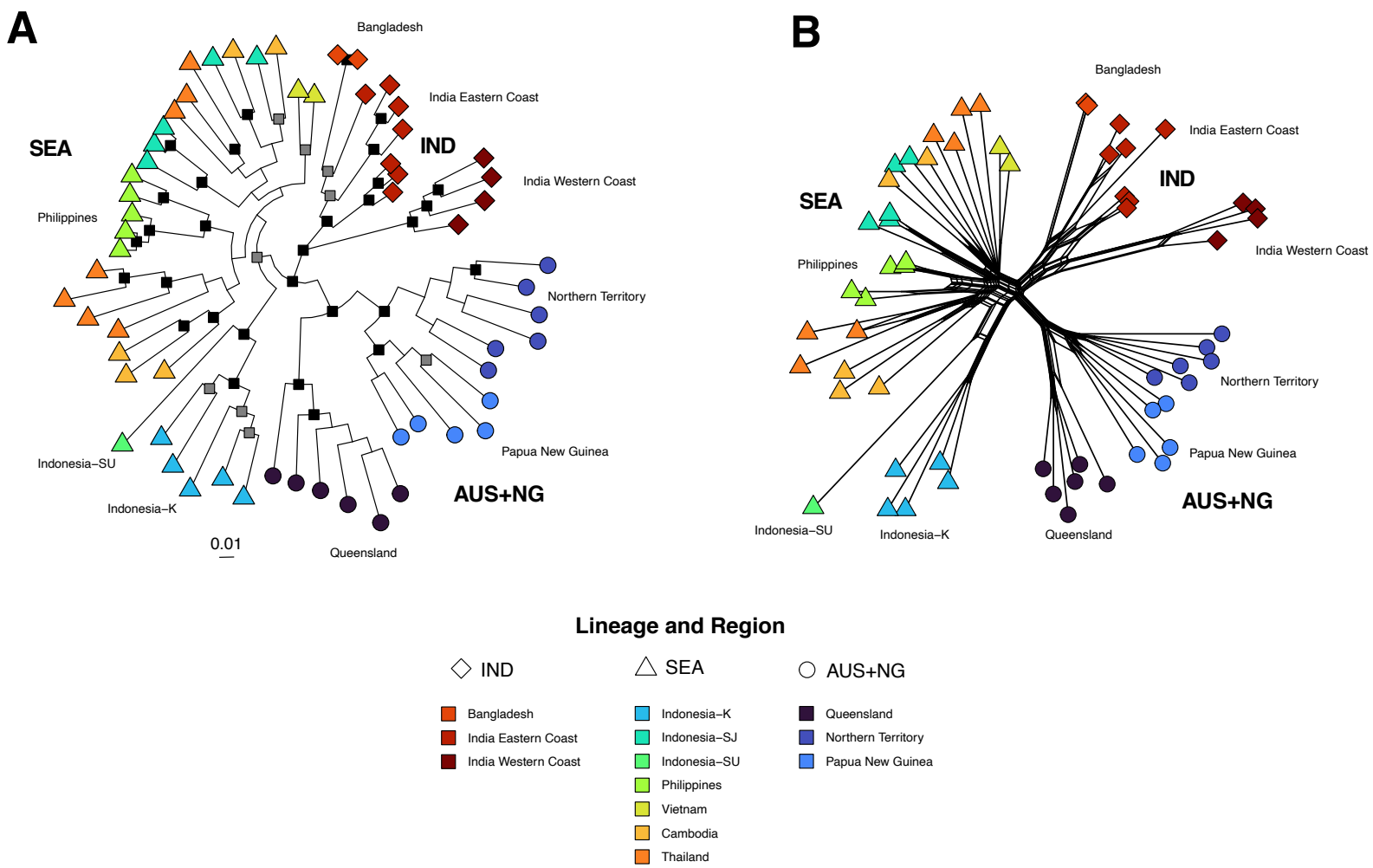
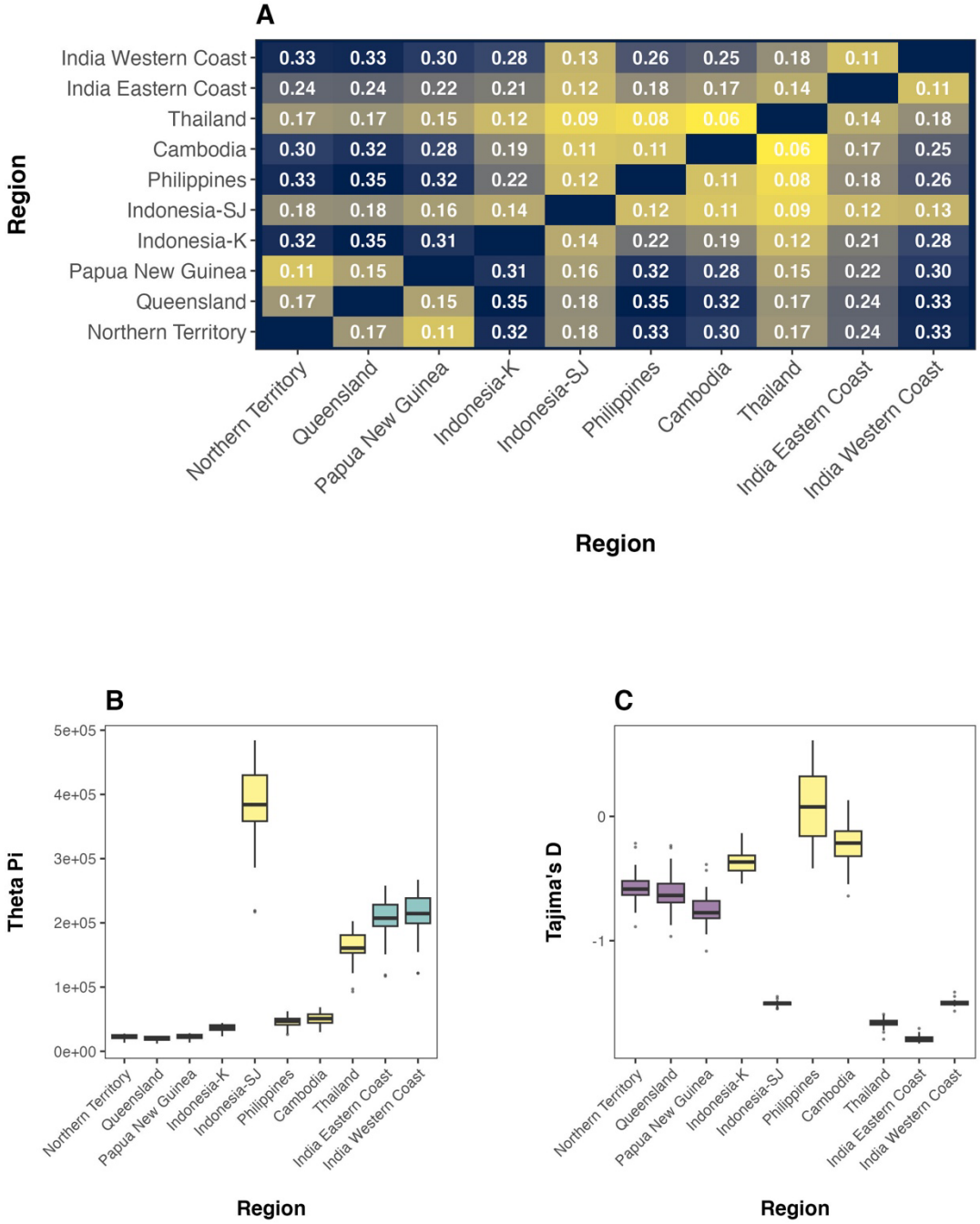


Figure 4 (A) Maximum-Likelihood phylogeny of genome-wide SNPs of 60 individual Indo-Pacific *Lates*. Nodes with < 75% bootstrap support are collapsed to polytomies. Nodes with ≥ 75% bootstrap support but < 90% bootstrap support are indicated as grey squares and nodes with ≥ 90% bootstrap support as black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

619



620

621

622 **Figure 5** Population genetic statistics of Indo-Pacific *Lates* from WGS data. **(A)** Pairwise F_{st}

623 values between regions. **(B)** Nucleotide diversity of sampling locations in the WGS data set. **(C)**

624 Tajima's D of sampling locations in the WGS data set. Indonesia is represented by two sampling

625 locations. Indonesia-SU, Vietnam and Bangladesh were excluded due to low sample sizes ($n \leq$

626 2).

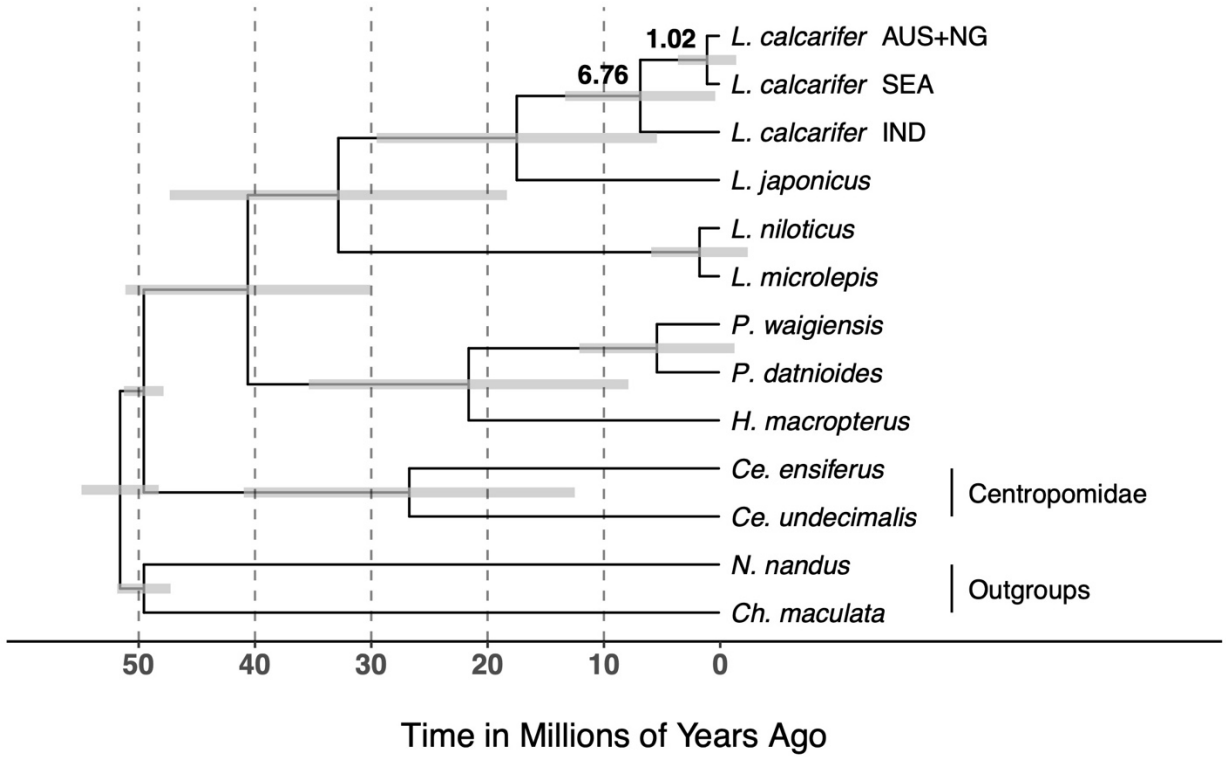
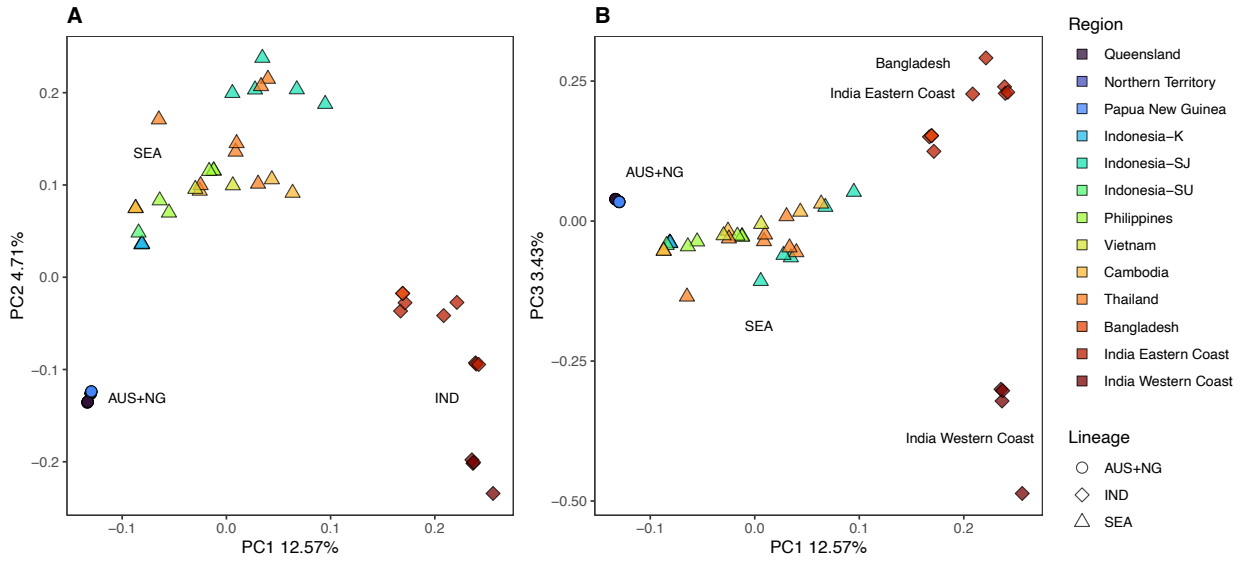
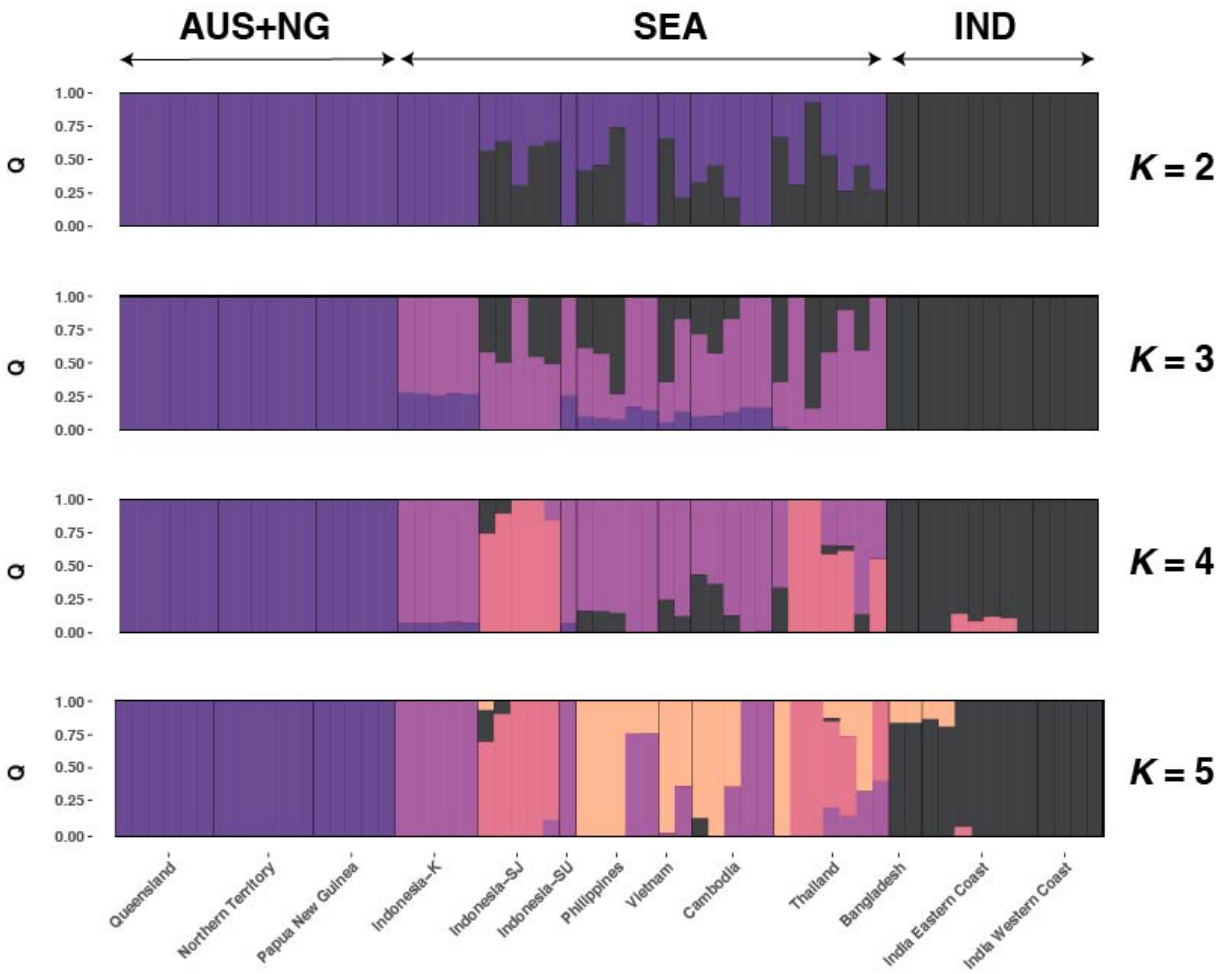


Figure 6 Time-calibrated phylogeny of main mitochondrial lineages of Indo-Pacific *Lates*. Node age is from the posterior medians with 95% Highest Posterior Density (HPD) indicated by horizontal bars.

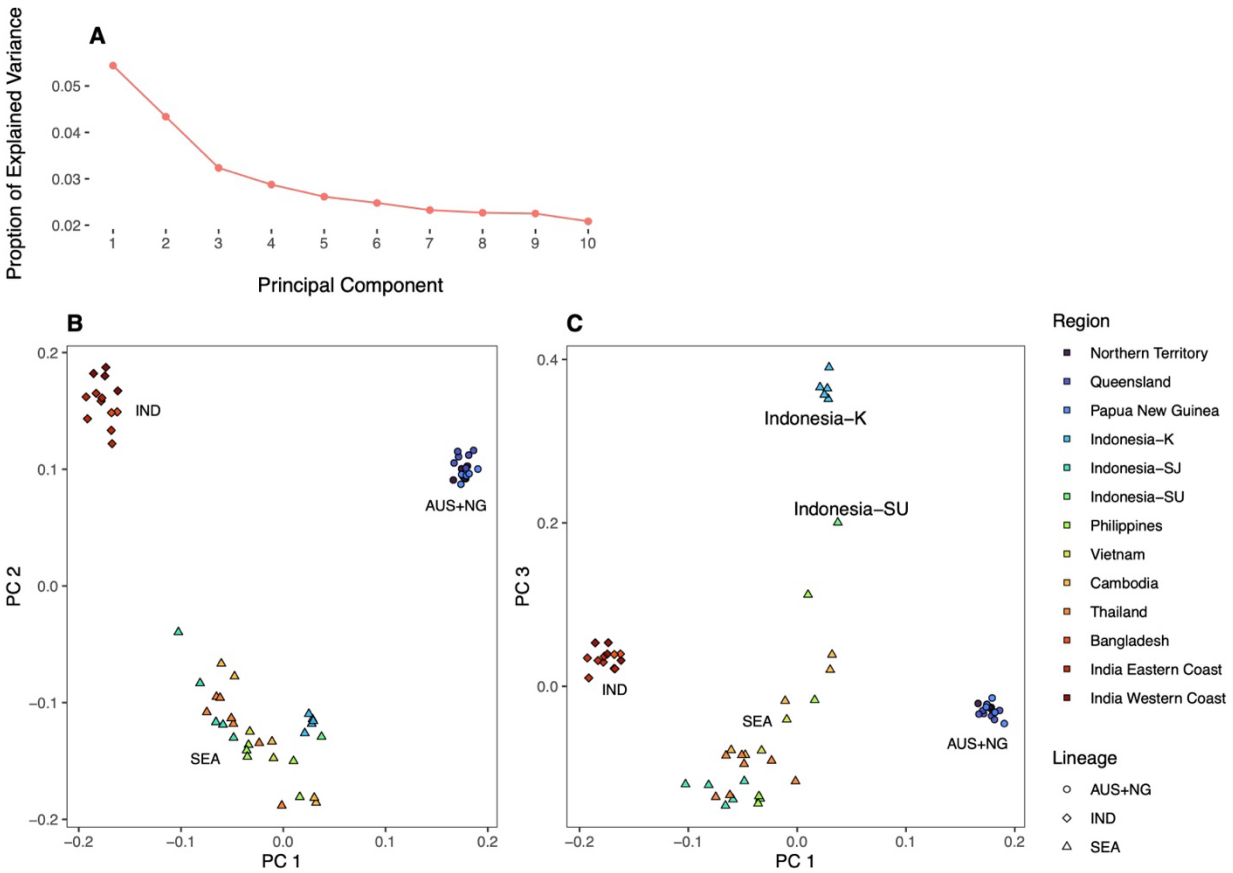


Supplemental Figure S1 Principal Component (PC) analysis of the whole-genome sequence data set of 60 Indo-Pacific *Lates* utilizing 7,321,852 genotype likelihoods. The first PC is on the *x* – axis in both figures with the second PC on *y* – axis in (A) and third PC in (B).



641

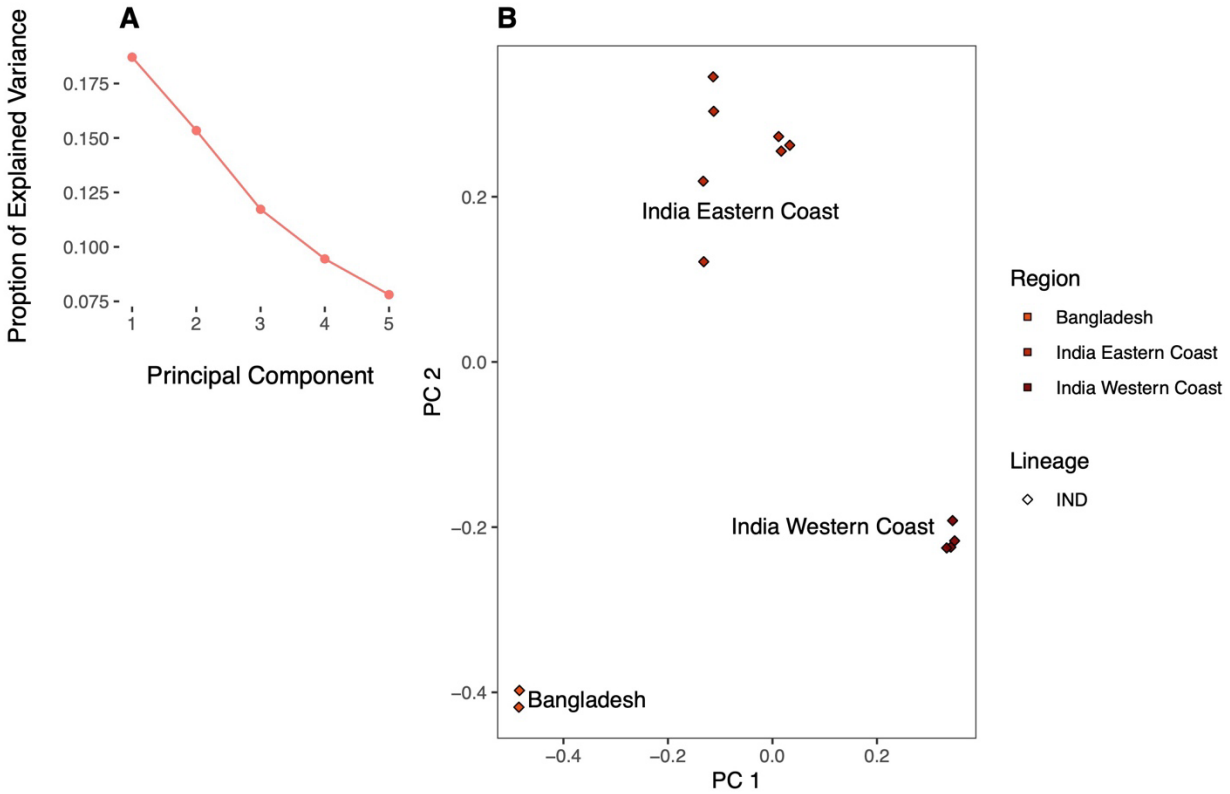
642 **Supplemental Figure S2** Genome-wide admixture analysis of Indo-Pacific *Lates* from WGS
643 data (n = 60) from 7,321,852 genotype likelihoods for K = 2 - 5 genetic clusters. Individuals are
644 plotted on the x – axis and ancestry coefficient (Q) on the y – axis.



Supplemental Figure S3 PCAdapt analyses for identification of adaptive genetic loci of all Indo-Pacific *Lates* analyzed in this study. **(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.

652

653

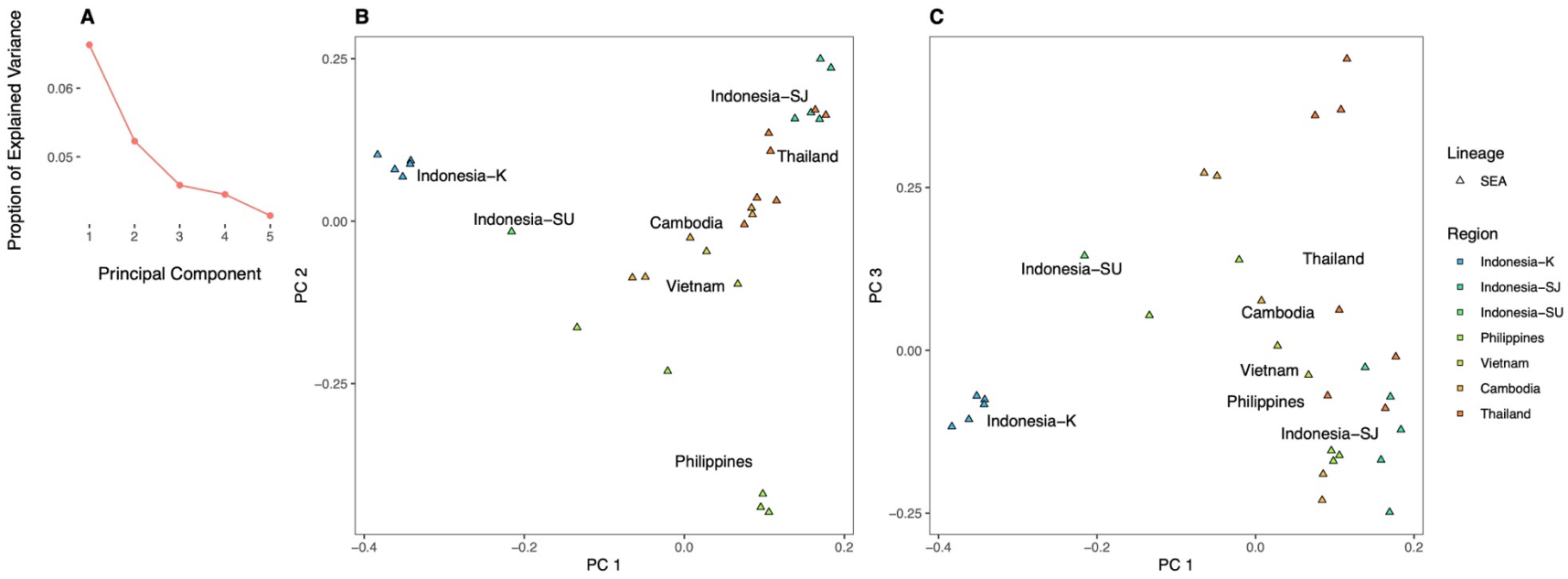


654

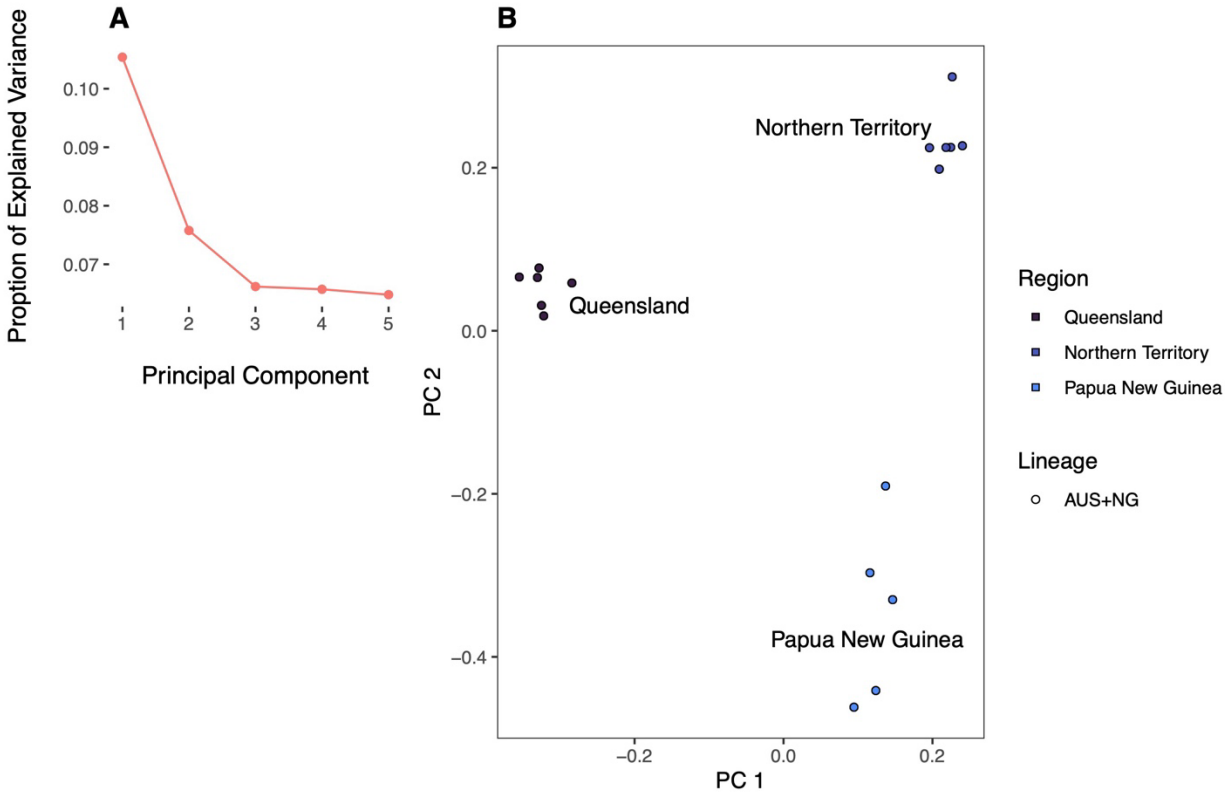
655 **Supplemental Figure S4** PCAdapt analyses for identification of adaptive genetic loci from the

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

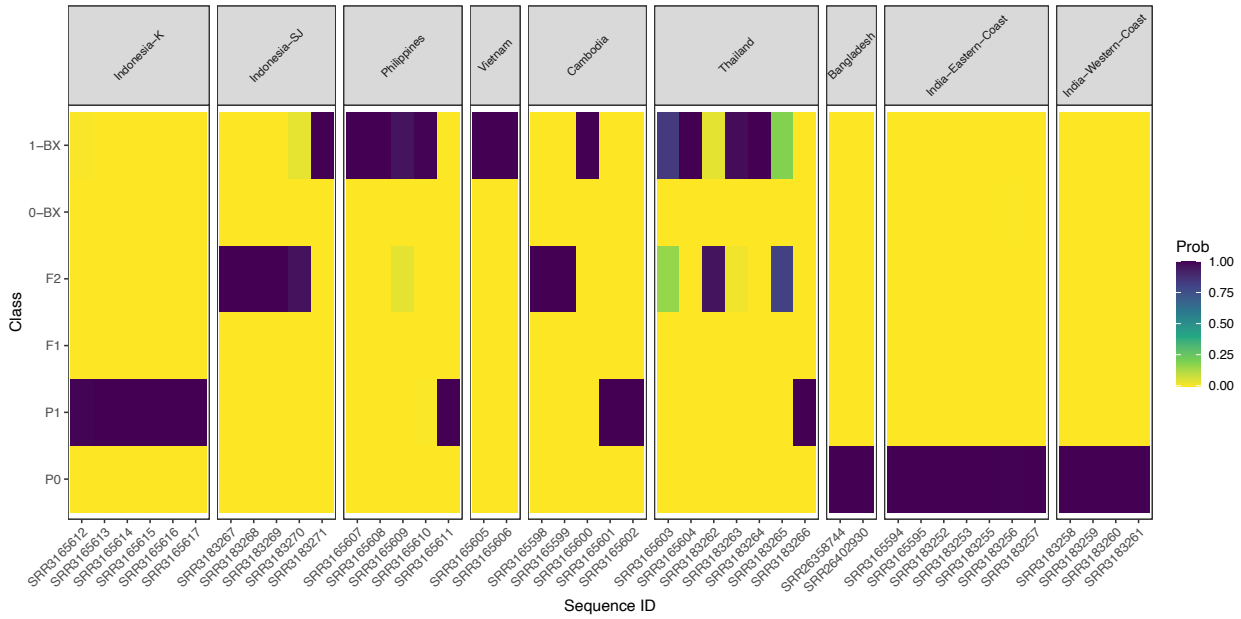
657 **(B)** Plot of PC 1 & PC 2.



Supplemental Figure S5 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 S6 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.



Supplemental Figure S7 NewHybrids output from analysis of IND and SEA fish. For display, Indonesia-SU was combined with Indonesia-K. Individual sequence numbers, Sequence ID, are indicated on the *x* – axis. Parental populations, (P0, P1), recent hybrids (F1, F2) and backcrosses (0-BX, 1-BX) are indicated on the *y* – axis. Probability of assignment is indicated by a color gradient (Prob). Two individuals from Cambodia with IND lineage mtDNA are indicated to be recent hybrids (SRR3165998, SRR3165999).

680 **Tables**

681 **Table 1** Sample origins by region of whole-genome sequence data set for Indo-Pacific *Lates*.

682 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

683

684

685 **Table 2** Taxa and sequences included in time-calibrated phylogeny of Indo-Pacific *Lates*
686 mitochondrial lineages. Lates *calcarifer* is represented by three lineages, Australia + New
687 Guinea (AUS+NG), Southeast Asia (SEA) and Indian subcontinent (IND) as described in the
688 text. The family is given, with the taxon, corresponding BOLD or NCBI accession and the
689 geographic origin of the sample if known.
690

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

691
692

References

Abed, J.M., Ali, A.H., Yaseen, A.T., Al-Faisal, A., Mutlak, F., Jassim, F.K., Jerry, D.R., Jawad, L.A., 2023. Barramundi (*Lates calcarifer*) from Iraq: a new record for the Arabian Gulf, with a highlight on it genetic origins and description of two skeletal deformities. New Zealand Journal of Zoology 1–15. https://doi.org/10.1080/03014223.2023.2255147

Agapow, P., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., Purvis, A., 2004. The Impact of Species Concept on Biodiversity Studies. The Quarterly Review of Biology 79, 161–179. https://doi.org/10.1086/383542

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., Hale, M.C., 2024. Genomic structural variation in Barramundi Perch *Lates calcarifer* and potential roles in speciation and adaptation. G3 Genes|Genomes|Genetics jkae141. https://doi.org/10.1093/g3journal/jkae141

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, T.Y., Nazir, A., Lim, H.C., Shiao, J.C., 2024. Genetic diversity and population structure of cultured and wild Barramundi *Lates calcarifer* (Latidae) in Taiwan. J. Ichthyol. https://doi.org/10.1134/S0032945224700450

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

Gruber, B., Unmack, P.J., Berry, O.F., Georges, A., 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources 18, 691–699. https://doi.org/10.1111/1755-0998.12745

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., 2018. 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

Jerry, D.R., Smith-Keune, C., 2014. The Genetics of Asian Seabass, in: The Biology and Culture of Asian Seabass. CRC Press, Boca Raton, Fla, pp. 137–177.

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

Mathew, G., 2009. Taxonomy, identification and biology of Seabass (*Lates calcarifer*), in: Cage Culture of Seabass. Central Marine Fisheries Research India, pp. 38–43.

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.

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