Molecular Phylogenetics and Evolution 205 (2025) 108296

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Historical biogeography of the *Mugil cephalus* species complex and its rapid global colonization

Philipp Thieme [a,b,*], Celine Reisser [a], Corinne Bouvier [a], Fabien Rieuvilleneuve [a], Philippe Béarez [c], Richard R. Coleman [d], Jean Jubrice Anissa Volanandiana [e], Esmeralda Pereira [f], Mauro Nirchio–Tursellino [g], María Inés Roldán [h], Sandra Heras [h], Nathalia Tirado-Sánchez [i], Eric Pulis [j], Fabien Leprieur [a], Jean-Dominique Durand [a]

[a] *MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, cc093, Place E. Bataillon, 34095 Montpellier Cedex 05, France*
[b] *Deutsches Meeresmuseum, Katharinenberg 14-20, 18439 Stralsund, Germany*
[c] *UMR 7209 AASPE, CNRS-MNHN, 43 rue Buffon, 75005 Paris, France*
[d] *Department of Marine Biology & Ecology, Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA*
[e] *Institut Halieutique et des Sciences Marines (IH SM), University of Toliara, BP 141 - Route du Port, Av. De France, Tulear 601, Madagascar*
[f] *MARE–Centro de Ciências do Mar e do Ambiente/ARNET-Rede de Investigação Aquática, Universidade de Évora, Largo Dos Colegiais N.2, 7004-516 Évora, Portugal*
[g] *Universidad Técnica de Machala, Facultad de Ciencias Agropecuarias, Escuela de Medicina Veterinaria. Machala, El Oro, Ecuador*
[h] *Laboratori d'Ictiologia Genètica, Campus Montilivi, Universitat de Girona, 17003 Girona, Spain*
[i] *Charles Darwin Foundation, Av. Charles Darwin s/n, Puerto Ayora, Galápagos, Ecuador*
[j] *Northern State University, 1200 S Jay Street, Aberdeen, SD 57401, USA*

ARTICLE INFO

Keywords:
Speciation
Plio-Pleistocene transition
trans-Atlantic species
Cryptic species
Shotgun sequencing
SNAPP
DECX

ABSTRACT

Our understanding of speciation processes in marine environments remains very limited and the role of different reproductive barriers are still debated. While physical barriers were considered important drivers causing reproductive isolation, recent studies highlight the importance of climatic and hydrological changes creating unsuitable habitat conditions as factors promoting population isolation. Although speciation in marine fishes has been investigated from different perspectives, these studies often have a limited geographical extant. Therefore, data on speciation within widely distributed species are largely lacking. Species complexes offer valuable opportunities to study the initial stages of speciation. Herein we study speciation within the *Mugil cephalus* species complex (MCSC) which presents a unique opportunity due to its circumglobal distribution.

We used a whole-genome shotgun analysis approach to identify SNPs among the 16 species within the MCSC. We inferred the phylogenetic relationships within the species complex followed by a time-calibration analysis. Subsequently, we estimated the ancestral ranges within the species complex to explore their biogeographical history.

Herein, we present a fully resolved and well-supported phylogeny of the MCSC. Its origin is dated at around 3.79 Ma after which two main clades emerged: one comprising all West Atlantic and East Pacific species and the other all East Atlantic and Indo-Pacific species. Rapid dispersal following an initial founder colonization from the West to the East Atlantic led to the population of all major realms worldwide in less than 2 Myr. Physical and climatic barriers heavily impacted the ancestral distribution ranges within the MCSC and triggered the onset of speciation.

1. Introduction

Speciation is an intriguing evolutionary process as it fundamentally shapes biodiversity in all environments (Coyne and Orr, 2004). Understanding the mechanisms that lead to the emergence of new species has been extensively studied both theoretically and experimentally (Turelli et al., 2001; White et al., 2020). However, most of our knowledge comes from terrestrial systems and insights into speciation

* Corresponding author at: Deutsches Meeresmuseum, Katharinenberg 14-20, 18439 Stralsund, Germany.
E-mail address: phil.thieme2016@gmail.com (P. Thieme).

https://doi.org/10.1016/j.ympev.2025.108296
Received 6 November 2024; Received in revised form 12 January 2025; Accepted 26 January 2025
Available online 28 January 2025

processes in aquatic environments remain limited (Bernardi, 2013; Potkamp and Fransen, 2019; Faria et al., 2021). Even within aquatic systems, progress has been unevenly distributed, with a greater focus on freshwater environments. These systems generally provide more restricted conditions to study spatial, ecological, and sexual components of speciation in unison (Bernardi, 2013). Marine ecosystems on the other hand have fewer physical barriers that limit gene flow and marine species are often associated with large distribution ranges and high dispersal potentials (Norris and Hull, 2012; Bernardi, 2013; Faria et al., 2021).

The geographical aspect of speciation, particularly allopatric speciation, has long been considered the main driver of reproductive isolation (Palumbi, 1992; Potkamp and Fransen, 2019; Faria et al., 2021). The rise of a physical boundary can strongly limit gene flow between populations leading to reproductive isolation and speciation in marine environments (e.g., Bernardi et al., 2003; Gaither and Rocha, 2013). Well-known examples include the disappearance of the Central American seaway due to the formation of the Isthmus of Panama and the rise of the Sundaland with Pleistocene sea level regressions in the Indo-Malaysian region (Chappell and Thom, 1977; Voris, 2000). Both events led to the establishment of geminate species in their respective regions (Briggs, 1995; Bernardi et al., 2003; Hubert et al., 2012; Bernardi, 2013; Cowman and Bellwood, 2013; O'Dea et al., 2016; Crandall et al., 2019). More recent, sea level fluctuations during the late Pleistocene glacial periods likely played a significant role in speciation of coastal species. For instance, sea level regressions have been shown to fragment habitats leading to population bottlenecks that disrupted gene flow and promoted reproductive isolation as observed in closely related but genetical distinct species with overlapping geographical distributions (Shen et al., 2011; Ludt and Rocha, 2015). Additionally, evidence suggests that sea level regressions exposed previously submerged sea mounts, which may have served as temporary dispersal pathways or "stepping stones", facilitating range expansions before subsequent isolation drove speciation (Macieira et al., 2015).

However, physical barriers are relatively rare in marine environments (Steeves et al., 2005). In contrast, soft barriers such as climatic (Gee, 2004) or hydrological (Cowman and Bellwood, 2013) barriers, which create ecological discontinuity due to unsuitable habitat conditions (Pyron and Burbrink, 2010), appear to be much more common and significant (Fransen, 2002; Hickerson and Meyer, 2008; Luiz et al., 2012; Cowman and Bellwood, 2013). Water currents, large open ocean areas, and other abiotic factors (e.g., temperature or salinity) can impact the distribution of populations which may lead to disruption and isolation (Hickerson and Meyer, 2008; Pyron and Burbrink, 2010). Local adaptations linked to temperature have been found in several fish species (Teske et al., 2019; Cayuela et al., 2021; Fuentes-Pardo et al., 2023; Fuentes-Pardo et al., 2024), reinforcing the notion that the distribution of marine species largely conforms to their thermal limits (Sunday et al., 2012). However, soft barriers in general remain permeable (Drew and Barber, 2012) and different life history traits related to species dispersal abilities (e.g., pelagic larvae stages) or reproductive strategies (benthic vs pelagic spawning) may prevent the complete isolation of populations (Riginos et al., 2014). Therefore, the impact of both soft and hard barriers should be interpreted with caution (Cowman and Bellwood, 2013).

Consequently, it has been questioned whether allopatric speciation should be considered the sole driver of speciation in marine ecosystems (Rocha and Bowen, 2008; Puebla, 2009; Norris and Hull, 2012). In addition to the geographical aspect of speciation, it has been proposed that the ecological aspect, i.e. reproductive isolation due to divergent (in allopatry) or disruptive (in sympatry) natural selection, also contributes to the speciation process in marine species (Puebla, 2009). While different components such as feeding competition or character displacements were evaluated, examples and data especially for disruptive selection in marine environments remain scarce (Puebla, 2009; Faria et al., 2021).

Oceans may not be considered homogenous environments and species with large distribution ranges, even those with extensive dispersal abilities, are subject to selection processes which can lead to gene flow limitations (Faria et al., 2021). Gaither et al. (2016) summarized that of the 107 circumtropical species originally described by Briggs (1960) molecular data suggest that 35 represent species complexes with restricted distribution areas. Species complexes offer a valuable opportunity to study the initial stages of species formation (Pinheiro et al., 2018), particularly given that they often include cryptic species, where morphological variation is minimal (Bickford et al., 2007). To date, studies addressing speciation in species complexes within marine environments have tended to focus on restricted geographic areas and/or primarily examined the spatial scale of speciation modes (Leray et al., 2010; Amor et al., 2014; Cornils and Held, 2014; Kolbasova et al., 2023). Within the list of circumtropical distributed species complexes (Gaither et al., 2016), the *Mugil cephalus* species complex (MCSC) is arguably the most conspicuous as it contains at least 16 cryptic species (Durand et al., 2012a; Whitfield et al., 2012; Durand and Borsa, 2015; Durand et al., 2017; Hasan et al., 2021).

The family Mugilidae is a ubiquitous fish taxon present in all major coastal areas from tropical to temperate waters around the world (Crosetti and Blaber, 2015). For a long time, *Mugil cephalus* Linnaeus 1758 was considered the most common and widespread mugilid species occurring in all oceans from about 42°N to 42°S (Whitfield et al., 2012; Crosetti and Blaber, 2015). Thomson (1997) even synonymized 48 nominal species from around the world with *Mugil cephalus* due to their highly conserved and uniform eidonomy and anatomy. However, early genetic analyses already suggested the presence of genetically distinct species (Crosetti et al., 1993; Rossi et al., 1998), which were later supported by more comprehensive genetic studies (Durand et al., 2012a; Whitfield et al., 2012; Durand and Borsa, 2015; Durand et al., 2017). At present, 16 species are recognized within the MCSC (Durand et al., 2017; Hasan et al., 2022), which are distributed along the coastlines of all major continents (Fig. 1), including *Mugil liza*, the only species within the MCSC which is distinguishable from the other species by morphological characters (Thomson, 1997). While most species exhibit distinct distribution ranges, there are areas of contact between some species, and even partial sympatry in others (Shen et al., 2015). Interestingly, species of the MCSC are absent or are not reported from many tropical areas (Fig. 1) such as eastern Africa, southwest Central America (Gilbert, 1993), or the southern Indo-Australian Archipelago (Delrieu-Trottin et al., 2020).

Although 16 species can currently be distinguished by molecular methods (Durand et al., 2017; Hasan et al., 2021), these species share many life history traits raising questions about their evolutionary origin, possible speciation events and speciation barriers. Species within the MCSC in general show an extraordinary adaptability to a variety of biotic and abiotic conditions (Crosetti and Blaber, 2015). These species exploit a unique trophic niche as primarily detritivorous feeders mainly consuming microphytobenthos, but also feeding on phyto- and zooplankton (Harrison and Howes, 1991; Carlier et al., 2007; Cardona, 2015; Crosetti and Blaber, 2015). Within coastal and estuarine environments these species inhabit various habitats as they can tolerate a variety of salinities, including freshwater and hypersaline ecosystems, as well as varying turbidity regimes, sediment types, dissolved oxygen levels, and temperatures (Whitfield et al., 2012; Nordlie, 2015; Whitfield, 2015). Another common life history trait is their reproduction cycle. During spawning season, adult specimens migrate to their offshore marine spawning areas (Whitfield et al., 2012). Reproduction in open ocean environments appears to be necessary, because stable salinity levels are a requirement for successful spawning and egg development (Van der Horst, 1981; Walsh et al., 1989; Walsh et al., 1991; Nordlie, 2015; Whitfield, 2015). Following spawning events, adults return to their coastal habitats, while fertilized eggs and preflexion larvae remain passively drifting in the offshore marine environment. They only actively migrate towards the surf zone in a postflexion stage before recruiting to nursery areas in coastal or estuarine

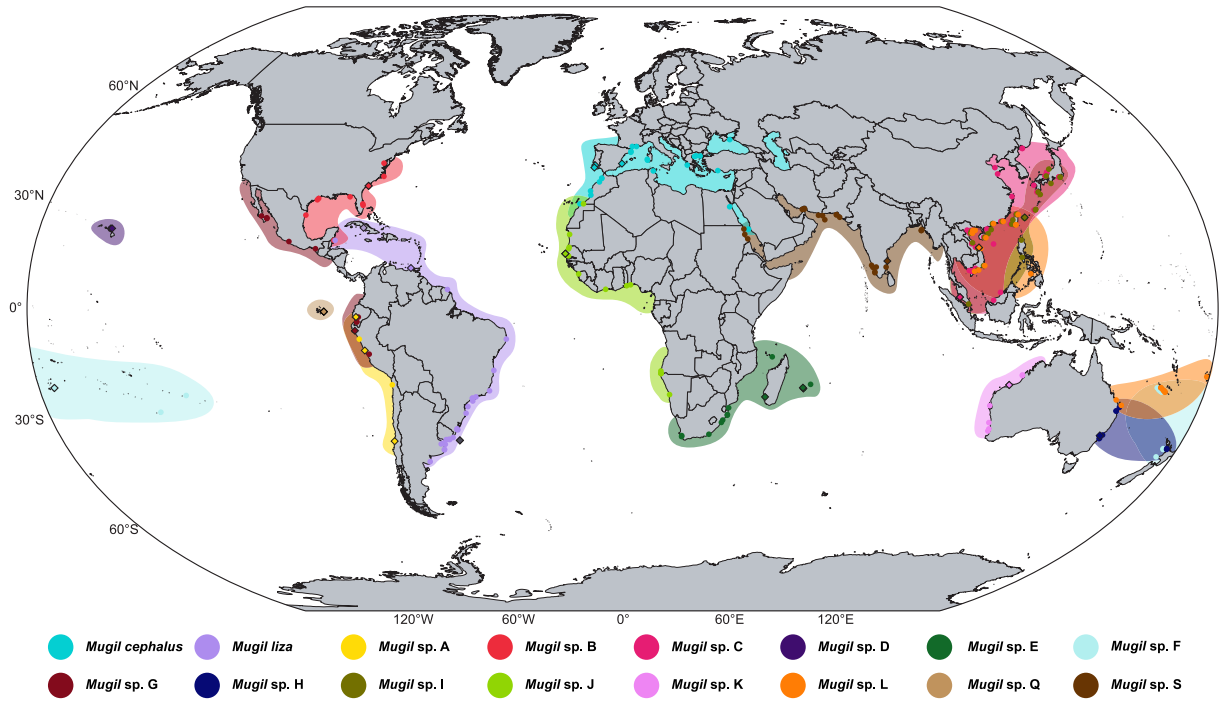

Fig. 1. Geographic distribution of the 16 different species currently recognized within the *Mugil cephalus* species complex based on barcoding data retrieved Genbank and BOLD (Supplementary Table 1). Additionally, origin of samples used in this study are depicted as diamonds; points marked as *Mugil* cf. *liza* correspond to samples 156_S19 and 157_S23.

habitats (Whitfield, 2015).

Adjustments in the spawning strategies of the single MCSC species seem to be an important factor preventing hybridization between them. Shen et al. (2015) revealed that segregation in spawning grounds among otherwise sympatric species in the Northwest Pacific constitutes a major prezygotic barrier. Furthermore, their work showed the presence of a partial reproduction asynchrony between the studied species. Both factors were likely influenced by the isolation of populations during Pleistocene glacial events, which led to speciation on one hand, but also exposed them to different temperature and current regimes on the other (Shen et al., 2011). Available data from different species across multiple ecoregions suggest that reproduction asynchrony occurs between species and to a lesser extent also within species (see Table 2 in Whitfield et al., 2012; Whitfield, 2015). A strong indication for the timing of spawning migrations seems to be temperature related: spawning occurs more frequently in summer in temperate regions, whereas in subtropical regions spawning peaks during winter (Whitfield et al., 2012). Whitfield (2015) suggested an optimal temperature range of 20–26 °C in coastal area sea surface temperatures (SST) for breeding activity, although spawning migrations also occur outside this range in areas with constantly high SSTs. In addition, other abiotic factors such as photoperiod and changes in meteorological tides have been suggested to influence the gonad development and onset of spawning migrations (Kuo et al., 1974; Whitfield et al., 2012; Castellanos-Juárez et al., 2024).

Overall, the MCSC constitutes an ideal case study due to its worldwide distribution and conserved morphology that can offer valuable insights into key aspects of the speciation process. In this study, we analyse the evolutionary history of the MCSC using a genome-wide shotgun analysis approach to infer for the first time a time-calibrated phylogeny. Additionally, we applied an ancestral range reconstruction analysis based on the resulting phylogenetic framework to gain new insights into the origin of the species complex. Lastly, we employed Genome-Environment Association (GEA) analyses to identify the presence of genetic variation linked to variations in sea surface temperature, which we hypothesise to be an important abiotic factor driving the speciation process in the MCSC.

2. Material & methods

2.1. Taxon sampling & DNA extraction and sequencing

Specimens used in this study were collected from fish markets across various global localities between 2010 and 2022 (Supplementary Table 2). Fin clips were taken, preserved in 100 % ethanol for genetic analyses and stored at −20 °C until DNA extraction. All specimens were morphologically identified as *Mugil liza* or *M. cephalus* and species affiliation was characterized using nucleotide variation present in the Cytochrome Oxidase I (COI) marker as detailed in Durand et al. (2017).

A total of 38 samples (Fig. 1) were available for this analysis representing 15 species of the MCSC (no sample of *Mugil* sp. S was available) and one sample of *M. bananensis* (Pellegrin 1927) as outgroup (Supplementary Table 2). Total genomic DNA was extracted from tissue samples using the Maxwell® 16 Bench-top extraction system (Promega Corporation, Madison, USA), following manufacturer's instructions. Small amounts of tissue preserved in ethanol were incubated 2 h at 56 °C in 300 μL of lysis buffer and 30 μL of proteinase K from Maxwell® Buccal Swab LEV DNA kits (Promega Corporation, Madison, USA). The totality of the solution was then placed in the Maxwell® cartridge of the extraction kit and extracted DNA was purified by magnetic beads technology before being eluted in 100 μL of elution buffer (Promega Corporation, Madison, USA), and quantified by spectrophotometry NanoDrop™One (Thermo Scientific, Waltham, MA, USA). Genomic DNA was sent to Montpellier GenomiX Facilities (MGX, Montpellier, France) where libraries were prepared and sequenced on an S4 chip on Illumina NovaSeq6000. Afterwards, raw Illumina sequence reads of *Mugil* sp. S (Shekhar et al., 2022; https://www.ncbi.nlm.nih.gov/ sample: SRR15214610) were added to the dataset resulting in a total of 39 samples.

2.2. Read filtering, and alignment to mitochondrial and nuclear genome

The quality of raw sequences was checked using FastQC and reads were trimmed with Fastp V0.23.2 (Chen, 2023), removing the poly-g

tails, reads with PHRED quality score lower than 28, reads containing N bases, and reads smaller than 50 bp. Filtered reads were aligned onto the *M. cephalus* reference mitochondrial and the whole genome (ref NCBI, Shekhar et al., 2022) using BWA (Li and Durbin, 2009) with default parameters, and BAM files were filtered with Sambamba V0.8.0 (Tarasov et al., 2015) to only keep properly paired reads and reads with a mapping quality MAPQ greater than 10. PCR duplicates were identified and removed using Picard Toolkit (2019). Variants were called using Freebayes (Garrison and Marth, 2012) with default parameters. For the mitochondrial dataset, diploid genotypes were corrected and haplotyped for the allele with the highest depth. We then extracted the FASTA sequences of all individuals from the mitochondrial VCF dataset with bcftools v1.13 (Danecek et al., 2021) and aligned them with other *M. cephalus* mitochondrial genomes (Supplementary Table 3) to validate their lineage. Two of the analysed samples (72_S2 and 73_S7 representing *Mugil* sp. I) had an outlier position in the mitochondrial tree and were discarded from the rest of the analyses (Supplementary Fig. 1). All datasets analysed herein therefore only contained a maximum of 37 samples representing 15 of 16 species within the MCSC and one outgroup species.

Before further filtering of the whole genome data, we created two datasets, one for phylogenetic inference (PI) and the other for the GEA analysis. For the GEA dataset, the outgroup sample was removed first. Only SNP variants were retained using the "TYPE" option of the "vcffilter" function implemented in *vcflib* v1.0.0-rc1 (Garrison et al., 2022). We then used the "genotype-filter" option of the "vcffilter" function to remove loci that did not fit the specified read depth range (PI: min. = 5, max. = 500; GEA: min. 4, max. 500). Variants were trimmed to be left-aligned and parsimonious and successive SNPs that were clumped during variant calling were split with the "normalize" and "decompose_blocksub"-options implemented in the software package *vt* (Tan et al., 2015). Loci with a minor allele frequency below 0.01 and sites with either a proportional rate (PI: 10 %) or an absolute number (GEA: 4) of missing genotypes were excluded using vcftools V0.1.16 (Danecek et al., 2011). Only biallelic sites were retained using bcftools v1.13 (Danecek et al., 2021) with the −M2 and −m2 options of the "view" function. Afterwards, a custom R-script was used to filter invariant sites as recognized by RAxML and IQtree from the PI dataset.

2.3. Phylogenetic inference and species delimitation

After trimming and filtering, two PI datasets (PI1: with invariant sites, PI2: without invariant sites) were converted into concatenated sequence alignments in FASTA-format and Phylip-format using the python script vcf2phylip.py (Ortiz, 2019). We first conducted the phylogenetic inference based on the concatenated sequence alignment of PI1 using the maximum likelihood algorithm implemented in RAxML v8.2.12 (Stamatakis, 2014) with the nucleotide evolution model GTRGAMMA. For branch support evaluation, 1,000 rapid bootstrap searches were conducted. Phylogenetic inference was then repeated with PI2, adjusting the nucleotide evolution model to ASC_GTRGAMMA to account for ascertainment bias using the ascertainment bias correction by Paul O. Lewis. Additionally, PI2 was used to perform phylogenetic inference in IQtree 2.2.0 (Minh et al., 2020) using the best-fit model (Kalyaanamoorthy et al., 2017) according to the Bayesian Information Criterion (UNREST + FO + ASC + R2) and with 1,000 optimized ultrafast bootstrap replicates (Hoang et al., 2018). The best-scoring trees were visualized using the software FigTree v1.4.4, respectively (available at https://tree.bio.ed.ac.uk/software/figtree/).

Afterwards, species delimitation hypotheses were tested using the Bayes factor delimitation with SNP data as described by Leaché et al. (2014). We combined the functions of the packages SNAPP v1.5.2 (Bryant et al., 2012) and MODEL_SELECTION v.1.5.3 both implemented in BEAST 2.6.7 (Bouckaert et al., 2019) to evaluate a total of eight species delimitation models in path sampling analyses. Species delimitation models were set up either based on previous phylogenetic hypotheses (e.g., Mugil liza split into two species [runB]: Cousseau et al., 2005) or based on results from phylogenetic inference presented in this study. To accommodate computational constraints due to the high number of species, we created a subset of SNPs sampled from PI1 using the Ruby script snapp_prep.rb (https://github.com/mmatschiner/snapp_prep). A total of 10,000 randomly sampled SNPs was retained with a minimum distance of 5,000 bp between loci. Afterwards, XML files were modified to match the species delimitation hypotheses and adjusted for species delimitation analysis according to Leaché et al. (2014). For each model, a path sampling analysis with 48 steps was performed of which each step was run for 200,000 generations, sampled every 100 generations, with 1,000 pre-burnin generations, and discarding the first 10 % of entries as burn-in. We then compared marginal estimate likelihoods (MLE) and calculated (log) Bayes factors (BF) by subtraction of the (log) MLE values for two models and multiplying the difference by two following the framework of Kass and Raftery (1995) to evaluate competing models.

2.4. Time calibration

We estimated divergence times based on the PI1 dataset using the package SNAPP v1.6.1 (Bryant et al., 2012) implemented in BEAST v2.7.4 (Bouckaert et al., 2019) which can directly infer species trees from SNP data as well as estimate precise divergence times in combination with a molecular clock model (Stange et al., 2018). For preparing the input data and analysis settings we utilized the Ruby script snapp_prep.rb (https://github.com/mmatschiner/snapp_prep). Due to disproportionately long calculation times and lack of convergence with larger-sized datasets, we used the options provided by the script to randomly sample SNPs for the divergence time estimation. We sampled 10,000 SNPs with a minimum distance of 5,000 bp in between each selected locus. Two secondary calibration points were used, i.e., a lognormal distribution (in real space) with a mean of 26 Ma and a standard deviation of 0.22 for the last common ancestor of our outgroup and the MCSC and a lognormal distribution (in real space) with a mean of 4.5 Ma and a standard deviation of 0.25 for the last common ancestor of the MCSC, based on results from Neves et al. (2020). The phylogenetic tree from the RAxML analysis of PI1 was used as a starting tree. The "tree-nodeswapper"-operator was removed from the BEAST input file to fixate the tree topology to that of the provided starting tree to, one, enhance computational efficiency and, two, prevent changes in tree topology caused by subsampling.

We used BEAST v2.7.4 (Bouckaert et al., 2019) to run Markov Chain Monte Carlo (MCMC) analyses for 2,500,000 generations, sampling every 100 generations. This analysis resulted in 25,000 trees in the posterior distribution with the first 10 % discarded as burn-in. To assess the convergence of runs and check the effective sample size (ESS) we used Tracer v1.7.2 (Rambaut et al., 2018). TreeAnnotater v2.7.6 (Drummond and Rambaut, 2007) was then employed to build the maximum clade credibility tree containing the mean node heights.

2.5. Biogeographical reconstruction

We used a maximum likelihood approach to estimate the ancestral states of origin and geographic range evolution of the MCSC. The Dispersal Extinction Cladogenesis eXtended model as implemented in DECX (Beeravolu and Condamine, 2016; https://github.com/champost/DECX) was used to model transitions between predefined, discrete biogeographical areas along the branches of a provided, time-calibrated phylogeny. Different scenarios including multiple combinations of time-stratified adjacency assumptions and varying transition rates between these areas, were tested to reconstruct the biogeographical history of the MCSC. A set of fourteen discrete biogeographical areas was defined based on the realms established by Spalding et al. (2007; Fig. 2). Coastal regions not directly adjacent to each other but originally grouped in one realm were separated. For example, the Tropical Atlantic was divided

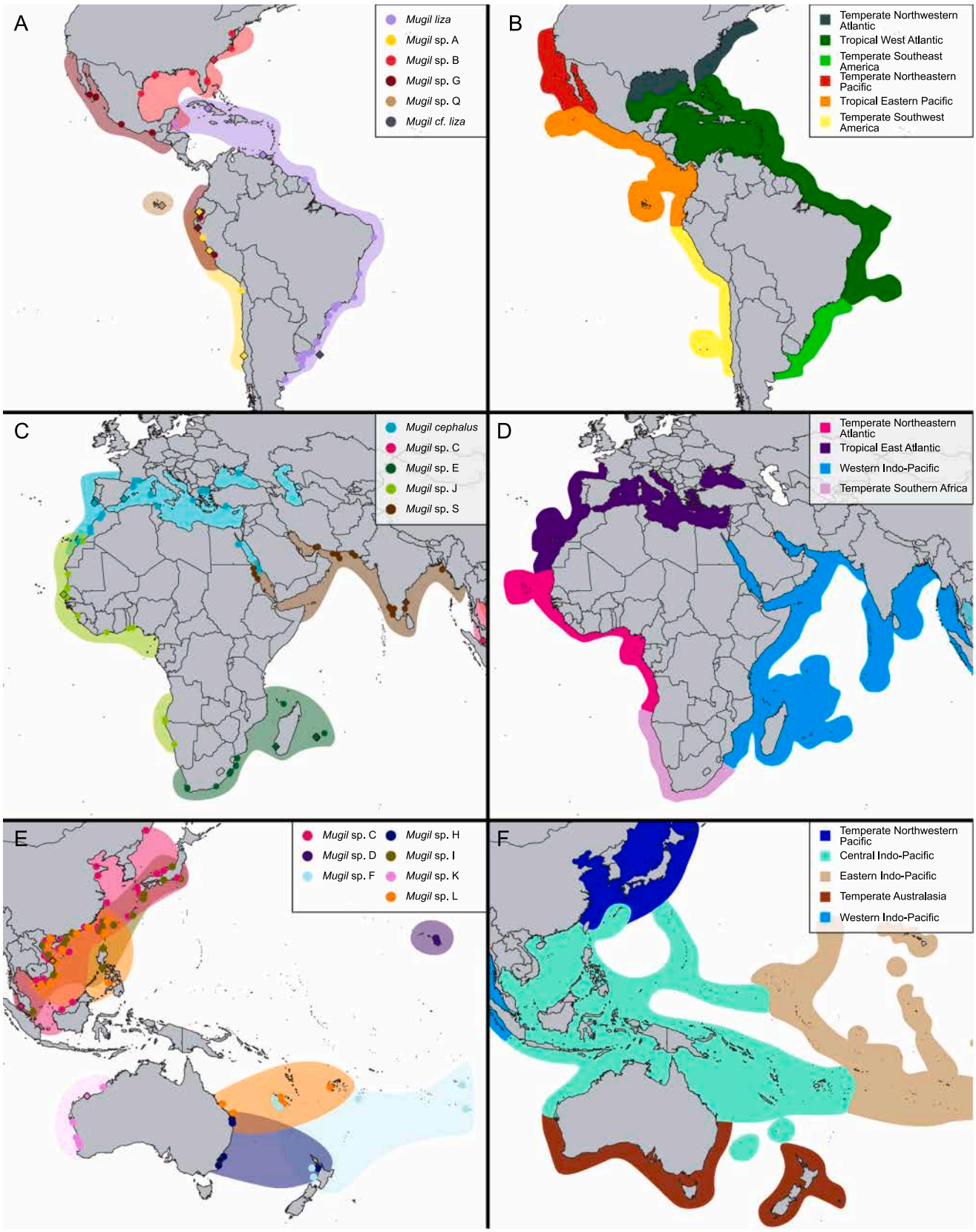

Fig. 2. Map sections of (A & B) the East Pacific and West Atlantic, (C & D) the East Atlantic and Western Indo-Pacific, and (E & F) the West and Central Pacific detailing (A, C, E) the distribution of the respective Mugil cephalus species and (B, D, F) the distinct biogeographical areas (based on coastal realms defined by Spalding et al., 2007) used for the modelling of the biogeographical history of the Mugil cephalus species complex.

into Tropical West and Tropical East Atlantic (Fig. 2B and D). Additionally, realms connected through areas not inhabited by any species of the species complex were split up. For instance, Temperate South America was split into Temperate Southwest and Temperate Southeast America (Fig. 2B). Further, we used the time-calibrated phylogenetic tree we obtained from the BEAST analysis. Models were calculated without and with time-stratified adjacency assumptions. For the latter, we created two biogeographic scenarios reflecting major geological events that changed the connectivity of our selected biogeographic areas during the evolutionary history of the MCSC. For the first time-stratified assumption (TS1) we defined two time slices, i.e. 0–3 million years ago (Ma) and 3–27 Ma, reflecting the closure of the Isthmus of Panama. TS2 assumes that the closure of the Isthmus of Panama was a gradual process with a phase of limited dispersal between East Pacific and West Atlantic, which we reflected in three time slices, i.e., 0–2.7 Ma, 2.7–3.3 Ma, 3.3–27 Ma. Additionally, we tested the application of rate matrices (none, equal rates, current dependent rates) that changed the transition rates between adjacent areas. In the equal rates transition matrix (ER), all transition rates between adjacent areas were uniformly set to 1, assuming equal probability of dispersal among connected regions. Conversely, in the current dependent transition matrix (CD), the transition rates between adjacent areas varied. These rates were adjusted if the areas were connected solely via ocean currents or if connectivity between areas changed across the defined time slices, reflecting the impact of historical changes in marine pathways and current systems on species dispersal. Finally, the maximum number of areas inhabited by the most recent common ancestors (MRCA) was varied across the different scenarios ranging from one (MXA1) to three (MXA3). We evaluated model performance through pairwise comparison of their final log-likelihood (LnL) values as the number of free parameters, i.e., dispersal (*d*) and extinction (*e*), did not vary between the models.

2.6. Genotype environment association

To identify genetic variations associated with sea surface temperature differences, we performed a partial redundancy analysis (RDA) on the genotypes of the individuals, while controlling for geographical location. We divided our GEA dataset into three new datasets corresponding to different geographical regions and phylogenetic groups: 1) all samples, 2) only West Atlantic and East Pacific samples, i.e., *Mugil* sp. A, B, G, Q and *Mugil liza*, 3) only Indo-Pacific and West Pacific samples, i. e., *Mugil* sp. C, D, E, F, H, K, L, and S, and performed the analysis within each group. We used Moran's Eigenvector Maps (MEM; Dray et al., 2006) as they can model spatial structure resulting from multiple processes in heterogeneous landscapes (Manel et al., 2010).

We used the cartesian latitudinal and longitudinal coordinates of each sample to compute a pairwise Euclidean distance matrix, which reflects the spatial arrangement of sampling sites (Supplementary Table 2). To represent geographic isolation of each site, distance-based Moran's eigenvector maps (db-MEMs; Dray et al., 2006), were calculated using the R package adespatial v0.3-23 (Dray et al., 2023). These db-MEMs were then incorporated as conditional variables in a RDA conducted with the R package vegan (Oksanen et al., 2022). Using the MEMs as conditional variables in a partial redundancy analysis can help identify and visualize spatial patterns in molecular genetic data and detect cryptic spatial genetic structure that may result from isolation by distance (IBD), isolation by resistance (IBR), isolation by environment (IBE), or a combination of the three (Hein et al., 2021). Four sea surface conditions, i.e., maximum temperature, temperature range, long term maximum temperature and long-term temperature range, were retrieved for each sample from the Bio-ORACLE dataset v2.2 (Assis et al., 2018; Supplementary Table 2). We used the pairs.panels function of the R package pych (Revelle, 2024) to test for co-variation of environmental variables and we kept variables showing a correlation less than R = 0.7 with the others.

The significance of the model, of each RDA axis, and of each environmental variable was tested with the anova.cca function using 1000 steps and variance inflation was verified using the vif.cca functions in vegan. Outlier SNPs were defined according to their loading coordinates on the significant axes and belonging to the extreme 1 ‰ of the global loadings' distribution. We finally used SNPEff with default parameters (Cingolani et al., 2012) to predict the impact of the outlier SNPs on potential surrounding genes of the *M. cephalus* genome.

3. Results

3.1. Sequencing, SNP Calling, and SNP filtering

A total of 1,004,650,984 raw reads (min = 7,137,438; max = 172,473,770) were generated by Illumina sequencing which averages 26,438,183 reads per sample. After adding an additional sample (Shekhar et al., 2022: SRR15214610 – Mugil sp. S) followed by adapter trimming and initial filtering an average of 622,535,712 reads remained for further processing. For the mitochondrial dataset, a total of 2,122 variant sites were identified with Freebayes. After filtering and haplotyping, we retained 1,931 variant sites. For the nuclear datasets, a total of 70,784,927 variant sites were identified using Freebayes, of which 62,744,404 were identified as SNPs. On average 6.2 % of data were missing per sample while the average read depth per sample was 10.9. After filtering and keeping only bi-allelic SNPs 871,972 variant sites were retained for the PI1 dataset with an average of 10.8 % missing data, 159,027 variant sites were retained in the PI2 dataset with an average of 13.2 % missing data, and in the RAD dataset 1,485,742 variant sites were retained with an average of 8.1 % missing data.

3.2. Phylogenetic relationships

The maximum likelihood tree inferred with RAxML of the concatenated sequence alignment of PI1 is shown in Fig. 2. The analysis produced a fully resolved and highly-supported phylogenetic tree with two nodes (85 %–98 %) and three nodes (60 %–70 %) between samples of the same species not fully supported (Fig. 3). The maximum likelihood tree inferred from the concatenated sequence alignment of PI2 with IQtree showed the same topology we retrieved from PI1 and while the tree is fully resolved and overall well-supported, the same nodes between samples of the same species as retrieved from PI1 are the ones without full support (Supplementary Fig. 3). The results from the RAxML analysis of PI2 showed consistent results with the other two analyses and only one node (between samples of *Mugil* sp. H) is not fully supported (Supplementary Fig. 4).

Generally, two clades within the MCSC can be distinguished (Fig. 3): Clade I comprises the Western Atlantic *Mugil liza* and *Mugil* sp. B and the Eastern Pacific species *Mugil* sp. A, G & Q. Clade II comprises the Eastern Atlantic *M. cephalus* and *Mugil* sp. J, the South African *Mugil* sp. E, the Northern Indic *Mugil* sp. S, the Western Australian *Mugil* sp. K, the Hawaiian *Mugil* sp. D, and the Western Pacific species *Mugil* sp. C, F, H & L. Samples of *Mugil liza* indicate the presence of two lineages, one in the Caribbean Sea (107_S4 and 110_S9) and another in southern Brazil (156_S19 and 157_S23). Within the second clade, the Eastern Atlantic *Mugil cephalus* and *Mugil* sp. J branch first from the other species. Two subgroups, the first comprising *Mugil* sp. E, L & S (Indo-Pacific subgroup) and the second *Mugil* sp. C, D, F, H & K (West Pacific subgroup) can be observed within the remaining taxa. The analysis of the mitochondrial genomes further suggests that *Mugil* sp. I, which was not included in the final PI datasets is closely linked to *Mugil* sp. C (Supplementary Fig. 2).

Species delimitation analyses based on a subset of the whole genome SNP data supports the division of *Mugil cephalus* into multiple species (Table 1). Lumping of species into one (sensu Thomson, 1997: runF) or two species (according to the two clades identified herein: runE) is clearly rejected by high Bayes Factor values (46526.52 & 31218.43, respectively). Splitting *Mugil liza* into two species (sensu Cousseau et al., 2005) was also rejected (run B: BF = 469.72), although combining

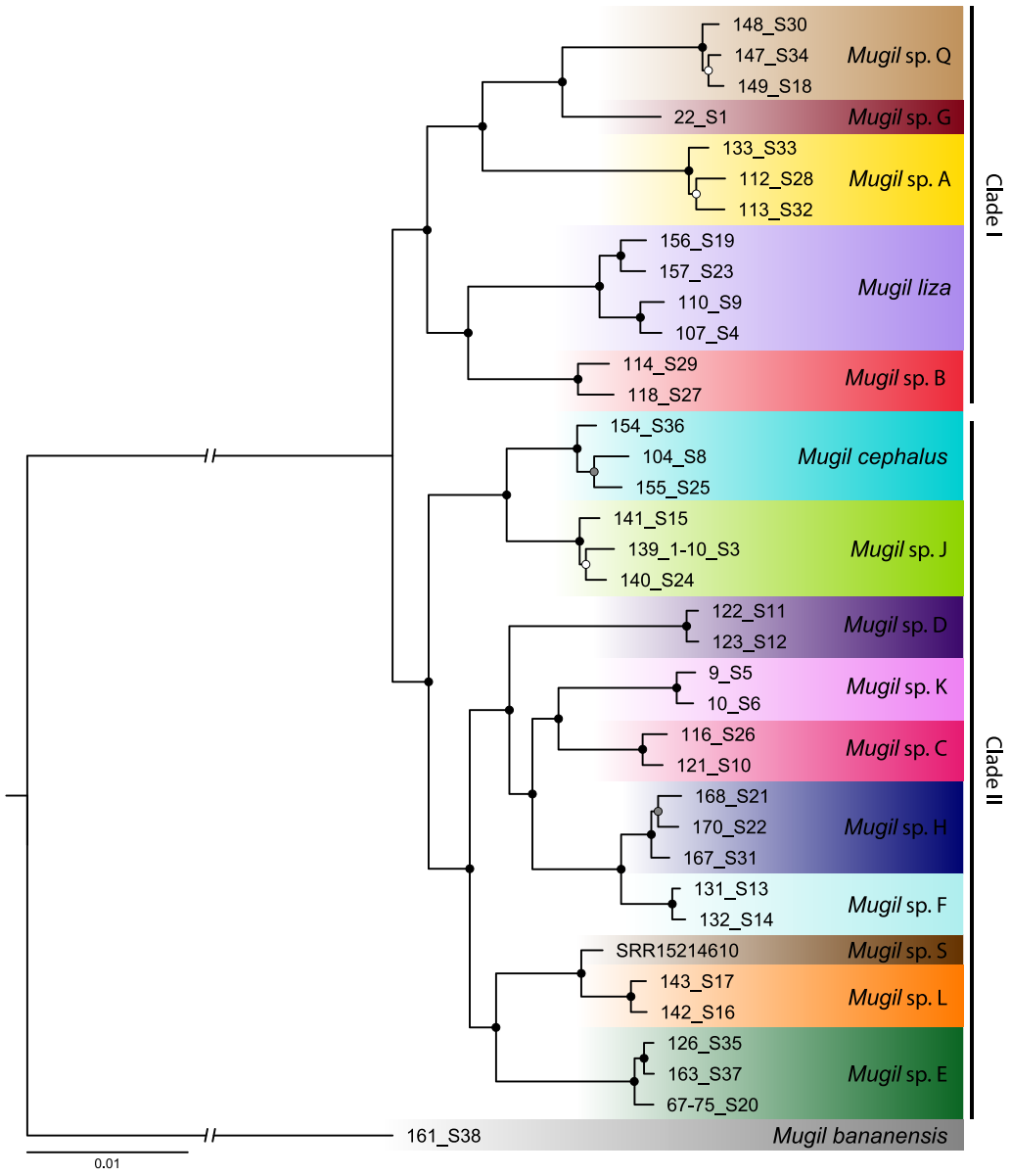

Fig. 3. Phylogenetic tree based on maximum likelihood analysis of the *Mugil cephalus* species complex inferred from whole-genome SNP data (PI1 dataset). Node support depicted by coloured circles representing posterior probabilities: black, 100 %; grey; 75–99 %; white, <75 %.

separated *Mugil liza* species with lumped *Mugil* sp. F and *Mugil* sp. H and lumped *Mugil* sp. L and *Mugil* sp. S based on a similar genetic distances between the latter species and within *Mugil liza* (runG: BF = 45.17) resulted in the second-best supported hypothesis. The analysis of the current taxonomic classification (sensu Durand et al., 2017 & Hasan et al., 2021) resulted in the highest (log) marginal likelihood estimate value and therefore the hypothesis best explained by whole genome SNP data.

3.3. Divergence time estimations

Divergence times were estimated running a SNAPP analysis of a random subsample of PI1 (Fig. 4). The analysis ran for 2,500,000 generations and reached stationarity for all evaluated parameters with all ESS values greater than 180. The origin of the MCSC is estimated to be at the end of the early Pliocene (3.79 Ma, 95 % highest posterior density (HPD): 2.58–4.99 Ma). Within Clade I, the West Atlantic *Mugil liza* and *Mugil* sp. B diverged from the East Pacific species at the beginning of the Piacenzian (3.61 Ma, HPD: 2.46–4.77 Ma). Separation of *M. liza* and *Mugil* sp. B (3.32 Ma, HPD: 2.27–4.41 Ma) and divergence of *Mugil* sp. A from *Mugil* sp. G & Q (3.21 Ma, HPD: 2.23–4.30 Ma) occurred shortly after. The two lineages retrieved within the *Mugil liza* samples show a divergence dated to the Chibanian (0.22 Ma, HPD: 0.01–0.36 Ma). Within Clade II, *M. cephalus* and *Mugil* sp. J diverged from the South African, Indian Ocean and West Pacific species in the late Pliocene (3.12 Ma, HPD: 2.13–4.13 Ma). Divergence between the Indo-Pacific subgroup and the West Pacific subgroup of Clade II occurred shortly after (2.96 Ma, HPD: 2.06–3.96 Ma). *Mugil* sp. E diverged from the northern Indic *Mugil* sp. S and the West Pacific *Mugil* sp. L by the end of the Pliocene (2.72 Ma, HPD: 1.91–3.68 Ma) while the latter two species were estimated to have diverged in the Late Pleistocene (0.11 Ma, HPD: 0.001–0.24 Ma). Divergence within the West Pacific species occurred within the Gelasian (between 2.4 Ma and 1.9 Ma) and only *Mugil* sp. F and *Mugil* sp. H separated later, presumably within the Chibanian (0.55 Ma, HPD: 0.36–0.78 Ma).

Table 1
Path sampling results for the eight species delimitation models ranked by (log) marginal likelihood estimate (MLE) values. 15 out of the 16 currently recognized species were evaluated (Mugil sp. I not included). (Log) Bayes factor (BF) calculations are made against the current taxonomy model (sensu Hasan et al., 2021) and used for model evaluation.

Model	Number of assigned species in dataset	MLE	Rank	BF
runA: Current taxonomy	15	−84814.11	1	–
runB: split *Mugil liza* in two clades	16	−85048.97	4	469.72
runC: lumped based on five geographical regions, i.e., East Pacific, West Atlantic, East Atlantic, Indo-Pacific, and West Pacific	5	−92664.35	6	15700.48
runD: lumped *M.* sp. F & H, *M.* sp. L & S, and *M. cephalus* & sp. J	12	−85202.50	5	776.78
runE: lumped Clade I and Clade II	2	−100423.32	7	31218.43
runF: lumped all samples as one species	1	−108077.37	8	46526.52
runG: lumped *M.* sp. H & F and *M.* sp. L & S	14	−84836.70	2	45.17
runH: split *Mugil liza* in two clades and lumped *M.* sp. H & F and *M.* sp. L & S	15	−84872.49	3	116.76

3.4. Biogeographic reconstructions

The biogeographic history of the MCSC was analysed under the DECX model. We tested different scenarios with varying parameters which included different time-stratification assumptions, adjusted transition rate matrices, and varying maximum numbers of areas inhabited by the most recent common ancestors. A full list of results, i.e., LnL-values utilized to assess model performance are compiled in Table 2.

The best-scoring model combined the following model parameters: TS1, ER, and M2. The estimated parameters suggest a higher dispersal rate with respect to the extinction rate ($d = 0.6483$ and $e = 0.084$). Results of the ancestral distribution range estimation under the best-scoring model are shown in Fig. 4. The origin of the MCSC is located in the Atlantic, more specifically in the tropical Atlantic. The model suggests that an ancient vicariance event separated Clade I and Clade II. Within Clade I, another vicariance event split the East Pacific species from the West Atlantic species, while the latter experienced a sympatric speciation event within the Tropical West Atlantic followed by a range extension by *Mugil liza*. Separation of *Mugil* sp. A and the MRCA of *Mugil* sp. G & Q cannot unambiguously be allocated to one mode of speciation. The biogeographical history of Clade II is much more complex. Initially, the MRCA of Clade II experienced a rapid range extension towards the Indo-Pacific while a MRCA of *M. cephalus* and *Mugil* sp. J split from the remaining species due to vicariance. The MRCA of *Mugil* sp. E, L, & S remained in a range composed of the Temperate Southern Africa and Western Indo-Pacific. Allopatric speciation led to the split of *Mugil* sp. E, followed by a range extension of the MRCA of *Mugil* sp. L & S and subsequent vicariance. Model estimations of the distribution ranges of the MRCAs of the remaining species were less clear. They indicate an initial range shift towards the West Pacific and an early separation of *Mugil* sp. D potentially as a result of a founder event. This was followed by an allopatric separation of the MRCA of *Mugil* sp. F & H and the MRCA of *Mugil* sp. C & K, which subsequently split due to a vicariance event. Further, this scenario suggests that the MRCA of the MCSC and the outgroup supposedly was located in the Tropical West and East Atlantic.

3.5. Genome environment association

Of the six environmental variables we considered, four showed high correlations and were discarded to only conserve the “long-term temperature range” and “long-term maximum temperature” variables (Supplementary Fig. 5). We identified five, two and three MEMs for group I, II and III respectively, and used them in the partial RDA as

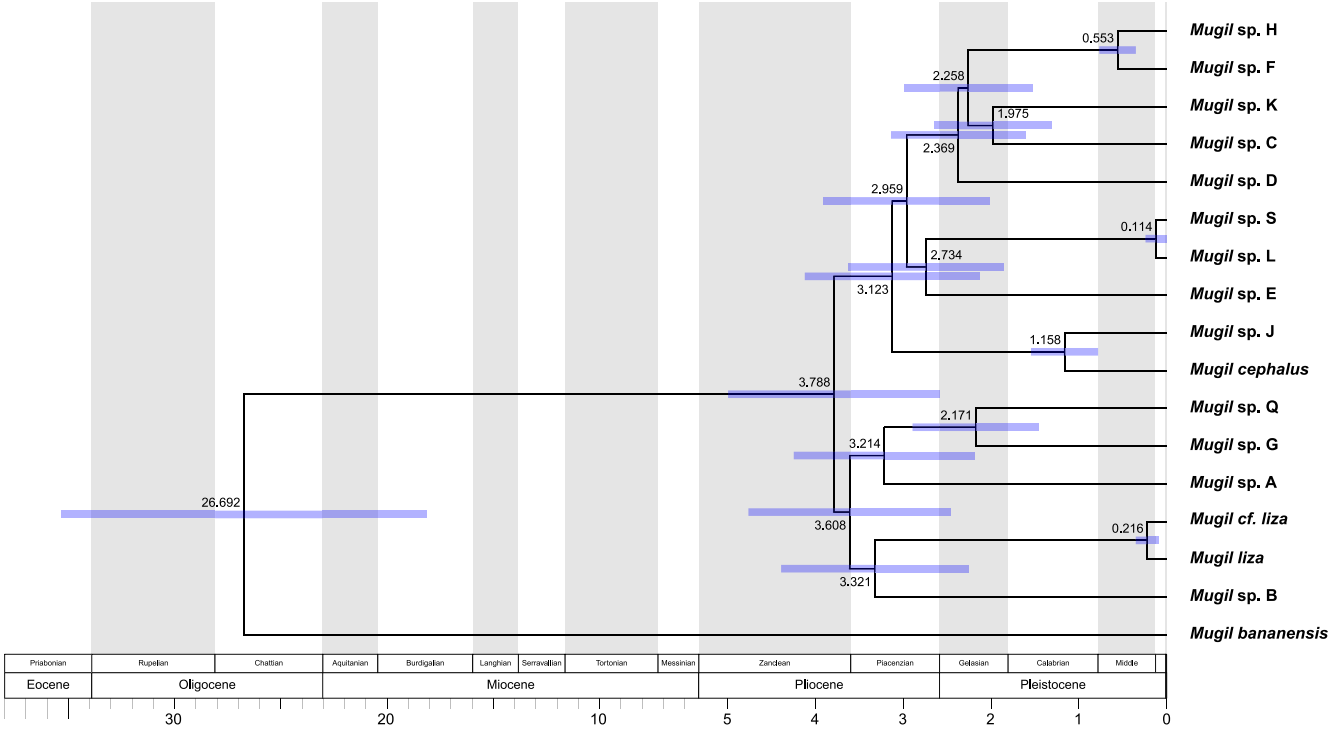

Fig. 4. Time-calibrated species tree of the *Mugil cephalus* species complex from SNAPP analysis. Nodelabels show mean divergence times and blue bars indicate the 95% highest posterior density (HPD) intervals.

Table 2
Overview of tested scenarios with varying time-stratification assumptions (TS), transition rate matrices (TRM), and maximum number of areas inhabited by the MRCA (MXA). Final Log-Likelihood (LnL) values are used for model comparison.

TS	TRM	MXA	Final LnL	Dispersal rate	Extinction rate
0	none	1	–	–	–
0	none	2	−59.726	0.548319	0.10015
0	none	3	−66.6982	0.148523	0.0850943
0	ER	1	–	–	–
0	ER	2	−58.0242	0.810613	0.128884
0	ER	3	−64.816	0.202604	0.0908958
0	CD	1	–	–	–
0	CD	2	−61.0313	2.85629	0.181192
0	CD	3	−65.9051	0.251332	0.0253442
1	none	1	–	–	–
1	none	2	−59.361	0.42508	0.0549699
1	none	3	−64.9768	0.149038	0.0587276
1	ER	1	–	–	–
1	ER	2	−57.9681	0.648344	0.0840069
1	ER	3	−63.1486	0.194557	0.0556959
1	CD	1	–	–	–
1	CD	2	−61.5222	0.795762	0.0220783
1	CD	3	−62.7138	0.311556	0.02423
2	none	1	–	–	–
2	none	2	−59.7711	0.422631	0.0566522
2	none	3	−65.3346	0.14742	0.0600114
2	ER	1	–	–	–
2	ER	2	−58.4092	0.664373	0.0898868
2	ER	3	−63.5713	0.194251	0.0591492
2	CD	1	–	–	–
2	CD	2	−61.7618	0.845581	0.0238781
2	CD	3	−62.8835	0.312423	0.023835

conditional terms against the two environmental variables.

Within group I, the RDA model explained 2.72 % of the genomic variance and was statistically significant (Table 3). The first RDA axes showed significance and the long-term maximum temperature significantly contributed to the genomic variance of the dataset (Table 3). A total of 185 SNPs were extracted from the top 1 ‰ distribution on RDA1. These SNPs were located within or upstream/downstream (5 kb) of 83 genes, among which many lnc-RNAs, pseudogenes, as well as a secretory phospholipase A2 receptor, a free fatty acid receptor and an immunoglobulin lambda (Supplementary Table 4).

Within group II, the model explained 8.99 % of the genomic variance and was statistically significant (Table 3). The first RDA axes as well as the long-term maximum temperature showed a significant correlation to the genomic dataset. A total of 301 SNPs were extracted from the 1 ‰ distribution of RDA1. These SNPs were located within or upstream/downstream (5 kb) of 96 genes, among which many lnc-RNAs, pseudogenes, as well as a secretory phospholipase A2 receptor, a free fatty acid receptor, an immunoglobulin lambda, and a Butyrophilin-like protein (Supplementary Table 4).

Within group III, the model explained 5.74 % of the genomic variance of the dataset and was statistically significant with a p-value = 0.014. RDA1 contributed significantly to the variance of the genomic dataset (p-value = 0.034) and both the long-term maximum temperature and the long-term temperature range correlated to the genetic variance (Table 3). In total, 453 SNPs were extracted from the top 1 ‰ distribution on RDA1 and were located within or upstream/downstream (5 kb) of 183 genes (Supplementary Table 4). Among these, several pseudogenes and non-coding RNAs were identified, as well as small rRNA subunit (5S, 5.8S), 18S and 28S rRNA, genes involved in immunity (immunoglobulins receptors, PARP12, PLA2R, mannose specific lectin like, C type lectins, T-cell receptors), genes involved in reproduction (Pibf1), genes involved in fatty acid metabolism (TMEM68, FFAR) and genes involved in translational functions (tRNAs, *Granulito*).

Table 3
Results of the partial RDA performed on the three groups, indicating the adjusted r2 value of the model, as well as the p-values associated with the model, the RDA axes and the two environmental variables tested.

	Group I	Group II	Group III
r2adj (%)	2.72	8.99	5.74
Model p-value	0.001	0.012	0.014
RDA1 p-value	0.001	0.015	0.034
RDA2 p-value	0.062	0.284	0.119
TempMaxLT_ss p-value	0.001	0.006	0.019
TempRangeLT_ss p-value	0.340	0.239	0.294

All three groups shared a total of 12 genes correlated with long-term maximum temperature (Supplementary Fig. 6, Supplementary Table 4). These include secretory phospholipase A2 receptor-like proteins, free fatty acid receptor 2-like proteins as well as lnc-RNAs. Groups II and III, representing East Pacific and West Atlantic as well as Indo-Pacific and West Pacific species, respectively, share eleven additional genes.

4. Discussion

4.1. Phylogenetic relationships & taxonomy

Analyses on the genetic structure of *Mugil cephalus* began over three decades ago (Crosetti et al., 1993; Crosetti et al., 1994; Rossi et al., 1998). The pronounced structure observed within the mitochondrial DNA suggested that *M. cephalus* is not a single species but rather a globally distributed species complex. Early morphometric analyses identified differences between various geographical locations (Corti and Crosetti, 1996). Despite these findings, a comprehensive taxonomical revision of the Mugilidae by Thomson (1997) did not recognize or incorporate this evidence, leaving the taxonomic status of the complex unresolved. The latter study maintained *M. cephalus* as a single species with a circumglobal distribution and synonymized it with 48 previously described species from around the world. However, subsequent genetic analyses revealed significant differences among *M. cephalus* populations (Rocha-Olivares et al., 2000; Rocha-Olivares et al., 2005; Fraga et al., 2007; Aurelle et al., 2008) leading to renewed calls for a taxonomical revision (Heras et al., 2009; Jamandre et al., 2009). Extensive genetic analyses eventually identified 14 distinct lineages (Durand et al., 2012a; Durand et al., 2012b; Whitfield et al., 2012), which later were recognized as distinct species based on the unified species concept of De Queiroz (2007) and prompted the assignment of arbitrary species names (Durand and Borsa, 2015). Further research uncovered additional species increasing the total number of recognized species within the MCSC to 16 (Durand et al., 2017; Viet Tran et al., 2017; Hasan et al., 2021). All of these studies relied solely on mitochondrial data and the resulting phylogenetic reconstructions did not provide sufficiently resolved and supported results to thoroughly investigate species evolution within the complex.

The whole-genome SNP data presented herein provided significantly enhanced phylogenetic resolution enabling the accurate identification of the phylogenetic relationships among the different species (Fig. 4). Within the MCSC we identified two main clades estimated to have diverged approximately 3.79 Ma (HPD: 2.58–4.99 Ma). Clade I includes five species that are present in the West Atlantic and East Pacific, while Clade II encompasses eleven species (including *Mugil* sp. I according to mitochondrial genome data; Supplementary Fig. 2) ranging from the East Atlantic to the West Pacific. The species delimitation analysis (Table 1) conducted in this study supports the recognition of at least 15 distinct species within the Mugil cephalus species complex (MCSC), consistent with the findings of (Durand et al., 2017) and (Hasan et al., 2021). However, these results should be interpreted with caution, as the limited number of specimens analysed per species may overlook intraspecific genetic variation or fail to capture the full geographic range of certain lineages. Future studies incorporating larger sample sizes and more comprehensive geographic coverage will be crucial to validate these findings and refine our understanding of species boundaries within the complex.

The phylogenetic reconstruction suggested a considerable genetic divergence between *Mugil liza* samples from the Caribbean Sea (107_S4 and 110_S9) and those from southern Brazil (156_S19 and 157_S23). This finding is consistent with previous studies indicating morphological disparity between samples from Argentina and the Caribbean Sea (Cousseau et al., 2005). However, despite the morphometric evidence supporting the division of *Mugil liza* into two species, molecular studies did not corroborate this (Heras et al., 2009; Heras et al., 2016). Our species delimitation methods, based on whole-genome SNP data, align with previous genetic findings and do not support the split of *Mugil liza* into two species (Table 1). However, our time-calibration analysis suggests that divergence among the analysed samples occurred as early as the Chibanian, around ~0.22 Ma (Fig. 5). To validate these results, we recommend conducting a more comprehensive whole-genome analysis encompassing samples from across the entire distribution range of *Mugil liza*.

Our results support the classification of the different lineages as species, much like Durand and Borsa (2015) already proposed, based on the unified species concept by De Queiroz (2007). Thomson (1997) was only able to separate *M. liza* from the remaining *Mugil* species he synonymized with *M. cephalus* by morphological and morphometric data, however, as seen in other species complexes primarily identified by genetic analyses, morphological variation is low and they are referred to as cryptic species or cryptic species complexes (Bickford et al., 2007). While the MCSC species appear superficially undistinguishable according to the results of Thomson (1997), no detailed morphological analysis has been conducted after specimens were assigned to specific species. A comprehensive taxonomical revision of the MCSC as suggested over 15 years ago (Heras et al., 2009) remains essential. This revision should integrate morphological and morphometric analyses alongside ecological and biogeographical data. Such efforts must focus on specimens that have been unequivocally matched to specific species through DNA barcoding as presented in this and in previous studies (e.g., Durand et al., 2017), to enable a thorough and accurate description of the species within the complex.

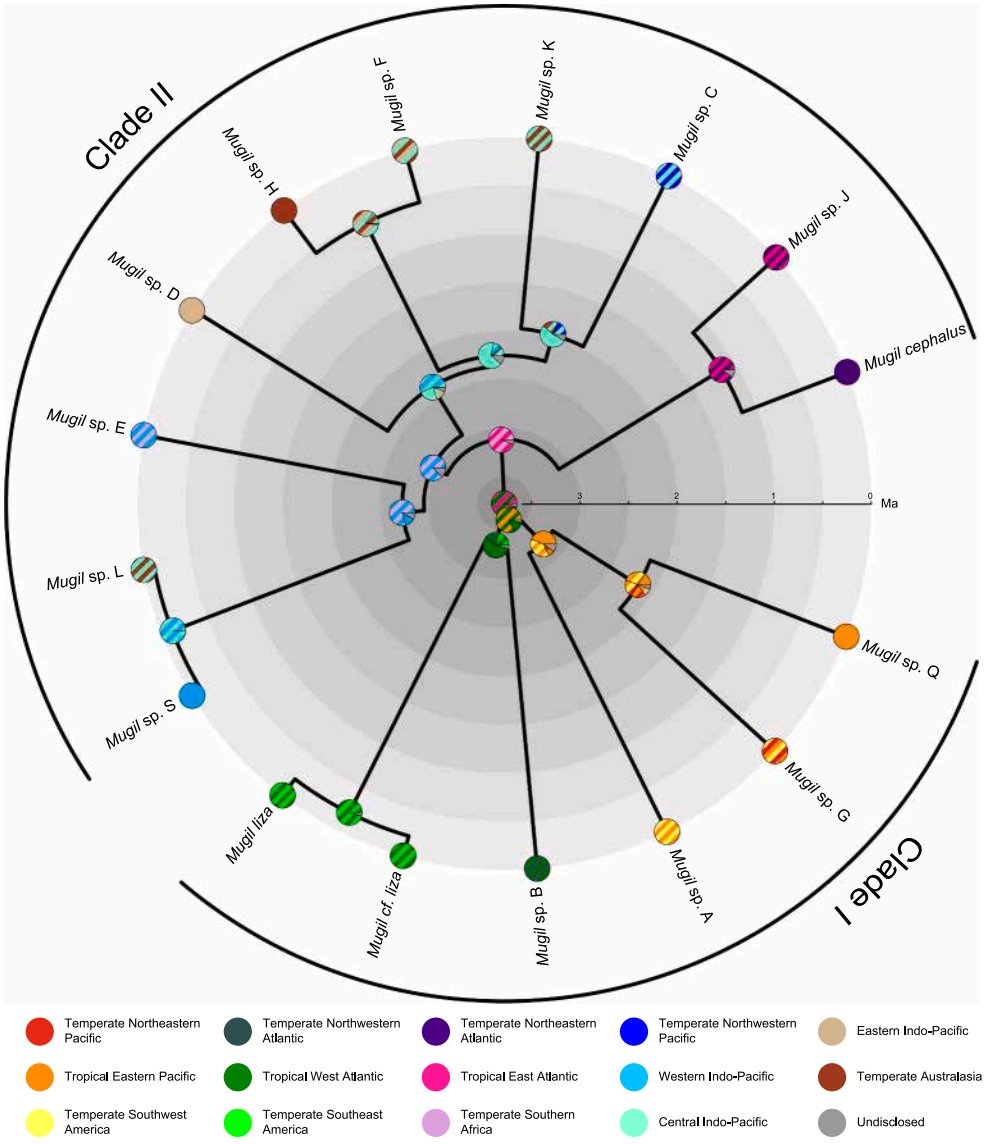

Fig. 5. Historical biogeography of the Mugil cephalus species complex according to the best-scoring model. Estimations of ancestral distribution areas were performed with a time-stratified model in Dispersal-Extinction-Cladogenesis. The legend at the bottom indicates biogeographic areas (illustrated in Fig. 2) used in this study according to Spalding et al. (2007). Coloured circles for each node depict inferred ancestral ranges, with probability for each single biogeographical area or combination of multiple biogeographical areas (indicated by striped sections) represented by the proportion of the circle (grey areas represent proportion for areas or area combinations not further disclosed by DECX).

4.2. Historical biogeography of the Mugil cephalus species complex

While there has been significant progress in unravelling the structure of the MCSC over the past 15 years, it remained unclear how these species dispersed and retained their unique morphological characteristics. Previous phylogenetic analyses suggested that the origin of the species complex dates back to around 4.5 Ma (Neves et al., 2020) indicating that its species would have dispersed around the globe in a very short time frame during a period when major contemporary physical barriers already existed. Our whole genome data suggests that the origin of the species complex is likely younger than indicated by Neves et al. (2020) at around 3.79 Ma (HPD: 2.58–4.99 Ma). Furthermore, ancestral range reconstruction revealed that all major marine realms were inhabited within less than two million years (Figs. 5 and 6). Such a rapid dispersal and speciation along with the retention of morphological, ecological, and behavioural traits is unique and not reported for any other fish species or species complex.

Ancestral range reconstruction traced the origin of the MCSC back to the Atlantic Ocean, although not one ecoregion could be identified exclusively (Fig. 5). Species within the complex are known to mainly inhabit coastal waters (Crosetti and Blaber, 2015) making it unlikely that the MRCA of the species complex maintained a distribution range spanning from the Tropical West to the Tropical East Atlantic. While the Tropical West Atlantic is a known hotspot for *Mugil* diversity with nine different species of the genus found in that area (Whitfield and Durand, 2023), previous phylogenetic inferences (Santini et al., 2015; Neves et al., 2020) suggest that the MCSC is more closely related to *M. capurii* (Perugia 1892) and *M. bananensis*, which are only present in the Tropical East Atlantic indicating that their MRCA probably was already established in that region.

Briggs (1995) noted that *trans*-Atlantic species are a common phenomenon in the shallow fish fauna of the Atlantic representing about 27 % of species along the tropical East Atlantic coast (Briggs and Bowen, 2013). However, many of these species likely form cryptic species complexes without persistent gene flow (Briggs and Bowen, 2013). When resolved phylogenies are available, they often show that these species originated in the West Atlantic and dispersed eastward (Muss et al., 2001; Floeter et al., 2008; Beldade et al., 2009; Boehm et al.,

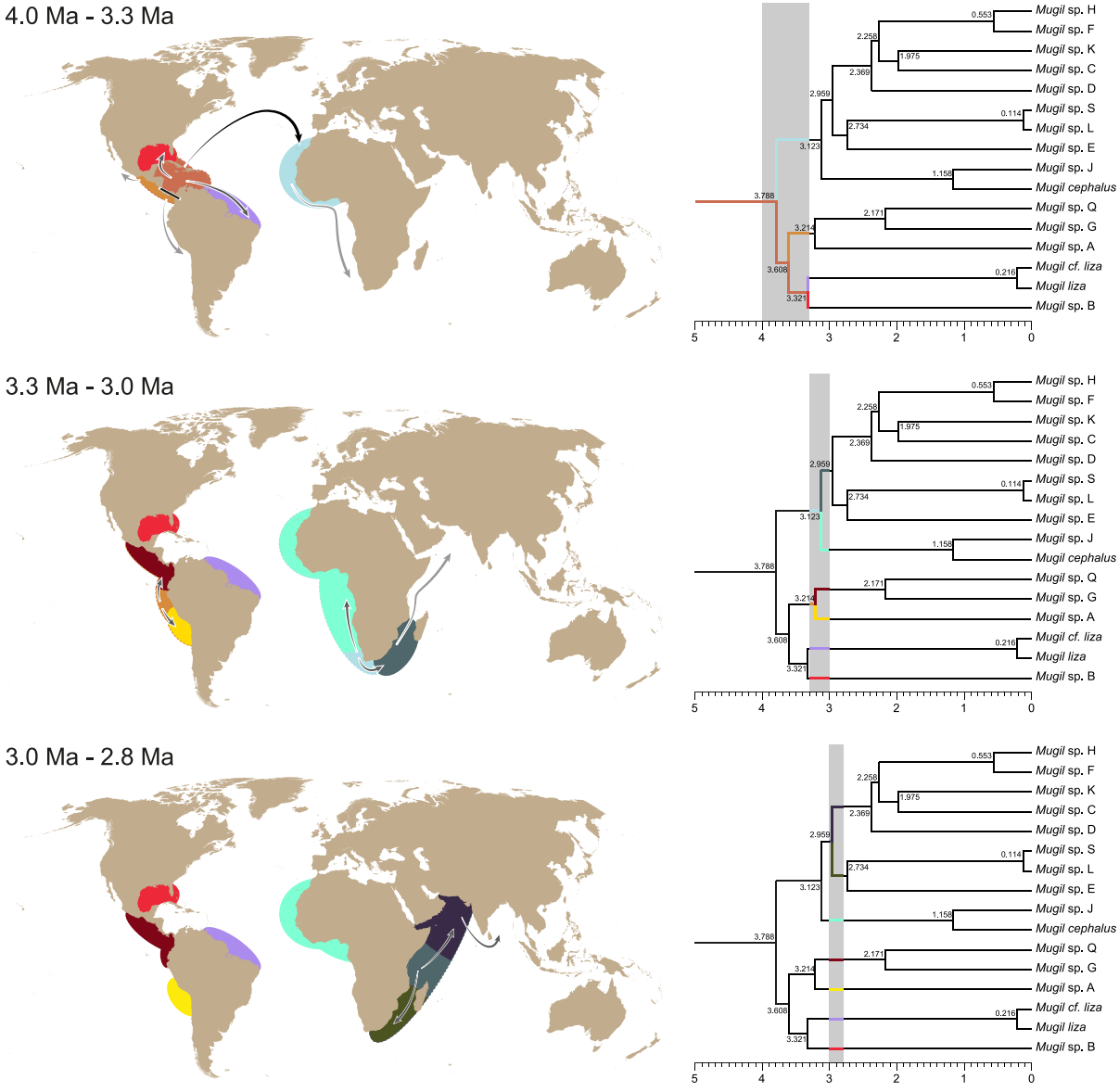

Fig. 6. Potential dispersal and speciation history of the *Mugil cephalus* species complex based on whole-genome SNP data presented in this study. Black bars: physical barrier, black arrow: founder event; grey arrow: distribution range disruption; light grey arrow: distribution range extensions; dotted light grey arrow: distribution range shift.

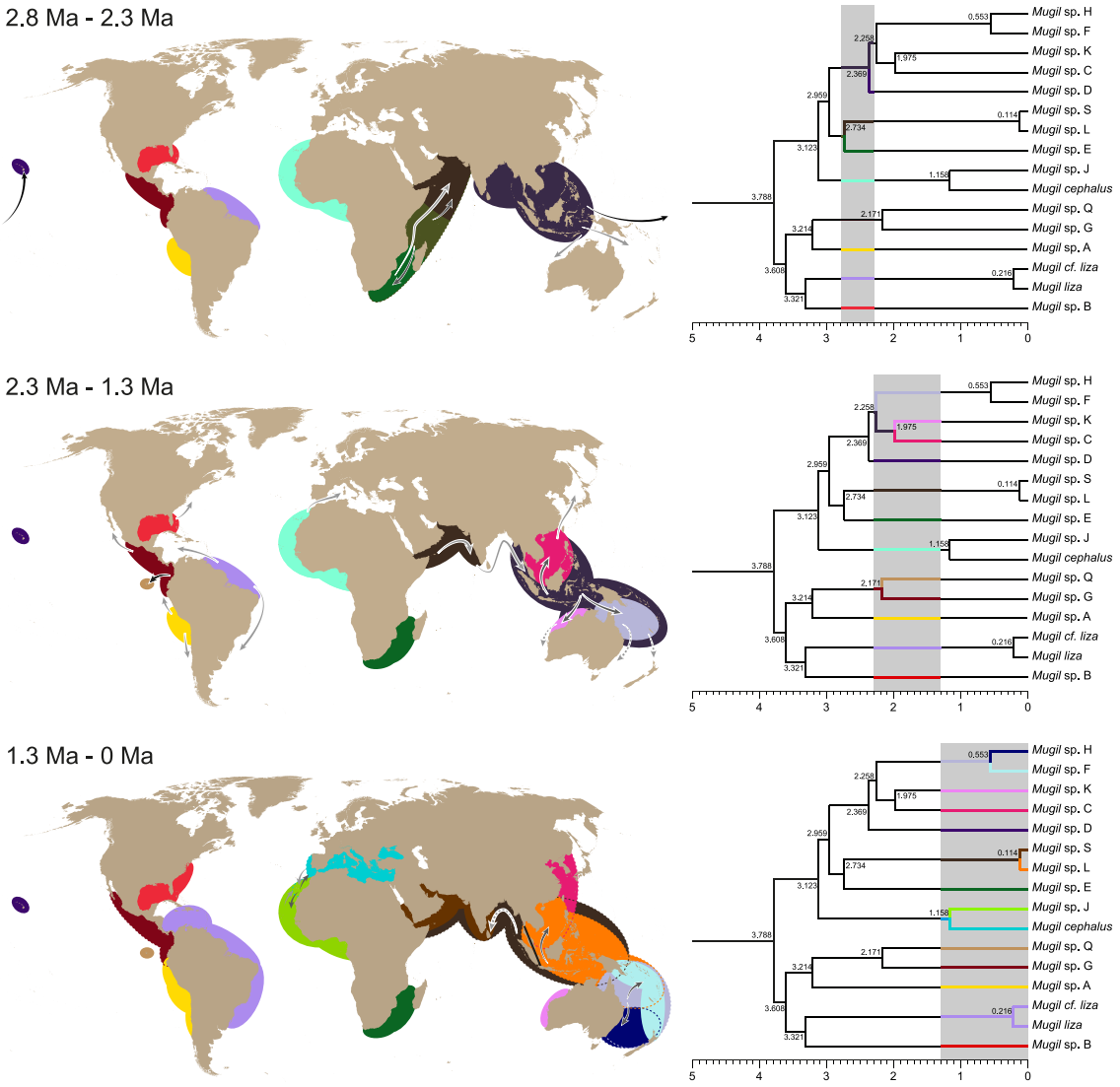

Fig. 6. (*continued*).

2013). Boehm et al. (2013), who analysed the biogeographic history of the *Hippocampus erectus* complex, concluded that a founder colonization from the tropical West Atlantic to the East Atlantic around 3.35 Ma followed by subsequent *trans*-Atlantic isolation prompted the speciation of *H. hippocampus* (Linnaeus 1758). They attributed this to the reorganization of North Atlantic currents especially the Gulf Stream and changes in the warm-temperate North Atlantic Gyre system (Kaneps, 1979; Haug and Tiedemann, 1998; Auderset et al., 2019). Although *H. erectus* Perry 1810 is a rather stationary species with seasonal migrations to avoid cold temperatures in northern West-Atlantic areas (Teixeira and Musick, 2001; Boehm et al., 2015), Boehm et al. (2013) hypothesised that a passive transportation of early life stages through the Gulf Stream or rafting via free-floating habitats likely facilitated the dispersal to the East Atlantic. A similar *trans*-Atlantic dispersal of *Mugil cephalus* occurring around 3.79 Ma (HPD: 2.58–4.99 Ma), might have followed analogous pathways. *Mugil cephalus* adults migrate off-shore for spawning with larvae remaining passively in open-water environments before recruiting to estuaries as juveniles (Whitfield et al., 2012). Casazza and Ross (2008) observed juvenile *Mugil* within the Gulf Stream associated with *Sargassum* and in open water, suggesting that both passive transport and rafting could have contributed to a *trans*-Atlantic dispersal of *M. cephalus*. Although ancestral range reconstruction and current phylogenetic placement do not provide a clear origin of the MCSC, the aforementioned mechanisms for *trans*-Atlantic dispersal support the hypothesis of an initial founder colonization from the West to the East Atlantic (Figs. 5 and 6).

Overcoming *trans*-Atlantic barriers kickstarted the successful dispersal of the MCSC. Two clades subsequently evolved in parallel, Clade I in the Western Atlantic and Eastern Pacific and Clade II from the Eastern Atlantic to the Indo-Pacific (Fig. 5). Within both clades, rapid range expansions with subsequent speciation occurred during the late Pliocene and beginning Pleistocene (Figs. 5 and 6). Especially within Clade II species dispersed rapidly towards the Indian Ocean and the West Pacific. Biogeographical reconstruction suggests two independent dispersal phases: The first one occurred between ~3.1 Ma and ~2.96 Ma involving the MRCA of the West Pacific subgroup which expanded quickly into the West Pacific but then disappeared from the Indian Ocean (Fig. 6). The second phase happened shortly after the divergence of the Indo Pacific/West Pacific MRCAs between ~2.96 and ~2.73 Ma. Pinsky et al. (2020) summarized that marine species are sensitive to environmental changes and can rapidly respond to shifting climatic conditions. The late Pliocene and the Plio-Pleistocene Transition are known for drastic climatic changes affecting all major marine realms, which likely influenced distribution range changes in the MCSC (e.g.,

O'Brien et al., 2014; Karas et al., 2017; He et al., 2021; Taylor et al., 2021).

Interestingly, in both clades one species each is endemic to an island group: *Mugil* sp. Q to Galapagos and *Mugil* sp. D to Hawaii. The origin of *Mugil* sp. Q can be linked to a founder colonization from the Tropical Eastern Pacific coast to Galapagos (Fig. 6). Previous studies have shown a high degree of endemism in the benthic and reef-related marine fauna of the Galapagos islands (Kelley et al., 2019) with founder colonizations from the mainland contributing to the regions diversity (e.g., Palmerín-Serrano et al., 2021).

The origin of *Mugil* sp. D, however, is much more challenging to reconstruct. Although Hawaii is recognized as a source for marine biodiversity export (Bowen et al., 2013), the biogeographical origin of Hawaiian species is poorly understood. McDowall (2003) showed that freshwater fish species from Hawaii have close relationships with sister-taxa in the Indo-Pacific, suggesting a possible marine dispersal from Central Pacific archipelagos south of Hawaii. The origin of *Mugil* sp. D might have followed a similar pathway, including an eastward dispersal along Central-Pacific island groups and a subsequent northward movement to Hawaii (Fig. 6). Although *Mugil* sp. D diverged early on from other West Pacific *M. cephalus* species and no closely related species are found in the southern Central Pacific, the presence of *Mugil* sp. F from Australia to French Polynesia suggests that *M. cephalus* species have the potential to disperse across widely separated islands. Details on the rapid dispersal of *Mugil* sp. D from the central West Pacific to Hawaii remain unclear. However, further data on the distribution and dispersal potential of South Pacific *Mugil* sp. F could provide a better understanding of its biogeographic history.

4.3. New insights into speciation in marine environments

4.3.1. Evidence for allopatric speciation through physical barriers

The rise of physical barriers and therefore the geographic isolation of and subsequent suppression of gene flow between lineages has been considered to be an important driver of speciation in marine ecosystems (Faria et al., 2021). However, absolute geographical barriers such as rising land masses are rare in marine environments, especially during the Pliocene and Pleistocene. Shen et al. (2011) previously identified an instance where a physical barrier led to the separation of *Mugil* sp. C and I. Their biogeographic analysis based on genetic markers suggested that parts of the population of their unique MRCA were repeatedly trapped in the Japan Sea during glacial periods. These populations experienced multiple demographic crashes which is still visible in the genetic data of *Mugil* sp. C (Shen et al., 2011). After a secondary contact both species now occur in sympatry but are isolated due to asynchronous reproductive seasons (Shen et al., 2015).

Within the MCSC, two other speciation events can be explained by the rise of physical barriers in the respective ecoregion: (1) separation of East Pacific and West Atlantic species at the origin of Clade I, and (2) separation of *Mugil* sp. L and *Mugil* sp. S (Fig. 6). In the first case, the time frame of the separation of the East Pacific lineage and the West Atlantic lineage coincides with the rise of the Isthmus of Panama. Fossil data from shallow water species indicate that the formation of the isthmus was completed at around 3.6 Ma (Coates et al., 1992), while a recent review points out that gene flow between marine populations ceased at around 3.2 Ma (O'Dea et al., 2016) despite surface water exchange between the East Pacific and West Atlantic probably continuing until about 2.76 Ma (Haug and Tiedemann, 1998). While our time-calibration estimates (Fig. 4) align with the latest median separation times of other fish species (O'Dea et al., 2016), additional climatic effects may have contributed to an earlier separation of both lineages while both waterbodies were still connected.

The separation of *Mugil* sp. L and *Mugil* sp. S is much more recent, around 114 ka. The current distribution of *Mugil* sp. S in the northern Indian Ocean (Fig. 1) and *Mugil* sp. L in the Northwest and Southwest Pacific as well as the biogeographical reconstruction (Fig. 5) suggest that both species separated in an area between Western Indo-Pacific and the Central Indo-Pacific. During the Pleistocene, extreme sea level fluctuations occurred within the Indo-Australian Archipelago, which restricted connections between the Indian Ocean and the Central West Pacific (Coates and Stallard, 2013). Voris (2000) estimated that during 62 % of the past 150,000 years (~93,000 years) sea levels were at least 30 m below today's levels, leading to a disconnection of the northern Indian Ocean and the northern and central West Pacific. Although multiple disconnections and reconnections between the two areas occurred, the physical barrier seemingly persisted long enough for the two lineages to evolve independently. Gaither and Rocha (2013) examined tropical reef fish distribution in the Western Indo-Pacific and Central Indo-Pacific and found other sister-species with a similar distinct distribution although many species show signs of much earlier separation. Furthermore, many species experienced secondary contact in the Indo-Australian Archipelago, a region where current data indicates that neither *Mugil* sp. L nor *Mugil* sp. S are present (Delrieu-Trottin et al., 2020).

4.3.2. Soft barriers promoting speciation

Species within the MCSC are highly adaptable and opportunistic (Crosetti and Blaber, 2015; Whitfield, 2015). Environmental factors such as salinity, turbidity, or dissolved oxygen levels, which otherwise act as limitations to the geographical ranges of other species, have little to no effect on these species (Nordlie, 2015; Whitfield, 2015). Their ability to feed on a broad spectrum of microphytobenthos, benthic invertebrates, or phyto- and zooplankton, allows them to exploit a unique trophic niche in their environments (Cardona, 2015). One environmental factor that has been shown to affect important aspects of their life history, such as larval development, or onset of reproductive periods, is temperature (Whitfield, 2015). It stands to reason that temperature has had a significant impact on the evolutionary history of the species complex.

Within the MCSC, we identified two instances where physical barriers directly impacted speciation. Although some species occur in sympatry today, differences in their reproductive cycles suggest that they experienced physical separation during their evolution, potentially caused by a temperature barrier (Shen et al., 2015). The following section details two examples demonstrating how changes in ocean currents and ocean temperature may have influenced ancestral species distribution ranges and therefore initiated speciation events.

The first example is the separation of the northeast Atlantic species (*Mugil* sp. J and *Mugil cephalus*) from the Indo-Pacific/West Pacific species which occurred around 3.12 Ma (Figs. 4 and 6). Ancestral range reconstruction indicates that the MRCA of these two clades was distributed within the Tropical East Atlantic and Temperate South Africa (Fig. 5). Today, the southwest African coastline is heavily influenced by the Benguela Upwelling System (BUS) which transports cold and nutrient-rich waters from deeper zones to the surface. A portion of the Benguela Current, the Benguela Coastal Current, flows close to the Namibian coastline transporting cool water northwards and meeting the southward-flowing, warm-water Angola Current at the Angolan coast (Leduc et al., 2014). The BUS supposedly started during the Miocene with a period of reduced upwelling during the early Pliocene and a steadily progressing temperature decline starting in the late Pliocene (Marlow et al., 2000; Etourneau et al., 2009; Rommerskirchen et al., 2011; Leduc et al., 2014; Rosell-Melé et al., 2014; Petrick et al., 2015; Mohanty et al., 2024). Sea surface temperature reconstructions from different localities within the BUS showed that upwelling in the mid-Pliocene (3.5 Ma to 3.0 Ma) initially occurred in the southern areas of the BUS as seasonal events (Leduc et al., 2014; Rosell-Melé et al., 2014; Petrick et al., 2015; Mohanty et al., 2024). Temperature differences of 5 °C or more were estimated between upwelling (winter) and non-upwelling (summer) periods accompanied by an overall decrease in SSTs (Leduc et al., 2014). The intensification of the BUS during the late Pliocene led to a more pronounced Angola-Benguela Frontal Zone

followed by its northward displacement (Rosell-Melé et al., 2014; Mohanty et al., 2024). This period also saw the occurrence of seasonal temperature fluctuations along the northern Namibian coastline (Leduc et al., 2014; Petrick et al., 2015), as well as to structural modifications of the Namibian coastline (Etourneau et al., 2009). These factors potentially had a significant impact on the life history of the MRCA of the above-mentioned clades. More pronounced temperature fluctuations due to seasonal upwelling could influence the onset of reproductive migrations. While reproductive cycles in temperate regions today are initiated by rising temperatures, falling temperatures contribute to the beginning of the reproduction cycle in subtropical populations (Whitfield et al., 2012). Although estimated annual mean SSTs during the late Pliocene remained above 24 °C in the BUS, seasonal upwelling generated a pronounced temperature gradient in the southern BUS (Leduc et al., 2014). This suggests that populations of the MRCA probably remained in sympatry along the BUS but potentially experienced a (gradual) shift in the initiation phases of their reproductive cycles along a northern to southern BUS gradient. Full geographical isolation of the MRCAs of the East Atlantic and the Indo-Pacific/West Pacific clades probably occurred only at the transition from the Pliocene to the Pleistocene, when the cooling temperatures and coastal landscape changes influenced their distribution ranges and both clades were already well established (Fig. 5; Etourneau et al., 2009; Rosell-Melé et al., 2014).

Today, a similar development can be observed in North Atlantic herring *Clupea harengus* Linnaeus 1758: two ecotypes can be identified, spring and autumn spawners, which differ in their reproductive cycles (Haegele and Schweigert, 1985). Fuentes-Pardo et al. (2024) showed that gene flow still persists between these ecotypes with different spawning seasons representing a state of incomplete reproductive isolation. Furthermore, Teske et al. (2019) implied temperature-related divergence in the absence of physical barriers in the Knysna sandgoby *Psammogobius knysnaensis* Smith 1935 along the South Africa coast, indicating that reproductive isolation can also occur along a thermal gradient.

The second example focusses on the divergence events along the East African coast. According to the biogeographic reconstruction, two subsequent events occurred in this region: (1) the separation of the MRCAs of the Indo-Pacific subgroup and the West Pacific subgroup around 2.96 Ma, and (2) the divergence of *Mugil* sp. E from the MRCA of *Mugil* sp. S and L around 2.73 Ma (Figs. 4–6). Predating the first event, the mid-Pliocene Warm Period (~3.3 Ma–3.0 Ma) was generally characterized by stable and warm SSTs (Haywood et al., 2016). However, data from the southeastern coast of Africa indicate unique paleoclimatic changes within the Indian Ocean. First, a gradual decline of SSTs between 3.7 Ma and 3.2 Ma by ~3°C led to cooler temperatures during the mid-Pliocene Warm Period (Taylor et al., 2021). This temperature decline coincides with the range extension of the MRCA of both subgroups observed after the divergence event described in the previous example (Figs. 5 and 6). This suggests that high SSTs in the southwest Indian Ocean coastal areas acted as dispersal barriers preventing the expansion into the Indian Ocean before ~3.2 Ma. Furthermore, paleoclimatic data indicate that this temperature decline was followed by a clear and gradual warming of SSTs from ~3.2 Ma to 2.9 Ma, increasing by ~2.2 °C with SSTs reaching above 30 °C (Taylor et al., 2021). Similar to conditions before ~3.2 Ma, this temperature rise likely caused a progressive disruption of the distribution range, acting as a barrier between the MRCA of the Indo-Pacific and the West Pacific subgroup.

The scarcity of paleoclimatic data for the northern Indian Ocean presents a challenge for the reconstruction of the biogeographic history of the MRCA of the West Pacific subgroup prior to its arrival in the Central West Pacific. However, the availability of additional data from the southwest Indian Ocean offers a valuable opportunity to gain deeper insights into the history of the Indo-Pacific subgroup. Following a period of gradual warming during the late Pliocene, a steep SST decrease by ~3.6 °C between 2.8 Ma and 2.2 Ma, presumably triggered by the Northern Hemisphere Glaciation, reopened the pathway to the Indian Ocean again. This event facilitated the rapid dispersal of the MRCA of the Indo-Pacific subgroup into the northern Indian Ocean (Fig. 6; Taylor et al., 2021). Beginning with the transition to the Pleistocene at ~2.7 Ma, coinciding with the divergence of *Mugil* sp. E from *Mugil* sp. L and S, periodic temperature fluctuations have been observed in the southwest and northwest Indian Ocean (Herbert et al., 2010; Taylor et al., 2021). However, while SSTs in the southwest Indian Ocean show a continuous cooling reaching a minimum of 24.4 °C ~ 2.5 Ma (Taylor et al., 2021), the long-term mean SSTs in the northwest Indian Ocean remain relatively stable between 2.7 Ma and 2.2 Ma, fluctuating around 27 °C (Herbert et al., 2010). The divergence within the Indo-Pacific subgroup could be explained by two primary hypotheses: 1) an increasing temperature gradient between both areas promoted distribution range shifts followed by distribution range disruption, or 2) decreasing SSTs in the southwest Indian Ocean combined with periodic temperature fluctuations caused reproductive cycles of populations along the east African coastline to become increasingly asynchronous. Similar to the patterns observed in the first example, these temperature-driven shifts in reproductive timing could have contributed to reproductive isolation without the need for physical barriers, gradually leading to divergence. The assessment of the first hypothesis is challenging due to the lack of paleoclimatic data on SST evolution from the tropical western Indian Ocean, the region that connects both areas. This gap in data makes it difficult to confirm whether such a gradient could have driven range shifts and disruptions leading to divergence. Similarly, data supporting the second hypothesis is also scarce. However, current observations of the reproductive activity provide some indirect support: *Mugil* sp. E from South Africa and *Mugil* sp. S from India exhibit asynchronous spawning seasons today (Whitfield, 2015). However, more detailed information is needed on the reproductive cycles in general, but also their plasticity, in the different species and how these are influenced by thermal gradients present in their distribution areas (e.g., Teske et al., 2019). Further, distinct spawning seasons could have evolved unrelated to the above-mentioned events due to climatic changes occurring in following time periods.

4.3.3. Genetic indications for temperature-related speciation

The examples discussed illustrate how coinciding paleoclimatic events, particularly changes in SSTs have likely influenced diversification within the MCSC. These temperature changes impacted distribution range expansion, shifts and disruptions resulting in periods of isolation with unique temperature influences throughout the evolutionary history of the species. Given that an important aspect of the life history of extant *Mugil cephalus* species, i.e., the reproduction cycle, is at least partially related to temperature, it is reasonable to expect that a genetic signature correlated with current temperature factors might be present.

Although we recognize that our results need to be taken carefully regarding the low number of individuals per clades, we showed that temperature variables can explain SNP genotype variation in our samples, considering their geographical location. The model applied to the whole species complex was significant, so we subdivided our dataset into the phylogenetic groups identified above to explore it further. Only the partial RDA of the third subset comprising the Indo-Pacific and West Pacific species was significant. RDA1 carried the significant variance of the genomic data, which correlated with both long-term temperature range and long-term maximum temperature (Table 3). We were able to identify 453 SNPs that are associated with temperature variation and which affect genes that are known to be influenced by temperature. Among these are copies of the 5S and 5.8S rRNA which have been shown to be impacted when species are experiencing heat stress (Rubin and Hogness, 1975; Darriere et al., 2022). Further, immunity and inflammation related genes were also impacted by SNPs, among whichHA1F and PLAR play a role in inflammation responses (Zhou et al., 2021) as well as Butyrophilin-like proteins which are mainly affecting the activity of T lymphocytes (Malinowska et al., 2017). Genes related to the lipid

metabolism like PLA2R for lipid mediator production, or FFAR2 for fat development were also identified and are known to be impacted by temperature changes, especially toward high temperatures (Zhao et al., 2021).

This analysis represents a first approach revealing important facets for future studies. Our results illustrate potential temperature-related variations in the genome due to specific temperature regimes experienced by the different species. Correlation of genomic data and temperature range and the retrieval of affected genes related to heat stress show differences between temperate and subtropical species. A broader sampling covering the complete temperature and distribution range of each species will contribute to a better understanding of genes affected by temperature-related diversification processes.

4.4. Implications for future speciation events

The mid-Pliocene Warm Period is frequently cited as an analogue to current and future climate conditions (Haywood et al., 2016). Spanning from ~3.3 Ma to ~3.0 Ma, this interval was characterized by a warmer and more stable climate within the Pliocene, for which a general cooling trend can be witnessed from its early to late period. The transition towards the Pleistocene is characterized by multiple severe sea surface temperature decreases and more temperature fluctuations (Haywood et al., 2016). Within the mid and late Pliocene, ancestral lineages of the MCSC experienced large distribution range expansions followed by divergence events due to changing climatic conditions. The slightly cooler SSTs during the late Pliocene in comparison to early and mid-Pliocene seemingly initiated and promoted their dispersal. This indicates that SSTs above the late Pliocene annual mean SSTs from warm water regions are inadequate habitats for *M. cephalus* and act as dispersal barriers. The current distribution of species within the MCSC supports this hypothesis. Modern warm water regions, such as the southern Indo-Australian Archipelago or the coastal region of eastern Africa, exhibit SSTs comparable to those of the mid-Pliocene Warm Period (Dowsett and Robinson, 2009). However, there are no records of *M. cephalus* species occurring in these areas (Fig. 1; Delrieu-Trottin et al., 2020). This has already caused a disruption in the distribution range of *Mugil* sp. L, with isolated populations now observed in the South China Sea as well as the Coral Sea (Fig. 1). This fragmentation may potentially lead to reproductive isolation and subsequent speciation within *Mugil* sp. L. A similar process can be predicted for *Mugil* sp. G, which is distributed along the East Pacific coast from Southern California to Peru. However, *Mugil* sp. G has not been observed in the tropical East Pacific, which could either be due to insufficient sampling in that area or could indicate that this species has also undergone distribution range disruption resulting in two isolated populations remaining north and south of the Equator.

In addition to distribution range disruptions, changing sea surface temperatures could impact all MCSC species, particularly in higher latitudes. Rising SSTs may lead to distribution range expansions or shifts. One example could be *Mugil cephalus* from the Northeastern Atlantic and Mediterranean Sea. Predicted warming of SSTs in the temperate Northern Atlantic and relatively stable SSTs in the subtropical Northern Atlantic (Dowsett et al., 2009) may allow this species to expand further north into the North Sea while retaining its current distribution. On the other hand, species in the West Atlantic, such as *Mugil liza* in the Caribbean Sea and *Mugil* sp. B in the Gulf of Mexico might be negatively affected by rising temperatures as predicted by mid-Pliocene Warm Period comparisons (O'Brien et al., 2014). SSTs that potentially exceed the thermal limit of both species in the respective areas could lead to distribution range shifts further south and north, respectively.

5. Conclusion

Following extensive analyses based on mitochondrial DNA data examining the structure of the *Mugil cephalus* species complex, genomic data provided in this study confirms the presence of at least 15 species currently grouped under the *M. cephalus* name. Phylogenetic reconstruction based on whole genome SNP data identified two main clades, the West Atlantic/East Pacific clade and the East Atlantic/Indo-Pacific clade. Our time calibration analysis dates the origin of the *Mugil cephalus* species complex within the early to mid-Pliocene transition and suggests a rapid colonisation of all major marine realms within less than 2 Myr. An initial founder colonization from the West to the East Atlantic may have been facilitated by changes in the North Atlantic current system. Within both clades we found support of physical barriers promoting speciation. Ancestral range reconstructions indicate that climatic changes, such as fluctuations in sea surface temperatures during the late Pliocene and early Pleistocene, likely impacted the distribution ranges and reproductive cycles of the species complex, playing a significant role in the early phases of speciation. However, to fully understand the evolutionary processes underlying speciation within the *Mugil cephalus* species complex, regional studies combining extensive genetic sampling and paleoclimatic reconstructions with a modelling approach (e.g., Yannic et al., 2020; De Jode et al., 2023) are needed.

CRediT authorship contribution statement

Philipp Thieme: Writing – original draft, Visualization, Software, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Celine Reisser:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Corinne Bouvier:** Writing – review & editing, Investigation. **Fabien Rieuvillneuve:** Writing – review & editing, Validation, Funding acquisition, Conceptualization. **Philippe Béarez:** Writing – review & editing, Resources. **Richard R. Coleman:** Writing – review & editing, Resources. **Jean Jubrice Anissa Volanandiana:** Writing – review & editing, Resources. **Esmeralda Pereira:** Writing – review & editing, Resources. **Mauro Nirchio–Tursellino:** Writing – review & editing, Resources. **María Inés Roldán:** Writing – review & editing, Resources. **Sandra Heras:** Writing – review & editing, Resources. **Nathalia Tirado-Sánchez:** Writing – review & editing, Resources. **Eric Pulis:** Writing – review & editing, Resources. **Fabien Leprieur:** Writing – review & editing, Validation, Funding acquisition, Conceptualization. **Jean-Dominique Durand:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization.

Funding

P. Thieme was funded by a Walter-Benjamin scholarship granted by the German Research Foundation (TH 2675/1-1); E. Pereira was supported by the Foundation for Science and Technology (SFRH/BD/121042/2016) and via projects UIDB/04292/2020 (https://doi.org/10.54499/UIDB/04292/2020) and LA/P/0069/2020 (https://doi.org/10.54499/LA/P/0069/2020).

Data availability

Data files, including data sets and scripts utilized in this study, can be found in the GitHub data repository: https://github.com/PhilThieme/WGS-Mugil-cephalus.

Acknowledgement

We are grateful to Sally Reader, Australia Museum, for contributing tissue samples of *Mugil* sp. H. We are thankful to Isaure de Buron, College Charleston SC, and the staff of the South Carolina Department of Natural Resources, Charleston SC, for providing tissue samples of *Mugil* sp. B. We thank Stéphanie D'Agata, Institut Européen Universitaire de la Mer, France, for contributing tissue samples of *Mugil* sp. F. We are grateful to Adeline Collet, OCEA Consult', for sharing tissue samples of *Mugil* sp. E. We are also thankful to Natalia Namarchiori for providing additional

tissue samples of *Mugil liza*. We are also grateful to María B. Cousseau for assistance in the collection of samples. Furthermore, we thank Iago Bonnici and Fabien Condamine for their assistance in the ancestral distribution range estimation using DECX.

Appendix A. Supplementary data

Supplementary Fig. 1: Phylogenetic tree of the Mugil cephalus species complex based on the mitochondrial genome (without the D-Loop) including all 39 samples sequenced herein (samples of Mugil sp. I are marked with blue dots) as well as seven mitochondrial genomes (marked with red dots) retrieved from Genbank (Supplementary Table 3). Supplementary Fig. 2: Phylogenetic tree of the Mugil cephalus species complex based on the mitochondrial genome without sequenced samples of Mugil sp. I. Supplementary Fig. 3: Phylogenetic tree of the Mugil cephalus species complex inferred with IQtree2 from PI2. Supplementary Fig. 4: Phylogenetic tree of the Mugil cephalus species complex inferred with RAxML from PI2. Supplementary Fig. 5: Pairwise panels of geographical and surface-temperature data evaluated in the GEA (Supplementary Table 2) showing co-variation between environmental factors. Supplementary Fig. 6: Venn diagram depicting the number of shared genes between the three groups of the GEA analysis. Supplementary Table 1: List of barcoded specimens (based on COI), with geographical locations and references, visualized in Fig. 1. Supplementary Table 2: List of samples used in whole-genome shotgun analysis including species association according to DNA barcoding results as well as respective BIN, alternative artificial species references used in Shen et al. (2011, 2015), their geographical origin, and sea surface data, i.e., maximum temperature, temperature range, long term maximum temperature and long-term temperature range, retrieved for each sample from the Bio-ORACLE dataset v2.2 (Assis et al. 2018). Supplementary Table 3: List of mitochondrial genomes referenced in Supplementary Figs. 1 and 2. Supplementary Table 4: Results of the partial RDA analysis, listing the outlier SNPs for group I and III, as well as the lists of genes impacted by those SNPs (shared genes between all three groups are marked in red). Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2025.108296.

References

Amor, M.D., Norman, M.D., Cameron, H.E., Strugnell, J.M., 2014. Allopatric speciation within a cryptic species complex of Australasian octopuses. PLoS One 9, e98982. https://doi.org/10.1371/journal.pone.0098982.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A., De Clerck, O., 2018. Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling. Glob. Ecol. Biogeogr. 27, 277–284. https://doi.org/10.1111/geb.12693.

Auderset, A., Martínez-García, A., Tiedemann, R., Hasenfratz, A.P., Eglinton, T.I., Schiebel, R., Sigman, D.M., Haug, G.H., 2019. Gulf Stream intensification after the early Pliocene shoaling of the Central American Seaway. Earth Planet. Sci. Lett. 520, 268–278. https://doi.org/10.1016/j.epsl.2019.05.022.

Aurelle, D., Barthelemy, R.-M., Quignard, J.-P., Trabelsi, M., Faure, E., 2008. Molecular phylogeny of Mugilidae (Teleostei: Perciformes). Open Marine Biol. J. 2, 29–37. https://doi.org/10.2174/1874450800802010029.

Beeravolu, C.R., Condamine, F.L., 2016. An extended maximum likelihood inference of geographic range evolution by dispersal, local extinction and cladogenesis. bioRxiv 038695. https://doi.org/10.1101/038695.

Beldade, R., Heiser, J., Robertson, D.R., Gasparini, J.L., Floeter, S., Bernardi, G., 2009. Historical biogeography and speciation in the Creole wrasses (Labridae, *Clepticus*). Mar. Biol. 156, 679–687. https://doi.org/10.1007/s00227-008-1118-5.

Bernardi, G., 2013. Speciation in fishes. Mol. Ecol. 22, 5487–5502. https://doi.org/10.1111/mec.12494.

Bernardi, G., Findley, L., Rocha-Olivares, A., 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. Evolution 57, 1599–1609. https://doi.org/10.1111/j.0014-3820.2003.tb00367.x.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155. https://doi.org/10.1016/j.tree.2006.11.004.

Boehm, J.T., Woodall, L., Teske, P.R., Lourie, S.A., Baldwin, C., Waldman, J., Hickerson, M., Rocha, L., 2013. Marine dispersal and barriers drive Atlantic seahorse diversification. J. Biogeogr. 40, 1839–1849. https://doi.org/10.1111/jbi.12127.

Boehm, J.T., Waldman, J., Robinson, J.D., Hickerson, M.J., 2015. Population genomics reveals Seahorses (*Hippocampus erectus*) of the Western Mid-Atlantic coast to be residents rather than vagrants. PLoS One 10, e0116219. https://doi.org/10.1371/journal.pone.0116219.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., 2013. The origins of tropical marine biodiversity. Trends Ecol. Evol. 28, 359–366. https://doi.org/10.1016/j.tree.2013.01.018.

Briggs, J.C., 1960. Fishes of worldwide (circumtropical) distribution. Copeia 1960, 171–180. https://doi.org/10.2307/1439652.

Briggs, J.C., Bowen, B.W., 2013. Marine shelf habitat: biogeography and evolution. J. Biogeogr. 40, 1023–1035. https://doi.org/10.1111/jbi.12082.

Briggs, J.C., 1995. Chapter 7 Neogene. In: Briggs, J.C. (Ed.), Dev Palaeontol Stratigr. Elsevier, pp. 147–189. doi: 10.1016/S0920-5446(06)80057-9.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Cardona, L., 2015. Food and feeding of Mugilidae. In: Crosetti, D., Blaber, S.J. (Eds.), Biology, Ecology and Culture of Grey Mullets (mugilidae). CRC Press, Boca Raton, USA, pp. 165–195.

Carlier, A., Riera, P., Amouroux, J.-M., Bodiou, J.-Y., Escoubeyrou, K., Desmalades, M., Caparros, J., Grémare, A., 2007. A seasonal survey of the food web in the Lapalme Lagoon (northwestern Mediterranean) assessed by carbon and nitrogen stable isotope analysis. Estuar. Coast. Shelf Sci. 73, 299–315. https://doi.org/10.1016/j.ecss.2007.01.012.

Casazza, T.L., Ross, S.W., 2008. Fishes associated with pelagic Sargassum and open water lacking Sargassum in the Gulf Stream off North Carolina. Fish. Bull. 106, 348–363.

Castellanos-Juárez, M., Mendoza-Carranza, M., Pacheco-Almanzar, E., Álvarez-Hernández, S.H., Ibáñez, A.L., 2024. The regularity of the striped mullet *Mugil cephalus* spawning in accordance with Gulf of Mexico tides. Environ. Biol. Fishes 1–11. https://doi.org/10.1007/s10641-024-01542-1.

Cayuela, H., Dorant, Y., Mérot, C., Laporte, M., Normandeau, E., Gagnon-Harvey, S., Clément, M., Sirois, P., Bernatchez, L., 2021. Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish. Mol. Ecol. 30, 1624–1641. https://doi.org/10.1111/mec.15835.

Chappell, J., Thom, B.G., 1977. Sea Levels and Coasts. Sunda and Sahul, pp. 275–291.

Chen, S., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2, e107.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly 6, 80–92. https://doi.org/10.4161/fly.19695.

Coates, A.G., Jackson, J.B.C., Collins, L.S., Cronin, T.M., Dowsett, H.J., Bybell, L.M., Jung, P., Obando, J.A., 1992. Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. Geol. Soc. Am. Bull. 104, 814–828. https://doi.org/10.1130/0016-7606(1992)104<0814:Cotiop>2.3.Co;2.

Coates, A.G., Stallard, R.F., 2013. How old is the Isthmus of Panama? Bull. Mar. Sci. 89, 801–813. https://doi.org/10.5343/bms.2012.1076.

Cornils, A., Held, C., 2014. Evidence of cryptic and pseudocryptic speciation in the *Paracalanus parvus* species complex (Crustacea, Copepoda, Calanoida). Front. Zool. 11, 19. https://doi.org/10.1186/1742-9994-11-19.

Corti, M., Crosetti, D., 1996. Geographic variation in the grey mullet: a geometric morphometric analysis using partial warp scores. J. Fish Biol. 48, 255–269. https://doi.org/10.1111/j.1095-8649.1996.tb01117.x.

Cousseau, M.B., González Castro, M., Figueroa, D.E., Gosztonyi, A.E., 2005. Does *Mugil liza* Valenciennes 1836 (Teleostei: Mugiliformes) occur in Argentinean waters? Rev. Biol. Mar. Oceanogr. 40, 133–140.

Cowman, P.F., Bellwood, D.R., 2013. Vicariance across major marine biogeographic barriers: temporal concordance and the relative intensity of hard versus soft barriers. Proc. R. Soc. B 280, 20131541. https://doi.org/10.1098/rspb.2013.1541.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer, Sunderland.

Crandall, E.D., Riginos, C., Bird, C.E., Liggins, L., Treml, E., Beger, M., Barber, P.H., Connolly, S.R., Cowman, P.F., DiBattista, J.D., Eble, J.A., Magnuson, S.F., Horne, J. B., Kochzius, M., Lessios, H.A., Liu, S.Y.V., Ludt, W.B., Madduppa, H., Pandolfi, J.M., Toonen, R.J., Network, C.M.o.t.D.o.t.I.-P., Gaither, M.R., 2019. The molecular biogeography of the Indo-Pacific: testing hypotheses with multispecies genetic patterns. Glob. Ecol. Biogeogr. 28, 943–960. https://doi.org/10.1111/geb.12905.

Crosetti, D., Blaber, S.J., 2015. Biology, Ecology and Culture of Grey Mullets (Mugilidae). CRC Press, Boca Raton, USA.

Crosetti, D., Avise, J., Placidi, F., Rossi, A., Sola, L., 1993. Geographic variability in the grey mullet *Mugil cephalus*: preliminary results of mtDNA and chromosome analyses. In: Genetics in Aquaculture. Elsevier, pp. 95-101. doi: 10.1016/B978-0-444-81527-9.50014-2.

Crosetti, D., Nelson, W., Avise, J., 1994. Pronounced genetic structure of mitochondrial DNA among populations of the circumglobally distributed grey mullet (*Mugil cephalus*). J. Fish Biol. 44, 47–58. https://doi.org/10.1111/j.1095-8649.1994.tb01584.x.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

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.
Darriere, T., Jobet, E., Zavala, D., Escande, M.L., Durut, N., de Bures, A., Blanco-Herrera, F., Vidal, E.A., Rompais, M., Carapito, C., Gourbiere, S., Sáez-Vásquez, J., 2022. Upon heat stress processing of ribosomal RNA precursors into mature rRNAs is compromised after cleavage at primary P site in Arabidopsis thaliana. RNA Biol. 19, 719–734. https://doi.org/10.1080/15476286.2022.2071517.
De Jode, A., Le Moan, A., Johannesson, K., Faria, R., Stankowski, S., Westram, A.M., Butlin, R.K., Rafajlović, M., Fraïsse, C., 2023. Ten years of demographic modelling of divergence and speciation in the sea. Evol. Appl. 16, 542–559. https://doi.org/10.1111/eva.13428.
De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083.
Delrieu-Trottin, E., Durand, J.-D., Limmon, G., Sukmono, T., Kadarusman, Sugeha, H., Chen, W.-J., Busson, F., Borsa, P., Darhuddin, H., Sauri, S., Fitriana, Y., Zein, M., Hocdé, R., Pouyaud, L., Keith, P., Wowor, D., Steinke, D., Hanner, R., Hubert, N., 2020. Biodiversity inventory of the grey mullets (Actinopterygii: Mugilidae) of the Indo-Australian Archipelago through the iterative use of DNA-based species delimitation and specimen assignment methods. Evolut. Appl. 13, 1451–1467. doi: 10.1111/eva.1292.
Dowsett, H.J., Chandler, M.A., Robinson, M.M., 2009. Surface temperatures of the Mid-Pliocene North Atlantic Ocean: implications for future climate. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 367, 69–84. https://doi.org/10.1098/rsta.2008.0213.
Dowsett, H.J., Robinson, M.M., 2009. Mid-Pliocene equatorial Pacific sea surface temperature reconstruction: a multi-proxy perspective. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 367, 109–125. https://doi.org/10.1098/rsta.2008.0206.
Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guénard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., Siberchicot, A., 2023. adespatial: Multivariate Multiscale Spatial Analysis.
Dray, S., Legendre, P., Peres-Neto, P.R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Model. 196, 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015.
Drew, J.A., Barber, P.H., 2012. Comparative phylogeography in Fijian coral reef fishes: a multi-taxa approach towards marine reserve design. PLoS One 7, e47710. https://doi.org/10.1371/journal.pone.0047710.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 1–8. https://doi.org/10.1186/1471-2148-7-214.
Durand, J.-D., Borsa, P., 2015. Mitochondrial phylogeny of grey mullets (Acanthopterygii: Mugilidae) suggests high proportion of cryptic species. C. R. Biol. 338, 266–277. https://doi.org/10.1016/j.crvi.2015.01.007.
Durand, J.-D., Chen, W.-J., Shen, K.-N., Fu, C., Borsa, P., 2012a. Genus-level taxonomic changes implied by the mitochondrial phylogeny of grey mullets (Teleostei: Mugilidae). C. R. Biol. 335, 687–697. https://doi.org/10.1016/j.crvi.2012.09.005.
Durand, J.-D., Shen, K.N., Chen, W.J., Jamandre, B.W., Blel, H., Diop, K., Nirchio, M., Garcia de Leon, F.J., Whitfield, A.K., Chang, C.W., Borsa, P., 2012b. Systematics of the grey mullets (Teleostei: Mugiliformes: Mugilidae): molecular phylogenetic evidence challenges two centuries of morphology-based taxonomy. Mol. Phylogenet. Evol. 64, 73–92. https://doi.org/10.1016/j.ympev.2012.03.006.
Durand, J.-D., Hubert, N., Shen, K.-N., Borsa, P., 2017. DNA barcoding grey mullets. Rev. Fish Biol. Fish. 27, 233–243. https://doi.org/10.1007/s11160-016-9457-7.
Etourneau, J., Martinez, P., Blanz, T., Schneider, R., 2009. Pliocene–Pleistocene variability of upwelling activity, productivity, and nutrient cycling in the Benguela region. Geology 37, 871–874. https://doi.org/10.1130/G25733A.1.
Faria, R., Johannesson, K., Stankowski, S., 2021. Speciation in marine environments: diving under the surface. J. Evol. Biol. 34, 4–15. https://doi.org/10.1111/jeb.13756.
Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J., Smith-Vaniz, W.F., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E., Gasparini, J.L., 2008. Atlantic reef fish biogeography and evolution. J. Biogeogr. 35, 22–47. https://doi.org/10.1111/j.1365-2699.2007.01790.x.
Fraga, E., Schneider, H., Nirchio, M., Santa-Brigida, E., Rodrigues-Filho, L., Sampaio, I., 2007. Molecular phylogenetic analyses of mullets (Mugilidae, Mugiliformes) based on two mitochondrial genes. J. Appl. Ichthyol. 23, 598–604. https://doi.org/10.1111/j.1439-0426.2007.00911.x.
Fransen, C.H., 2002. Taxonomy, phylogeny, historical biogeography, and historical ecology of the genus *Pontonia* (Crustacea: Decapoda: Caridea: Palaemonidae). Zoologische Verhandelingen 336, 1–433.
Fuentes-Pardo, A.P., Farrell, E.D., Pettersson, M.E., Sprehn, C.G., Andersson, L., 2023. The genomic basis and environmental correlates of local adaptation in the Atlantic horse mackerel (*Trachurus trachurus*). Evol. Appl. 16, 1201–1219. https://doi.org/10.1111/eva.13559.
Fuentes-Pardo, A.P., Stanley, R., Bourne, C., Singh, R., Emond, K., Pinkham, L., McDermid, J.L., Andersson, L., Ruzzante, D.E., 2024. Adaptation to seasonal reproduction and environment-associated factors drive temporal and spatial differentiation in northwest Atlantic herring despite gene flow. Evol. Appl. 17, e13675. https://doi.org/10.1111/eva.13675.
Gaither, M.R., Bowen, B.W., Rocha, L.A., Briggs, J.C., 2016. Fishes that rule the world: circumtropical distributions revisited. Fish Fish. 17, 664–679. https://doi.org/10.1111/faf.12136.
Gaither, M.R., Rocha, L.A., 2013. Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. J. Biogeogr. 40, 1638–1648. https://doi.org/10.1111/jbi.12126.
Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. doi: 10.48550/arXiv.1207.3907.
Garrison, E., Kronenberg, Z.N., Dawson, E.T., Pedersen, B.S., Prins, P., 2022. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. PLoS Comput. Biol. 18, e1009123. https://doi.org/10.1371/journal.pcbi.1009123.
Gee, J.M., 2004. Gene flow across a climatic barrier between hybridizing avian species, California and Gambel's quail (*Callipepla californica* and *C. gambelii*). Evolution 58, 1108–1121. https://doi.org/10.1111/j.0014-3820.2004.tb00444.x.
Gilbert, C.R., 1993. Geographic Distribution of the Striped Mullet (*Mugil cephalus* Linnaeus) in the Atlantic and Eastern Pacific Oceans. Fla Sci. 56, 204–210.
Haegele, C.W., Schweigert, J.F., 1985. Distribution and characteristics of herring spawning grounds and description of spawning behavior. Can. J. Fish. Aquat. Sci. 42, s39–s55. https://doi.org/10.1139/f85-261.
Harrison, I.J., Howes, G.J., 1991. The pharyngobranchial organ of mugilid fishes; its structure, variability, ontogeny, possible function and taxonomic utility. Bulletin of the British Museum, Natural History. Zoology 57, 111–132.
Hasan, M., Hasan, A., Béarez, P., Shen, K.-N., Chang, C.-W., Tran, T., Golani, D., Al-Saboonchi, A., Siddiqui, P., Durand, J.-D., 2022. Planiliza lauvergnii (Eydoux & Souleyet, 1850), a senior synonym of Planiliza affinis (Günther, 1861) with a re-evaluation of keeled back mullets (Mugiliformes: Mugilidae). Zootaxa 5194. doi: 10.11646/ZOOTAXA.5194.4.2.
Hasan, A., Siddiqui, P.J.A., Amir, S.A., Durand, J.-D., 2021. DNA Barcoding of Mullets (Family Mugilidae) from Pakistan Reveals Surprisingly High Number of Unknown Candidate Species. Diversity 13, 232. https://doi.org/10.3390/d13060232.
Haug, G.H., Tiedemann, R., 1998. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. Nature 393, 673–676. https://doi.org/10.1038/31447.
Haywood, A.M., Dowsett, H.J., Dolan, A.M., 2016. Integrating geological archives and climate models for the mid-Pliocene warm period. Nat. Commun. 7, 10646. https://doi.org/10.1038/ncomms10646.
He, Y., Wang, H., Liu, Z., 2021. Development of the Leeuwin Current on the northwest shelf of Australia through the Pliocene-Pleistocene period. Earth Planet. Sci. Lett. 559, 116767. https://doi.org/10.1016/j.epsl.2021.116767.
Hein, C., Abdel Moniem, H.E., Wagner, H.H., 2021. Can We compare effect size of spatial genetic structure between studies and species using Moran eigenvector maps? Front. Ecol. Evol. 9. https://doi.org/10.3389/fevo.2021.612718.
Heras, S., Roldán, M.I., Castro, M.G., 2009. Molecular phylogeny of Mugilidae fishes revised. Rev. Fish Biol. Fish. 19, 217–231. https://doi.org/10.1007/s11160-008-9100-3.
Heras, S., Maltagliati, F., Fernández, M.V., Roldán, M.I., 2016. Shaken not stirred: a molecular contribution to the systematics of genus Mugil (Teleostei, Mugilidae). Integr. Zool. 11, 263–281. https://doi.org/10.1111/1749-4877.12173.
Herbert, T.D., Peterson, L.C., Lawrence, K.T., Liu, Z., 2010. Tropical ocean temperatures over the past 3.5 million years. Science 328, 1530–1534. https://doi.org/10.1126/science.1185435.
Hickerson, M.J., Meyer, C.P., 2008. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. BMC Evol. Biol. 8, 322. https://doi.org/10.1186/1471-2148-8-322.
Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.
Hubert, N., Meyer, C.P., Bruggemann, H.J., Guérin, F., Komeno, R.J.L., Espiau, B., Causse, R., Williams, J.T., Planes, S., 2012. Cryptic diversity in indo-pacific coral-reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. PLoS One 7, e28987. https://doi.org/10.1371/journal.pone.0028987.
Institute, B., 2019. Picard Toolkit. Broad Institute, Broad Institute, GitHub repository.
Jamandre, B.W., Durand, J.D., Tzeng, W., 2009. Phylogeography of the flathead mullet *Mugil cephalus* in the north-west Pacific as inferred from the mtDNA control region. J. Fish Biol. 75, 393–407. https://doi.org/10.1111/j.1095-8649.2009.02332.x.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
Kaneps, A.G., 1979. Gulf stream: velocity fluctuations during the Late Cenozoic. Science 204, 297–301. https://doi.org/10.1126/science.204.4390.297.
Karas, C., Nürnberg, D., Bahr, A., Groeneveld, J., Herrle, J.O., Tiedemann, R., Demenocal, P.B., 2017. Pliocene oceanic seaways and global climate. Sci. Rep. 7, 39842.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10.1080/01621459.1995.10476572.
Kelley, D., Page, K., Quiroga, D., Salazar, R., 2019. The origins and ecology of the Galapagos Islands. In: In the Footsteps of Darwin: Geoheritage, Geotourism and Conservation in the Galapagos Islands. Springer International Publishing, Cham, pp. 67–93. https://doi.org/10.1007/978-3-030-05915-6_3.
Kolbasova, G., Schmidt-Rhaesa, A., Syomin, V., Bredikhin, D., Morozov, T., Neretina, T., 2023. Cryptic species complex or an incomplete speciation? Phylogeographic analysis reveals an intricate Pleistocene history of *Priapulus caudatus* Lamarck, 1816. Zool. Anz. 302, 113–130. https://doi.org/10.1016/j.jcz.2022.11.013.
Kuo, C.-M., Nash, C.E., Shehadeh, Z.H., 1974. The effects of temperature and photoperiod on ovarian development in captive grey mullet (*Mugil cephalus* L). Aquaculture 3, 25–43. https://doi.org/10.1016/0044-8486(74)90096-9.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species Delimitation using Genome-Wide SNP Data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018.
Leduc, G., Garbe-Schönberg, D., Regenberg, M., Contoux, C., Etourneau, J., Schneider, R., 2014. The late Pliocene Benguela upwelling status revisited by means of multiple temperature proxies. Geochem. Geophys. Geosyst. 15, 475–491. https://doi.org/10.1002/2013GC004940.
Leray, M., Beldade, R., Holbrook, S.J., Schmitt, R.J., Planes, S., Bernardi, G., 2010. Allopatric divergence and speciation in coral reef fish: the three-spot dascyllus,

Dascyllus trimaculatus, species complex. Evolution 64, 1218–1230. https://doi.org/10.1111/j.1558-5646.2009.00917.x.
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.
Ludt, W.B., Rocha, L.A., 2015. Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. J. Biogeogr. 42, 25–38.
Luiz, O.J., Madin, J.S., Robertson, D.R., Rocha, L.A., Wirtz, P., Floeter, S.R., 2012. Ecological traits influencing range expansion across large oceanic dispersal barriers: insights from tropical Atlantic reef fishes. Proc. R. Soc. B 279, 1033–1040. https://doi.org/10.1098/rspb.2011.1525.
Macieira, R.M., Simon, T., Pimentel, C.R., Joyeux, J.-C., 2015. Isolation and speciation of tidepool fishes as a consequence of Quaternary sea-level fluctuations. Environ. Biol. Fishes 98, 385–393. https://doi.org/10.1007/s10641-014-0269-0.
Malinowska, M., Tokarz-Deptuła, B., Deptuła, W., 2017. Butyrophilins: an important new element of resistance. Central Eur. J. Immunol. 42, 399–403. https://doi.org/10.5114/ceji.2017.72806.
Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenberg, M.S., Scribner, K.T., Bonin, A., Fortin, M.J., 2010. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. 19, 3760–3772. https://doi.org/10.1111/j.1365-294X.2010.04717.x.
Marlow, J.R., Lange, C.B., Wefer, G., Rosell-Melé, A., 2000. Upwelling intensification as part of the pliocene-pleistocene climate transition. Science 290, 2288–2291. https://doi.org/10.1126/science.290.5500.2288.
McDowall, R.M., 2003. Hawaiian biogeography and the islands' freshwater fish fauna. J. Biogeogr. 30, 703–710. https://doi.org/10.1046/j.1365-2699.2003.00851.x.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
Mohanty, R.N., Clemens, S.C., Gupta, A.K., 2024. Dynamic shifts in the southern Benguela upwelling system since the latest Miocene. Earth Planet. Sci. Lett. 637, 118729. https://doi.org/10.1016/j.epsl.2024.118729.
Muss, A., Robertson, D.R., Stepien, C.A., Wirtz, P., Bowen, B.W., 2001. Phylogeography of Ophioblennius: the role of ocean currents and geography in reef fish evolution. Evolution 55, 561–572. https://doi.org/10.1111/j.0014-3820.2001.tb00789.x.
Neves, J.M., Almeida, J.P., Sturaro, M.J., Fabré, N.N., Pereira, R.J., Mott, T., 2020. Deep genetic divergence and paraphyly in cryptic species of *Mugil* fishes (Actinopterygii: Mugilidae). Syst. Biodivers. 18, 116–128. https://doi.org/10.1080/14772000.2020.1729892.
Nordlie, F.G., 2015. Adaptation to salinity and osmoregulation in Mugilidae. In: Crosetti, D., Blaber, S.J. (Eds.), Biology, Ecology and Culture of Grey Mullets (mugilidae). CRC Press, Boca Raton, USA, pp. 303–333.
Norris, R.D., Hull, P.M., 2012. The temporal dimension of marine speciation. Evol. Ecol. 26, 393–415. https://doi.org/10.1007/s10682-011-9488-4.
O'Brien, C.L., Foster, G.L., Martínez-Botí, M.A., Abell, R., Rae, J.W.B., Pancost, R.D., 2014. High sea surface temperatures in tropical warm pools during the Pliocene. Nat. Geosci. 7, 606–611. https://doi.org/10.1038/ngeo2194.
O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., De Queiroz, A., Farris, D.W., Norris, R.D., 2016. Formation of the isthmus of panama. Sci. Adv. 2, e1600883. https://doi.org/10.1126/sciadv.1600883.
Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C., Weedon, J., 2022. Vegan: Community Ecology Package.
Ortiz, E.M., 2019. vcf2phylip v2.0: Convert a VCF Matrix Into Several Matrix Formats for Phylogenetic Analysis. doi: 10.5281/zenodo.2540861.
Palmerín-Serrano, P.N., Tavera, J., Espinoza, E., Angulo, A., Martínez-Gómez, J.E., González-Acosta, A.F., Domínguez-Domínguez, O., 2021. Evolutionary history of the reef fish *Anisotremus interruptus* (Perciformes: Haemulidae) throughout the Tropical Eastern Pacific. J. Zool. Syst. Evol. Res. 59, 148–162. https://doi.org/10.1111/jzs.12392.
Palumbi, S.R., 1992. Marine speciation on a small planet. Trends Ecol. Evol. 7, 114–118. https://doi.org/10.1016/0169-5347(92)90144-Z.
Petrick, B., McClymont, E.L., Felder, S., Rueda, G., Leng, M.J., Rosell-Melé, A., 2015. Late Pliocene upwelling in the Southern Benguela region. Palaeogeogr. Palaeoclimatol. Palaeoecol. 429, 62–71. https://doi.org/10.1016/j.palaeo.2015.03.042.
Pinheiro, F., Dantas-Queiroz, M.V., Palma-Silva, C., 2018. Plant species complexes as models to understand speciation and evolution: a review of South American studies. Crit. Rev. Plant Sci. 37, 54–80. https://doi.org/10.1080/07352689.2018.1471565.
Pinsky, M.L., Selden, R.L., Kitchel, Z.J., 2020. Climate-driven shifts in marine species ranges: scaling from organisms to communities. Ann. Rev. Mar. Sci. 12, 153–179. https://doi.org/10.1146/annurev-marine-010419-010916.
Potkamp, G., Fransen, C.H.J.M., 2019. Speciation with gene flow in marine systems. Contrib. Zool. 88, 133–172. https://doi.org/10.1163/18759866-20191344.
Puebla, O., 2009. Ecological speciation in marine v. freshwater fishes. J. Fish Biol. 75, 960–996. https://doi.org/10.1111/j.1095-8649.2009.02358.x.
Pyron, R.A., Burbrink, F.T., 2010. Hard and soft allopatry: physically and ecologically mediated modes of geographic speciation. J. Biogeogr. 37, 2005–2015. https://doi.org/10.1111/j.1365-2699.2010.02336.x.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian Phylogenetics using tracer 1.7. Syst. Biol. 67, 901–904. https://doi.org/10.1093/sysbio/syy032.
Revelle, W.R., 2024. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois.
Riginos, C., Buckley, Y.M., Blomberg, S.P., Treml, E.A., 2014. Dispersal capacity predicts both population genetic structure and species richness in reef fishes. Am. Nat. 184, 52–64. https://doi.org/10.1086/676505.
Rocha, L., Bowen, B., 2008. Speciation in coral-reef fishes. J. Fish Biol. 72, 1101–1121. https://doi.org/10.1111/j.1095-8649.2007.01770.x.
Rocha-Olivares, A., Garber, N., Stuck, K., 2000. High genetic diversity, large inter-oceanic divergence and historical demography of the striped mullet. J. Fish Biol. 57, 1134–1149. https://doi.org/10.1111/j.1095-8649.2000.tb00476.x.
Rocha-Olivares, A., Garber, N.M., Garber, A.F., Stuck, K.C., 2005. Structure of the mitochondrial control region and flanking tRNA genes of *Mugil cephalus*. Hidrobiologica 15, 139–149.
Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L., Schefuss, E., 2011. Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. Paleoceanography 26. https://doi.org/10.1029/2010PA002074.
Rosell-Melé, A., Martínez-Garcia, A., McClymont, E.L., 2014. Persistent warmth across the Benguela upwelling system during the Pliocene epoch. Earth Planet. Sci. Lett. 386, 10–20. https://doi.org/10.1016/j.epsl.2013.10.041.
Rossi, A., Capula, M., Crosetti, D., Campton, D., Sola, L., 1998. Genetic divergence and phylogenetic inferences in five species of Mugilidae (Pisces: Perciformes). Mar. Biol. 131, 213–218. https://doi.org/10.1007/s002270050313.
Rubin, G.M., Hogness, D.S., 1975. Effect of heat shock on the synthesis of low molecular weight RNAs in drosophila: accumulation of a novel form of 5S RNA. Cell 6, 207–213. https://doi.org/10.1016/0092-8674(75)90011-2.
Santini, F., May, M.R., Carnevale, G., Moore, B.R., 2015. Bayesian inference of divergence times and feeding evolution in grey mullets (Mugilidae). bioRxiv, 019075. doi: 10.1101/019075.
Shekhar, M.S., Katneni, V.K., Jangam, A.K., Krishnan, K., Prabhudas, S.K., Jani Angel, J. R., Sukumaran, K., Kailasam, M., Jena, J., 2022. First Report of chromosome-level genome assembly for Flathead Grey Mullet, *Mugil cephalus* (Linnaeus, 1758). Front. Genet. 13. https://doi.org/10.3389/fgene.2022.911446.
Shen, K.-N., Jamandre, B.W., Hsu, C.-C., Tzeng, W.-N., Durand, J.-D., 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus*. BMC Evol. Biol. 11, 83. https://doi.org/10.1186/1471-2148-11-83.
Shen, K.-N., Chang, C.-W., Durand, J.-D., 2015. Spawning segregation and philopatry are major prezygotic barriers in sympatric cryptic *Mugil cephalus* species. C. R. Biol. 338, 803–811. https://doi.org/10.1016/j.crvi.2015.07.009.
Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. Bioscience 57, 573–583. https://doi.org/10.1641/b570707.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stange, M., Sánchez-Villagra, M.R., Salzburger, W., Matschiner, M., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea Catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. Syst. Biol. 67, 681–699. https://doi.org/10.1093/sysbio/syy006.
Steeves, T.E., Anderson, D.J., Friesen, V.L., 2005. The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. J. Evol. Biol. 18, 1000–1008. https://doi.org/10.1111/j.1420-9101.2005.00906.x.
Sunday, J.M., Bates, A.E., Dulvy, N.K., 2012. Thermal tolerance and the global redistribution of animals. Nat. Clim. Chang. 2, 686–690. https://doi.org/10.1038/nclimate1539.
Tan, A., Abecasis, G.R., Kang, H.M., 2015. Unified representation of genetic variants. Bioinformatics 31, 2202–2204. https://doi.org/10.1093/bioinformatics/btv112.
Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2015. Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.
Taylor, A.K., Berke, M.A., Castañeda, I.S., Koutsodendris, A., Campos, H., Hall, I.R., Hemming, S.R., LeVay, L.J., Sierra, A.C., O'Connor, K., 2021. Plio-Pleistocene continental hydroclimate and Indian ocean sea surface temperatures at the southeast African margin. Paleoceanogr. Paleoclimatol. 36, 18. https://doi.org/10.1029/2020PA004186.
Teixeira, R., Musick, J.A., 2001. Reproduction and food habits of the lined seahorse, *Hippocampus erectus* (Teleostei: Syngnathidae) of Chesapeake Bay, Virginia. Rev. Bras. Biol. 61, 79–90. https://doi.org/10.1590/S0034-71082001000100011.
Teske, P.R., Sandoval-Castillo, J., Golla, T.R., Emami-Khoyi, A., Tine, M., von der Heyden, S., Beheregaray, L.B., 2019. Thermal selection as a driver of marine ecological speciation. Proc. R. Soc. B 286, 20182023. https://doi.org/10.1098/rspb.2018.2023.
Thomson, J.M., 1997. The Mugilidae of the world. Mem Queensl Mus 41, 457–562.
Turelli, M., Barton, N.H., Coyne, J.A., 2001. Theory and speciation. Trends Ecol. Evol. 16, 330–343. https://doi.org/10.1016/S0169-5347(01)02177-2.
Van der Horst, T., 1981. Spawning time and spawning groungs of mullet with special reference to *Liza dumerili* (Steindachner, 1869). S. Afr. J. Sci. 77, 73–78.
Viet Tran, T.T., Ke Phan, L., Durand, J.-D., 2017. Diversity and distribution of cryptic species within the *Mugil cephalus* species complex in Vietnam. Mitochondrial DNA Part A 28, 493–501. https://doi.org/10.3109/24701394.2016.1143467.
Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27, 1153–1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x.

Walsh, W., Swanson, C., Lee, C.S., Banno, J., Eda, H., 1989. Oxygen consumption by eggs and larvae of striped mullet, *Mugil cephalus*, in relation to development, salinity and temperature. J. Fish Biol. 35, 347–358. https://doi.org/10.1111/j.1095-8649.1989.tb02987.x.

Walsh, W.A., Swanson, C., Lee, C.-S., 1991. Combined effects of temperature and salinity on embryonic development and hatching of striped mullet, *Mugil cephalus*. Aquaculture 97, 281–289. https://doi.org/10.1016/0044-8486(91)90270-H.

White, N.J., Snook, R.R., Eyres, I., 2020. The Past and Future of Experimental Speciation. Trends Ecol. Evol. 35, 10–21. https://doi.org/10.1016/j.tree.2019.08.009.

Whitfield, A.K., 2015. Ecological role of Mugilidae in the coastal zone. In: Crosetti, D., Blaber, S.J. (Eds.), Biology, Ecology and Culture of Grey Mullets (mugilidae). CRC Press, Boca Raton, USA, pp. 324–348.

Whitfield, A.K., Durand, J.-D., 2023. An overview of grey mullet (Mugilidae) global occurrence and species-rich ecoregions, with indications of possible past dispersal routes within the family. J. Fish Biol. https://doi.org/10.1111/jfb.15450.

Whitfield, A., Panfili, J., Durand, J.-D., 2012. A global review of the cosmopolitan flathead mullet *Mugil cephalus* Linnaeus 1758 (Teleostei: Mugilidae), with emphasis on the biology, genetics, ecology and fisheries aspects of this apparent species complex. Rev. Fish Biol. Fish. 22, 641–681. https://doi.org/10.1007/s11160-012-9263-9.

Yannic, G., Hagen, O., Leugger, F., Karger, D.N., Pellissier, L., 2020. Harnessing paleo-environmental modeling and genetic data to predict intraspecific genetic structure. Evol. Appl. 13, 1526–1542. https://doi.org/10.1111/eva.12986.

Zhao, T., Ma, A., Huang, Z., Liu, Z., Sun, Z., Zhu, C., Yang, J., Li, Y., Wang, Q., Qiao, X., Chen, Z., 2021. Transcriptome analysis reveals that high temperatures alter modes of lipid metabolism in juvenile turbot (*Scophthalmus maximus*) liver. Comp. Biochem. Physiol. D: Genomics Proteomics 40, 100887. https://doi.org/10.1016/j.cbd.2021.100887.

Zhou, C.-Q., Ka, W., Yuan, W.-K., Wang, J.-L., 2021. The effect of acute heat stress on the innate immune function of rainbow trout based on the transcriptome. J. Therm. Biol 96, 102834. https://doi.org/10.1016/j.jtherbio.2021.102834.