Demographic history shapes genetic variability in cryptic fish species of high ecologic and economic relevance

Jessika M. M. Neves[1,*], Zachary J. Nolen[2], Nidia N. Fabré[1], Tamí Mott[1,†], Ricardo J. Pereira[2,†,*]

[1]Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, 57072-900, Maceió, Alagoas, Brazil.

[2]Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.

[†]These authors contributed equally to this work

*corresponding authors: jessika.neves@icbs.ufal.br, ricardojn.pereira@gmail.com

Running title: Demographic history of cryptic fish species

ABSTRACT

Human overexploitation of natural resources has placed conservation and management as one of the most pressing challenges in modern societies, particularly regarding highly vulnerable marine ecosystems. Although a large effort has been made to design Marine Protected Areas (MPAs) worldwide, it is still unclear how many species actually exist in these MPAs, what is the genetic connectivity between areas with different protective regimes, and what is their relative genetic diversity. We answer these questions using morphologically cryptic species of the genus *Mugil* that are sympatric in the largest MPA in the Tropical Southwestern marine province. Population structure analyses show the existence of five highly divergent species (F_{ST} > 0.855) and no genetic divergence between two estuaries with different protection status (F_{ST} = 0.005). Sympatric individuals are assigned to single clusters and show strong concordance among hundreds of independent gene trees, consistent with full reproductive isolation and no ancestral nor ongoing hybridization. Differences of genetic diversity within species suggest that effective population sizes differ up to two-fold, probably reflecting differences in the magnitude of population expansions during the evolutionary history of these species, rather than recent impact of fisheries. Together, our results suggest that designing MPAs with areas of integral protection in between areas where fisheries are permitted could be an effective way to manage cryptic species that cannot have species-specific quotas. More generally, this work shows a cost-efficient approach that is transferable to other marine or terrestrial organisms of special concern, helping to implement science-based regulations for management and conservation.

Keywords: Conservation, Evolution, Diversity, Demography, Cryptic Species, Grey Mullets.

1 INTRODUCTION

Human overexploitation of natural resources has placed conservation and management as one of the most pressing challenges in modern societies. The heavy targeting by fisheries is one of the causes of more than 36% of decline in marine species and 81% of freshwater species during the last four decades (WWF, 2016). This has not only caused a direct decrease in fish stocks, but also an imbalance on ecosystem functioning, as fish species play an important role on fluxes of energy and matter between trophic levels and the environment (Holmlund & Hammer, 1999). Thus, understanding the temporal changes on the distribution of fish species, in their population size, and in the genetic connectivity between populations is fundamental for establishing strategies to protect biological diversity and the evolutionary processes that sustain it (Cook & Sgrò, 2019; Crandall, Bininda-Emonds, Mace, & Wayne, 2000; Moritz, 2002).

Addressing these gaps is particularly challenging in morphologically cryptic species because they are difficult to distinguish based on external morphology alone (Bickford et al., 2006), preventing straightforward assessment of biodiversity, and threatening fisheries sustainability (Garcia-Vazquez, Machado-Schiaffino, Campo, & Juanes, 2012; Lyon, Tonkin, Moloney, Todd, & Nicol, 2018). When such assessments are possible, difficulties in species identification further limit the implementation of protective measures by legislators and fishery managers. Studies integrating information from fast evolving mitochondrial genes are revealing an exponentially growing number of cryptic species, including in previously known tropical hotspots of biodiversity (Benzaquem, Oliveira, Da Silva Batista, Zuanon, & Porto, 2015; Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009) and in marine ecosystems (Asgharian, Sahafi, Ardalan, Shekarriz, & Elahi, 2011; Brandão et al., 2016). Although many cryptic species have allopatric distributions, many others are partially or fully sympatric

(McBride, Van Velzen, & Larsen, 2009; Moritz et al., 2018). This suggests that genetic and/or ecological differences maintain species boundaries, and thus that cryptic species play a meaningful role in the ecosystem. What is less clear is whether cryptic species are evolving in complete genetic isolation in sympatry or whether leaky reproductive barriers can lead to interspecific gene flow in parts of their genomes (Wu, 2001), possibly even reverting species formation (Seehausen, van Alphen, & Witte 1997; Taylor et al., 2006).

Despite major advances provided by mitochondrial studies in identifying cryptic species and mapping their distribution (Ward, Hanner, & Hebert, 2009), mitochondrial markers cannot be used to assess demographic processes relevant for conservation (Galtier, Nabholz, Glémin, & Hurst, 2009). Among other constraints, mitochondrial diversity can be biased by processes disproportionally affecting females (Prugnolle & de Meeus, 2002), by natural selection (Ballard & Whitlock, 2004), or by mitochondrial introgression between closely related species (Toews & Brelsford, 2012). It is thus necessary to use hundreds or thousands of independent nuclear markers to properly understand the evolutionary history of cryptic species, and better inform sustainable conservation measures (Allendorf, 2017; Grewe et al., 2015; Grundler, Singhal, Cowan, & Rabosky, 2019). Recent advances in sequencing technology and statistical methods now offer unprecedented opportunities for the field of conservation biology, providing new insights into the relative abundance of species (Hansson & Westerberg, 2008), genetic connectivity (Pedersen et al., 2018), and even in their adaptive genetic variation (Librado et al., 2017). Understanding these processes remains an important task in marine ecosystems where overfishing was found to be correlated with lower genetic diversity, which in turn can affect species' adaptation capacity (Pinsky & Palumbi, 2014).

To understand the evolutionary processes underlying the genetic diversity of cryptic marine fish of high ecological and commercial importance, we focus on a group of sympatric

and species of the genus *Mugil*, commonly known as mullets. These species have a global distribution in tropical, subtropical and temperate waters (González-Castro & Ghasemzadeh, 2016), and are heavily targeted by traditional and industrial fisheries (Pacheco-Almanzar, Ramírez-Saad, Velázquez-Aragón, Serrato, & Ibáñez, 2017; Whitfield, Panfili, & Durand, 2012), reaching about 140k tons in 2013 worldwide (Crosetti, 2016). *Mugil* species live in fresh and brackish waters during most of their life cycle, migrating to the sea to reproduce (Nordlie, 2016). Thus, they play a fundamental role in transferring energy between estuaries and coastal systems (Lebreton, Richard, Parlier, Guillou, & Blanchard, 2011), helping in the maintenance of biological productivity and, consequently, the yield of other fisheries. Several *Mugil* species occur sympatrically in the Coral Coast, the largest Marine Protected Area (MPA) of the Tropical Southwestern marine province (Souza, Batista, & Fabré, 2012). Although sympatric species from this MPA are highly divergent (between 29 and 6 mya; Neves, Almeida, et al., 2020) they are extremely challenging to identify based on external morphology alone (Neves, Perez, Fabré, Pereira, & Mott, *in review*), leading to strong disagreements among biologists regarding the number and distribution of species, and leading to difficulty in establishing species-specific regulations by legislators and fisheries. Even though the Coral Coast MPA is organized by zones with different management actions (ICMBio, 2013), *Mugil* species are targeted by traditional fisheries throughout the MPA without species-specific quotas. Nevertheless, data from the most commercially valuable species (*M. curema* and *M. liza*) have shown decreases on census sizes in the last 25 years (Mendonça & Bonfante, 2011; Sant'Ana et al., 2017; Vieira, Román-Robles, Rodrigues, Ramos, & dos Santos, 2019). It is yet unclear whether current patterns of genetic diversity reflect such recent human driven dynamics or older demographic history that could differ among cryptic species.

Here, we use double-digest restriction of genomic DNA associated with high throughput sequencing to genotype thousands of markers across nominal species of *Mugil* that are sympatric in the Coral Coast MPA. We use population structure analyses to test how many species exist in the MPA and whether there are barriers to gene flow between two estuaries with different levels of protection. We use phylogenomic methods to estimate a species tree and infer the degree of incomplete lineage sorting among species. Finally, we use demographic analyses to estimate relative effective population size of each species and test for demographic changes during their evolution. Our results provide new insights into the evolution of these cryptic species and provide guidelines for sustainable management.

2 MATERIALS AND METHODS

2.1 Sampling

2.1.1 Specimen collection

We used 94 muscle tissue samples of *Mugil* species that occur sympatrically in the Coral Coast MPA (Fig. 1): 14 *Mugil liza,* 16 *Mugil brevirostris*, 17 *Mugil rubrioculus*, 17 *Mugil curema* and 30 *Mugil curvidens*, all assigned based on a diagnostic COI barcoding gene (Table S1). Importantly, a previous study integrating mitochondrial and morphometric data has shown some disagreement between these alternative categorization methods, suggesting either low informativity of morphologic traits used by taxonomic keys, or interspecific hybridization (Neves, Perez, et a*l., in re*view). This is the case of two individuals included in this present study that were mitochondrially classified as *M. curema* but that have the morphological traits of *M. incilis.*

While most specimens were collected in a partial reserve of the MPA where traditional fisheries exploit *Mugil* without restrictions (Santo Antonio estuary), 15 of the 30 individuals of *Mugil curvidens* were collected in an estuary 38km away where fisheries have some restrictions to protect manatees populations (Manguaba estuary; Fig. S1). We include this population to test if populations from different estuaries function as a single panmitic population, or if some barrier to gene flow exists between locations.

2.1.2 Genomic sequencing

Tissue samples were sent to DArT™ (Diversity Array Technology), who performed DNA extraction, tested two combinations of enzymes (PstI/HpaII and PstI/SphI), performed high-throughput sequencing, assembled the loci, and called genotypes for the 94 individuals.

Based on a subset of 8 samples from 4 species, we chose to genotype all samples for the enzymes PstI/SphI, since this combination showed the highest reproducibility (Table S2).

Each individual is characterized by an array of SNPs, where 0 is homozygous for the major allele, 1 is heterozygous, 2 is homozygous for the minor allele, and - represents missing data. Each SNPs is characterized by: reproducibility (proportion of technical replicate assay pairs for which the marker score is consistent), call rate (the proportion of individuals scored for that locus), and polymorphism information content (PIC: index for evaluating the informative extent of a SNP marker, varying between zero for no allelic variation and 1.0 for maximum allele variation).

Because focal species have diverged between ~29 to ~6 million years ago (Neves, Almeida, et al., 2020) and some have diverged in chromosome number and structure (Galetti Jr., Aguilar, & Molina, 2000; Nirchio, Cipriano, Cestari, & Fenocchio, 2005; Nirchio et al., 2017; Rossi, Gornung, Sola, & Nirchio, 2005), it is possible that the restriction enzymes will not cut the same genomic regions across species. We test for biases on the distribution of missing data by plotting the missing data per individual and the call rate per species, using the package dartR (Gruber, Unmack, Berry, & Georges, 2018) in R software.

2.1.3 Data filtering

We used the package dartR (Gruber et al., 2018) to filter the data and to produce the input files for all downstream analyses. We retained the SNPs with the following criteria: 1) SNPs with reproducibility above 97%, to avoid genotyping error; 2) one SNP per locus favoring higher PIC values, to assure statistical independence among SNPs; 3) SNPs that are in Hardy-Weinberg equilibrium, to satisfy a central assumption of population genetic

163 methods; and 4) loci with trimmed sequence tags that are distinct enough to avoid paralogous
164 loci.
165 Because missing data was not equally distributed among species (Fig. S2), for
166 performing comparative analyses across species we built data sets with three stringency levels
167 for the 6 *Mugil* species (hereafter "6sp"): "0MD", the most stringent filtering without missing
168 data using a call rate of 100%; "20MD", a conservative filtering allowing for a maximum of
169 20% missing data using a call rate of 80%; and "40MD", a less stringent filtering allowing for
170 a maximum of 40% missing data using a call rate of 60%. Because *M. liza* contained most of
171 the missing data (Fig. S2), we repeated this process excluding *M. liza* ("5sp"), without
172 missing data. This resulted in 4 data sets for studies of genetic variability between species:
173 6sp_0MD, 6sp_20MD, 6sp_40MD, 5sp_0MD.
174 For studies of genetic variability within species, we have produced datasets following
175 the criteria described above for no missing data. This resulted in 5 species specific datasets:
176 liza_0MD, brevirostris_0MD, rubrioculus_0MD, curema_0MD, and curvidens_0MD. The
177 individuals of *M. incilis* were found to belong to *M. curema* and therefore were included in
178 that species dataset (see Results; Fig. 2).
179 For analyses that also require invariable sites ("inva"), we produced a data set including
180 5 species supported by our genetic data, where all variable and invariable sites are present
181 (skipping criterium 2 above), and randomly assigning the heterozygous sites; 0MD_inva.
182 All filtered data sets and respective analyses are summarized in Table S3.
183
184 **2.2 Analyses**
185 **2.2.1 Population structure**

In order to assess how many evolutionary lineages compose our sampling, we performed two analyses of population structure using the 4 comparative data sets, without *a priori* information on grouping of the individuals.

First, we performed a Principal Coordinates Analysis (PCoA) to visualize how the genetic variance is distributed among samples based on presence or absence of alleles, using the package dartR (Gruber et al., 2018) in the R software. Second, we estimated the number of genetic clusters and tested if there is ongoing hybridization between them, based on Hardy-Weinberg equilibrium. We used STRUCTURE v.2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to assign individuals to one or more K ancestral populations, varying K between 1 and 10, with 5 replicates for each K. We considered 10k iterations as burn-in, 10k MCMC steps, independent allelic frequencies, and no prior on the assignment of individuals. We chose the most likely K based on log-likelihood values (Pritchard et al., 2000). The graphic output was built using Clumpak (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). We repeated this second test for the data set with *Mugil curvidens* (curvidens_0MD) to test whether there is population structure between the two estuaries.

For both analyses, we expect that individuals will be assigned either to 6 clusters, as suggested by the taxonomic key based on external morphology, or in 5 clusters, as suggested by the mitochondrial clades (Table S1; Neves, Perez, et a*l., in re*view). If species hybridize in sympatry, we expect that individuals sampled in Santo Antonio estuary will be assigned to more than one cluster.

2.2.2 Phylogenetic relationships

We evaluated phylogenetic relationships among species based on our genomic data, both using phylogenomic and population genomic methods, that differ in their assumptions.

First, we built a Maximum Likelihood (ML) phylogenetic tree describing the relationships between all the 94 individuals, without an *a priori* classification of these into species. We used the comparative data sets of the 6 species (6sp_0MD, 6sp_20MD, and 6sp_40MD) to produce a fasta alignment containing a concatenation of all loci, with a random allele in heterozygous sites, using dartR. We used RAxML v.8 (Stamatakis, 2014) through CIPRES gateway (Miller, Pfeiffer, & Schwartz, 2010) to perform 1,000 bootstrap replicates (bs), with GTR+GAMMA model and calculate a consensus tree according to majority rule. We visualized the ML tree using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), rooting the tree with *M. liza* (Neves, Almeida, et al., 2020). We expect individuals of the same species to form monophyletic clades, either with individuals morphologically classified as *M. incilis* nested within the clade of *M. curema*, as found in mitochondrial DNA, or forming their own clade, if previous results are driven by mitochondrial introgression.

Second, we built a coalescent species tree describing the relationships between the species included in our sampling. Using SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012), we estimated a gene tree for each independent SNP and a single species tree allowing for incomplete lineage sorting. We used the most stringent data sets without missing data (6sp_0MD) to generate a nexus alignment where individuals were grouped *a priori* according to the 5 genetic clusters inferred by our population structure analyses (Fig. 2). This analysis was performed in Beauti and Beast software v.2.5.2 (Bouckaert et al., 2019), using the following parameters: 1 million MCMC, recorded every 1,000 generations, and two replicated runs to assess convergence. All the other parameters were set as default. LogCombiner v.2.5.2 was used to combine the two generated trees, and TreeAnnotator v.2.5.2 was used to generate a maximum clade credibility tree and access

posterior probabilities values (PP) of every node. The species tree was visualized in DensiTree v.2.0 and maximum clade credibility tree was visualized in FigTree.

2.2.3 Genetic variability

We estimated genetic variability between and within species. To assure an unbiased comparison among species, we used the comparative data set without missing data (6sp_0MD).

First, we estimated genetic differentiation between all pairwise comparisons of species, using the fixation index F_{ST} (Wright, 1943), as implemented in dartR. We grouped the individuals into species and considered the two sampling locations of *M. curvidens* as separated populations. Second, we estimated several indices of genetic variability within each species, as a proxy for relative differences in effective population size (Ne). Using the dartR package and the dataset 0MD_inva, which allowed more than one SNP per loci, we estimated expected heterozygosity (He) and observed heterozygosity (Ho). Using the DNAsp software v.6.12.03 (Rozas et al., 2017) and the dataset including invariable sites (0MD_inva), we estimated nucleotide diversity (θ and π), number of singletons, and departures from demographic stability with Tajima's D. Because $\pi = 4\ Ne\ \mu$, where μ is the mutation rate per nucleotide site per generation, ratios of diversity indexes calculated from the same loci directly reflect relative Ne. For visualization purposes, all the indices of diversity within species (He, Ho, θ and π) were divided by the diversity of *M. brevirostris*, which showed the lowest diversity, thus reflecting Ne of each species relative to the Ne of *M. brevirostris*.

2.2.4 Demographic history

We inferred the demographic history that better explains the observed patterns of genetic variability within each of the five species specific datasets.

We use the diffusion approximation methods implemented in DaDi (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) to simulate patterns of intraspecific genetic variability under 4 demographic models of increasing complexity: 1) a *neutral* model, assuming a constant population size; 2) a *two-epoch* model, describing an instantaneous change in population size (Ne1) at a certain time (T1); 3) a *bottlegrowth* model, describing an instantaneous size change similar to the previous one, but followed by a period of continuous population size change to the present size (Ne2); and 4) a *three-epoch* model, describing two instantaneous size changes at times T1 and T2. To apply these models to our observed data, we converted our data sets into a VCF file using the radiator package (Gosselin, 2017), and further into the DaDi's format (.snp) using the python script available at https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2DaDi.py. To explore the parameter space properly while assuring convergence, we used the following parameters: initial values of 1 for Ne and of 0.5 for T, lower boundaries for Ne and T set at 0.001, and upper boundaries for Ne set at 100. We used several rounds of parameter optimizations, using the highest likelihood parameters of the previous round as the initial parameters of the next (Portik et al., 2017). We replicated this approach three times per species and model combination to ensure convergence on model selection and parameter estimation. To select the most likely demographic model for each species while accounting with the different number of parameters of the 4 models, we used the Akaike's information criterion (AIC) (Link & Barker, 2006). Standard deviation for each parameter was calculated through Fisher information matrix (Coffman, Hsieh, Gravel, & Gutenkunst, 2016).

3 RESULTS

3.1 Data filtering

Our raw data (Fig. S2) was composed by 55,507 loci of ~69bp with 71,585 SNPs. We observed a large amount of missing data (53%) that is not homogeneously distributed across species. By plotting the call rate of all loci by species (Fig. S3) we consistently observed bi-modal distributions, showing that loci are either always or never called across individuals of the same species, and thus genotyping is highly consistent for individuals of the same species. The most divergent species, *Mugil liza*, shows the largest amount of missing data (85%), followed by *M. curema* (72%). This phylogenetic signal on the distribution of missing data is consistent with the chosen enzymes not cutting highly divergent genomes, rather than a methodological error.

Considering the 6 nominal species (i.e. in the 6sp data sets), we found 7,762 SNPs with 40%MD, 3,508 SNPs with 20%MD and 987 SNPs with 0%MD (Fig. S4). By removing *Mugil liza* (i.e. in the 5sp_0MD), the number of SNPs increases more than two-fold (2,332 SNPs).

When considering each species separately, the number of SNPs is still considerably large, despite allowing for no missing data: 555 SNPs for *M. liza,* 756 for *M. brevirostris*, 1,389 for *M. rubrioculus,* 918 for *M. curema*, and 3,240 for *M. curvidens*.

When considering all sites with 0% MD, including invariable sites and multiple SNPs per locus, our alignment has 82,148 bp and 1,293 SNPs.

3.2 Analyses

3.2.1 Population structure

The first four dimensions of the PCoA using the comparative data set without missing data (6sp_0MD) explained 95.3% of the data variability (Fig. 2A). In general, individuals

303 from the same species strongly overlap with each other and there is no overlap of individuals
304 from different species. The exceptions are the two localities of *M. curvidens*, which
305 completely overlap with each other, and the 2 individuals of *M. incilis*, which overlap with *M.*
306 *curema*. PC1 explains 41% of the variance, reflecting a strong separation of *M. liza* and all
307 other species. PC2 explains 27.5% the variance, with *M. curvidens* and *M. brevirostris* in each
308 extreme. PC3 explains 18% the variance, with *M. curema* and *M. curvidens* in each extreme.
309 PC4 explains 8.8% the variance, further separating *M. brevirostris* and *M. rubrioculus*.
310 Allowing for missing data did not change these results (Fig. S5).
311 In agreement, our STRUCTURE analyses show that the sampled individuals are
312 assigned to five well-differentiated clusters, with the highest likelihood values at K=5 (Fig.
313 2B; Fig. S6). The clusters correspond perfectly with the five mitochondrial lineages, with the
314 2 individuals morphologically classified as *M. incilis* being assigned to the *M. curema* cluster.
315 The two populations of *M. curvidens* are assigned to the same cluster. Our results show no
316 sign of ongoing hybridization, as every individual is entirely assigned to a single cluster.
317 These results remained constant when performing the same analysis without *M. liza* (2,332
318 SNPs, Fig. 7A) or when only considering *M. curvidens* (3,240 SNPs, Fig. S7B).
319
320 **3.2.2 Phylogenetic relationships**
321 Our ML tree (Fig. 3A) recovered five well supported clades (bs= 100), with a consistent
322 topology to what was previously described using a fragment of the mitochondrial COI gene
323 (Neves, Perez, et al., in review). The *M. curvidens* individuals sampled in the Manguaba and
324 Santo Antonio estuaries form a single clade, suggesting no population-level divergence. Also,
325 in agreement with mtDNA, the two individuals morphologically identified as *M. incilis* nested
326 within the clade of *M. curema*, being sister of *M. curvidens*. *Mugil brevirostris* is sister of *M.*

rubrioculus, having the shortest branch lengths between species. The analysis adding missing data showed the same topology (Fig. S8).

Our estimated species tree (Fig. S9) showed the same topology and relative branch lengths of the ML tree, with the two most recent splits between species being well supported (pp > 0.99) and the basal split between *M. brevirostris* + *M. rubrioculus* and *M. curema* + *M. curvidens* being less supported (pp = 0.49). The large majority of SNPs follow the same topology of the species tree. There is no gene tree discordance at the two most recent splits, and there is a small fraction of genes that follow two alternative topologies at basal splits (Fig. 3B).

3.2.3 Genetic variability

Our measures of genetic variability between populations showed extremely high F_{ST} values between the 5 species identified by our population structure analyses (Table 1), showing that most SNPs are fixed between species. The minimum differentiation was observed between *M. brevirostris* and *M. rubrioculus* (F_{ST} = 0.855), and the maximum was between *M. liza* and *M. brevirostris* (F_{ST} = 0.951). Again, the individuals morphologically assigned to *M. incilis* are genetically identical to *M. curema*, with a F_{ST} of 0.025, despite the low sample size of *M. incilis* that could lead to an inflation of estimated F_{ST}. The two sampling localities of *M. curvidens* are also genetically identical, with a F_{ST} of 0.005.

Our estimated levels of genetic variability within species showed consistent patterns across summary statistics (Table S4). *M. brevirostris* consistently showed the lowest values across all diversity indexes, while *M. liza* showed the highest diversity, except for observed heterozygosity (Fig. 4A). All species show a negative Tajima's D, reflecting an excess of singletons across all populations (Fig. 4B). According to Tajima's D estimations, only *M.*

curvidens show a significant deviation from the neutral expectations of a population with a constant Ne ($p < 0.05$).

3.2.4 Demographic history

All species rejected the *neutral* model of constant population size, in favor of two similar models showing a recent range expansion: *two-epoch* and *bottlegrowth* (Fig. 4C). The comparisons of observed and simulated SFS under each demographic scenario (Fig. S10) showed that we observe an excess of singletons and a deficit of low frequency SNPs across all species, when assuming a constant effective population size. The fit of the SFS is resolved by including an instantaneous change in effective population size (Ne1) at time T1 for all species, with an additional continuous change in population size (Ne2) for *M. brevirostris*, *M. curema*, and *M. curvidens*. The AIC weights (Fig. 4C, Table S5) support these two simpler models showing a change in effective population size and show no support for a more complex model of two changes in Ne (*three-epoch* model). The parameter estimations for the model with the best fit always show an increase of current Ne across species relative to ancestral Ne (Table S6); between 2-fold for *M. liza* and 10-fold for *M. brevirostris* and *M. curvidens*.

4 DISCUSSION

Numerous efforts have been made to protect and manage species of high economic and ecologic importance. Yet, designing sustainable and science-based measures for managing wild populations requires a fundamental knowledge on the number of species, their genetic connectivity, relative abundance, and demographic history. This is particularly challenging in morphologically cryptic species such as the *Mugil* fishes, which are heavily targeted by fisheries and play an essential role in the ocean carbon cycling. Here, we use a genomic approach to resolve these evolutionary questions and inform measures for managing these species in the largest Marine Protected Area (MPA) in the Tropical Southwestern Atlantic province.

4.1 Five species of *Mugil* occur in sympatry and show high vagility

The number of species of *Mugil* that are sympatric in the Coral Coast MPA in Brazil has been debated (Fig. 1) (Barletta & Dantas, 2016; Menezes, De Oliveira, & Siccha-Ramirez, 2015). A previous study reported that all specimens morphologically assigned to *M. incilis* had the mitochondrial lineage of *M. curema*, among other discordances (Table S1; Neves, Perez, et a*l., in re*view), suggesting the absence of this species in the area or mitochondrial introgression. Testing the number of independent evolutionary lineages in this highly cryptic species complex required extensive sampling of the nuclear genome, as presented in this study.

Our population structure analyses conclusively show that all sampled specimens belong to five well defined genetic clusters (Fig. S6), irrespective of the amount of missing data allowed (from 2,332 to 987 SNPs; Fig. S7A), and of the assumptions of the clustering method (Fig. 2). These clusters align perfectly with the barcoding gene COI (Table S1), showing that

this gene can reliably distinguish between the five cryptic species that cannot always be differentiated using external morphological data. Importantly, the two individuals morphologically assigned to *M. incilis* are entirely assigned to *M. curema*, ruling out the hypothesis of mitochondrial introgression through hybridization, and conclusively reducing the number of species in the study area to 5: *M. liza*, *M. brevirostris*, *M. rubrioculus*, *M. curema* and *M. curvidens*.

Our results also show that there is no population structure in *M. curvidens* sampled in the two estuaries. This result is robust to a 2-fold increase of the number of SNPs by considering this species alone (Fig. S7B). This shows that the heterogenous habitat in these 38 km are not enough to restrict gene flow and that the areas with different conservation regimes harbor a single meta population of *M. curvidens*. Although this result is perhaps not surprising because adults of *Mugil* species migrate long distances along the shoreline (Livi, Sola, & Crosetti, 2011), and because dispersal also occurs passively through pelagic eggs and larvae (Livi et al., 2011; Whitfield et al., 2012), this hypothesis had not been tested at the genomic level in *Mugil*. Such level of genetic connectivity across heterogeneous habitats contrasts with genomic studies in other marine species (Hauser & Carvalho, 2008; Riginos & Nachman, 2001; Selkoe et al., 2016), and has relevant implications for conservation, as discussed below.

4.2 *Mugil* species show complete reproductive isolation despite full sympatry

According to a mitochondrial study (Neves, Almeida, et al., 2020), these focal species of *Mugil* have diverged between ~29 mya, when *M. liza* split from the remaining species, and ~6 mya, when *M. brevirostris* and *M. rubrioculus* diverged. Despite such a long time since initial divergence, it has not been tested if these lineages are reproductively isolated from each other in areas of sympatry.

Our assignment tests show that every individual is assigned to a single cluster, irrespective of the stringency of the filtering (Fig. 2B, Fig. S7A). Importantly, every individual that showed a mismatch between the morphological and mitochondrial assignments (Table S1) is also assigned to a single cluster (probability of assignment > 0.99). This shows that sympatric species of *Mugil* show no ongoing hybridization in sympatry.

Our maximum likelihood phylogenetic analysis using concatenated loci (Fig. 3A) shows a topology and branch lengths that is in large agreement with that previously estimated from mitochondrial loci (Neves, Almeida, et al., 2020), irrespective of the missing data allowed (Fig. S8). This tree shows that *M. brevirostris* and *M. rubrioculus* are the species pair that diverged most recently, followed by the split between *M. curema* (which includes the "*M. incilis*" individuals discussed above) and *M. curvidens* (which includes both sampling localities). Interestingly, most SNPs are fixed among all species pairwise comparisons, reflected in the extremely high values of fixation indices among the 5 species or clades (Table 1), and consistent with old divergence without gene flow.

By using a coalescent approach that co-estimates independent gene trees of each independent SNP and the species tree that each is embedded in (Razkin et al., 2016), we can infer the number of topologies that disagree with the species tree. Our results show that the vast majority of the gene trees agree with the species tree (Fig. 3B), particularly at the two most recent species splits described above. A small fraction of two alternative topologies reflect disagreements in the deeper nodes of the tree that a previous time calibrated tree have dated between ~29 and ~23 mya (Neves, Almeida, et al., 2020). Given the age of these nodes, no evidence for shared variants between more closely related taxa (Table 1), and no ongoing admixture between any species (Fig. 2B), we interpret this small fraction of disagreement as reflective of incomplete lineage sorting rather than gene flow.

Together, our results show that these five species of *Mugil* have experienced a long period of independent evolutionary history, with reduced levels of incomplete lineage sorting and no ongoing hybridization. Although these species lack any geographic barriers, use the same macrohabitat and are morphologically similar, several studies of *Mugil* species have shed some light on possible reproductive barriers that might contribute to the strong reproductive isolation reported here. Although these species occur mostly in sympatry, previous studies have shown that they differ in spawning time and habitat preferences (Albieri, Araújo, & Uehara, 2010; Mai, dos Santos, Lemos, & Vieira, 2018), possibly constituting a behavioral isolation barrier. Diet studies have shown that sympatric species of *Mugil*, although all limno-benthophagous, can present differential particle size selection to avoid competition (LeLoc'h, Durand, Diop, & Panfili, 2015), possibly constituting an ecological isolating barrier. Furthermore, cytogenetic studies have shown that some of these species differ in the number and arrangement of chromosomes (Harrison, Nirchio, Oliveira, Ron, & Gaviria, 2007; Nirchio et al., 2017; Rossi et al., 2005), possibly constituting a genetic isolating barrier. Although the relative contribution of these barriers has not been tested, our results show that they result in complete reproductive isolation between sympatric *Mugil* species.

4.3 Species differ in their relative abundance and demographic expansion

Information regarding population size change is fundamental for understanding the evolutionary history of a species and to delineate conservation strategies (Dussex & Robertson, 2018; Ramakrishnan, Hadly, & Mountain, 2005). Determining the effective population size (Ne) is a good alternative for inferring census population size because it can be estimated from genetic data and can inform us about the demographic history of the

464 species over their evolutionary time, helping to evaluate the probabilities of extinction or
465 maintenance of the populations (Dussex & Robertson, 2018; Wang, Santiago, & Caballero,
466 2016).

467 Using genomic data we show that, according to every diversity index, *M. brevirostris*
468 has the lowest diversity, reflecting the smallest effective population size of the five species
469 (Table S4). *Mugil liza* has the highest diversity in He, θ and π, corresponding to twice the Ne
470 of *M. brevirostris* (Fig. 4A). Considering these indices, *M. rubrioculus*, *M. curema* and *M.*
471 *curvidens* have intermediate intraspecific diversity, corresponding to ~1.3 to 1.7 times the Ne
472 of *M. brevirostris*.

473 Our demographic analyses are consistent with a strong demographic expansion over the
474 evolutionary history of all species. We found a large number of singletons that are reflected in
475 negative Tajima's D (Fig. 4B). Considering Tajima's D, only *M. curvidens* shows significant
476 departures from the neutral expectation of a constant population size ($p < 0.05$). However, our
477 demographic modelling conclusively shows that none of the five species have had stable
478 effective population sizes and are indeed expanding. According to the weighted AIC values
479 (Fig. 4C), the neutral demographic model was rejected for every species in favor of two
480 similar models allowing a single change of Ne over time: a sudden change of Ne (*two-epoch*
481 model) for *M. liza* and *M. rubrioculus*, and an incremental change of Ne (*bottlegrowth* model)
482 for *M. brevirostris*, *M. curema*, and *M. curvidens*. Although for all species the observed
483 patterns of variability are strongly different from those expected under constant Ne, they fit
484 very similarly to the variability simulated under these two models (Fig. S10), reflecting in
485 similar uncorrected AIC values (Table S5) and thus caution is needed in trying to distinguish
486 between such similar models. Regardless of the model, our demographic analyses always
487 infer a large expansion of Ne relative to the ancestral, between 2.7-fold for *M. liza* to 10-fold

488 increase for *M. curvidens* and *M. brevirostris* (Table S6). The most complex model (three-
489 epoch) did not improve the fitting of the observed data after penalizing for the extra parameter
490 and therefore was not supported for any species.
491 Taken together, our results show that all focal species of *Mugil* have experienced a
492 demographic expansion over their evolutionary history, probably reflecting post-glacial
493 demographic expansions that characterized most terrestrial and marine organisms (e.g.
494 Alexandri et al., 2012; Kuchta & Tan, 2005; Silva, Horne, & Castilho, 2014). Although Ne is
495 not similar across species, this measure reflects effective population size averaged over
496 evolutionary time and thus is affected by historical demography. Our results show that the
497 magnitude of the demographic expansion was likely much stronger in *M. curvidens* and *M.*
498 *brevirostris*, suggesting that historically these species had a much smaller effective population
499 size than they have today, strongly determining their current genetic diversity. Genomic
500 studies in other fish species of high ecologic and economic interest have found similar
501 demographic scenarios of a bottleneck followed by an expansion (e.g. in the Atlantic herring
502 (Barrio et al., 2016)), or a simpler expansion model (e.g. in the North American Lake
503 Whitefish (Rougeux, Bernatchez, & Gagnaire, 2017)). This suggests that current patterns of
504 genetic diversity in wild populations are highly determined by historical changes in Ne
505 associated with the Quaternary ice age, and thus future studies of genetic diversity in
506 protected areas must consider the effect of these important historical events.
507
508 **4.4 Implications for conservation**
509 Human overexploitation of natural resources has placed conservation and management
510 of natural resources as one of the most pressing challenges in modern societies, particularly
511 for highly vulnerable marine organisms. Although governmental institutions and legislators

512 are willing to implement new science-based regulations, we lack a fundamental knowledge on
513 the number of species and an estimation of the relative abundance of each species, both
514 largely unknown in tropical regions, particularly in cryptic species. Mugilidae are the main
515 fish target in tropical artisanal fisheries and these exploit a higher species diversity relative to
516 industrial fisheries (Batista, Fabré, Malhado, & Ladle, 2014). Our results on *Mugil* fishes
517 from the largest MPA in Brazil provide important guidelines for conservation and for the
518 sustainable management of tropical mullets' fisheries.

519 Our results revealed the presence of five *Mugil* species, showing that *M. incilis* is absent
520 from this MPA, in agreement with a previous study based on morphological and
521 mitochondrial data (Menezes et al., 2015). This implies that conservation efforts inside the
522 Costa dos Corais MPA should focus only on five *Mugil* species and that *M. incilis* has a more
523 restricted distribution than previously thought in northern waters.

524 Currently, fisheries in northeastern Brazil report capture of all *Mugil* species under the
525 same category (9,219.5 tons in 2007; IBAMA, 2007) and thus it remains unclear how the five
526 sympatric species of this MPA are differently targeted by fisheries. Since commercial value is
527 largely determined by body size, *M. liza* is expected to be the most targeted species, followed
528 by *M. curema* and *M. rubrioculus.* Although *M. brevirostris* and *M. curvidens* are not
529 particularly targeted, bycatch due gillnets of varied mesh sizes (Torres et al., 2007) leads to an
530 impact in these species. Irrespective of their unknown current census population sizes, our
531 results suggest that these species do not have similar effective population sizes. Surprisingly,
532 the most targeted species, *M. liza,* shows the largest Ne, and the least targeted species, *M.*
533 *brevirostris,* shows the smallest Ne. However, these results must not be interpreted as
534 supportive of unregulated fisheries, as we show that none of these species has had stable
535 population sizes over time, and thus current levels of genetic diversity likely reflect

demographic change over thousands of years, well before human impact in this region. Recent studies using data of fishing effort and catch per unit effort showed a reduction of stocks of *M. liza* and *M. curema* in the South-eastern Brazil due to fisheries exploitation (Mendonça & Bonfante, 2011; Sant'Ana et al., 2017; Vieira et al., 2019). Future studies of fishing effort through landing monitoring and historical data (Verba, Pennino, Coll, & Lopes, 2020) or paleogenomic studies using time series (Dehasque et al., 2020; Fages et al., 2019) may infer how such populations have responded to the impact of fishing at a more proximal temporal scale. Such studies are particularly challenging in morphologically cryptic species such as *Mugil* sp., yet our finding of a large agreement between species assignment using thousands of genomic markers and using a 569 bp fragment of a COI gene show that this marker provides a cost-effective tool for future monitoring studies that can be scaled up to monitor fisheries bycatch.

Lack of population structure between the two estuaries suggests that short distance (38km) is not enough to prevent gene flow in *M. curvidens*, despite the ecological heterogeneity of the habitat in between. This species has the most restrictive distribution (Barletta & Dantas, 2016), and presumably the lowest dispersal abilities of our focal species. Thus, finding no population differentiation in this species (F_{ST} = 0.005) is a conservative proxy for the genetic structure expected in the other co-distributed species that have broader distribution and higher dispersal capabilities. This finding implies that designing MPAs with areas of integral reserve interposed by areas where fisheries are allowed could maintain gene flow along the ecosystem and thus maintain stable populations of these species. Taking into consideration the practical difficulties of identifying *Mugil* species at fishing landings and establishing fishing quotas, such a design of MPAs can be an effective way to preserve species that cannot be accurately identified in the landing area without genetic data, and

where the establishment of species-specific quotas is not a viable solution. With this information it is possible to direct a science-based management plan to guarantee the maintenance of *Mugil* and co-distributed species in a sustainable way.

ACKNOWLEDGMENTS

This work is part of the Long Term Ecological Research – Brazil site PELD-CCAL (Projeto Ecológico de Longa Duração – Costa dos Corais, Alagoas) funded by the Brazilian National Council for Scientific and Technological Development – CNPq (#441657/2016-8), the Brazilian Coordination for the Improvement of Higher Education Personnel - PELD/CAPES (#23038.000452/2017-16) and the Research Support Foundation of the State of Alagoas – FAPEAL (#60030.1564/2016). This network was also funded by Bayerisches Hochschulzentrum für Lateinamerika - BAYLAT (#914-20.1.1). JMMN thanks to scholarship provided by Coordination of Improvement of Higher Level Personnel - CAPES (#88887.137815/2017-00 and #88881.189448/2018-01), TM thanks to CNPq fellowship (#309904/2015–3 and # 312291/2018-3) and FAPEAL fellowship (#60030.000406/2017). We thank C. Peart and S. Tusso for the discussion on bioinformatics, and J. Verba and M. Querejeta for comments in an early version of this manuscript. We are very grateful to the fishermen that provided the specimens, and also to the whole team from LABI, LAEPP and LACOM who helped processing these.

REFERENCES

Albieri, R. J., Araújo, F. G., & Uehara, W. (2010). Differences in reproductive strategies between two co-occurring mullets Mugil curema Valenciennes 1836 and Mugil liza Valenciennes 1836 (Mugilidae) in a tropical bay. *Tropical Zoology*, *23*(1), 51–62.

Alexandri, P., Triantafyllidis, A., Papakostas, S., Chatzinikos, E., Platis, P., Papageorgiou, N., … Triantaphyllidis, C. (2012). The Balkans and the colonization of Europe: The post-

586 glacial range expansion of the wild boar, Sus scrofa. *Journal of Biogeography*, *39*(4),
587 713–723. doi: 10.1111/j.1365-2699.2011.02636.x

588 Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to
589 genomes. *Molecular Ecology*, *26*(2), 420–430. doi: 10.1111/mec.13948

590 Asgharian, H., Sahafi, H. H., Ardalan, A. A., Shekarriz, S., & Elahi, E. (2011). Cytochrome c
591 oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf
592 and analysis using meta-data flag several cryptic species. *Molecular Ecology Resources*,
593 *11*(3), 461–472. doi: 10.1111/j.1755-0998.2011.02989.x

594 Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria.
595 *Molecular Ecology*, *13*(4), 729–744. doi: 10.1046/j.1365-294X.2003.02063.x

596 Barletta, M., & Dantas, D. (2016). Biogeography and Distribution of Mugilidae in the
597 Americas. In D. Crosseti & S. J. M. Blaber (Eds.), *Biology, Ecology and Culture of Grey*
598 *Mullets (Mugilidae)* (pp. 42–62). doi: 10.1201/b19927-4

599 Barrio, A. M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., … Andersson,
600 L. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by
601 genome sequencing. *ELife*, *5*, 1–32. doi: 10.7554/eLife.12081

602 Batista, V. S., Fabré, N. N., Malhado, A. C. M., & Ladle, R. J. (2014). Tropical Artisanal
603 Coastal Fisheries: Challenges and Future Directions. *Reviews in Fisheries Science &*
604 *Aquaculture*, *22*(1), 1–15. doi: 10.1080/10641262.2013.822463

605 Benzaquem, D. C., Oliveira, C., Da Silva Batista, J., Zuanon, J., & Porto, J. I. R. (2015).
606 DNA barcoding in pencilfishes (Lebiasinidae: Nannostomus) reveals cryptic diversity
607 across the Brazilian Amazon. *PLoS ONE*, *10*(2), 1–14. doi:
608 10.1371/journal.pone.0112217

609 Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., … Das, I.
610 (2006). Cryptic species as a window on diversity and conservation. *Trends in Ecology*
611 *and Evolution*, *22*(3), 148–155. doi: 10.1016/j.tree.2006.11.004

612 Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina,
613 A., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for
614 Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), 1–28. doi:
615 10.1371/journal.pcbi.1006650

616 Brandão, J. H. S. G., Bitencourt, J. de A., Santos, F. B., Watanabe, L. A., Schneider, H.,
617 Sampaio, I., & Affonso, P. R. A. D. M. (2016). DNA barcoding of coastal ichthyofauna
618 from Bahia, northeastern Brazil, South Atlantic: High efficiency for systematics and
619 identification of cryptic diversity. *Biochemical Systematics and Ecology*, *65*, 214–224.
620 doi: 10.1016/j.bse.2016.02.012

621 Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012).
622 Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a
623 full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. doi:
624 10.1093/molbev/mss086

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. C. (2009). Stability predicts genetic diversity in the Brazilian atlantic forest hotspot. *Science*, *323*(February), 785–789.

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, *33*(2), 591–593. doi: 10.1093/molbev/msv255

Cook, C. N., & Sgrò, C. M. (2019). Poor understanding of evolutionary theory is a barrier to effective conservation management. *Conservation Letters*, *12*(2), 1–10. doi: 10.1111/conl.12619

Crandall, K. A., Bininda-Emonds, O. R. R., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, *15*(7), 290–295. doi: 10.1016/S0169-5347(00)01876-0

Crosetti, D. (2016). Current State of Grey Mullet Fisheries and Culture. In D. Crosseti & S. J. M. Blaber (Eds.), *Biology, Ecology and Culture of Grey Mullets (Mugilidae)* (pp. 398–450). doi: 10.1201/b19927-17

Dehasque, M., Ávila-Arcos, M. C., Díez-del-Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E. D., … Foote, A. D. (2020). Inference of natural selection from ancient DNA. *Evolution Letters*, *4*(2), 94–108. doi: 10.1002/evl3.165

Durand, J. D., & Whitfield, A. K. (2016). Biogeography and Distribution of Mugilidae in the Western, Central and Southern Regions of Africa. In D. Crosseti & S. J. M. Blaber (Eds.), *Biology, Ecology and Culture of Grey Mullets (Mugilidae)* (pp. 102–115). doi: doi:10.1201/b19927-7

Dussex, N., & Robertson, B. C. (2018). Contemporary effective population size and predicted maintenance of genetic diversity in the endangered kea (Nestor notabilis). *New Zealand Journal of Zoology*, *45*(1), 13–28. doi: 10.1080/03014223.2017.1325381

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435. doi: 10.1016/j.cell.2019.03.049

Galetti Jr., P. M., Aguilar, C. T., & Molina, W. F. (2000). An overview of marine fish cytogenetics. *Hydrobiologia*, *420*, 55–62.

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. *Molecular Ecology*, *18*(22), 4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Garcia-Vazquez, E., Machado-Schiaffino, G., Campo, D., & Juanes, F. (2012). Species misidentification in mixed hake fisheries may lead to overexploitation and population bottlenecks. *Fisheries Research*, *114*, 52–55. doi: 10.1016/j.fishres.2011.05.012

González-Castro, M., & Ghasemzadeh, J. (2016). Morphology and Morphometry Based Taxonomy of Mugilidae. In D. Crosseti & S. J. M. Blaber (Eds.), *Biology, Ecology and*

663 *Culture of Grey Mullets (Mugilidae)* (pp. 1–21). doi: 10.1201/b19927-2

664 Gosselin, T. (2017). *Radiator: RADseq Data Exploration, Manipulation and Visualization*
665 *Using R*. Retrieved from https://github.com/thierrygosselin/radiator

666 Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., …
667 Davies, C. R. (2015). Evidence of discrete yellowfin tuna (Thunnus albacares)
668 populations demands rethink of management for this globally important resource.
669 *Scientific Reports*, *5*, 1–9. doi: 10.1038/srep16916

670 Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to
671 facilitate analysis of SNP data generated from reduced representation genome
672 sequencing. *Molecular Ecology Resources*, *18*(3), 691–699. doi: 10.1111/1755-
673 0998.12745

674 Grundler, M. R., Singhal, S., Cowan, M. A., & Rabosky, D. L. (2019). Is genomic diversity a
675 useful proxy for census population size? Evidence from a species-rich community of
676 desert lizards. *Molecular Ecology*, *28*(7), 1664–1674. doi: 10.1111/mec.15042

677 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009).
678 Inferring the joint demographic history of multiple populations from multidimensional
679 SNP frequency data. *PLoS Genetics*, *5*(10), e1000695. doi:
680 10.1371/journal.pgen.1000695

681 Hansson, B., & Westerberg, L. (2008). Heterozygosity-fitness correlations within inbreeding
682 classes: Local or genome-wide effects? *Conservation Genetics*, *9*(1), 73–83. doi:
683 10.1007/s10592-007-9309-z

684 Harrison, I. J., Nirchio, M., Oliveira, C., Ron, E., & Gaviria, J. (2007). A new species of
685 mullet (Teleostei: Mugilidae) from Venezuela, with a discussion on the taxonomy of
686 Mugil gaimardianus. *Journal of Fish Biology*, *71*(SUPPL. A), 76–97. doi:
687 10.1111/j.1095-8649.2007.01520.x

688 Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly
689 hypotheses slain by beautiful facts. *Fish and Fisheries*, *9*(4), 333–362. doi:
690 10.1111/j.1467-2979.2008.00299.x

691 Holmlund, C. M., & Hammer, M. (1999). Ecosystem services generated by fish populations.
692 *Ecological Economics*, *29*(2), 253–268. doi: 10.1016/S0921-8009(99)00015-4

693 IBAMA. (2007). *Estatística da Pesca 2007 Brasil: Grandes Regiões e Unidades da*
694 *Federação*. Brasília.

695 ICMBio. (2013). *Plano de manejo da APA Costa dos Corais*. Retrieved from
696 https://www.icmbio.gov.br/apacostadoscorais/images/stories/plano_de_manejo/PM_AP
697 ACC_2013_JANEIRO.pdf

698 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015).
699 CLUMPAK: a program for identifying clustering modes and packaging population
700 structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. doi:

10.1111/1755-0998.12387

Kuchta, S. R., & Tan, A. N. M. (2005). Isolation by distance and post-glacial range expansion in the rough-skinned newt, Taricha granulosa. *Molecular Ecology*, *14*(1), 225–244. doi: 10.1111/j.1365-294X.2004.02388.x

Lebreton, B., Richard, P., Parlier, E. P., Guillou, G., & Blanchard, G. F. (2011). Trophic ecology of mullets during their spring migration in a European saltmarsh: A stable isotope study. *Estuarine, Coastal and Shelf Science*, *91*(4), 502–510. doi: 10.1016/j.ecss.2010.12.001

LeLoc'h, F., Durand, J. D., Diop, K., & Panfili, J. (2015). Spatio-temporal isotopic signatures (δ13C and δ15N) reveal that two sympatric West African mullet species do not feed on the same basal production sources. *Journal of Fish Biology*, *86*(4), 1444–1453. doi: 10.1111/jfb.12650

Librado, P., Gamba, C., Gaunitz, C., Sarkissian, C. Der, Pruvost, M., Albrechtsen, A., … Orlando, L. (2017). Ancient Genomic Changes Associated With Domestication of the Horse. *Science*, *356*, 442–445.

Link, W. A., & Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, *87*(10), 2626–2635. doi: 10.1890/0012-9658(2006)87[2626:MWATFO]2.0.CO;2

Livi, S., Sola, L., & Crosetti, D. (2011). Phylogeographic relationships among worldwide populations of the cosmopolitan marine species, the striped gray mullet (Mugil cephalus), investigated by partial cytochrome b gene sequences. *Biochemical Systematics and Ecology*, *39*(2), 121–131. doi: 10.1016/j.bse.2011.01.006

Lyon, J. P., Tonkin, Z., Moloney, P. D., Todd, C., & Nicol, S. (2018). Conservation implications of angler misidentification of an endangered fish. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *28*(6), 1396–1402. doi: 10.1002/aqc.2980

Mai, A. C. G., dos Santos, M. L., Lemos, V. M., & Vieira, J. P. (2018). Discrimination of habitat use between two sympatric species of mullets, Mugil curema and Mugil liza (Mugiliformes: Mugilidae) in the rio Tramandaí Estuary, determined by otolith chemistry. *Neotropical Ichthyology*, *16*(2), 1–8. doi: 10.1590/1982-0224-20170045

McBride, C. S., Van Velzen, R., & Larsen, T. B. (2009). Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry. *Molecular Ecology*, *18*(17), 3639–3651. doi: 10.1111/j.1365-294X.2009.04309.x

Mendonça, J., & Bonfante, T. (2011). Assessment and management of white mullet Mugil curema (Valencienne, 1836) (Mugilidae) fisheries of the south coast of São Paulo state, Brazil. *Brazilian Journal of Biology*, *71*(3), 663–672. doi: 10.1590/s1519-69842011000400010

Menezes, N. A., De Oliveira, C., & Siccha-Ramirez, R. (2015). Taxonomic review of the species of Mugil (Teleostei: Perciformes: Mugilidae) from the Atlantic South Caribbean and South America, with integration of morphological, cytogenetic and molecular data.

Zootaxa, *3918*(1), 1–38. doi: 10.11646/zootaxa.3918.1.1

Miller, M. a., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010*. doi: 10.1109/GCE.2010.5676129

Moritz, C. C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, *51*(2), 238–254. doi: 10.1080/10635150252899752

Moritz, C. C., Pratt, R. C., Bank, S., Bourke, G., Bragg, J. G., Doughty, P., … Oliver, P. M. (2018). Cryptic lineage diversity, body size divergence, and sympatry in a species complex of Australian lizards (Gehyra). *Evolution*, *72*(1), 54–66. doi: 10.1111/evo.13380

Neves, J. M. M., Almeida, J. P. F. A., 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). *Systematics and Biodiversity*, *18*(2), 116–128. doi: 10.1080/14772000.2020.1729892

Neves, J. M. M., Perez, A., Fabré, N. N., Pereira, R. J., & Mott, T. (2020). Integrative taxonomy reveals extreme morphologic conservatism in species of Mugil fishes in a protected area. *Journal of Zoological Systematics and Evolutionary Research*.

Nirchio, M., Cipriano, R., Cestari, M., & Fenocchio, A. (2005). Cytogenetical and morphological features reveal significant differences among Venezuelan and Brazilian samples of Mugil curema (Teleostei: Mugilidae). *Neotropical Ichthyology*, *3*(1), 107–110. doi: 10.1590/S1679-62252005000100006

Nirchio, M., Oliveira, C., Siccha-Ramirez, Z. R., de Sene, V. F., Sola, L., Milana, V., & Rossi, A. R. (2017). The mugil curema species complex (pisces, mugilidae): A new karyotype for the pacific white mullet mitochondrial lineage. *Comparative Cytogenetics*, *11*(2), 225–237. doi: 10.3897/CompCytogen.v11i2.11579

Nordlie, F. G. (2016). Adaptation to Salinity and Osmoregulation in Mugilidae. In *Biology, Ecology and Culture of Grey Mullet (Mugilidae)* (pp. 293–323). doi: 10.1201/b19927-14

Pacheco-Almanzar, E., Ramírez-Saad, H., Velázquez-Aragón, J. A., Serrato, A., & Ibáñez, A. L. (2017). Diversity and genetic structure of white mullet populations in the Gulf of Mexico analyzed by microsatellite markers. *Estuarine, Coastal and Shelf Science*, *198*, 249–256. doi: 10.1016/j.ecss.2017.09.015

Pedersen, C. E. T., Albrechtsen, A., Etter, P. D., Johnson, E. A., Orlando, L., Chikhi, L., … Heller, R. (2018). A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology and Evolution*, *2*(3), 491–498. doi: 10.1038/s41559-017-0453-7

Pinsky, M. L., & Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, *23*(1), 29–39. doi: 10.1111/mec.12509

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita,

778 M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical
779 forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263.
780 doi: 10.1111/mec.14266

781 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
782 Multilocus Genotype Data. *Genetics*, *155*, 945–959.

783 Prugnolle, F., & de Meeus, T. (2002). Inferring sex-biased dispersal from population genetic
784 tools: a review. *Heredity*, *88*, 161–165. doi: 10.1038/sj/hdy/6800060

785 Ramakrishnan, U., Hadly, E. A., & Mountain, J. L. (2005). Detecting past population
786 bottlenecks using temporal genetic data. *Molecular Ecology*, *14*(10), 2915–2922. doi:
787 10.1111/j.1365-294X.2005.02586.x

788 Razkin, O., Sonet, G., Breugelmans, K., Madeira, M. J., Gómez-Moliner, B. J., & Backeljau,
789 T. (2016). Species limits, interspecific hybridization and phylogeny in the cryptic land
790 snail complex Pyramidula: The power of RADseq data. *Molecular Phylogenetics and*
791 *Evolution*, *101*, 267–278. doi: 10.1016/j.ympev.2016.05.002

792 Riginos, C., & Nachman, M. W. (2001). Population subdivision in marine environments: the
793 contributions of biogeography, geographical distance and discontinuous habitat to
794 genetic differentiation in a blennioid fish, Axoclinus nigricaudus. *Molecular Ecology*,
795 *10*(6), 1439–1453. doi: 10.1046/j.1365-294X.2001.01294.x

796 Rossi, A. R., Gornung, E., Sola, L., & Nirchio, M. (2005). Comparative molecular
797 cytogenetic analysis of two congeneric species, Mugil curema and M. liza (Pisces,
798 Mugiliformes), characterized by significant karyotype diversity. *Genetica*, *125*(1), 27–
799 32. doi: 10.1007/s10709-005-2668-9

800 Rougeux, C., Bernatchez, L., & Gagnaire, P. A. (2017). Modeling the multiple facets of
801 speciation-with-gene-flow toward inferring the divergence history of lake whitefish
802 species pairs (Coregonus clupeaformis). *Genome Biology and Evolution*, *9*(8), 2057–
803 2074. doi: 10.1093/gbe/evx150

804 Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-
805 Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism
806 analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. doi:
807 10.1093/molbev/msx248

808 Sant'Ana, R., Gerhard Kinas, P., Villwock de Miranda, L., Schwingel, P. R., Castello, J. P., &
809 Paes Vieira, J. (2017). Bayesian state-space models with multiple CPUE data: the case of
810 a mullet fishery. *Scientia Marina*, *81*(3), 1–10. doi: 10.3989/scimar.04461.11a

811 Seehausen, O., van Alphen, J. J. M., & Witte, F. (1997). Cichlid Fish Diversity Threatened by
812 Eutrophication That Curbs Sexual Selection. Science 277:1808–1811.

813 Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., …
814 Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied
815 marine connectivity. *Marine Ecology Progress Series*, *554*, 1–19. doi:
816 10.3354/meps11792

817 Silva, G., Horne, J. B., & Castilho, R. (2014). Anchovies go north and west without losing
818 diversity: Post-glacial range expansions in a small pelagic fish. *Journal of Biogeography*,
819 *41*(6), 1171–1182. doi: 10.1111/jbi.12275

820 Souza, C. D. de, Batista, V. da S., & Fabré, N. N. (2012). Caracterização da pesca no extremo
821 sul da área de proteção ambiental Costa dos Corais, Alagoas, Brasil. *Boletim Do Instituto*
822 *de Pesca*, *38*(2), 155–169.

823 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis
824 of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. doi:
825 10.1093/bioinformatics/btu033

826 Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J.
827 L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a
828 three-spined stickleback (Gasterosteus aculeatus) species pair. *Molecular Ecology*, *15*(2),
829 343–355. doi: 10.1111/j.1365-294X.2005.02794.x

830 Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear
831 discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. doi: 10.1111/j.1365-
832 294X.2012.05664.x

833 Torres, C. M., Travassos, P., Figueiredo, M. B., Hazin, F., Campos, D. F., & Andrade, F.
834 (2007). Caracterização da pesca de tainhas no município de Porto de Pedras, estado de
835 Alagoas, Brasil. *Revista Brasileira de Engenharia de Pesca*, *2*, 6–17.

836 Verba, J. T., Pennino, M. G., Coll, M., & Lopes, P. F. M. (2020). Assessing drivers of tropical
837 and subtropical marine fish collapses of Brazilian Exclusive Economic Zone. *Science of*
838 *the Total Environment*, *702*, 134940. doi: 10.1016/j.scitotenv.2019.134940

839 Vieira, J., Román-Robles, V., Rodrigues, F., Ramos, L., & dos Santos, M. L. (2019). Long-
840 Term Spatiotemporal Variation in the Juvenile Fish Assemblage of the Tramandaí River
841 Estuary (29°S) and Adjacent Coast in Southern Brazil. *Frontiers in Marine Science*,
842 *6*(269). doi: 10.3389/fmars.2019.00269

843 Wang, J., Santiago, E., & Caballero, A. (2016). Prediction and estimation of effective
844 population size. *Heredity*, *117*(4), 193–206. doi: 10.1038/hdy.2016.43

845 Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes,
846 FISH-BOL. *Journal of Fish Biology*, *74*(2), 329–356. doi: 10.1111/j.1095-
847 8649.2008.02080.x

848 Whitfield, a. K., Panfili, J., & Durand, J. D. (2012). A global review of the cosmopolitan
849 flathead mullet Mugil cephalus Linnaeus 1758 (Teleostei: Mugilidae), with emphasis on
850 the biology, genetics, ecology and fisheries aspects of this apparent species complex.
851 *Reviews in Fish Biology and Fisheries*, *22*(3), 641–681. doi: 10.1007/s11160-012-9263-
852 9

853 Wright, S. (1943). Isolation by distance. *Genetics*, *28*, 114–138. doi:
854 10.1109/IGARSS.2006.90

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865. doi: 10.1046/j.1420-9101.2001.00335.x

WWF. (2016). *Living Planet Report 2016. Risk and resilience in a new era.* Gland, Switzerland.

Table 1. Genetic differentiation (F_{ST}) of *Mugil* species estimated with 6sp_0MD (987 SNPs).

	M. liza	***M. brevirostris***	***M. rubrioculus***	***M. incilis***	***M. curema***	***M. curvidens*_SA**
M. brevirostris	0.951					
M. rubrioculus	0.94	0.855				
M. incilis	0.936	0.939	0.907			
M. curema	0.944	0.933	0.909	0.025		
***M. curvidens*_SA**	0.941	0.922	0.891	0.891	0.89	
***M. curvidens*_MB**	0.947	0.931	0.902	0.911	0.902	0.005

SA – individuals sampled at Santo Antonio estuary;
MB – individuals sampled at Manguaba estuary.

DATA ACCESSIBILITY

All individuals sampled for this study were deposited in the ichthyology collection of the Alagoas university; Voucher number, sampling locations, morphological data and COI Genbank accession numbers are listed in Table S1. All the raw genomic data, filtered data sets, and infiles for all the analyses were deposited in Dryad doi:10.5061/dryad.ngf1vhhr6.

AUTHOR CONTRIBUTION

The project was conceived by TM and RJP. The data was produced by JMMN and analyzed by JMMN and ZJL. The results were interpreted by JMMN, ZJN, TM, and RJP. The first version of the manuscript was drafted by JMMN and RJP, and all remaining coauthors contributed to the final version.

FIGURE LEGENDS

Figure 1. Distribution of the six species of *Mugil* fishes that are potentially sympatric at the Coral Coast MPA (following Barletta & Dantas (2016) and Durand & Whitfield (2016)).

Dashed pink line indicates the area where the occurrence of *M. incilis* has been debated (Menezes et al., 2015). Dashed black line indicates that individuals of all species except *M. brevirostris* may occur. (*M. incilis* photo by A. Carvalho).

Figure 2. Population structure analyses performed with 94 *Mugil* individuals and the dataset with 0% missing data (987 SNPs). A) PCoA analysis; B) Structure analysis. In both analyses, individuals morphologically identified as *Mugil incilis* belong to the same evolutionary lineages as *M. curema* individuals, and there is no sign of population genetic structure between *M. curvidens* sampled at the two estuaries (SA – Santo Antonio; MB – Manguaba).

Figure 3. Phylogenetic history of *Mugil* species. A) Maximum Likelihood phylogeny; numbers on nodes represent bootstrap support. B) Coalescent species tree; cloudogram representing the individual trees of 987 independent SNPs with colors related to alternative topologies; the rare cases of deep coalescence were cut off.

Figure 4. Demographic history of the 5 *Mugil* species. A) Relative effective population size according to diversity indexes relative to *M. brevirostris*: expected heterozygosity (He), observed heterozygosity (Ho), θ and π. B) Departures from a demographically stable population according to demographic indexes: singletons and Tajima´s D. C) Weighted support for alternative demographic models according to Akaike's information criterion (AIC).

SUPPLEMENTARY FIGURES

Figure S1. Sampling sites (red dots). Tropical Southwestern Atlantic marine province is highlighted in inset map.

Figure S2. Raw data. Lines represent the 94 *Mugil* individuals, and columns represent the 71,595 SNPs before filtering. SA – individuals sampled at Santo Antonio river; MB – individuals sampled at Manguaba river.

Figure S3. Histogram of call rate of all loci in each *Mugil* species. The bi-modal distributions show that loci are either always or never called across all individuals of the same species.

Figure S4. SNPs after filtering considering three levels of maximum missing data (MD) per locus. From the top to the bottom: 40% missing data resulted in 7,762 SNPs, 20% missing data resulted in 3,508 SNPs, and 0% missing data resulted in 987 SNPs. SA – individuals sampled at Santo Antonio river; MB – individuals sampled at Manguaba river.

Figure S5. PCoA analysis considering a maximum of 20 and 40% of missing data.

Figure S6. Log-likelihood plot of Structure analysis using the dataset with 0% missing data (987 SNPs). The Log-likelihoods were highly consistent across all replicates, showing a peak at K=5.

Figure S7. Structure analyses considering two partitions of the data. A) Analysis performed with 80 *Mugil* individuals (excluding *M. liza* lineage) and a dataset with 0% of missing data (2,332 SNPs). It is evidenced the presence of four homogeneous lineages, where the individuals morphologically identified as *M. incilis* belong to the same evolutionary lineages as *M. curema* individuals. B) Analysis performed only with 30 individuals *M. curvidens* individuals sampled at two different rivers (Santo Antonio, SA, and Manguaba, MB) considering 0% of missing data (3,240 SNPs). There is no sign of population genetic structure between the two rivers at K=2.

Figure S8. Maximum likelihood phylogenetic tree performed with 1,000 bootstrap replicates (bs) and GTRGAMMA model. The majority consensus tree is represented for a dataset considering a maximim of 20% (left) and 40% of missing data (right). Numbers near the nodes represent node support bs.

Figure S9. Bayesian phylogenetic analysis performed with the dataset with 0% missing data (987 SNPs), using SNAPP. The cloudogram on the left represents the topologies estimated for all independent SNPS, with the most common topology in blue, and alternative topologies in green and red. The maximum credibility tree on the right shows a consensus species tree, showing the posterior probabilities values at the nodes.

Figure S10. Observed (blue) and simulated (red) site frequency spectra for each species, considering 4 demographic models.