Phylogenomics reveals extensive introgression and a case of mito-nuclear discordance in the killifish genus *Kryptolebias*

Waldir M. Berbel-Filho[1,2]*, George Pacheco[3], Andrey Tatarenkov[4], Mateus G. Lira[6], Carlos Garcia de Leaniz[2], Carlos M. Rodríguez López[7], Sergio M. Q. Lima[6] and Sofia Consuegra[2]

[1] Department of Biology, University of Oklahoma, Norman, OK, USA (*present address*)

[2] Department of Biosciences, College of Science, Swansea University, Swansea, UK.

[3] Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, 8600 Silkeborg, Denmark.

[4] Department of Ecology and Evolutionary Biology, University of California, Irvine, USA.

[5] Núcleo de Ecologia Aquática e Pesca da Amazônia, Universidade Federal do Pará, Belém, Brazil.

[6] Laboratório de Ictiologia Sistemática e Evolutiva, Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande, Natal, Brazil.

[7] Environmental Epigenetics and Genetics Group, Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY, USA.

*Corresponding author: waldirmbf@gmail.com

Highlights

- A genomic-based phylogeny is presented for the killifish genus *Kryptolebias,* a genus with a unique diversity of mating systems (e.g., self-fertilization, mixed-mating, outcrossing), covering more species/lineages and genomic loci than previous reconstructions.
- Nuclear phylogeny and introgression analyses revealed the presence of a previously unknown lineage hidden in a case of mito-nuclear discordance with *K. hermaphroditus*.
- The new lineage *Kryptolebias* sp. ‘ESP’ possesses high heterozygosity and extensive history of introgression with *K. hermaphroditus*.

Abstract

Introgression is a widespread evolutionary process leading to phylogenetic inconsistencies among distinct parts of the genomes, particularly between mitochondrial and nuclear-based phylogenetic reconstructions (e.g., mito-nuclear discordances). Here, we used mtDNA and genome-wide nuclear sites to provide the first phylogenomic-based hypothesis on the evolutionary relationships within the killifish genus *Kryptolebias*. In addition, we tested for evidence of past introgression in the genus given the multiple reports of undergoing hybridization between its members. Our mtDNA phylogeny generally agreed with the relationships previously proposed for the genus. However, our reconstruction based on nuclear DNA revealed an unknown lineage - *Kryptolebias* sp. 'ESP' – as the sister group of the self-fertilizing mangrove killifishes, *K. marmoratus and K. hermaphroditus*. All individuals sequenced of *Kryptolebias* sp. 'ESP' had the same mtDNA haplotype commonly observed in *K. hermaphroditus*, demonstrating a clear case of mito-nuclear discordance. Our analysis further confirmed extensive history of introgression between *Kryptolebias* sp. 'ESP' and *K. hermaphroditus*. Population genomics analyses indicate no current gene flow between the two lineages, despite their current sympatry and history of introgression. We also confirmed introgression between other species pairs in the genus that have been recently reported to form hybrid zones. Overall, our study provides a phylogenomic reconstruction covering most of the *Kryptolebias* species, reveals a new lineage hidden in a case of mito-nuclear discordance, and provides evidence of multiple events of ancestral introgression in the genus. These findings underscore the importance of investigating different genomic information in a phylogenetic framework, particularly in taxa where introgression is common as in the sexually diverse mangrove killifishes.

Keywords: Hermaphroditism; Mating systems; Mangrove; Mangrove rivulus, Self-fertilization; Rivulidae.

1. Introduction

Estimating the evolutionary relationships among species is a crucial goal of evolutionary biology. With the unprecedented availability of large numbers of loci brought by the genomics era, it has become increasingly clear that organisms generally have a more complex evolutionary history than previously acknowledged, with biological processes such as recombination, incomplete lineage sorting, introgression, and genome rearrangements (Mallet et al., 2016; Nakhleh, 2013) contributing to different phylogenetic signals among topologies generated from different sets of data for the same group of organisms (Bravo et al., 2019).

Although phylogenetic incongruence may have appeared as a problem at first (Jeffroy et al., 2006; Maddison, 1997), evolutionary biologists now embrace heterogeneity of phylogenetic signals (Bravo et al., 2019; Hahn and Nakhleh, 2016), recognizing that phylogenetic incongruences offer a unique opportunity to investigate the biological phenomena underlying discordance. Among these, reticulate evolution through introgression is today commonly accepted as a widespread evolutionary process contributing to phylogenetic discordance (Bravo et al., 2019; Mallet, 2005; Nakhleh, 2013; Taylor and Larson, 2019). A striking example of how introgression can affect phylogenetic congruence is mito-nuclear discordance, which arises when phylogenetic reconstructions based on mitochondrial or nuclear loci for the same group of organisms substantially differ in their topologies (Bonnet et al., 2017). Although other biological factors (e.g., incomplete lineage sorting, selection on mtDNA, sex-biased dispersal) are also known to generate mito-nuclear discordances, introgression is commonly pointed out as a major source of mito-nuclear phylogenetic incongruences (Toews and Brelsford, 2012).

Differences in mating systems (i.e., defined as the proportion of selfing versus outcrossing in organisms with hermaphrodites (Barrett, 2014)) are expected to influence the extent and

direction of hybridization (Pickup et al., 2019) and in the long-term the degree of introgression. For instance, prior selfing (i.e., eggs are self-fertilized before the window for outcrossing) is expected to provide a strong barrier for hybridization given the limited reproductive opportunity for outcrossing (Brys et al., 2016). The variety of mating systems (e.g., predominantly-selfing, mixed-mating, obligately outcrossing) found in the killifish genus *Kryptolebias* provides an ideal opportunity to investigate: i) how different mating systems may affect the extent of hybridization (Berbel-Filho et al., 2021); and ii) the role that introgression between lineages with different mating systems may have on phylogenetic congruence between mitochondrial and nuclear genomes.

Kryptolebias is a rivulid genus of killifishes (Order Cyprinodontiformes) (Costa, 2011b; Murphy et al., 1999; Thompson et al., 2021), currently composed of seven valid non-seasonal oviparous species (Costa, 2004; Costa, 2011a; Vermeulen and Hrbek, 2005). Previous phylogenetic analyses (based on mtDNA and/or few nuclear genes) proposed two distinct clades within *Kryptolebias*. The 'freshwater' clade composed of narrowly distributed freshwater species living in shallow streams and pools in South America: *K. campelloi* (Costa 1990), *K. sepia* Vermeulen & Hrbek 2005, *K. gracilis* Costa 2007, *K. brasiliensis* (Valenciennes 1821). The second clade, known as the 'mangrove killifishes clade', is composed of three androdioecious species (i.e., populations consisting of males and hermaphrodites) living on mangrove forests along the tropical and subtropical western Atlantic basin: *K. marmoratus* (Poey 1880), *K. hermaphroditus* Costa 2011, and *K. ocellatus* (Hensel 1868)) (Berbel-Filho et al., 2020; Costa et al., 2010; Murphy et al., 1999; Tatarenkov et al., 2017; Tatarenkov et al., 2009; Vermeulen and Hrbek, 2005) (Figure 1). *Kryptolebias marmoratus* and *K. hermaphroditus* are the only two known

vertebrate species capable of self-fertilization (selfing) (Avise and Tatarenkov, 2015; Costa et al., 2010; Tatarenkov et al., 2009).

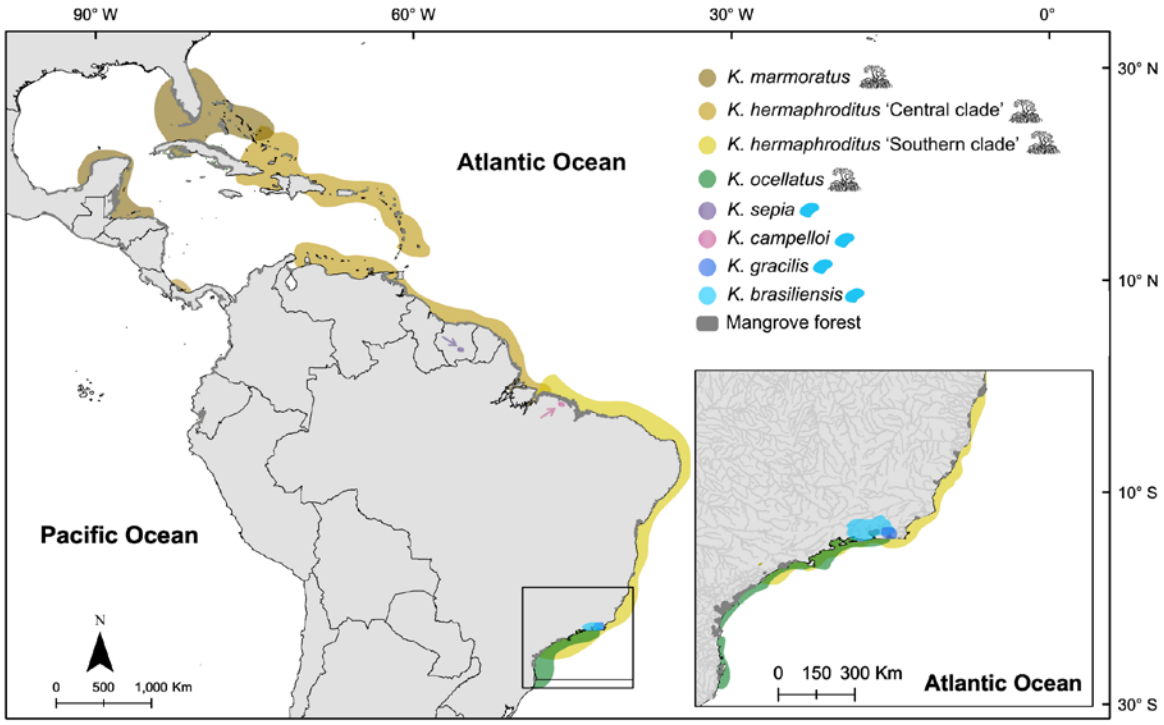


Figure 1. Approximated geographic distribution of known *Kryptolebias* species and lineages. Geographic distributions for the species/lineages were on the literature (Berbel-Filho et al., 2020; Costa, 1990, 2004; Costa, 2007; Costa, 2006, 2016; Lira et al., 2021; Tatarenkov et al., 2017; Vermeulen and Hrbek, 2005) as well as online databases for sampling records (GBIF; www.gbif.org) and museum collections (CRIA - SpeciesLink; http://splink.cria.org.br/). Symbols next to species name represent species inhabiting mangrove (mangrove tree) or freshwater (blue spot) habitats.

In the selfing mangrove killifish species (*K. marmoratus* and *K. hermaphroditus*), most of the eggs laid externally are already fertilized via selfing (Harrington, 1971; Lomax et al., 2017),

leaving a limited window of opportunity for outcrossing either by intra or heterospecific males. Despite this expected limitation, recent studies identified cases of undergoing hybridization involving the selfing *Kryptolebias* species. Tatarenkov et al. (2018) (and later expanded by Tatarenkov et al. (2021)) reported hybridization between highly-selfing lineages of *K. marmoratus* and *K. hermaphroditus* 'Central clade' (a lineage closely related to *K. hermaphroditus* present in the southern portions of the Caribbean, Central America, and northern South America, with taxonomic status still under debate (Lira et al., 2021; Tatarenkov et al., 2017)). In Southeast Brazil, two hybrid zones formed by interbreeding between *K. hermaphroditus* (predominantly-selfing, (Berbel-Filho et al., 2019)) and *K. ocellatus* (exclusively outcrossing, (Berbel-Filho et al., 2020)) represented the first known case of hybridization between species with different mating systems in vertebrates (Berbel-Filho et al., 2021). These unlikely hybridization cases called for further research on the role of past introgression in the diversification of the genus *Kryptolebias*.

Although *K. hermaphroditus* populations are mostly composed of selfing hermaphrodites, outcrossing occasionally happens (Berbel-Filho et al., 2019), most likely involving rare males and hermaphrodites (Furness et al., 2015). Despite historical sampling, particularly in Southeast Brazil (Costa, 2011a), males of *K. hermaphroditus* were only reported recently (Berbel-Filho et al., 2016; Costa, 2016). Costa (2016) reported a relatively high frequency of *K. hermaphroditus* males (e.g., three out of 20 individuals) in a single population in the Brazilian state of Espírito Santo. In rivulids, the male color pattern is the most conspicuous character to diagnose species (Costa 2003). This is particularly true for the mangrove killifish clade, in which hermaphrodites are remarkably similar morphologically (Costa, 2011a, 2016). Therefore, Costa (2016) argued that the color patterns observed (in two different 'morphs') in *K. hermaphroditus* males from this Espírito Santo and other locality in the Rio de Janeiro state represented an important diagnostic trait to the

morphological identification of species in the group (Costa, 2009). However, the pattern of coloration of the *K. hermaphroditus* males reported by Costa (2016) differed substantially from the male color reported for *K. hermaphroditus* in Berbel-Filho et al. (2016) as well as other males reported for the species (Supplementary Figure S1; Amorim et al., 2022). Particularly the 'dark morph' (Costa 2016), which exhibited a dark body flank with broad black margin along the whole caudal fin, while the 'light morph' and the other *K. hermaphroditus* males found in other populations exhibited an orange pattern of pigmentation along its body and often had faded black margins in the caudal fin (Supplementary Figure S1). The relatively high frequency of males, the presence of male two color morphs and the disparity in their coloration, together with the multiple evidence for hybridization in mangrove killifishes prompted further research on the identity and evolutionary history of the unusual Espírito Santo locality in Brazil.

The reports of undergoing hybridization (Berbel-Filho et al., 2021; Tatarenkov et al., 2018, 2021), as well as the recent advances in the knowledge of natural history and distribution of *Kryptolebias* species (Berbel-Filho et al., 2019; Berbel-Filho et al., 2020; Costa, 2016; Guimarães-Costa et al., 2017; Lira et al., 2021; Sarmento-Soares et al., 2014; Tatarenkov et al., 2017) highlight the need of an updated hypothesis regarding the evolutionary relationships within the genus *Kryptolebias*. Using a phylogenomic approach together with a higher number of loci and taxonomic sampling than previous phylogenetic reconstructions, our study aimed to provide the first phylogenomic-based hypothesis for the species relationships in *Kryptolebias*. In addition, we aimed to investigate the hypothesis that reticulation and past introgression events contributed to the diversification of *Kryptolebias* lineages. We reveal a previously unknown lineage/species with strong evidence of ancestral introgression hidden in a case of mito-nuclear discordance. Our

findings highlight how the use of a phylogenomic approach can shed light on the phylogenetic history of groups with common history of interspecific hybridization and challenging taxonomy.

2. Material and Methods

2.1. Mitochondrial DNA dataset

We generated a cytochrome oxidase 1 (*cox1*) dataset of 423 sequences from 50 sampling localities and five out of the seven species (with exception of *K. sepia* and *K. campelloi*) formally described as *Kryptolebias* species. Given the high nuclear divergence found in the Espírito Santo locality in the 'Southern clade' with nuclear data (see results below), we incorporated sequences for 18 individuals from this population generated here. Three additional *cox1* sequences for species in the 'freshwater' clade, namely *K. gracilis* and *K. brasiliensis* were also generated here, while the remaining samples were extracted from previously published data. The samples processed for this study followed primers and PCR protocols described in Tatarenkov et al. (2017). Both forward and reverse DNA strands were sequenced and assembled using Geneious v. 9.1.8 (www.geneious.com). A detailed list of samples used in the mtDNA analyses is presented in Supplementary Table S1.

2.2. Mitochondrial Phylogeny and haplotype network

Our dataset containing 423 Kryptolebias individuals was reduced to unique 49 cox1 haplotypes of 591bp. A cox1 sequence from Atlantirivulus santensis (Köhler 1906) (GenBank accession number GU701924.1) was used as an outgroup for the phylogenetic reconstructions. We identified the best partition scheme and substitution models using ModelFinder in IQ-Tree2 v. 2.1.0 (Kalyaanamoorthy et al., 2017; Minh et al., 2020). We used the suggested partition scheme to infer a maximum likelihood reconstruction and inferred uncertainty with 1000 standard non-parametric bootstrap iterations. Given the evidence of mito-nuclear discordance within the selfing mangrove killifishes clade (see results below), we isolated the 27 haplotypes within this clade and used POPART v. 1.7 (https://popart.otago.ac.nz/) to generate a TCS haplotype network (Clement et al., 2002).

2.3. Nuclear DNA dataset

We combined newly-generated and previously-published data to generate a nuclear DNA dataset across *Kryptolebias* species. First, we sampled populations of *K. ocellatus* (sensu Costa 2011), *K. hermaphroditus* (sensu Costa 2011), *K. brasiliensis* and *K. gracilis* during a field trip in Southeast Brazil between August and September 2017. We collected the fish using hand nets. *Kryptolebias ocellatus* and *K. hermaphroditus* are syntopic in their type-locality (GUA in Supplementary Fig. S3) (Berbel-Filho et al., 2020). Sampling was conducted under license ICMBio/SISBIO 57145-1/2017 and approved by Swansea University Ethics Committee reference SU-Ethics-Student-250717/245.

We build a genotype-by-sequencing library (GBS) for a total of 96 fin clips samples. DNA was extracted the using Qiagen DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. GBS libraries were prepared as described in Kitimu et al.

(2015). In brief, extracted DNA was digested using the restriction enzymes EcoRI and HpaII and ligated to sequencing adapters. Those enzymes were selected based on successful sampling of many restriction sites in *K. hermaphroditus* populations (Berbel-Filho et al., 2019). An aliquot of 200 ng of genomic DNA were digested using a EcoRI (cutsite: GAATTC) and HpaII (cutsite: CCGG). Digested DNA was ligated to individually barcoded adapters with a HpaII cut site overhang and a common EcoRI Y adapter. Ligation products were individually cleaned to remove excess of adapters using Agencourt AMPure XP purification system (#A63880, Beckman Coulter, Brea, CA, USA) at a v/v ratio of 0.85 following the manufacturer's instructions. A single library was produced by pooling 20 ng of digested DNA from each restriction/ligation product and amplified in eight separate PCR reactions which were pooled after amplification, size-selected (range 200–350 bp) and sequenced in a single lane of an Illumina NextSeq500 sequencer.

Out of the original 96 samples included in the library, 61 (36 *K. ocellatus* and 25 *K. hermaphroditus*) had been analysed in Berbel-Filho et al. (2021), while the remaining 35 (13 *K. gracilis*, 11 *K. brasiliensis* and 11 individuals from the Espírito Santo population) were generated specifically for the present study. We extracted 11 *K. hermaphroditus* and 9 *K. ocellatus* GBS samples from Guaratiba (type-locality for both species) previously published in Berbel-Filho et al. (2021). Furthermore, to expand our taxonomic sampling of *Kryptolebias* species, we incorporated raw whole-genome sequencing data for *K. marmoratus* individuals from Florida, Belize, Honduras, and San Salvador Island obtained by Lins et al. (2018). Additional raw whole-genome sequencing data for localities from Guantanamo Bay in Cuba (from Lins et al. (2018)) and Panama (from Choi et al. (2020)), representing individuals from the 'Central clade' lineage were also included. These samples represent a lineage closely related to *K. hermaphroditus* present in the Greater Antilles, Lesser Antilles, southern Central America, and northern portions of South

America (Lira et al., 2021; Tatarenkov et al., 2017). The formal taxonomic status of the 'Central clade' lineage (i.e., either as a distinct species or a differentiated lineage of *K. hermaphroditus*) is still under debate (Lira et al., 2021; Tatarenkov et al., 2021; Tatarenkov et al., 2017). To simultaneously highlight its proximity and divergence with *K. hermaphroditus*, we refer to mtDNA and SNPs data from individuals of the 'Central clade' as "*K. hermaphroditus* 'Central clade'" throughout the manuscript. Data from *K. hermaphroditus* individuals from Southeast Brazil is referred as "*K. hermaphroditus* 'Southern clade'" following the classifications in Lira et al. (2021) and Tatarenkov et al. (2017). Similarly to the mtDNA dataset, we were unable to get hold of samples from the freshwater species *K. sepia* and *K. campelloi*. Those species are only known from very limited geographical distributions in creeks in the Amazon Forest (Costa 1990; Vermeulen and Hrbek, 2005) (Figure 1). No further reports for neither of those species have been found since the sampling reported in the original description (2003 for *K. sepia* in Vermeulen et al., (2006); 1974 for *K. campelloi* in Costa (1990)). We also incorporated raw whole-genome sequencing data for *Nematolebias whitei* (Myers 1942) from Thompson et al., (2022), to be used as an outgroup in the phylogenetic reconstruction based on concatenated nuclear sites.

2.3.1. Nuclear DNA data processing

We used GBSX v1.3 (Herten et al., 2015) to demultiplex the paired-end reads data from the GBS library allowing for one mismatch in the barcodes (-mb 1), no mismatch in the enzyme cut-site (-me 0) and ensuring that no common sequencing adapter was to be removed (-ca false). We then filtered (-qtrim r; -minlength 25) and merged the GBS reads by individuals using BBmap tools (Bushnell, 2014). All samples (both from GBS and whole-genome sequencing) were mapped to the assembled *Kryptolebias hermaphroditus* reference genome (Choi et al., 2020) using either BWA v0.7.17 (for the phylogenetic reconstruction – Dataset I) or Bowtie 2 v2.3.5 (for analyses

within *Kryptolebias* - Datasets II to VIII) using default parameters (Langmead and Salzberg, 2012) and generated filtered and indexed individual BAM files using samtools v 1.10.0 (Li et al., 2009). The different aligners were used due to their different mapping algorithms. While Bowtie2 tends to have faster throughput, it does that at the expense of mapping a lower number of reads when compared to BWA (Hatem et al., 2013). This can dramatically decrease the number of sites shared across samples, especially when a distantly related sample is incorporated in the dataset (i.e., an outgroup). For this reason, we used BWA v.07.17 (Li and Durbin, 2009) as the aligner in the dataset incorporating the *N. whitei* outgroup sample (Dataset I), while the remaining datasets including only *Kryptolebias* samples (Datasets II to VIII) had Bowtie2 v2.3.5 (Langmead and Salzberg, 2012) as an aligner. For the whole-genome samples extracted from the literature, we removed sequencing adapters using AdapterRemoval v. 2.2.2 (Schubert et al., 2016), mapped samples to *K. hermaphroditus* genome using either BWA v0.7.17 (for phylogenetic reconstruction – Dataset I) or Bowtie 2 v2.3.5 (for analysis within *Kryptolebias* – Datasets II to VIII) and filtered and indexed individual BAM files using samtools v1.10.0 within a pipeline in Paleomix v1.3.2 (Schubert et al., 2014). We limited our dataset to samples with ≥ 500k reads. These resulted in a dataset of 48 (out of 61) *Kryptolebias* samples. A detailed list of the samples and sampling localities is provided in Supplementary Table S2.

2.3.2. Variant calling

For all datasets analysed (details provided in Table S4), we inferred genotypes using ANGSD v0.9.32 (Korneliussen et al., 2014). Due to the methylation sensitivity of HpaII, we constrained our variant calling to a maximum of 5% of missing data per loci across all individuals. We used ANGSD with the following parameters: minimum mapping quality (-minMapQ 30), minimum base quality (-minQ 20), missing data (-minInd 95%), Global Depth (-setMaxDepth 600

* number of individuals), minimum genotype posterior probability (-postCutoff 0.95), single and double-tons were accordingly removed based on minimum minor allele frequencies (-MinMaf), anomalous reads (-remove_bads 1; SAM flag above 255), adjusted mapping quality for excessive mismatches (-C 50), performed BAQ computation (-baq 1), minimum coverage for genotype calling (-geno_minDepth 3), use of SAMtools genotype likelihood model (-GL 1), and estimated posterior genotype probabilities assuming a uniform prior (-doPost 2). In addition, we used the ANGSD SNP calling method (-SNP_pval 1e-6), where a Likelihood Ratio Test is used to compare between the null (maf = 0) and alternative (estimated maf) hypotheses by using a X^2 distribution with one degree of freedom.

2.4. Nuclear DNA phylogeny

Our first dataset (Dataset I) consisted of the full set of GBS sites (all sites recovered in our library passing the filtering scheme - both constant and variable) from 48 *Kryptolebias* individuals and *Nematolebias whitei* as an outgroup. This dataset consisted of 174,282 nuclear sites. We ran the ModelFinder algorithm (Kalyaanamoorthy et al., 2017) implemented in IQ-Tree2 v. 2.1.0 (Minh et al., 2020) to infer the best optimal substitution model for the concatenated dataset. Then, we ran IQ-Tree2 to infer the maximum likelihood (ML) tree for the concatenated alignment and to assess the support of internal branches using the Shimodaira-Hasegawa-like procedure support (SH-aLRT) (Guindon et al., 2010), the Bayesian-like transformation of SH-aLRT support (aBayes) (Anisimova et al., 2011), and the ultrafast bootstrap support (UFBoot) (Hoang et al., 2018) with 1,000 replicates.

To visualize the most common phylogenetic signal between *Kryptolebias* species while considering uncertainty that may derive from reticulation, we ran a NeighborNet analysis based on uncorrected p-distances among individuals from Dataset II (115,397 GBS sites across 48

*Kryptolebia*s samples with coverage between 4.77X and 444.95X (mean 153.48X) and missing data per loci ranging from 0% to 1.83% (mean 0.25%); Supplementary Table S3) was conducted in SplitsTree v. 4.18.2 (Huson and Bryant, 2006).

2.5. Phylogenetic networks and ancestral introgression analysis

The bifurcating nature of phylogenetic trees may not accurately describe the phylogenetic history of a particular group, especially when introgression events are common (Olave and Meyer, 2020). Given the reduced representation nature of our GBS library and the high genetic divergence between the freshwater (*K. brasiliensis and K. gracilis*) and the remaining *Kryptolebias* species, we limited our introgression analysis to the species composing the 'mangrove killifishes clade,' namely *K. marmoratus*, *K. hermaphroditus* (Central and Southern clades), *Kryptolebias* sp. 'ESP' and *K. ocellatus*.

We used two approaches to assess reticulation and ancestral introgression events in *Kryptolebias*. First, to evaluate the incidence of reticulation events across the *Kryptolebias* phylogenetic tree, we used the *julia* package PhyloNetworks v. 0.14.2 (Solís-Lemus et al., 2017). This package uses concordant factor tables to infer networks using pseudolikelihood under the multispecies network coalescent model. SNP-based concordant factors were inferred using the program SNPs2CFs v.1.4 (Olave and Meyer, 2020). SNPs2CFs requires phased and unliked SNPs data. We called SNPs for 33 individuals of the mangrove killifishes clade, following the parameters described in the variant calling section. This call resulted in a dataset containing a total of 9,532 SNPs (Dataset III). To phase the data, we limited our dataset to SNPs located only in the 24 chromosomes of the *K. hermaphroditus* reference genome, filtering out all SNPs located in unplaced scaffolds. To minimize linkage amongst SNPs, we further filtered out our dataset to SNPs separated by a minimum distance of five thousand base-pairs, resulting in a dataset containing

5,813 SNPs with an average distance of 110,809 base-pairs among SNPs (Dataset IV). We phased this dataset using Beagle v. 5.2 (Browning et al., 2021) and generated concordance factors using SNPs2CFs. The full number of quartets in our dataset is too large to be processed in PhyloNetworks (Solís-Lemus et al., 2017). Therefore, we limited our sample to 1,000 alleles per species quartet (n.quartets = 1,000 on SNPs2CFs), resulting in a total 5,000 quartets. We estimated phylogenetic networks with a hmax (maximum number of hybridization events) value ranging from zero to seven. Our starting network (hmax=0) was represented by a concatenated ML tree ran on IQ-Tree2 using the full set of GBS loci (both constant and variable) for this dataset (Dataset V – 1,631,872 nuclear sites with no missing data: Supplementary Figure S5). The resulting network for each hmax value was used for every subsequent run. We plotted the log-pseudolikelihood and selected the networks that resulted in substantial pseudolikelihood improvements.

To further evaluate the evidence of ancestral introgression in *Kryptolebias*, we used the software Dsuite v. 0.4 r28 (Malinsky et al., 2021) to calculate Patterson's D statistics (ABBA-BABA test) and f4-ratios (an estimate of admixture fraction). ABBA-BABA rely on comparisons between bi-allelic SNPs for four taxa (e.g., T1, T2, T3, O) which are related to each other by a rooted tree (e.g. (((T1, T2), T3), O)). 'A' and 'B' represents the ancestral and derived alleles, respectively. Under a no gene flow scenario, the patterns of ABBA (sharing of alleles between T2 and T3) and BABA (sharing of alleles between T1 and T3) are expected to occur with equal frequencies, while significant deviation from equal frequencies is consistent with introgression between T3 and either T2 (ABBA) or T1 (BABA). We formally tested for three possible past introgression events within the mangrove killifishes clades in *Kryptolebias* based on either previous or current evidence: i) between *K. marmoratus* and *K. hermaphroditus* 'Central clade' as

suggested by Tatarenkov et al. (2018, 2021) with the following tree topology: (((Kher_South, Kher_Central), Kmar), KspESP)); ii) between *K. hermaphroditus* 'Southern clade' and *Kryptolebias* sp. 'ESP' (see results) with the following tree topology (((Kmar, Kher_South), KspESP), Koce)); and between (iii) *K. hermaphroditus* 'Southern clade' and *K. ocellatus*, given the ongoing hybridization found in Berbel-Filho et al. (2021) with the following tree topology: (((Kher_South, KspESP), Koce), Kbra). 'Kmar', 'Kher_Central', 'Kher_South', 'KspESP', 'Koce', and 'Kbra' refer to *K. marmoratus, K. hermaphroditus* 'Central clade', *K. hermaphroditus* 'Southern clade', *Kryptolebias* sp. 'ESP', *K. ocellatus*, and *K. brasiliensis,* respectively. For the first two introgression tests (i and ii), we used Dataset III while for test iii we called SNPs for a dataset containing one representative per species (for maximize the number of sites given the inclusion of *K. brasiliensis* as an outgroup) of the mangrove killifish clade and an individual of *K. brasiliensis* as an outgroup (Dataset VI – 10,648 SNPs with no missing data).

2.6. Genetic structure of the mangrove killifishes clade

Our results indicated (see below) the existence of a previously unknown lineage of *Kryptolebias* in a single coastal sampling site in of Espírito Santo State in Brazil (referred above as *Kryptolebias* sp. 'ESP') (Supplementary Figure S4). We further explored the nuclear genomic structure *Kryptolebias* sp. 'ESP' in comparison to the other lineages in the mangrove killifishes clade (see Fig. 3) using Dataset III. To estimate individual ancestries, we used ngsAdmix v. 3.2 (Skotte et al., 2013) with K values ranging between 2–10 for 100 replicates using default parameters, except for tolerance for convergence (-tol 1×10^{-6}), log likelihood difference in 50 iterations (-tolLike 50 1×10^{-3}), and a maximum number of EM iterations (-maxiter 10,000). We used StructureSelector (Li and Liu, 2018) to estimate the most likely number of genetic clusters. A pairwise genetic distance matrix between individual's matrix was computed directly from the

genotype likelihoods using ngsDist v1.0.2 (Vieira et al., 2015) and was then used for Multidimensional Scaling (MDS) using the R function *cmdscale*. To calculate heterozygosity, we called the full set of GBS sites (both constant and variable) for a dataset containing all 33 individuals of the mangrove killifish clade (Dataset VII - 863,662 nuclear sites for 33 individuals with 5% of missing data). We used ANGSD to compute the unfolded global estimate of the Site Frequency Spectrum (SFS) using one individual of *K. brasiliensis* as the source of ancestral sequence. Heterozygosity was calculated as the proportion of heterozygous sites by the total number of sites per individual.

2.7. Introgression between Kryptolebias sp. 'ESP' and K. hermaphroditus 'Southern clade'

To gain further insights into the structure of the hybrid zone between *Kryptolebias* sp. 'ESP' lineage and *K. hermaphroditus* 'Southern clade' (as indicated in our results), we called SNPs for a dataset containing only *Kryptolebias* sp. 'ESP' and *K. hermaphroditus* 'Southern clade' individuals (Dataset VIII – 5,688 SNPs for 18 individuals with no missing data). With this dataset, we addressed the patterns of allele distribution (e.g., number of fixed and/or shared alleles) between the two lineages.

3. Results

3.1. Mitochondrial phylogeny

The phylogenetic reconstruction based on 49 unique *cox1* haplotypes extracted from 423 *Kryptolebias* individuals was largely consistent with previously suggested phylogenetic relationships in the genus (Berbel-Filho et al., 2020; Costa, 2004; Costa, 2007; Costa et al., 2010; Kanamori et al., 2016; Murphy et al., 1999; Tatarenkov et al., 2017; Tatarenkov et al., 2009; Vermeulen and Hrbek, 2005) (Fig. 2a). The freshwater species *K. gracilis* and *K. brasiliensis*

371 formed a clade which is sister group of the mangrove killifishes clade, composed of *K. ocellatus*,
372 *K. hermaphroditus and K. marmoratus*. The latter two formed a well-supported clade within the
373 mangrove killifishes clade (the selfing mangrove killifishes). As previously indicated, there were
374 two *K. hermaphroditus* mtDNA clades, one comprising samples from San Salvador Island, the
375 Caribbean and northern portions of South America (the 'Central clade' in Tatarenkov et al. 2017
376 and Lira et al. 2021), and another composed of samples from Northeast and Southeast Brazil (the
377 'Southern clade' in Tatarenkov et al. 2017 and Lira et al. 2021). All the 23 individuals from the
378 Espírito Santo locality exhibited a single *cox1* haplotype (Hap22, Fig 2), which is widespread in
379 many *K. hermaphroditus* populations along approximately 2,600km of the Northeast and
380 Southeast regions of the Brazilian coast (Lira et al., 2021) (Fig. 2b).

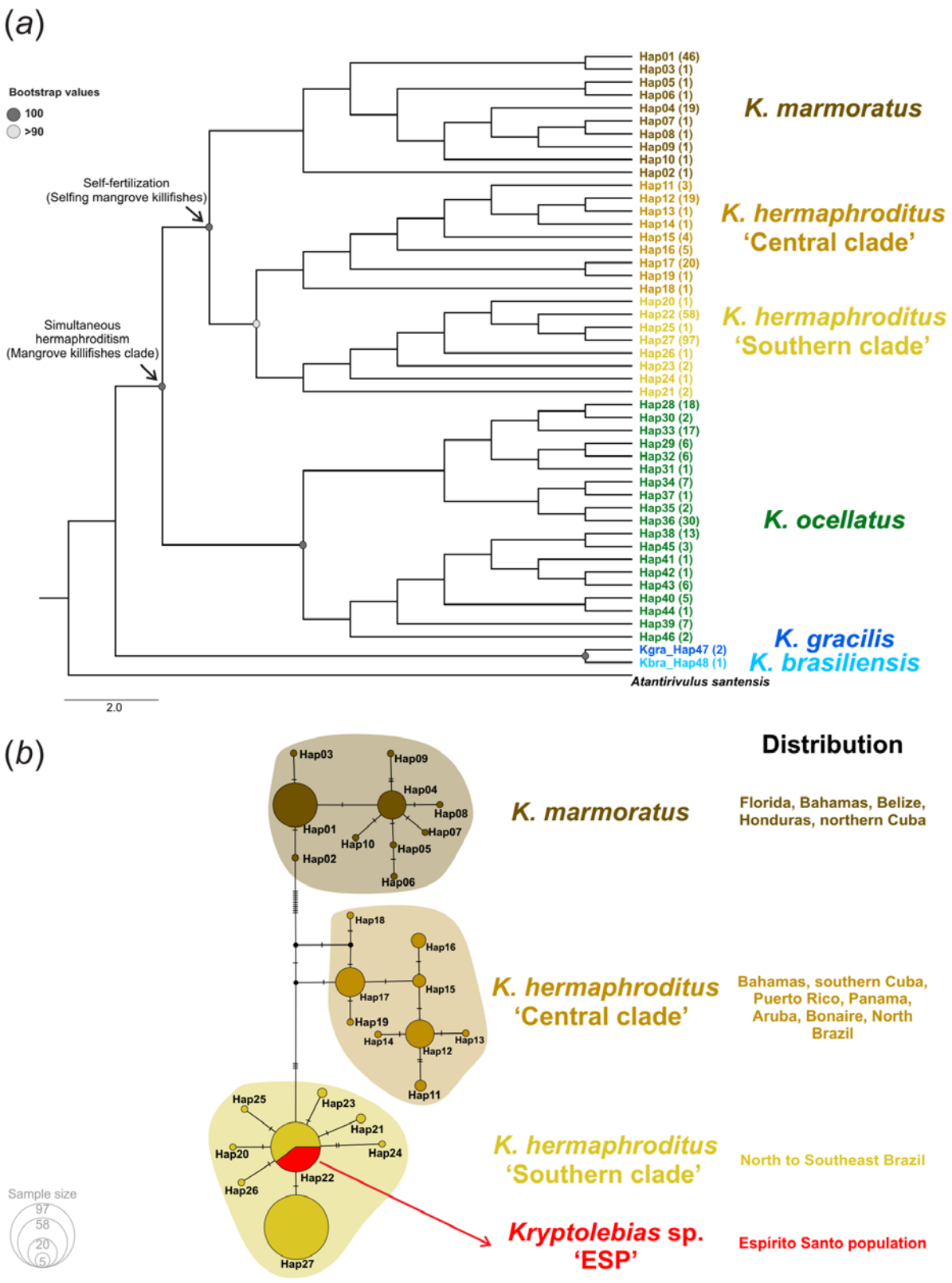
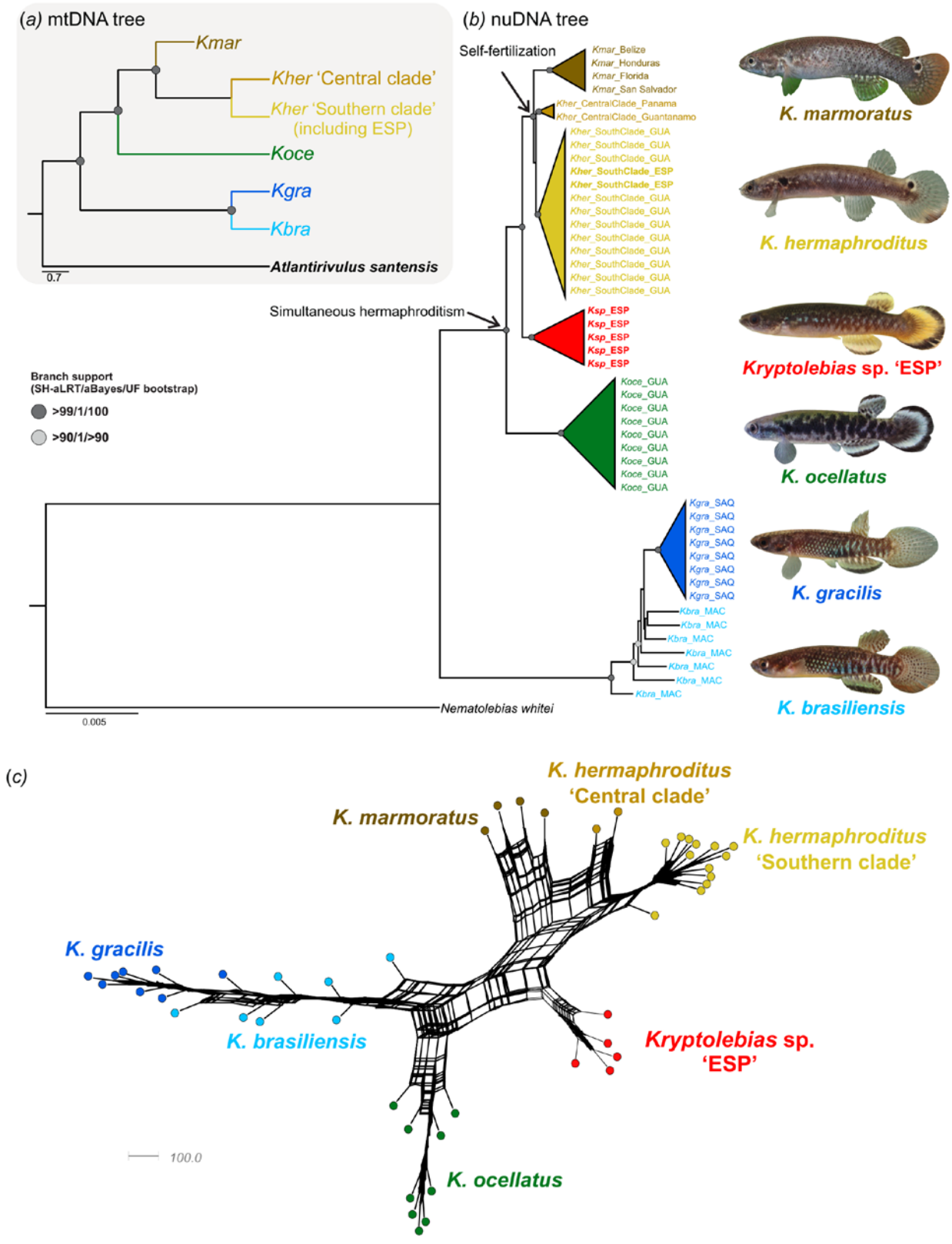

(a)
Bootstrap values
100
>90
Self-fertilization
(Selfing mangrove killifishes)
Simultaneous
hermaphroditism
(Mangrove killifishes clade)
Hap01 (46)
Hap03 (1)
Hap05 (1)
Hap06 (1)
Hap04 (19)
Hap07 (1)
Hap08 (1)
Hap09 (1)
Hap10 (1)
Hap02 (1)
Hap11 (3)
Hap12 (19)
Hap13 (1)
Hap14 (1)
Hap15 (4)
Hap16 (5)
Hap17 (20)
Hap19 (1)
Hap18 (1)
Hap20 (1)
Hap22 (58)
Hap25 (1)
Hap27 (97)
Hap26 (1)
Hap23 (2)
Hap24 (1)
Hap21 (2)
Hap28 (18)
Hap30 (2)
Hap33 (17)
Hap29 (6)
Hap32 (6)
Hap31 (1)
Hap34 (7)
Hap37 (1)
Hap35 (2)
Hap36 (30)
Hap38 (13)
Hap45 (3)
Hap41 (1)
Hap42 (1)
Hap43 (6)
Hap40 (5)
Hap44 (1)
Hap39 (7)
Hap46 (2)
Kgra_Hap47 (2)
Kbra_Hap48 (1)
Atantirivulus santensis
2.0
K. marmoratus
K. hermaphroditus
'Central clade'
K. hermaphroditus
'Southern clade'
K. ocellatus
K. gracilis
K. brasiliensis
(b)
Distribution
Hap03
Hap09
Hap04
Hap08
Hap01
Hap07
Hap10
Hap05
Hap02
Hap06
K. marmoratus
Florida, Bahamas, Belize,
Honduras, northern Cuba
Hap18
Hap16
Hap17
Hap15
Hap19
Hap14
Hap13
Hap12
Hap11
K. hermaphroditus
'Central clade'
Bahamas, southern Cuba,
Puerto Rico, Panama,
Aruba, Bonaire, North
Brazil
Hap25
Hap23
Hap21
Hap20
Hap24
Hap22
Hap26
Hap27
K. hermaphroditus
'Southern clade'
North to Southeast Brazil
Kryptolebias sp.
'ESP'
Espírito Santo population
Sample size
97
58
20
5

Figure 2. Maximum-likelihood reconstruction for 49 unique *cox1* haplotypes extracted from 423 *Kryptolebias* individuals. (*a*) Full tree containing the relationships among the 49 haplotypes with tip labels colored by species/lineages names. Tip labels show haplotypes and number of individuals sequenced (in parenthesis). (*b*) Haplotype network for the 27 *cox1* haplotypes for individuals belonging to the selfing mangrove killifishes clade with their respective distribution. Details for all samples used for these analyses are provided in Table S1.

3.2. Nuclear Phylogeny

Our ML phylogenetic reconstruction based on 174,282 nuclear DNA sites (Dataset I) from 48 *Kryptolebias* individuals was generally concordant with the phylogenetic relationships proposed in the mtDNA tree, with two main exceptions (Fig. 3b). Although the freshwater species *K. gracilis* and *K. brasiliensis* grouped together, samples from the former formed a monophyletic group within a non-monophyletic composed of *K. brasiliensis* samples. The other exception consisted of a previously unknown and well-supported clade containing five individuals from the Espírito Santo locality. This clade (hereafter named as *Kryptolebias* sp. 'ESP') formed a sister clade to the selfing mangrove killifishes, consisting of *K. marmoratus* and *K. hermaphroditus* (both Central and Southern clades). Two additional individuals from the Espírito Santo locality clearly belonged to *K. hermaphroditus* 'Southern clade', suggesting this population consisted of two sympatric species. All 23 individuals sequenced for mtDNA from Espírito Santo locality (including the five *Kryptolebias* sp. 'ESP' individuals) had the same mtDNA haplotype typically observed in *K. hermaphroditus* populations in Northeast and Southeast Brazil (Hap22, Fig. 2), representing a clear case of mito-nuclear discordance in *Kryptolebias*. Our phylogenetic network reconstruction using SplitsTree largely agreed with the lineages found in our mtDNA and ML phylogenetic reconstruction (Figs 2a and 3b). However, it also indicated the highest levels of site tree

405 discordance in *Kryptolebias* are within the mangrove killifish clade (Fig. 3c). This finding suggests
406 events of introgression may have been common during the evolutionary history of this clade.


(a) mtDNA tree
Kmar
Kher 'Central clade'
Kher 'Southern clade' (including ESP)
Koce
Kgra
Kbra
Atlantirivulus santensis
0.7
(b) nuDNA tree
Self-fertilization
Kmar_Belize
Kmar_Honduras
Kmar_Florida
Kmar_San Salvador
Kher_CentralClade_Panama
Kher_CentralClade_Guantanamo
Kher_SouthClade_GUA
Kher_SouthClade_ESP
Simultaneous hermaphroditism
Ksp_ESP
Koce_GUA
Kgra_SAQ
Kbra_MAC
Nematolebias whitei
0.005
Branch support
(SH-aLRT/aBayes/UF bootstrap)
>99/1/100
>90/1/>90
K. marmoratus
K. hermaphroditus
Kryptolebias sp. 'ESP'
K. ocellatus
K. gracilis
K. brasiliensis
(c)
K. hermaphroditus 'Central clade'
K. marmoratus
K. hermaphroditus 'Southern clade'
K. gracilis
K. brasiliensis
Kryptolebias sp. 'ESP'
100.0
K. ocellatus

407

Figure 3. Phylogenetic reconstructions of the genus *Kryptolebias* using IQ-Tree 2 v. 2.0.1. (*a*) Schematic representation of the maximum-likelihood phylogenetic tree based on 49 unique mtDNA *cox1* haplotypes extracted from 423 *Kryptolebia*s individuals. Node circles represent nonparametric bootstrap values = 100. The full mtDNA phylogenetic reconstruction is provided on Figure 2. (*b*) Maximum-likelihood phylogenetic tree based on 174,842 GBS nuclear sites (Dataset I). Node circles represent SH-aLRT (%), aBayes, and ultrafast bootstrap (%) support values, respectively. Only nodes with high support (SH-aLRT ≥90, aBayes = 1, and ultrafast bootstrap > 90) are shown. Branch lengths are shown in substitutions per site. Intraspecific clades were collapsed to facilitate visualization. The full nuclear phylogenetic reconstruction is provided on Supplementary Fig. S4. (*c*) 95% confidence phylogenetic network (Neighbor-Net) constructed using SplitsTree based on all sites from Dataset II. All species, with exception of *K. marmoratus* and *K. hermaphroditus* (represented by hermaphrodites), are represented in the figure by male individuals.

3.3. Phylogenetic networks and ancestral introgression

Our PhyloNetworks analysis indicated that the largest improvement in pseudolikelihood scores across the number of reticulation events evaluated (ranging from 1 to 7) occurred between zero (the original tree) and one reticulation event (Fig. 4a). The network generated with one reticulation indicated ancestral introgression from *K. hermaphroditus* 'Southern clade' and *Kryptolebias* sp. 'ESP', with the latter as hybrid lineage inheriting 40% of its genomic content from the former, despite the fact these two lineages are relatively far from each other in the phylogenetic tree (Fig. 4b).

Introgression events between *K. hermaphroditus* 'Central clade' and *K. marmoratus* (D = 0.22; Z-score = 8.71; F_4 ratio = 0.22; $p < 0.001$) (Fig. 4c), between *K. hermaphroditus* 'Southern

clade' and *Kryptolebias sp.* 'ESP' (D = 0.62; Z-score = 50.15; F_4 ratio = 1.07; p < 0.001) (Fig. 4d) and between *K. hermaphroditus* 'Southern clade' and *K. ocellatus* (D = -0.76; Z-score = 62.77; F_4 ratio = -0.13; p < 0.001) (Fig. 4e) were all confirmed by our ABBA-BABA test, revealing extensive introgression events in several *Kryptolebias* lineages.

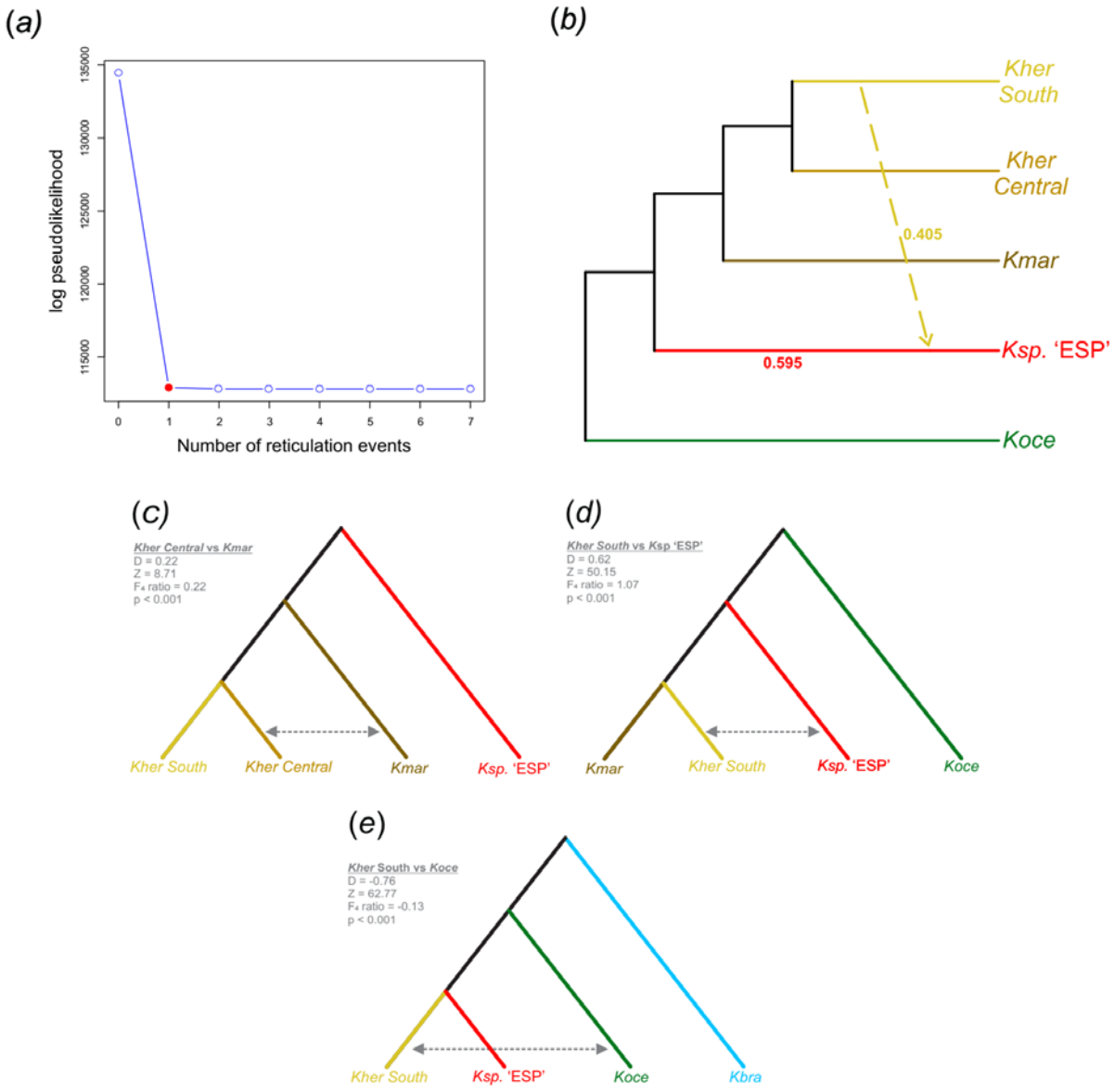


Figure 4. Phylogenetic networks and introgression analysis in the mangrove killifishes clade. Genetic structure analysis for mangrove killifish species based on 9,532 SNPs (Dataset III). (*a*)

Log-pseudolikelihood scores per number of reticulation events tested using PhyloNetworks v. 0.14.2 (Solís-Lemus et al., 2017). The red dot indicates the network chosen based on a large improvement of the pseudolikelihood score. (*b*) Phylogenetic network with one reticulation event. Numbers indicate inheritance proportions in the hybrid lineage. (*c*-*e*) ABBA-BABA results for introgression tests between: (*c*) *K. hermaphroditus* 'Central clade' ('*Kher Central'*) and *K. marmoratus* ('*Kmar*'); (*d*) *K. hermaphroditus* 'Southern clade' ('*Kher South'*) and *Kryptolebias* sp. 'ESP' ('*Ksp*. ESP'); (*e*) *K. hermaphroditus* 'Southern clade' ('*Kher South'*) and *K. ocellatus* ('*Koce*').

3.4. Genetic structure of the mangrove killifishes clade

Admixture analysis indicated the presence of five genetic clusters (Fig 4a), each representing the mangrove killifish clade lineages recovered by the phylogenetic reconstruction based on nuclear sites (Fig. 2b). All four metrics generated by StructureSelector further suggested five as the most likely number of genetic clusters (Supplementary Fig. S7; all clusters shown in Supplementary Fig. S8). As also indicated by our phylogenetic reconstruction, the Espírito Santo population is composed of two highly different lineages at the nuclear genome, with two of the individuals sequenced belonging to *K. hermaphroditus* 'Southern clade', and the remaining five belonging to the previously unknown lineage *Kryptolebias* sp. 'ESP', despite the fact that all those individuals (and other 18 individuals sequenced from the same population) have the same mtDNA haplotype (Hap22 in Fig. 2) commonly found in individuals from the *K. hermaphroditus* 'Southern clade' in Northeast and Southeast Brazil. Our admixture analysis further indicated that these two lineages are quite differentiated from each other, with no evidence of current admixture (or early hybrid generation individuals) between them (Fig. 5a). This result can also be observed in our MDS analysis (Fig. 5b), in which the clusters representing the species (with exception of *K.*

hermaphroditus Central and Southern clades – highlighting the proximity between these two lineages) occupied different portions of the eigenspace, with *Kryptolebias* sp. 'ESP' and *K. hermaphroditus* occupying opposite sides of the first dimension of genetic distance variation . In terms of genetic diversity, *Kryptolebias* sp. 'ESP' individuals had in average 4.25x higher proportion of heterozygous sites (average: 0.45) than the outcrossing species *K. ocellatus* (average: 0.10), and 11.7x higher than the selfing (and sympatric) *K. hermaphroditus* 'Southern clade' (average: 0.04) (Fig. 5c). Possibly due to long-term generation of selfing and/or low-coverage nature of sequencing, *K. marmoratus* and *K. hermaphroditus* 'Central clade' individuals had an extremely low proportion of heterozygous sites (< 0.001) across the GBS sites sampled in Dataset VII.

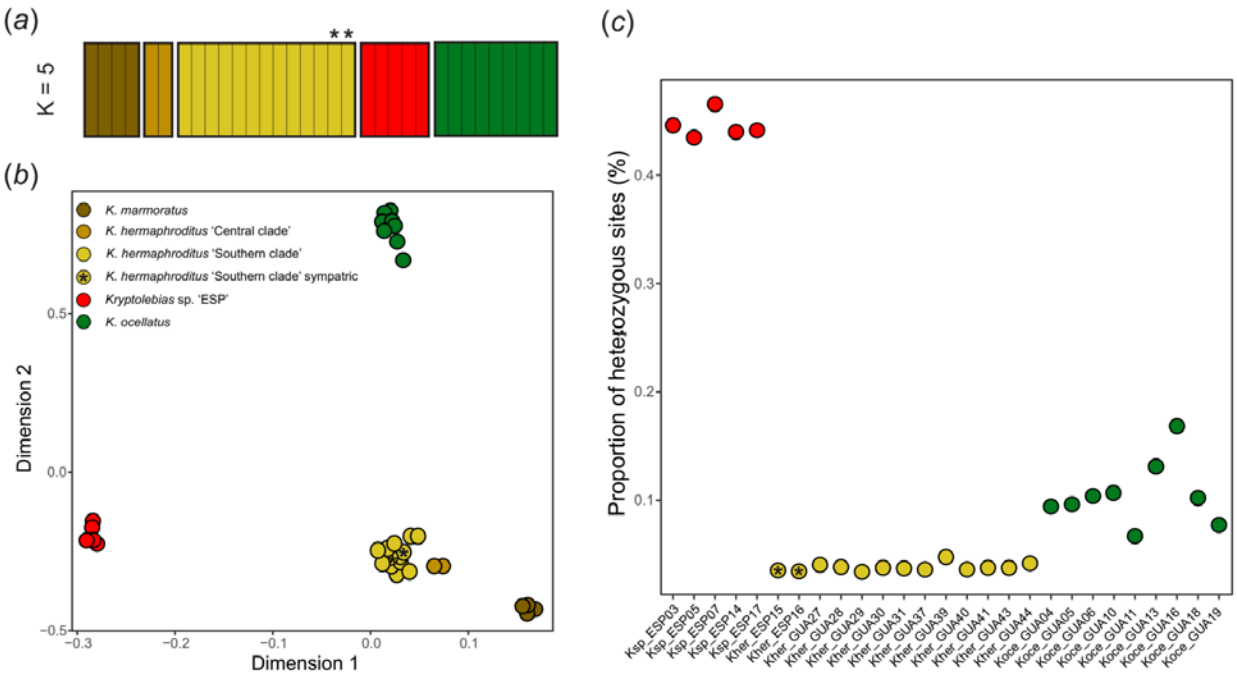


Figure 5. Genetic structure plots for the lineages in the mangrove killifish clade. (*a*) Admixture plot for K=5, indicated by StructureSelector (Li and Liu, 2018) as the most likely number of genetic clusters based on Dataset III (9,532 SNPs). (*b*) Multidimensional scaling plot based on the

pairwise genetic distances between individuals extracted from Dataset III. (*c*) Proportion of heterozygous sites per individual based on the site frequency spectrum for Dataset VII (863,662 nuclear sites). For ease of visualization, species based on data extracted from whole-genome sequences (*K. marmoratus and K. hermaphroditus* 'Central clade') were omitted from the plot given a very low number of heterozygous sites (see Results). All plots follow the color scheme described in (*b*). Across all plots, individuals marked with asterisks represent *K. hermaphroditus* 'Southern clade' sympatric to *Kryptolebias* sp. 'ESP'.

3.5. Introgression between Kryptolebias sp. 'ESP' and K. hermaphroditus 'South clade'

Out of 5,688 SNPs in Dataset VIII, 4,976 (87.48%) are fixed and homozygous across all 13 *K. hermaphroditus* 'Southern clade' individuals, reflecting the highly selfing nature of the species. Out of those, 4,186 (84.12%) SNPs are present in a heterozygous state in *Kryptolebias* sp. 'ESP', another strong indication that the genome of *K. hermaphroditus* 'Southern clade' introgressed into the genome of a previously unknown *Kryptolebias* lineage, resulting in *Kryptolebias* sp. 'ESP'. This latter lineage is highly heterozygous (4,714 out of 5,688 (82.87%) SNPs are heterozygote). Further indication that an unknown and highly differentiated species was originally involved in the introgression with individuals of the *K. hermaphroditus* 'Southern clade' is the fact that 74.68% (4,248 out of 5,688) of the SNPs contained alleles exclusive to individuals of *Kryptolebias* sp. 'ESP'.

4. Discussion

Our study provided the first phylogenomic-based hypothesis for the phylogenetic relationships within killifish genus *Kryptolebias,* involving five out of the seven currently valid species. Our results (based on both mtDNA and nuclear markers) largely agreed with previously proposed phylogenetic relationships within the genus, comprising two major monophyletic groups: one

498 grouping the freshwater fishes (*K. brasiliensis and K. gracilis*), while the other grouping the
499 'mangrove killifish clade', comprising *K. ocellatus*, *K. hermaphroditus* (Central and Southern
500 clades) and *K. marmoratus* (Berbel-Filho et al., 2020; Costa et al., 2010; Murphy et al., 1999;
501 Tatarenkov et al., 2017; Tatarenkov et al., 2009; Vermeulen and Hrbek, 2005). In addition, our
502 results revealed an extensive history of introgression in *Kryptolebias*. Our findings revealed yet a
503 highly differentiated (and previously unknown) lineage with elevated levels of heterozygosity and
504 a history of admixture with the predominantly-selfing and sympatric *K. hermaphroditus*.

505 *4.1. Kryptolebias* phylogenetic relationships and introgression

506 Previous attempts to reconstruct the phylogenetic relationships within *Kryptolebias* were
507 based either exclusively on mtDNA (Vermeulen and Hrbek, 2005), and/or were exclusively
508 focused on the 'mangrove killifish clade' (Berbel-Filho et al., 2020; Kanamori et al., 2016; Murphy
509 et al., 1999; Tatarenkov et al., 2017; Tatarenkov et al., 2009; Weibel et al., 1999). Our phylogenetic
510 reconstruction not only expanded the number of genomic loci used, but also widened the
511 taxonomic sampling of *Kryptolebias*, particularly by including the freshwater species: *K. gracilis*
512 and *K. brasiliensis*; representatives of recently uncovered lineages (*K. hermaphroditus* Central and
513 Southern clades); and a lineage revealed by the present study (*Kryptolebias* sp. 'ESP'). Overall,
514 our phylogenetic reconstruction generally agrees with the topologies previously proposed by
515 Vermeulen and Hrbek (2005) based on mtDNA loci, where two monophyletic *Kryptolebias* clades
516 were represented by species living in freshwater streams or brackish environments close to
517 mangrove forests (the freshwater clade and mangrove killifishes clade, respectively). Although we
518 have not sampled the Amazonian freshwater species *K. sepia* and *K. campelloi*, those species are
519 thought to be closely related to *K. brasiliensis* (Costa, 1990; Vermeulen and Hrbek, 2005), which
520 in our phylogeny grouped together with *K. gracilis* in the 'freshwater clade'. It is important to

highlight that our *K. brasiliensis* samples have not formed a monophyletic clade within the freshwater clade of our nuclear phylogeny, with *K. gracilis* being nested within *K. brasiliensis* (Fig. 3b) Those two species are morphologically very similar, with *K. brasiliensis* being distributed in broader area along lowland streams and creeks in the state of Rio de Janeiro (Costa 2007). Our study thus calls for further research on the taxonomic status of *K. brasiliensis* and *K. gracilis*, with the possibility of the former representing a species complex.

Our nuclear phylogeny revealed some differences from previous reconstructions within the mangrove killifishes clade. So far, all studies that tried to reconstruct the phylogenetic relationships within this clade, regardless of whether it was based on mtDNA (Berbel-Filho et al., 2020; Murphy et al., 1999; Tatarenkov et al., 2017; Tatarenkov et al., 2009; Vermeulen and Hrbek, 2005; Weibel et al., 1999) or nuclear markers (Kanamori et al., 2016), have supported the obligated outcrossing species *K. ocellatus* (Berbel-Filho et al., 2020) as the sister-species of the clade containing the selfing species (*K. marmoratus* and *K. hermaphroditus*). Our phylogenetic reconstruction revealed a clear case of mito-nuclear discordance (see discussion below) including a previously unknown lineage more closely related to the selfing mangrove killifishes than to *K. ocellatus*. This change in topology may have implications for understanding the evolution of mating systems within the genus. As all the other known *Kryptolebias* species are dioecious and inhabit freshwater habitats, the classical view based on the phylogenetic mapping of reproductive traits in *Kryptolebias* suggested that synchronous hermaphroditism has emerged in the common ancestor of all mangrove killifish species (*K. ocellatus*, *K. hermaphroditus* and *K. marmoratus*), with the self-fertilization evolving later in the common ancestor of the sister-species *K. hermaphroditus* and *K. marmoratus* (Avise and Tatarenkov, 2015; Costa et al., 2010). However, the phylogenetic positioning of *Kryptolebias* sp. ‘ESP’ as the sister-group of the selfing species raises the discussion

of whether self-fertilization may have evolved earlier in the genus. Notwithstanding, it is important to note that self-fertilization tends to reduce heterozygosity levels in half every generation (Avise, 2008). The fact that the individuals from *Kryptolebias* sp. 'ESP' examined here had over 11x higher level of heterozygosity when compared to *K. hermaphroditus*, together with the evidence of ancestral introgression from *K. hermaphroditus* 'Southern clade' into *Kryptolebias* sp. 'ESP', suggests that *Kryptolebias* sp. 'ESP' may not undergo self-fertilization. All non-male individuals captured in the sampling locality of *Kryptolebias* sp. 'ESP' (and previously by Costa 2016) had a typical external appearance of hermaphrodites of the selfing mangrove killifishes, suggesting this may be another androdioecious, but not self-fertilizing, species in the genus, similarly to *K. ocellatus* (Berbel-Filho et al., 2020*)*. Nonetheless, our limited sample size makes imperative the need for further life-history and behavioural evaluation of the mating system of *Kryptolebias* sp. 'ESP' individuals. Until then, the possibility that self-fertilization may have evolved earlier (in the common ancestor between *Kryptolebias* sp. 'ESP', *K. hermaphroditus*, and *K. marmoratus*) in *Kryptolebias* must be considered.

Another major goal of our study was to evaluate the possibility that reticulate evolution may have played a role in the diversification of the genus *Kryptolebias*. Our phylogenetic networks analyses showed complex history of reticulate evolution in the mangrove killifish clade (Fig. 3c), and revealed an ancient introgression event from *K. hermaphroditus* 'Southern clade' into *Kryptolebias* sp. 'ESP'. This reticulation event may explain the fact that all individuals of *Kryptolebias* sp. 'ESP' had high heterozygosity levels and the same mtDNA haplotype commonly found in *K. hermaphroditus* 'Southern clade' in Northeast Brazil. Contrary to the prediction that highly-selfing taxa (such as *K. hermaphroditus*) provide a low opportunity for hybridization and introgression (Pickup et al., 2019), this ancestral reticulation event suggests that hermaphrodites

of the *K. hermaphroditus* lineage played the maternal role in introgression events with a previously unknown lineage, now evident in the genome of *Kryptolebias* sp. 'ESP'. Tatarenkov et al. (2021) found bi-directional hybridization between two highly selfing strains of *K. marmoratus* and *K. hermaphroditus* 'Central clade', while Berbel-Filho et al. (2021) also found a single backcross between an F1 individual and *K. hermaphroditus* 'Southern clade' in Southeast Brazil. Taken together, these results suggest that although rare, opportunities to outcross and hybridize with the selfing *Kryptolebias* may occasionally occur. Backcrossing between F1 individuals and males of the *Kryptolebias* sp. 'ESP' lineage may have then further contributed to the movement of genomic DNA from *K. hermaphroditus* 'Southern clade' into *Kryptolebias* sp. ESP. The fact that we only found individuals with *K. hermaphroditus* 'Southern clade' mtDNA hints at the possibility of local extinction of individuals with the mtDNA of *Kryptolebias* sp. 'ESP', however we acknowledge that testing of this hypothesis requires further sampling in the area. The question of whether 'pure' *Kryptolebias* sp. 'ESP' individuals still exist or evidence for this (possibly extinct) lineage can only be found in extant admixed populations with *K. hermaphroditus* is open to investigation. Thus far, the closest sampling site around the Espírito Santo locality (only 5 km apart from Coqueiral Beach, where *Kryptolebias* sp. 'ESP' was found) had mtDNA haplotypes and hermaphrodite appearance commonly found in *K. hermaphroditus* (Lira et al., 2021).

The single reticulation event found in our phylogenetic networks analysis seems to contradict the finding of multiple introgression events using site patterns counts (ABBA-BABA tests) found between *Kryptolebias* lineages of the mangrove killifishes clade. However, these two types of introgression tests tend to recover introgression events at different time scales. ABBA-BABA test assumes that multiple substitutions at a particular site are rare or do not occur, as many substitutions at individual sites could affect the patterns of site discordance. This assumption tends

590 to not hold true for deeply diverged taxa (Hibbins and Hahn, 2022). Therefore, ABBA-BABA tests
591 are usually more suitable for testing more recent introgression events. Phylogenetic networks, on
592 the other hand, use discordance between gene trees and/or concordant factors, being thus less
593 impacted by the multiple substitutions at individual sites, making them more suitable for
594 estimating ancestral introgression events (Hibbins and Hahn, 2022). While we only found evidence
595 for one reticulation event from *K. hermaphroditus* 'Southern clade' into *Kryptolebias* sp. 'ESP.'
596 in our phylogenetic network using PhyloNetowrks, our phylogenetic network visualization using
597 SplitsTree revealed many tree discordances in the mangrove killifish clade, which is indicative of
598 potential introgression. Our genome-wide ABBA-BABA tests further confirmed that indication,
599 with evidence of introgression in the hybrid zones recently described between *K. marmoratus* and
600 *K. hermaphroditus* 'Central clade' (Tatarenkov et al., 2021), between *K. hermaphroditus*
601 'Southern clade' and *K. ocellatus* (Berbel-Filho et al. 2021), and finally a significant signal of
602 introgression in the hybridization found here between *K. hermaphroditus* 'Southern clade' and
603 *Kryptolebias* sp. 'ESP'. In addition to the assumption of low substitutions per site, it is also
604 important to acknowledge here another caveat of ABBA-BABA tests, which is the assumption of
605 no ancestral population structure in the ancestor between P1, P2 and P3 (Hibbins and Hahn, 2022).
606 If present, ancestral population structure can result to similar deviations of site patterns counts as
607 the ones caused by real introgression events (Eriksson and Manica, 2012). Despite those
608 limitations, overall, our findings suggest that introgressive events in *Kryptolebias* are common,
609 both at ancestral and/or recent time scales Considering the possibility that both lineages of *K.*
610 *hermaphroditus* (Central and Southern clades) may belong to the same species (Lira et al., 2021)),
611 our results indicate that *K. hermaphroditus* has been involved in at least three different
612 introgression events with other *Kryptolebias* species across its range. Although ABBA-BABA

tests cannot evaluate the direction of introgression, these findings challenge the idea that highly-selfing taxa provide low opportunities for hybridization and introgression in the long-term (Berbel-Filho et al., 2021; Pickup et al., 2019).

4.2. The mysterious *Kryptolebias sp.* 'ESP' lineage

Given the unusually high frequency of males and their unique external coloration, we sampled in the same locality (Coqueiral Beach, in Aracruz, Espírito Santo (Supplementary Figure S1)) described in Costa (2016) to grasp further insights on the taxonomic status of this population. In total, we sampled 46 *Kryptolebias* individuals from this sampling locality, three of them could be identified based on external coloration as *K. hermaphroditus* males according to Costa (2016). All other non-male individuals captured in this sampling locality had the typical external appearance of hermaphrodites of *K. hermaphroditus*, as also reported by Costa (2016). However, our results revealed that in this locality two clearly differentiated *Kryptolebias* lineages/species coexist. The fact that the *Kryptolebias* sp. 'ESP' lineage was only found in a case of mito-nuclear discordance with *K. hermaphroditus,* calls for attention to the possibility of a cryptic species of *Kryptolebias* in the region. In fact, Sarmento-Soares et al. (2014) found two *Kryptolebias* populations in other coastal streams in the state of Espírito Santo. Those individuals were initially identified as *K. ocellatus*. Later, Costa (2016) collected individuals from Coqueiral beach (the same locality sampled here) and identified them as *K. hermaphroditus*. Our analyses did not find any evidence that individuals from Coqueiral Beach are *K. ocellatus*, however they add another syntoptic *Kryptolebias* lineage currently coexisting with *K. hermaphroditus* in Coqueiral Beach. In addition, Costa (2016) used the male coloration found on males from this locality (together with another locality in Rio de Janeiro state) to generally describe the coloration of *K. hermaphroditus* males. While his description of male coloration is detailed and accurate, there are clear differences

in coloration between the *K. hermaphroditus* males found by Berbel-Filho et al. (2016), other males sampled in different localities within the species range (Amorim et al. (2022); Supplementary Figure S1) and the ones described in Costa (2016). Although we do not have nuclear data for male individuals from this population, the evidence presented here for *Kryptolebias* sp. 'ESP' prompts for further taxonomical evaluation on the identity of males from this population, especially whether the males described in Costa (2016) represented males of *Kryptolebias* sp. 'ESP' or *K. hermaphroditus* (or both).

Another striking feature of the hybrid zone between *Kryptolebias* sp. 'ESP' and *K. hermaphroditus* in Coqueiral Beach is the evidence that despite the two species have exchanged DNA in the past and are currently syntopic, there is no apparent evidence of current gene flow between them (e.g., no F1s or early hybrid generations). Although we acknowledge our small sampling size for this population, this scenario may suggest a strong mechanism of pre and/or postzygotic reproductive isolation between the two sympatric lineages. Alternatively, we cannot fully rule out the possibility that the two individuals of *K. hermaphroditus* in Coqueiral beach found here represent a recent case of secondary contact, given the high heterozygosity found in *Kryptolebias* sp. 'ESP' together with the evidence from recent phylogeographic studies indicating that *K. hermaphroditus* has been dispersing southwards along the mangrove forests in the Brazilian coast recently (Berbel-Filho et al., 2020; Lira et al., 2021; Tatarenkov et al., 2011; Tatarenkov et al., 2017). Overall, the evolutionary history of *Kryptolebias* sp. 'ESP', its potential distribution, as well its historical and current relationship with *K. hermaphroditus* is now open for investigation, with putative scenarios such as hybrid speciation involving *K. hermaphroditus* and a previously unknown (and possibly extinct) lineage, ancestral introgression involving a yet unknown lineage,

as fruitful lines of research to understand the origins of the mysterious *Kryptolebias* sp. 'ESP' lineage.

5. Conclusion

Introgression is the most common cause of mito-nuclear phylogenetic incongruences among taxa (Toews and Brelsford, 2012). Our phylogenetic reconstruction using mtDNA and genome-wide nuclear sites for the genus *Kryptolebias* generally agreed with previous reconstructions but yielded different topologies for the same set of species. Such discordance seems to have been caused by past introgression events in the genus. More importantly, our nuclear reconstruction recovered a cryptic *Kryptolebias* lineage hidden behind the case of mito-nuclear discordance. The striking example of mito-nuclear discordance found here (with an unknown lineage having the same mtDNA haplotype of the introgressing lineage) highlights the need of using multiple genomic regions (particularly with different genomic, levels of recombination and inheritance properties) when reconstructing phylogenetic histories and making taxonomic inferences in clades where introgression is relatively common, such as *Kryptolebias*.

CRediT authorship contribution statement

Waldir M. Berbel-Filho: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Data Curation Writing – original draft, Writing – review & editing, Funding acquisition, Visualization. **George Pacheco**: Methodology, Software, Formal Analysis. Writing – review & editing. **Andrey Tatarenkov:** Resources, Writing – review & editing. **Mateus G. Lira:** Resources, Writing – review & editing. **Carlos Garcia de Leaniz:** Resources, Funding acquisition, Supervision, Writing – review & editing. **Carlos M. Rodríguez-López**: Methodology, Software Writing – review & editing. **Sergio M. Q. Lima:** Resources, Funding

acquisition, Supervision, Writing – review & editing. **Sofia Consuegra**: Resources, Funding acquisition, Supervision, Writing – review & editing, Project Administration.

Declaration of competing interests

The authors declare no conflict of interest.

Acknowledgements

We are grateful to Dr. Helder Espírito-Santo for help during fieldwork. We thank Dr. Andrew Thompson for discussions on the phylogenetic reconstruction. We are also grateful to Frans Vermeulen for providing great pictures of live *K. marmoratus* individuals.

Funding

This work was supported by National Geographic/Waitt program [W461-16], The Fisheries Society of the British Isles (FSBI) through the small research grant program, and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [233161/2014-7]. C.M.R.L. was partially supported by AFRI competitive grants [grant no. 2019-67013-29168/project accession no. 1018617; grant no. 2021-67019-34606/project accession no. 1025891] and USDA National Institute of Food and Agriculture, United States Department of Agriculture, Hatch Program project no. KY011050/Accession No. 1020852. S. M. Q. L. receive CNPq Research Productivity Grants (313644/2018-7).

Data accessibility

The 21 additional *cox1* sequences generated for this study are available at GenBank (access numbers: OM962875-OM962895). Merged FastaQ files generated for this study can be found at accessed at NCBI (accession PRJNA815481). FastaQ files for the samples of *K. hermaphroditus* and *K. ocellatus* generated in Berbel-Filho et al. (2021) and used here can be accessed at NCBI

(accession PRJNA563625). All scripts used in the project are available at: https://github.com/waldirmbf/KryptolebiasGenomics.

Supplementary data

Table S1. Sampling size, localities, and respective geographical coordinates for the 423 individuals included in the mtDNA phylogenetic reconstruction. 'Reference' refers to the study where *cox1* sequences were extracted from. mtDNA haplotype refers to the haplotype numbers in Figure 2 found for each sampling locality.

Table S2. Sampling locations and sizes for samples included in analysis. 'ES' and 'RJ' denote the Brazilian states of Espírito Santo and Rio de Janeiro, respectively. 'Reference' refers to the source of samples for the genomic analysis. Geographical coordinates for the additional whole-genome samples included in the analysis were not provided in the original references. When present, names in parenthesis refer to samples belonging to the same sampling locations as the mtDNA samples described in Table S1. Asterisk denotes a sampling point which is either the species type-locality or within the type-locality area.

Table S3. Summary statistics for 61 samples included in the study. Parameters 'proportion of heterozygous sites', 'coverage', and missing data' (for Dataset II) are described in methods. Samples in red did not pass the reads threshold of ≥ 500,000 reads.

Table S4. Summary of nuclear DNA datasets generated for this study. 'N' represents the number of individuals used in each dataset. 'Sites' represent the total number of nucleotides (either variable or not across samples) covered in each dataset. 'SNPs' is an abbreviation for single-nucleotide polymorphisms. Scripts used to generate the datasets are provided at https://github.com/g-pacheco/KryptolebiasGenomics/wiki/08.-Nuclear-Genome-Datasets.

Figure S1. Coqueiral beach sampling locality and *Kryptolebias* spp. specimens. (*a*) Freshwater stream at Coqueiral beach in Aracruz, Espírito-Santo state, Brazil (19°56'3.44"S; 040° 7'48.13"W), where individuals of *Kryptolebias* sp. 'ESP' and *K. hermaphroditus* 'South Clade' have been collected. (*b*) Individual male of *K. hermaphroditus* (sensu Costa 2016) collected from the sampling locality. (*c) Kryptolebias hermaphroditus* 'Southern clade' male collected in Ceará-Mirim River, Extremoz, Rio Grande do Norte state, Brazil (05°40'25.88"S; 035°14'14.48"W) same male as described in Berbel-Filho et al. (2016) as the first male reported for the species. (*c) Kryptolebias hermaphroditus* 'South clade' (following the geographical criteria in Lira et al. (2021)) male (photographed fresh after sampling) collected in mangrove forest in Viseu, Viseu, Pará, Brazil (01°10'54.60"S; 46°9'31.90"W). The external differences between the previous *Kryptolebias hermaphroditus* 'South clade' males (also found in Amorim et al. (2022) and the ones sampled in Coqueiral beach by Costa (2016) and us (*a*), together with the syntopy between two lineages on that sampling point our results (see Discussion) calls for further attention on the taxonomic status of males as either *K. hermaphroditus* or *Kryptolebias* sp. 'ESP'.

Figure S2. Sampling sites for the 423 Kryptolebias individuals included in the mtDNA phylogenetic reconstruction. Site details are provided in Table S1.

Figure S3. Sampling sites for the individuals included in the genetic analysis. Site names and details are included in Table S2.

Figure S4. Maximum-likelihood reconstruction for 115,397 concatenated nuclear sites (Dataset I) from 48 *Kryptolebias* individuals. Node values represent SH-aLRT (%), aBayes, and ultrafast bootstrap (%) support values, respectively. In the tip labels, 'Kmar' (dark brown) denotes *K. marmoratus individuals*; 'Kher_CentralClade' (light brown) represent individuals from K. hermaphroditus 'Central clade'; 'Kher_SouthClade' (yellow) represent *K. hermaphroditus*

747 'Southern clade' individuals; 'KspESP' (red) refers to *Kryptolebias* sp. 'ESP' individuals; 'Koce'
748 (green) denotes *K. ocellatus* individuals; 'Kgra' (dark blue) denotes *K. gracilis* individuals; 'Kbra'
749 (light blue) represents *K. brasiliensis*. Details for all samples used are provided in Table S2. All
750 species, with exception of *K. marmoratus* and *K. hermaphroditus* (represented by hermaphrodites),
751 are represented by male individuals.

752 **Figure. S5.** Maximum-likelihood reconstruction for 1,631,872 concatenated nuclear sites from
753 five individuals (one representative per species) within the 'mangrove killifish' clade. Node values
754 represent standard bootstrap (%) support values. In the tip labels, '*Kmar*' denotes *K. marmoratus*;
755 '*Kher*_CentralClade' represent an individual from *K. hermaphroditus* 'Central clade';
756 '*Kher*_SouthClade' represent *K. hermaphroditus* 'Southern clade'; '*Ksp*_ESP' refers to
757 *Kryptolebias* sp. 'ESP'; '*Koce*' denotes *K. ocellatus* individuals. This tree was used as our starting
758 network (hmax=0) for out phylogenetic networks analyses using PhyloNetworks v. 0.14.2 (Solís-
759 Lemus et al., 2017).

760 **Figure S6.** Phylogenetic networks reconstructed in PhyloNetworks v. 0.14.2 (Solís-Lemus et al.,
761 2017) showing reticulation events in *Kryptolebias.* Phylogenies shown here represent the tree with
762 the lowest pseudolikelihood scores for each maximum number of reticulation events ('hmax').
763 Blue arrow indicates the direction of reticulation, while blue numbers represent the proportion of
764 genes inherited by each parent.

765 **Figure S7.** Estimated number of genetic clusters (K) based on different metrics retrieved from
766 StructureSelector (Li and Liu, 2018). Those four metrics (median of medians (MedMedK);
767 medians of means (MedMeanK); maximum of medians (MaxMedK); maximum of the means
768 (MaxMeaK)) are implemented Puechmaille (2016) to account for unevenness of sampling sizes
769 and hierarchical structure.

Figure S8. Individual ancestry plots for each K value (2 to 8) ran in nsgAdmix with 9,532 SNPs (Dataset III). Each column represents an individual and each color represents a different genetic cluster.

Figure legends

Figure 1. Approximated geographic distribution of known *Kryptolebias* species and lineages. Geographic distributions for the species/lineages were on the literature (Berbel-Filho et al., 2020; Costa, 1990, 2004; Costa, 2007; Costa, 2006, 2016; Lira et al., 2021; Tatarenkov et al., 2017; Vermeulen and Hrbek, 2005) as well as online databases for sampling records (GBIF; www.gbif.org) and museum collections (CRIA - SpeciesLink; http://splink.cria.org.br/). Symbols next to species name represent species inhabiting mangrove (mangrove tree) or freshwater (blue spot) habitats.

Figure 2. Maximum-likelihood reconstruction for 49 unique *cox1* haplotypes extracted from 423 *Kryptolebias* individuals. (*a*) Full tree containing the relationships among the 49 haplotypes with tip labels colored by species/lineages names. Tip labels show haplotypes and number of individuals sequenced (in parenthesis). (*b*) Haplotype network for the 27 *cox1* haplotypes for individuals belonging to the selfing mangrove killifishes clade with their respective distribution. Details for all samples used for these analyses are provided in Table S1.

Figure 3. Phylogenetic reconstructions of the genus *Kryptolebias* using IQ-Tree 2 v. 2.0.1. (*a*) Schematic representation of the maximum-likelihood phylogenetic tree based on 49 unique mtDNA *cox1* haplotypes extracted from 423 *Kryptolebia*s individuals. Node circles represent nonparametric bootstrap values = 100. The full mtDNA phylogenetic reconstruction is provided on Figure 2. (*b*) Maximum-likelihood phylogenetic tree based on 174,842 GBS nuclear sites (Dataset I). Node circles represent SH-aLRT (%), aBayes, and ultrafast bootstrap (%) support values, respectively. Only nodes with high support (SH-aLRT ≥90, aBayes = 1, and ultrafast bootstrap > 90) are shown. Branch lengths are shown in substitutions per site. Intraspecific clades were collapsed to facilitate visualization. The full nuclear phylogenetic reconstruction is provided

on Supplementary Fig. S4. (*c*) 95% confidence phylogenetic network (Neighbor-Net) constructed using SplitsTree based on all sites from Dataset II. All species, with exception of *K. marmoratus* and *K. hermaphroditus* (represented by hermaphrodites), are represented in the figure by male individuals.

Figure 4. Phylogenetic networks and introgression analysis in the mangrove killifishes clade. Genetic structure analysis for mangrove killifish species based on 9,532 SNPs (Dataset III). (*a*) Log-pseudolikelihood scores per number of reticulation events tested using PhyloNetworks v. 0.14.2 (Solís-Lemus et al., 2017). The red dot indicates the network chosen based on a large improvement of the pseudolikelihood score. (*b*) Phylogenetic network with one reticulation event. Numbers indicate inheritance proportions in the hybrid lineage. (*c*-*e*) ABBA-BABA results for introgression tests between: (*c*) *K. hermaphroditus* 'Central clade' ('*Kher Central'*) and *K. marmoratus* ('*Kmar*'); (*d*) *K. hermaphroditus* 'Southern clade' ('*Kher South'*) and *Kryptolebias* sp. 'ESP' ('*Ksp*. ESP'); (*e*) *K. hermaphroditus* 'Southern clade' ('*Kher South'*) and *K. ocellatus* ('*Koce*').

Figure 5. Genetic structure plots for the lineages in the mangrove killifish clade. (*a*) Admixture plot for K=5, indicated by StructureSelector (Li and Liu, 2018) as the most likely number of genetic clusters based on Dataset III (9,532 SNPs). (*b*) Multidimensional scaling plot based on the pairwise genetic distances between individuals extracted from Dataset III. (*c*) Proportion of heterozygous sites per individual based on the site frequency spectrum for Dataset VII (863,662 nuclear sites). For ease of visualization, species based on data extracted from whole-genome sequences (*K. marmoratus and K. hermaphroditus* 'Central clade') were omitted from the plot given a very low number of heterozygous sites (see Results). All plots follow the color scheme

described in (*b*). Across all plots, individuals marked with asterisks represent *K. hermaphroditus* ‘Southern clade’ sympatric to *Kryptolebias* sp. ‘ESP’.

References

Amorim, P.F., Katz, A.M., Ottoni, F.P., de Bragança, P.H.N., 2022. Genetic structure of the mangrove Killifish *Kryptolebias hermaphroditus* Costa, 2011 (Cyprinodontiformes: Aplocheiloidei) supports a wide connection among its populations. Zoological Studies 61.

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Systematic Biology 60, 685-699.

Avise, J., 2008. Clonality: the genetics, ecology, and evolution of sexual abstinence in vertebrate animals. Oxford University Press on Demand.

Avise, J.C., Tatarenkov, A., 2015. Population genetics and evolution of the mangrove rivulus *Kryptolebias marmoratus*, the world's only self-fertilizing hermaphroditic vertebrate. J Fish Biol 87, 519-538.

Barrett, S.C., 2014. Evolution of mating systems: outcrossing versus selfing. The Princeton Guide to Evolution, pp. 356-362.

Berbel-Filho, W.M., de Leaniz, C.G., Morán, P., Cable, J., Lima, S.M., Consuegra, S., 2019. Local parasite pressures and host genotype modulate epigenetic diversity in a mixed-mating fish. Ecology and evolution 9, 8736-8748.

Berbel-Filho, W.M., Espirito-Santo, H.M.V., Lima, S.M.Q., 2016. First record of a male of *Kryptolebias hermaphroditus* Costa, 2011 (Cyprinodontiformes: Cynolebiidae). Neotrop Ichthyol 14, e160024.

Berbel-Filho, W.M., Tatarenkov, A., Espirito-Santo, H.M., Lira, M.G., De Leaniz, C.G., Lima, S.M., Consuegra, S., 2020. More than meets the eye: syntopic and morphologically similar

842 mangrove killifish species show different mating systems and patterns of genetic structure
843 along the Brazilian coast. Heredity, 1-13.
844 Berbel-Filho, W.M., Tatarenkov, A., Pacheco, G., Espírito-Santo, H., Lira, M.G., Garcia de
845 Leaniz, C., Avise, J.C., Lima, S.M., Rodríguez-López, C.M., Consuegra, S., 2021. Against
846 the odds: hybrid zones between mangrove killifish species with different mating systems.
847 Genes 12, 1486.
848 Bonnet, T., Leblois, R., Rousset, F., Crochet, P.A., 2017. A reassessment of explanations for
849 discordant introgressions of mitochondrial and nuclear genomes. Evolution 71, 2140-2158.
850 Bravo, G.A., Antonelli, A., Bacon, C.D., Bartoszek, K., Blom, M.P., Huynh, S., Jones, G.,
851 Knowles, L.L., Lamichhaney, S., Marcussen, T., 2019. Embracing heterogeneity: coalescing
852 the Tree of Life and the future of phylogenomics. PeerJ 7, e6399.
853 Browning, B.L., Tian, X., Zhou, Y., Browning, S.R., 2021. Fast two-stage phasing of large-scale
854 sequence data. The American Journal of Human Genetics 108, 1880-1890.
855 Brys, R., Van Cauwenberghe, J., Jacquemyn, H., 2016. The importance of autonomous selfing in
856 preventing hybridization in three closely related plant species. Journal of Ecology 104, 601-
857 610.
858 Bushnell, B., 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National
859 Lab.(LBNL), Berkeley, CA (United States).
860 Choi, B.-S., Park, J.C., Kim, M.-S., Han, J., Kim, D.-H., Hagiwara, A., Sakakura, Y., Hwang, U.-
861 K., Lee, B.-Y., Lee, J.-S., 2020. The reference genome of the selfing fish Kryptolebias
862 hermaphroditus: Identification of phases I and II detoxification genes. Comparative
863 Biochemistry and Physiology Part D: Genomics and Proteomics 35, 100684.

Costa, W., 1990. Description d'une nouvelle espèce du genre Rivulus (Cyprinodontiformes, Rivulidae) de l'Amazone orientale. Revue française d'aquariologie herpétologie 17, 41-44.

Costa, W., 2004. Relationships and redescription of Fundulus brasiliensis (Cyprinodontiformes: Rivulidae), with description of a new genus and notes on the classification of the Aplocheiloidei. Ichthyol Explor Fres 15, 105-120.

Costa, W., 2009. Peixes aploqueilóideos da Mata Atlântica brasileira: história, diversidade e conservação/Aplocheiloid fishes of the Brazilian Atlantic Forest: history, diversity and conservation. Rio de Janeiro: Museu Nacional UFRJ, 172.

Costa, W.J., 2007. *Kryptolebias gracilis* n. sp.(Teleostei: Cyprinodontiformes: Rivulidae): a new killifish from the Saquarema Lagoon basin, southeastern Brazil. aqua 13, 7.

Costa, W.J.E.M., 2006. Redescription of *Kryptolebias ocellatus* (Hensel) and *K. caudomarginatus* (Seegers)(Teleostei: Cyprinodontiformes: Rivulidae), two killifishes from mangroves of south-eastern Brazil. Aqua: Journal of Ichthyology & Aquatic Biology 11, 5-13.

Costa, W.J.E.M., 2011a. Identity of *Rivulus ocellatus* and a new name for a hermaphroditic species of *Kryptolebias* from south-eastern Brazil (Cyprinodontiformes: Rivulidae). Ichthyol Explor Fres 22, 185-192.

Costa, W.J.E.M., 2011b. Phylogenetic position and taxonomic status of *Anablepsoides*, *Atlantirivulus*, *Cynodonichthys*, *Laimosemion* and *Melanorivulus* (Cyprinodontiformes: Rivulidae). Ichthyol Explor Fres 22, 233-249.

Costa, W.J.E.M., 2016. Colouration, taxonomy and geographical distribution of mangrove killifishes, the *Kryptolebias marmoratus* species group, in southern Atlantic coastal plains of Brazil (Cyprinodontiformes: Rivulidae). Ichthyol Explor Fres 27, 183-192.

886 Costa, W.J.E.M., Lima, S.M.Q., Bartolette, R., 2010. Androdioecy in *Kryptolebias* killifish and
887 the evolution of self-fertilizing hermaphroditism. Biol J Linn Soc 99, 344-349.
888 Clement, M., Snell, Q., Walker, P., Posada, D., Crandall, K. 2002. TCS: Estimating gene
889 genealogies. Parallel and Distributed Processing Symposium, International Proceedings, 2,
890 184.
891 Eriksson, A. and Manica, A., 2012. Effect of ancient population structure on the degree of
892 polymorphism shared between modern human populations and ancient hominins. Proceedings
893 of the National Academy of Sciences, 109,13956-13960.
894 Furness, A.I., Tatarenkov, A., Avise, J.C., 2015. A genetic test for whether pairs of hermaphrodites
895 can cross-fertilize in a selfing killifish. Journal of Heredity 106, 749-752.
896 Guimarães-Costa, A., Schneider, H., Sampaio, I., 2017. New record of the mangrove rivulid
897 *Kryptolebias hermaphroditu*s Costa, 2011 (Cyprinodontiformes: Cynolebiidae) in the Pará
898 state, northern Brazil. Check List 13, 2093.
899 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New
900 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
901 performance of PhyML 3.0. Systematic biology 59, 307-321.
902 Hahn, M.W., Nakhleh, L., 2016. Irrational exuberance for resolved species trees. Evolution 70, 7-
903 17.
904 Hatem, A., Bozdağ, D., Toland, A.E., Çatalyürek, Ü.V., 2013. Benchmarking short sequence
905 mapping tools. BMC Bioinformatics 14, 184.. doi:10.1186/1471-2105-14-184.
906 Harrington, R.W., 1971. How ecological and genetic factors interact to determine when self-
907 fertilizing hermaphrodites of *Rivulus marmoratus* change into functional secondary males,
908 with a reappraisal of the modes of intersexuality among fishes. Copeia, 389-432.

909 Herten, K., Hestand, M.S., Vermeesch, J.R., Van Houdt, J.K., 2015. GBSX: a toolkit for
910 experimental design and demultiplexing genotyping by sequencing experiments. BMC
911 Bioinformatics 16, 73.
912 Hibbins, M.S., Hahn, M.W., 2022. Phylogenomic approaches to detecting and characterizing
913 introgression. Genetics 220.
914 Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2:
915 improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35, 518-
916 522.
917 Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies.
918 Molecular Biology and Evolution 23, 254-267.
919 Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of
920 incongruence? Trends in Genetics 22, 225-231.
921 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., Jermiin, L.S., 2017.
922 ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14,
923 587-589.
924 Kanamori, A., Sugita, Y., Yuasa, Y., Suzuki, T., Kawamura, K., Uno, Y., Kamimura, K., Matsuda,
925 Y., Wilson, C.A., Amores, A., Postlethwait, J.H., Suga, K., Sakakura, Y., 2016. A genetic
926 map for the only self-fertilizing vertebrate. G3: Genes, Genomes, Genetics 6, 1095-1106.
927 Kitimu, S.R., Taylor, J., March, T.J., Tairo, F., Wilkinson, M.J., Rodríguez López, C.M., 2015.
928 Meristem micropropagation of cassava (*Manihot esculenta)* evokes genome-wide changes in
929 DNA methylation. Frontiers in Plant Science 6, 590.
930 Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation
931 sequencing data. BMC Bioinformatics 15, 356.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357-359.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079.

Li, Y.L., Liu, J.X., 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18, 176-177.

Lins, L.S.F., Trojahn, S., Sockell, A., Yee, M.C., Tatarenkov, A., Bustamante, C.D., Earley, R.L., Kelley, J.L., 2018. Whole-genome sequencing reveals the extent of heterozygosity in a preferentially self-fertilizing hermaphroditic vertebrate. Genome 61, 241-247.

Lira, M.G., Berbel-Filho, W.M., Espírito-Santo, H.M., Tatarenkov, A., Avise, J.C., Garcia de Leaniz, C., Consuegra, S., Lima, S.M., 2021. Filling the gaps: phylogeography of the self-fertilizing Kryptolebias species (Cyprinodontiformes: Rivulidae) along South American mangroves. J Fish Biol, 644-655.

Lomax, J.L., Carlson, R.E., Wells, J.W., Crawford, P.M., Earley, R.L., 2017. Factors affecting egg production in the selfing mangrove rivulus (*Kryptolebias marmoratus*). Zoology 122, 38-45.

Maddison, W.P., 1997. Gene trees in species trees. Systematic Biology 46, 523-536.

Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite-Fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour 21, 584-595.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20, 229-237.

955 Mallet, J., Besansky, N., Hahn, M.W., 2016. How reticulated are species? Bioessays 38, 140-149.
956 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A.,
957 Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference
958 in the genomic era. Molecular Biology and Evolution 37, 1530-1534.
959 Murphy, W.J., Thomerson, J.E., Collier, G.E., 1999. Phylogeny of the Neotropical killifish family
960 Rivulidae (Cyprinodontiformes, Aplocheiloidei) inferred from mitochondrial DNA
961 sequences. Mol Phylogenet Evol 13, 289-301.
962 Nakhleh, L., 2013. Computational approaches to species phylogeny inference and gene tree
963 reconciliation. Trends in Ecology & Evolution 28, 719-728.
964 Olave, M., Meyer, A., 2020. Implementing large genomic single nucleotide polymorphism data
965 sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of
966 cichlid fish. Systematic Biology 69, 848-862.
967 Pickup, M., Brandvain, Y., Fraïsse, C., Yakimowski, S., Barton, N.H., Dixit, T., Lexer, C.,
968 Cereghetti, E., Field, D.L., 2019. Mating system variation in hybrid zones: facilitation,
969 barriers and asymmetries to gene flow. New Phytol 224, 1035-1047.
970 Sarmento-Soares, L.M., Ingenito, L.F., Duboc, L., Martins-Pinheiro, R., Borçato, R., Silva, J.,
971 2014. Primeiros registros de *Kryptolebias ocellatu*s (Hensel)(Cyprinodontiformes, Rivulidae)
972 para riachos de Mata Atlântica no Espírito Santo. Boletim Sociedade Brasileira de Ictiologia
973 111, 15-19.
974 Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin,
975 M.D., Fernández, R., Kircher, M., McCue, M., 2014. Characterization of ancient and modern
976 genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX.
977 Nature Protocols 9, 1056-1082.

978 Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming,
979 identification, and read merging. BMC Research Notes 9.. doi:10.1186/s13104-016-1900-2
980 Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions
981 from next generation sequencing data. Genetics 195, 693-702.
982 Solís-Lemus, C., Bastide, P., Ané, C., 2017. PhyloNetworks: a package for phylogenetic networks.
983 Molecular Biology and Evolution 34, 3292-3298.
984 Tatarenkov, A., Earley, R.L., Taylor, D.S., Davis, W.P., Avise, J.C., 2018. Natural hybridization
985 between divergent lineages in a selfing hermaphroditic fish. Biol Letters 14, 20180118.
986 Tatarenkov, A., Earley, R.L., Taylor, D.S., Davis, W.P., Avise, J.C., 2021. Extensive hybridization
987 and past introgression between divergent lineages in a quasi-clonal hermaphroditic fish:
988 ramifications for species concepts and taxonomy. J Evolution Biol 34, 49-59.
989 Tatarenkov, A., Lima, S.M.Q., Avise, J.C., 2011. Extreme homogeneity and low genetic diversity
990 in *Kryptolebias ocellatus* from south-eastern Brazil suggest a recent foundation for this
991 androdioecious fish population. J Fish Biol 79, 2095-2105.
992 Tatarenkov, A., Lima, S.M.Q., Earley, R.L., Berbel-Filho, W.M., Vermeulen, F.B.M., Taylor,
993 D.S., Marson, K., Turner, B.J., Avise, J.C., 2017. Deep and concordant subdivisions in the
994 self-fertilizing mangrove killifishes (Kryptolebias) revealed by nuclear and mtDNA markers.
995 Biological Journal of the Linnean Society 122, 558-578.
996 Tatarenkov, A., Lima, S.M.Q., Taylor, D.S., Avise, J.C., 2009. Long-term retention of self-
997 fertilization in a fish clade. Proceedings of the National Academy of Sciences 106, 14456-
998 14459.
999 Taylor, S.A., Larson, E.L., 2019. Insights from genomes into the evolutionary importance and
1000 prevalence of hybridization in nature. Nature Ecology and Evolution 3, 170-177.

1001 Thompson, A.W., Wojtas, H., Davoll, M., Braasch, I., 2022. Genome of the Rio Pearlfish
1002 (Nematolebias whitei), a bi-annual killifish model for Eco-Evo-Devo in extreme
1003 environments. G3 Genes|Genomes|Genetics.
1004 Toews, D.P., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in
1005 animals. Molecular Ecology 21, 3907-3930.
1006 Vermeulen, F.B., Hrbek, T., 2005. *Kryptolebias sepia* n. sp. (Actinopterygii: Cyprinodontiformes:
1007 Rivulidae), a new killifish from the Tapanahony River drainage in southeast Surinam.
1008 Zootaxa 928, 1-20.
1009 Vieira, F.G., Lassalle, F., Korneliussen, T.S., Fumagalli, M., 2015. Improving the estimation of
1010 genetic distances from Next-Generation sequencing data. Biol J Linn Soc 117, 139-149.
1011 Weibel, A.C., Dowling, T.E., Turner, B.J., 1999. Evidence that an outcrossing population is a
1012 derived lineage in a hermaphroditic fish (Rivulus marmoratus). Evolution 53, 1217-1225.

1013