OXFORD

Original Article

New Genetic Insights About Hybridization and Population Structure of Hawksbill and Loggerhead Turtles From Brazil

Larissa S. Arantes, Sibelle T. Vilaça[®], Camila J. Mazzoni, and Fabrício R. Santos[®]

From the Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Avenida Antônio Carlos, 6627, Belo Horizonte - MG, Brazil, 31270-010 (Arantes and Santos); Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6–8, Berlin, Germany, D-14195 (Vilaça and Mazzoni); and Evolutionary Genetics Department, Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, Berlin, Germany, D-10315 (Vilaça and Mazzoni).

Address correspondence to F. R. Santos at the address above, or e-mail: fsantos@icb.ufmg.br.

Abstract

An extremely high incidence of hybridization among sea turtles is found along the Brazilian coast. This atypical phenomenon and its impact on sea turtle conservation can be elucidated through research focused on the evolutionary history of sea turtles. We assessed high-quality multilocus haplotypes of 143 samples of the 5 species of sea turtles that occur along the Brazilian coast to investigate the hybridization process and the population structure of hawksbill (*Eretmochelys imbricata*) and loggerhead turtles (*Caretta caretta*). The multilocus data were initially used to characterize interspecific hybrids. Introgression (F2 hybrids) was only confirmed in hatchlings of F1 hybrid females (hawksbill × loggerhead), indicating that introgression was either previously overestimated and F2 hybrids may not survive to adulthood, or the first-generation hybrid females nesting in Brazil were born as recent as few decades ago. Phylogenetic analyses using nuclear markers recovered the mtDNA-based Indo-Pacific and Atlantic lineages for hawksbill turtles, demonstrating a deep genetic divergence dating from the early Pliocene. In addition, loggerhead turtles that share a common feeding area and belong to distinct Indo-Pacific and Atlantic mtDNA clades present no clear genetic differentiation at the nuclear level. Finally, our results indicate that hawksbill and loggerhead rookeries along the Brazilian coast are likely connected by male-mediated gene flow.

Subject area: population structure
Keywords: introgression, sea turtles, multilocus sequencing, phylogenetics

Sea turtles have complex life cycles, with life stages associated with different environments affected directly by human activities. This close interaction exposes them to several threats, which led to a global population decline of most species during the XX century.

The main threats are related to fisheries bycatch, coastal urbanization, pollution (sewage, garbage, toxic substances), pathogens, and exploitation of eggs, meat, or other turtle products (Wallace et al. 2011). Thus, the monitoring of sea turtle populations followed by

actions to mitigate the anthropogenic impact are essential to their conservation worldwide.

Understanding the population dynamics of sea turtles is challenging due to their highly migratory behavior, long lives, and different levels of population structure associated with each life stage (Bowen and Karl 2007). Methodological advances in the last decades, such as satellite telemetry and molecular analyses, made important contributions to the comprehension of complex sea turtle behaviors, deepening the knowledge about the factors affecting the composition of foraging aggregations (Carreras et al. 2011; Proietti et al. 2014b), the migration route of pelagic juveniles also known as "lost years" (Putman and Mansfield 2015; Briscoe et al. 2016), the frequency of occurrence of multiple paternity (Moore and Ball 2002; González-Garza et al. 2015), the level of gene flow among populations (Bowen et al. 2005; Monzón-Argüello et al. 2011; Clusa et al. 2018), and opportunistic mating systems (Stewart and Dutton 2011). Novel technologies have also allowed investigating the arguable reproductive isolation between sea turtle species, as interspecific hybridization was detected among 5 out of 7 extant species of sea turtles (Vilaça et al. 2012). Some hybridization cases involve crosses of species of the Cheloniidae family that diverged at 63 million years ago (mya; Naro-Maciel et al. 2008), probably the most deeply divergent species group capable of producing viable hybrids in nature (Karl et al. 1995).

Hybrid zones of sea turtles may occur where there is an overlap of nesting areas and reproductive seasons of 2 or more species, which occurs in 2 coastal areas of Brazil (Soares et al. 2017). The hybridization process of sea turtles in the northeastern Brazilian coast is atypical, since the frequency of hybrids is much higher than in any other analyzed population worldwide. While hybridization cases have been sporadically reported around the world (Karl et al. 1995), the frequency of hybrid females along the Brazilian coast reach frequencies as high as 42% in some nesting sites (Lara-Ruiz et al. 2006; Reis et al. 2010b).

In the northern coast of the Bahia state in Brazil, where the largest in-country rookeries of hawksbill (*Eretmochelys imbricata*) and loggerhead turtles (*Caretta caretta*) are found, 42% of female turtles morphologically identified as *E. imbricata* exhibited mitochondrial sequences of *C. caretta* (Lara-Ruiz et al. 2006). A more recent study confirmed that the incidence of hybrids is as high as 31.58% of the assumed *E. imbricata* population (Soares et al. 2018). The majority of the surveyed hybrids appears to have 50% of alleles of each parental species, thus being considered as first-generation (F1), but backcrossing with both parental species was also detected, revealing the occurrence of introgression (Vilaça et al. 2012). Because some backcrossed nesting females were also found, hybridization was estimated to have started at least 2 generations ago (>40 years), during a period when populations were heavily depleted due to the anthropogenic impact (Vilaça et al. 2012).

In a nearby nesting site in the Sergipe state of Brazil, where olive ridley (*Lepidochelys olivacea*) and loggerhead turtles present a spatial and temporal overlapping distribution, 27% of individuals morphologically identified as loggerheads were shown to be hybrids between both species (Reis et al. 2010b), all of them classified as F1 hybrids (Vilaça et al. 2012). This large frequency of interspecific hybrids in Brazilian rookeries is an important conservation concern because it may result in outbreeding depression, which is a decrease of the fitness and/or reproductive viability of local populations (Allendorf et al. 2001; Maheshwari and Barbash 2011). Outbreeding depression can be observed in F1 hybrids, and

also in F2 or later generations due to the disruption of coadapted gene complexes as a result of meiotic recombination during gametogenesis in F1 hybrids (Goldberg et al. 2005). However, other studies suggest that interspecific hybridization may eventually represent an important source of variation as it may confer an advantageous effect on fitness, also called adaptive introgression (Hedrick 2013). In any case, it is extremely necessary to carefully investigate the consequences of hybridization for the populations where it occurs in high frequency, like the ones in Brazil.

Two previous studies (Soares et al. 2017, 2018) evaluated the potential outbreeding depression effects of hybridization via the comparison of several reproductive parameters between F1 hybrids and parental species in a nesting site located in Bahia. Even though emergence success was shown to be lower for hybrid nests, other parameters such as the hatchling production per clutch and clutch frequency were similar to parental species, suggesting that hybrids may persist in this region (Soares et al. 2017). The initial viability of hybrid hatchlings was also similar to non-hybrid hatchlings, revealing no significant evidence for hybrid breakdown at this early stage (Soares et al. 2018). However, once hybrid hatchlings achieve the sea, little is known about their survival until adulthood and reproductive fitness.

Other genetic studies including Brazilian sea turtles have investigated their demographic history (Bjorndal et al. 2006; Vargas et al. 2008; Molfetti et al. 2013), population structure (Reis et al. 2010a; Vilaça et al. 2013; Shamblin et al. 2014; Arantes et al. 2020), mixed stocks at foraging aggregations (Proietti et al. 2009, 2014b; Reis et al. 2010a; Vilaça et al. 2013) and interspecific hybridization (Lara-Ruiz et al. 2006; Reis et al. 2010b; Vilaça et al. 2012; Proietti et al. 2014a; Soares et al. 2017, 2018). For example, phylogeographic analyses using mtDNA showed significant genetic divergence among 3 Brazilian rookeries of *C. caretta*, suggesting the recognition of 3 different management units (Shamblin et al. 2014), while 2 separate demographic units were recognized for *E. imbricata* in Brazilian nesting areas (Vilaça et al. 2013). The mixed stocks found at foraging aggregations along the Brazilian coast demonstrated connectivity among distant ocean basins (Reis et al. 2010a), which is highly influenced by oceanic currents (Vilaça et al. 2013; Proietti et al. 2014b).

All the above-cited studies have used data from mitochondrial DNA (mtDNA), microsatellites and/or other few nuclear (nDNA) markers, but recent advances in high-throughput sequencing technology (NGS) have opened up new opportunities to assess genome-wide data in a cost-effective way. Indeed, recent population genomic methods have allowed the survey of selected subsets of genetic markers in many individuals simultaneously (Harrisson et al. 2014). Furthermore, genome-wide multilocus data is being increasingly used in many ecological and evolutionary studies, providing inferences about the life history, population dynamics and demographic patterns of species, with important conservation implications (Davey et al. 2011; Harrisson et al. 2014).

In this work, we designed a highly informative multilocus panel based on a genomic survey for the identification and characterization of interspecific hybrids between *E. imbricata* and *C. caretta*. Informative PCR amplified loci/haplotypes were selected from double-digest RADseq (ddRAD; Peterson et al. 2012) data produced for both species. For each selected locus, high-quality phased sequences were produced via Sanger sequencing and used for both characterization of hybrids and population studies. We compared our results with previous studies to test the efficiency of ddRAD-derived re-sequenced multilocus markers in increasing knowledge

about the hybridization phenomenon and population structure of sea turtles along the Brazilian coast.

Materials and Methods

Sampling

We analyzed 143 DNA samples from 5 species of sea turtles and hybrid individuals from 4 different hybrid classes (Vilaça et al. 2012) that occur along the Brazilian coast (Table 1). The DNA samples were derived from individuals collected between 1999 and 2011 by the Projeto TAMAR team, a consolidated and successful Brazilian Sea Turtle Conservation Program. These biological samples have already been surveyed in previous studies of our research group using mtDNA and few autosomal markers (Lara-Ruiz et al. 2006; Reis et al. 2010a; Vilaça et al. 2012, 2013). The species, localities and number of individuals analyzed are shown in Table 1. Detailed information of hybrid individuals (locality, morphology, collected individual, mtDNA haplotype based on control region and previous classification by Vilaça et al. 2012) is available in Supplementary Table S1. All *C. caretta* × *E. imbricata* (*Cc* × *Ei*), *E. imbricata* × *L. olivacea* (*Ei* × *Lo*) and *E. imbricata* × *C. caretta* × *Chelonia mydas* (*Ei* × *Cc* × *Cm*) hybrids were reported in Praia do Forte (Bahia)

Table 1. Sampling localities of sea turtles and hybrids and the number of individuals per locality (*N*)

Sea turtle species	Sample locality		*N*
Loggerhead turtle (*Caretta caretta*)	Foraging area	Elevação do Rio Grande	15
	Nesting area	Bahia	7
		Rio Grande do Norte	1
		Sergipe	10
		Rio de Janeiro	10
		Espírito Santo	10
Hawksbill turtle (*Eretmochelys imbricata*)	Foraging area	Fernando de Noronha	7
		Atol das Rocas	6
	Nesting area	Bahia	13
		Rio de Janeiro	1
		Sergipe	1
		Rio Grande do Norte	11
Green turtle (*Chelonia mydas*)	Foraging area	Fernando de Noronha	1
		Ilha do Arvoredo	1
Olive ridley turtle (*Lepidochelys olivacea*)	Nesting area	Sergipe	7
	Foraging area	Sergipe	2
Leatherback turtle (*Dermochelys coriacea*)	Foraging area	Ceará	1
		Pesca	1
Hybrid *Cc* × *Ei*	Nesting area	Bahia	17
	Foraging area	Ceará	2
		Atol das Rocas	1
		Sergipe	1
Hybrid *Ei* × *Lo*	Nesting area	Bahia	2
Hybrid *Cc* × *Lo*	Foraging area	São Paulo	1
	Nesting area	Sergipe	10
Hybrid *Ei* × *Cc* × *Cm*	Nesting area	Bahia	4

C. caretta × *E. imbricata* (*Cc* × *Ei*), *E. imbricata* × *L. olivacea* (*Ei* × *Lo*), *C. caretta* × *L. olivacea* hybrids (*Cc* × *Lo*), *E. imbricata* × *C. caretta* × *C. mydas* (*Ei* × *Cc* × *Cm*).

nesting site, and *C. caretta* × *L. olivacea* (*Cc* × *Lo*) hybrids were reported in Pirambu (Sergipe) nesting site. Other few hybrid individuals were reported in foraging aggregations or bycatch in fisheries.

Discovery and Standardization of Nuclear Markers

The initial selection of nDNA markers was performed using a reduced genomic dataset produced via ddRAD. As a preliminary analysis to help establishing a standardized ddRAD protocol for sea turtles (Driller et al. 2020), one individual of each parental species (*E. imbricata* and *C. caretta*) and one F1 *Ei* × *Cc* hybrid individual were used for ddRAD library construction. The sequencing library was generated by digesting the genomic DNA using the restriction enzymes *Nde*I and *Mlu*CI with subsequent ligation of Illumina adapters followed by a 10-cycle polymerase chain reaction (PCR) for completeness of sequencing adapters, as described in Peterson et al. (2012). Libraries were pooled and size-selected between 500 and 600 bp using the PippinPrep equipment (Sage Science). Sequencing was performed on an Illumina MiSeq machine using a 600-cycle kit with the 300 bp paired-end sequencing mode. Samples were demultiplexed and the 3 samples were run through the pyRAD (Eaton 2014) pipeline for homologous loci recognition and genotyping. Briefly, the 300 bp paired-end reads were merged using PEAR (Zhang et al. 2014) and aligned into single sequences, and those were clustered at 85% identity with a minimum coverage of 10 and a maximum of 5 heterozygous sites per locus. The selected loci were manually screened for interspecific variation between the 2 species, and their sequences were extracted from the pyRAD output file and aligned using MUSCLE (Edgar 2004). For subsequent primer design, we selected only loci found in both parental species and the hybrid individual, showing a maximum of 2 indels and at least 2 interspecific differences between *E. imbricata* and *C. caretta*, which were confirmed as heterozygous in the hybrid. Primers were designed using the Primer3 (Untergasser et al. 2012) algorithm implemented in Geneious 8.1 (Kearse et al. 2012) using default parameters (Supplementary Table S2). Thus, we initially selected 24 anonymous nDNA markers with interspecific variation to be further validated for population studies.

The validation of ddRAD-derived nDNA markers was made through PCR amplification and Sanger sequencing. PCR was done in a final volume of 20 µL using 200 µM dNTP, 0.5 units of Platinum™ Taq DNA polymerase (Life Technologies™), 1.5 mM of MgCl₂, 0.5 µM of forward and reverse primers and 10 ng of genomic DNA in 1X reaction buffer. PCR conditions were performed with one initial denaturation cycle of 95 °C for 5 min, 35 cycles of denaturation of 95 °C for 30 s, variable annealing temperatures (Supplementary Table S2) for 40 s, extension at 72 °C for 1 min, and a final extension at 72 °C for 7 min. PCR products were purified by precipitation using a solution of 20 mM polyethylene glycol and 2.5 mM NaCl.

The Sanger sequencing reaction was performed using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems™) following the manufacturer's standard protocol. Forward and reverse sequences were generated on the ABI 3130xl DNA sequencer (Applied Biosystems™). The SeqScape v2.6 software (Applied Biosystems™) was used to check the electropherogram quality. Heterozygous sites were verified for accuracy and coded as ambiguous sites according to IUPAC code. High-quality consensus sequences were aligned using the ClustalW algorithm in the MEGA 7 software (Kumar et al. 2016). The PHASE algorithm (Stephens et al. 2001) was used for gametic phase reconstruction of the heterozygous sequences with the assistance of Seq-PHASE

input/output interconversion tool (Flot 2010). The DnaSP v5 program (Librado and Rozas 2009) was used for haplotype assignment. Heterozygous indels found in some sequences of locus 966 were phased using the Indelligent web tool (Dmitriev and Rakitov 2008). Finally, high-quality phased sequences were verified again with the overlapping sequence chromatographs to edit for any inconsistencies.

Fifteen individuals of *E. imbricata*, 15 *C. caretta*, 2 *L. olivacea*, 2 green turtles (*Chelonia mydas*), and 2 leatherback turtles (*Dermochelys coriacea*), as well as 10 hybrids, were initially sequenced for the 24 selected loci. Based on the intra and interspecific variation found, 15 out of the 24 loci were selected to be analyzed in a greater number of individuals (Supplementary Table S2). The most variable loci (based on the number of polymorphisms and haplotypes) were selected for intraspecific analyses with the *C. caretta* (11 loci) and *E. imbricata* (14 loci) species (Table 2), while loci with greater interspecific divergence were used for hybrid and phylogenetic analysis (Supplementary Table S2). Thus, each dataset was developed specifically for intra- and interspecific studies of the target species using variation analyzed as high-quality phased haplotypes obtained by Sanger re-sequencing, minimizing the ascertainment bias.

All sequences generated in this study have been deposited in GenBank and this research is registered in the National System for Genetic Heritage and Associated Traditional Knowledge (SisGen) of Brazil under number A03A2C2.

Analyses of Hybrids

A phylogenetic network was built to represent the interspecific lineage admixture of hybrid individuals (Joly et al. 2015). We estimated genetic distances (Joly et al. 2015) using a distance matrix of alleles and converting it into a distance matrix of individuals using the program POFAD (Joly and Bruneau 2006). We used the MEGA 7 software (Kumar et al. 2016) to generate genetic distances using Kimura-2-parameters model for each of the 14 loci (dataset 1 in Supplementary Table S2) and then generated a combined-locus distance matrix using POFAD. The resulting matrix was used to build a phylogenetic network (neighborNet) using the software SplitsTree 4 (Huson and Bryant 2006).

Bayesian clustering analysis was done in the STRUCTURE software (Pritchard et al. 2000) for inference on population structure and assignment of individuals to populations using multilocus data. We assumed the admixture model where the individuals may have mixed ancestry in more than one of the K populations (species), allowing detection of the introgression level (Pritchard et al. 2000; Falush et al. 2003). Five loci were excluded from the STRUCTURE analysis because they present either a high-level of shared haplotypes

between species or a considerable level of missing data. Two individuals of *D. coriacea* were also excluded due to a large amount of missing data, likely due to the low level of homology in the selected primers originally designed from *E. imbricata* and *C. caretta* sequences. The final dataset was composed by haplotypic data inferred for 9 nDNA loci (dataset 2 in Supplementary Table S2) genotyped in individuals from rookeries and feeding areas from 4 sea turtle species. The robustness of our final dataset and the success of our nDNA markers to identify hybrids development is evidenced by the low missing data rate, which was only 0.3% for *E. imbricata*, *C. caretta* or hybrids. We also performed intraspecific analyses using the datasets, including 11 loci for *C. caretta* and 14 loci for *E. imbricata* (Table 2). Twenty independent runs for each *K* value (from *K* = 1 to *K* = 7) were performed with 200 000 Markov Chain Monte Carlo (MCMC) repeats after a 100 000 burn-in period. The independent and correlated allele frequencies were tested. The best *K* was assessed using Evanno's methodology (Evanno et al. 2005) through the online tool STRUCTURE Harvester (Earl and VonHoldt 2012). We combined the replicate result files and visualized the estimated membership coefficients using CLUMPAK (Kopelman et al. 2015).

The posterior probability of each individual belonging to different hybrid classes was analyzed in NewHybrids v. 1.1 Beta3 (Anderson and Thompson 2002). Separate datasets combining different hybrid crossings were tested, since the NewHybrids only consider hybridization events involving 2 diploid species (Anderson 2008). Therefore, individuals resulted from crosses involving likely more than 2 species (R0264 and R0265) could not be analyzed. The analysis was done using the Jeffrey option, no priors, with a burn-in period of 100 000 and 500 000 MCMC sweeps. The following genotype classes were considered: pure parental (Pure 1 and Pure 2), first and second-generation hybrids (F1 and F2 between 2 F1 hybrids) and backcrosses between F1 and pure parental (BC1 and BC2). The R package HybridDetective was used to plot NewHybrids analysis (Wringe et al. 2017).

Genetic Diversity and Population Structure

Population analyses were performed using the most variable nDNA markers for *C. caretta* (11 loci) and *E. imbricata* (14 loci) (Table 2). Diversity indexes were generated using the Arlequin v3.5 (Excoffier and Lischer 2010), DnaSP, and MEGA software. The summary statistics used were: number of haplotypes (H), haplotype diversity (k), and number of polymorphic sites (S). Principal Component Analysis (PCA) was performed using R package *adegenet* to evaluate the genetic diversity among the sampled individuals (Jombart and Ahmed 2011). The missing data were replaced by the mean allele frequency, and the PCA of standardized allele frequencies at the individual level

Table 2. Genetic diversity of 15 nuclear markers selected for *Caretta caretta* (11) and *Eretmochelys imbricata* (14) intraspecific analyses

		421	856	966	3061	9672	23712	30573	31476	42006	46208	67959	76958	109472	114650	267557
E. imbricata	N	39	39	39	35	-	39	39	34	39	39	35	35	39	39	39
	H	3	2	2	2	-	3	5	2	3	3	2	4	4	3	5
	S	3	1	3	1	-	2	3	4	2	3	2	4	3	2	4
	k	0.48	0.46	0.41	0.11	-	0.39	0.71	0.47	0.21	0.54	0.32	0.21	0.6	0.49	0.3
C. caretta	N	53	53	53	-	49	53	53	-	53	53	-	-	52	53	47
	H	2	3	6	-	4	3	2	-	5	4	-	-	4	3	7
	S	1	3	5	-	3	2	1	-	5	3	-	-	3	2	6
	k	0.17	0.12	0.47	-	0.44	0.07	0.05	-	0.67	0.23	-	-	0.54	0.51	0.59

Number of individuals (*N*), number of haplotypes (H), number of polymorphic sites (S), and haplotype diversity (k). Uninformative marker (-).

was calculated using multivariate methods without spatial components. The analyses were performed, including either 1) individuals collected in nesting and feeding areas along the Brazilian coast, or 2) only females sampled in Brazilian nesting areas.

To investigate the relationship of different lineages and to represent part of the worldwide genetic diversity within species, mitochondrial control region haplotypes were compiled from literature and depicted in a network analysis. For *C. caretta*, the haplotypes (776 bp) were obtained from Shamblin et al. (2014), Nishizawa et al. (2014), and from the database of The Archie Carr Center for Sea Turtle Research (http://accstr.ufl.edu/resources/mtdna-sequences). For *E. imbricata*, the control region haplotypes (739 bp) were obtained from LeRoux et al. (2012), Vilaça et al. (2012), Vargas et al. (2016), and Gaos et al. (2018). The haplotype networks were constructed using the Reduced Median algorithm with reduction threshold 9 followed by Median Joining algorithm (RM-MJ network; Bandelt et al. 1995) using the software Network 5.0 (http://www.fluxus-engineering.com). The delimitation and nomenclature of the mtDNA clades were based on previous studies (LeRoux et al. 2012; Shamblin et al. 2014; Vargas et al. 2016) and are available in Supplementary Table S3.

To investigate the intraspecific multilocus allelic variation, we built a phylogenetic network from a combined-locus genetic distance matrix. We used the dataset of 11 loci for *C. caretta* and 14 for *E. imbricata*, and performed the network reconstruction using POFAD and SplitsTree 4 software as described above.

Phylogenetic Analysis

A phylogenetic reconstruction between sea turtle species was inferred using multilocus data with a Bayesian method implemented in BEAST v2.4.3 (Bouckaert et al. 2014). The sequences of 14 anonymous loci (dataset 1 in Supplementary Table S2) were analyzed for 5 species of sea turtles. The selection of partitioned models of molecular evolution was made using the PartitionFinder2 software (Lanfear et al. 2017). The best-fit model was selected by the corrected Akaike Information Criterion (AICc) (Supplementary Table S4). The phylogenetic tree was inferred assuming a relaxed lognormal molecular clock under the birth-death model. This diversification model assumes that each species has a constant probability of speciating or going extinct along the lineage. It was employed considering that sequences from different species were used, and the species were sampled in different levels and presented very different

branch length (Drummond and Bouckaert 2014). Fossil and genetic evidence (Bowen et al. 1993; Duchene et al. 2012) provided reference dates to be used as priors for tree calibration with a lognormal distribution, as follows: 1) split between Dermochelidae and Cheloniidae family was set to 115 mya with a 95% confidence interval of 106–130 mya (Hirayama 1998) and 2) Carettini and Chelonini tribe was set to 65 mya with a 95% confidence interval of 50–90 mya (Moody 1974; Cadena and Parham 2015). The monophyly of the ingroup (Cheloniidae) was assumed a priori, by using *D. coriacea* as outgroup. The estimated date should be interpreted as maximum age constraints of the nodes.

Three independent MCMC chains were run for 200 000 000 generations and sampled every 5000 generations. Trace files were checked for chain convergence and sufficient effective sample sizes (ESS) in Tracer v. 1.6 (http://beast.bio.ed.ac.uk/Tracer), considering ESS > 200 as acceptable. The maximum clade credibility (MCC) tree was summarized after a 50% burn-in in TreeAnnotator from the 20 000 trees.

Results

Analyses of Hybridization

High-quality multilocus data standardized in this work was used to identify hybrids and estimate the introgression level in sea turtles. Some nDNA markers were more informative to characterize hybrids since they presented species-specific haplotypes that allowed us to identify the parental origin of the alleles with greater confidence. Considering the combined data from all 14 nDNA loci selected with interspecific differences, the POFAD analysis produced a reticulated network of all individuals. The 5 sea turtle species were recovered in different clusters and the hybrids were observed in an intermediate position between species involved in the hybridization process (Figure 1). This method allowed the characterization of the genomic admixture of hybrids using distance measures to estimate the contribution of parental genomes.

The Bayesian clustering analysis generated by STRUCTURE using correlated allele frequencies model showed that the number of clusters that best fit the data according to Evanno's statistics (Evanno et al. 2005) was 5 (Figure 2). The 4 sea turtle species included in this analysis were distinguished in different groups with high probability (99.9%) according to NewHybrids. *Caretta caretta* individuals were clustered in 2 different subgroups, one corresponding to individuals
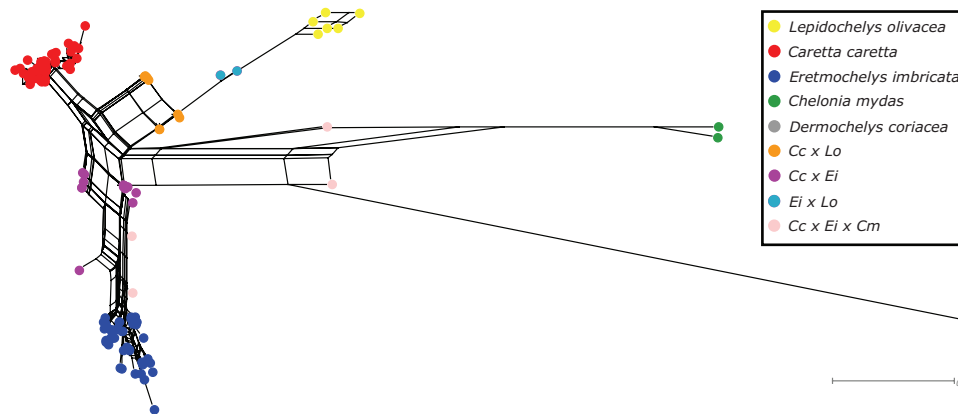
Figure 1. NeighborNet of individuals based on multilocus nuclear data for sea turtle species and hybrids that occur along the Brazilian coast. The hybrids are observed intermediately between species involved in the hybridization process. Details of sampling (*N* = 143) are described in Table 1. Tips of the neighborNet represent unique multilocus genotypes. *Cc*: *Caretta caretta*, *Ei*: *Eretmochelys imbricata*, *Lo*: *Lepidochelys olivacea*, *Cm*: *Chelonia mydas*.
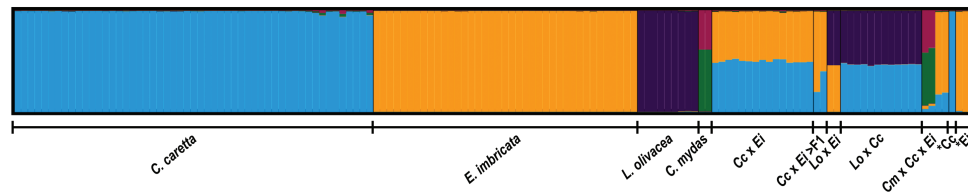
Figure 2. STRUCTURE bar plots representing *K* = 5 using correlated allele frequencies model using 9 nuclear markers. The *x* axis represents each individual analyzed and the *y* axis represents the estimated admixture proportions related to each parental species. The barplot was obtained with CLUMPAK. The asterisks (*) show misidentified individuals. *Cc*: *Caretta caretta*, *Ei*: *Eretmochelys imbricata*, *Lo*: *Lepidochelys olivacea*, *Cm*: *Chelonia mydas*, >F1: introgressed hybrid.

with mtDNA haplotypes commonly found in Brazilian rookeries, and another including foraging individuals sampled at Elevação do Rio Grande (ERG) with mtDNA haplotypes found in rookeries of the Caribbean, Mediterranean and Indo-Pacific oceans. The same intraspecific subdivision was obtained when *C. caretta* individuals were analyzed separately (Supplementary Figure S1), but it was not observed using an independent allele frequency model (Supplementary Figure S2). All individuals of *E. imbricata* from Brazil were attributed to a single population in STRUCTURE analysis using both multi-species (Figure 2) and intraspecific (Supplementary Figure S1) datasets.

Since the admixture model was assumed in STRUCTURE, the introgression level of hybrids could also be inferred. F1 hybrids clearly displayed intermediary genomic composition between parental species. All *Cc* × *Lo* and *Ei* × *Lo* hybrids were classified as F1 with a probability of 99.9%, according to NewHybrids analysis. The *Cc* × *Ei* hybrids identified as F1 presented a posterior probability of 99.9% (NewHybrids) of belonging to this category. The parental *C. caretta* population involved in hybridization cases is associated with mtDNA haplotypes typically found in Brazil.

Three individuals (R0069, R0072, and R0217) previously assigned as hybrids (>F1, F1, and >F1, respectively) by Vilaça et al. (2012) showed no evidence of admixture between species for all 9 nDNA markers analyzed. Individuals R0069 and R0072 were identified as *E. imbricata* and R0217 as *C. caretta* with high posterior probability. This could have resulted from sample misidentification, as we confirmed by re-sequencing the loci RAG1 and CMOS used by Vilaça et al. (2012), which reinforced that they are indeed "pure" individuals (Supplementary Table S5). We have also re-sequenced the control region of mtDNA for individual R0072, and, in contrast to the previous work, it presents a haplotype from *E. imbricata*. For the individual R0217 that was morphologically identified as *E. imbricata*, all the genetic data suggest that it is a pure *C. caretta*, probably due to misidentification during subsequent sample manipulation.

Previous work (Vilaça et al. 2012) identified 17 individuals as introgressed (>F1) hybrids, of which 15 were re-analyzed in this work with a multilocus nDNA approach. Using our nDNA dataset, we were able to recognize only 6 individuals with evidence of being >F1 generation hybrids (Supplementary Figure S3 and Supplementary Table S1). Remarkably, they were all hatchlings collected in nests and showing characteristics of more than one sea turtle species (Vilaça et al. 2012). Individual R0025 is a hatchling of a *Cc* × *Ei* hybrid female (R0024) and it was attributed to the category of backcrossing with an *E. imbricata* male with a probability of 99.8% (NewHybrids). Individual R0196 was sampled with morphological evidence of hybridization and it was also identified as backcrossing with an *E. imbricata* male with a probability of 96.8% (NewHybrids). The remaining 4 hatchlings (R0264, R0265, R0267, and R0268) are siblings derived from a single clutch. The genetic admixture of 3 species *E. imbricata* × *C. caretta* × *C. mydas* (*Ei* × *Cc* × *Cm*) was confirmed in 2 individuals (R0264 and R0265), although the posterior probability could not be estimated because NewHybrids only considers hybridization cases involving 2 species. The remaining siblings R0267 and R0268 were attributed to the category backcrossing with *E. imbricata*, with a posterior probability of 99.8% (NewHybrids). This result is in accordance with Vilaça et al. (2012) which hypothesized that these hatchlings could have resulted from the crossing between one *Cc* × *Ei* F1 hybrid female with at least one *C. mydas* male (evidenced by the R0264 and R0265) and another *E. imbricata* male (evidenced by the R0267 and R0268).

Population Analyses

Population analyses were performed independently for *C. caretta* and *E. imbricata* using the intraspecific nuclear variation. A total of 4492 bp were sequenced from 14 nDNA markers for *E. imbricata* and 3592 bp were sequenced from 11 nDNA markers for *C. caretta* (Table 2).

A PCA of multilocus data (Figure 3) was conducted to infer population structure assessing continuous axes of genetic variation of these species. First, we investigated nesting areas along the Brazilian coast for *C. caretta*. The first axis explained only 33.8% of the total variation, while the second axis explained 29.8% of the variation, showing no relevant structure between Brazilian populations (Figure 3C). When individuals sampled in feeding areas were included in the analysis, the first principal component (PC1) explained 69.7% of the total variation and divided the samples into 2 clusters (Figure 3D). The first one corresponds to all individuals from Brazilian rookeries and 9 individuals captured in the feeding area in southern Brazil—Elevação do Rio Grande (ERG)—that present mtDNA haplotypes commonly found in Brazilian rookeries (CC-A4 and CC-A24). The second one corresponds to individuals from ERG that present mitochondrial haplotypes (CC-A11, CC-A2, CC-A33, and CC-A34) found in rookeries of the Caribbean, Mediterranean, and Indo-Pacific oceans. The second principal component (PC2) represents 29.7% of the variation.

A PCA was performed for individuals of *E. imbricata* sampled in Brazilian rookeries. PC1 and PC2 explained 46.12% and 34.87% of the total variation, showing no correlation between genetic variation and geographic distribution (Figure 3A). When sea turtles from feeding areas were included, PC1 and PC2 explained 49.77% and 40.44% of the total variation, respectively (Figure 3B). Five individuals sampled at feeding areas of Fernando de Noronha and Atol das Rocas were slightly separated from other individuals. They presented mtDNA haplotypes either typically found in the Indo-Pacific Ocean basin (EiIP16 and EiIP33) or "orphan" haplotypes (EiA49, EiA75) which are differentiated from EiIP16 by one mutation step.

Population structure was analyzed comparing mtDNA and nDNA data of *C. caretta* (*N* = 53) and *E. imbricata* (*N* = 39) from
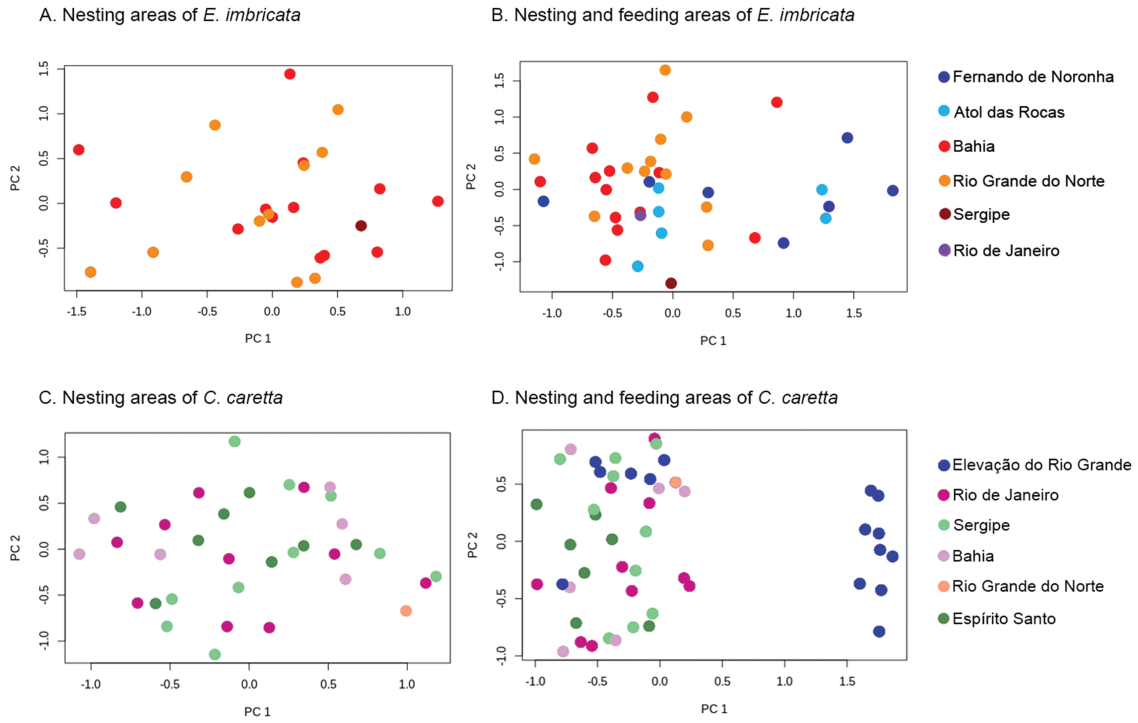
Figure 3. Principal component analysis of multilocus data for *C. caretta* (11 markers) and *E. imbricata* (14 markers), including only sea turtles sampled in Brazilian rookeries (**A** and **C**) and individuals collected in rookeries and feeding areas along the Brazilian coast (**B** and **D**). Color codes indicate the geographical location where the individuals were collected. (for color figure refer online version)
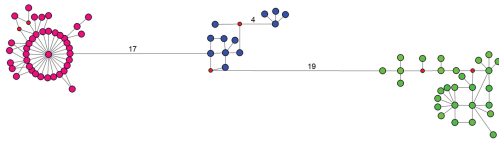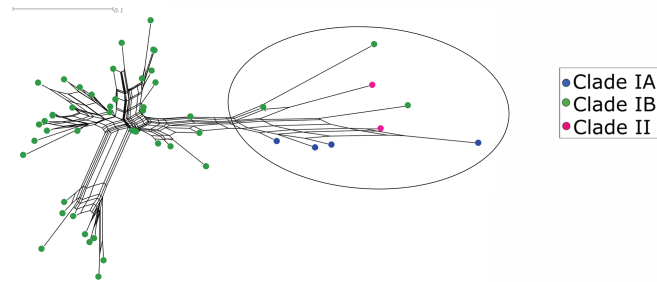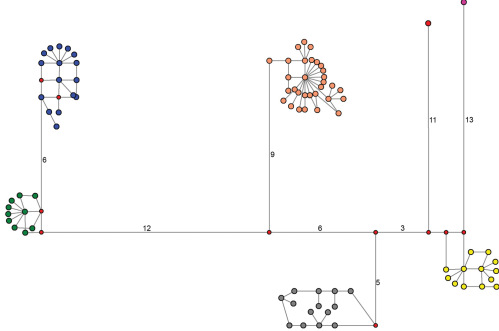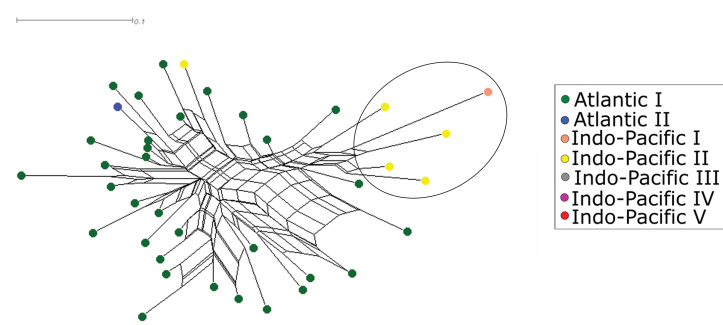

Figure 4. Haplotype network based on mtDNA control region data (**A** and **C**) and neighborNet of organisms based on multilocus nuclear data (**B** and **D**) for *C. caretta* and *E. imbricata*. The mitochondrial data were obtained from haplotypes based on control region previously published in literature. The nuclear data comprised 11 loci for 53 individuals of *C. caretta* and 14 loci for 39 individuals of *E. imbricata*. Tips of the neighborNet represent unique multilocus genotypes. The ellipses highlight the individuals of *C. caretta* more distantly related and supposed to have Indo-Pacific origin (**B**) and the individuals of *E. imbricata* that belong to the Indo-Pacific mtDNA clades and were grouped together (**D**).

the Brazilian coast. The 98 mtDNA haplotypes of *C. caretta* compiled from the literature were depicted in the network ([Figure 4A](#)), which showed 3 clades representing the main lineages within species

([Shamblin et al. 2014](#)). There was a large genetic divergence between mtDNA clades, a pattern not observed with our multilocus data ([Figure 4B](#)). The neighborNet showed that some individuals

from different mtDNA clades exhibited a close phylogenetic relationship when nDNA data was considered (highlighted with an ellipse in Figure 4B).

For *E. imbricata*, the relationship among 87 control region mtDNA haplotypes obtained from literature was depicted in a network shown in Figure 4C. They were clustered in 7 main clades, 2 reported in the Atlantic Ocean, and 5 in the Indo-Pacific Ocean. The neighborNet built with multilocus data did not present large genetic distances between individuals from Atlantic and Indo-Pacific (Figure 4D). However, 5 of 6 individuals collected in foraging aggregations in northern Brazilian coast that belong to Indo-Pacific mtDNA clades were clustered in an end of the neighborNet (highlighted with an ellipse in Figure 4D), suggesting they come from another gene pool. Only individual R0242 from the Indo-Pacific mtDNA clade appeared more closely related to individuals that belong to the Atlantic mtDNA clade.

Phylogenetic Analysis

The MCC tree obtained with multilocus data (Figure 5) showed the topology and dating congruent with previous phylogenetic studies of sea turtles (Bowen et al. 1993; Naro-Maciel et al. 2008; Duchene et al. 2012). The estimation of the time to most recent common ancestors (TMRCA) for the 5 species of sea turtles was 112.4 mya. The divergence between Carettini and Chelonini tribe (Cheloniidae) was estimated to have occurred at 65.9 mya. *Eretmochelys imbricata* separated from *C. caretta* and *L. olivacea* at 25 mya, followed by the split between *C. caretta* and *L. olivacea* at 21.6 mya.

The divergence between Atlantic and Indo-Pacific lineages of *E. imbricata* was estimated to have occurred at 5.93 mya, approximately the same date estimated using control region haplotypes (Vargas et al. 2016). Monophyly of clades based on mtDNA of *E. imbricata* (LeRoux et al. 2012; Vargas et al. 2016) was supported with nDNA in the Bayesian analysis using BEAST, except for one individual (R0242). This sea turtle belongs to mtDNA Indo-Pacific

clade II, but its nuclear composition showed that it is more similar to individuals from the Atlantic mtDNA clade. However, R0242 represents an early nDNA diverging branch in the Atlantic mtDNA clade, despite the low clade Bayesian posterior probability (0.41).

The earliest divergence between *C. caretta* lineages was estimated to have occurred 4.29 mya, similar to the date estimated using mitogenomes (Duchene et al. 2012) and control region haplotypes (Shamblin et al. 2014). One nDNA lineage gathers nesting and foraging individuals from Brazil that presents mtDNA haplotypes derived from CC-A4 and the other nDNA lineage presents individuals foraging in ERG that belong to 3 different mtDNA clades (IA, IB, and II). Thus, the mtDNA-based clades were only partially recovered with nuclear multilocus data, since individuals from different mtDNA clades were grouped in a single nDNA lineage.

Discussion

The Interspecific Hybridization Phenomenon Along the Brazilian Coast

The use of informative datasets based on high-quality haplotypes allowed expanding our comprehension about the hybridization process of sea turtles. In this study, we re-analyzed 15 out of 17 individuals previously identified as introgressed (>F1) hybrids by Vilaça et al. (2012), but we only confirmed 6 backcrossed individuals. All the introgressed (F2 hybrids) individuals were hatchlings. Seven F1 hybrids detected with our multilocus data were previously identified as introgressed (>F1) based on the information of only one genetic marker (Supplementary Table S1). For 3 individuals, the introgression signal was detected with a single microsatellite or RFLP marker, which are based on allele size differences and present a high level of homoplasy and genotyping artifacts such as null alleles and allele dropouts (Zhang and Hewitt 2003). Here we used nDNA multilocus re-sequencing to characterize high-quality haplotypes that supply a much higher level
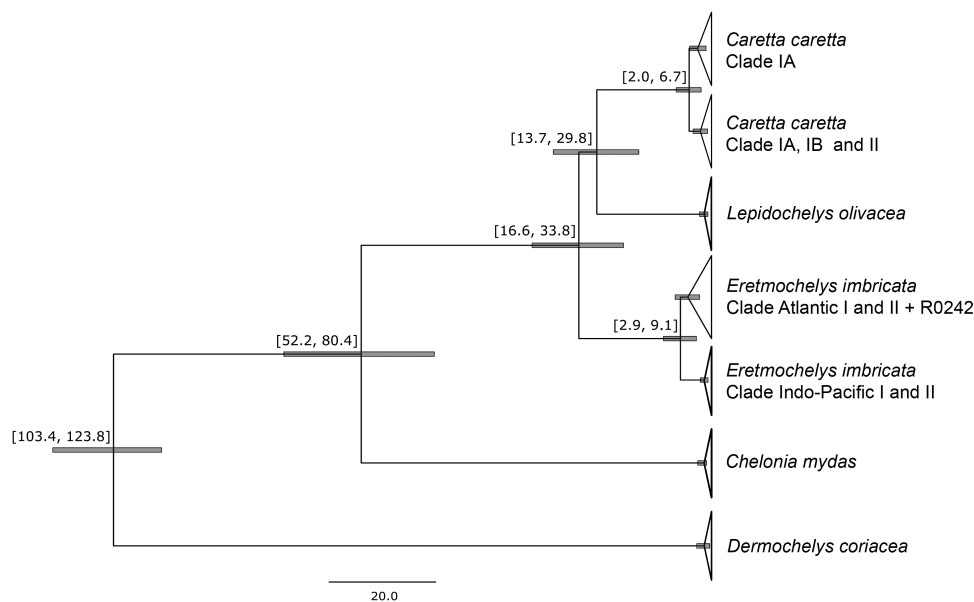
Figure 5. Dated Bayesian phylogeny of sea turtles from the Brazilian coast inferred from multilocus data. The horizontal axis indicates divergence times in million years before present. Horizontal bars and the numbers above branches correspond to the 95% highest posterior density (HPD) interval values estimated for all tree nodes with posterior probabilities above 0.8 calculated in BEAST. Clade names are based on mtDNA haplotypes as grouped by previous studies (LeRoux et al. 2012; Nishizawa et al. 2014; Shamblin et al. 2014; Vargas et al. 2016).

of resolution at both inter and intraspecific analyses (Schlötterer 2004). Besides, haplotypes randomly distributed throughout the genome generated by Sanger re-sequencing approach also provide the highest genotyping accuracy with low ascertainment bias. Indeed, our multilocus dataset displayed a higher resolution to distinguish different hybrid crossings and introgression levels as compared to the previous methods.

Even though the initial NGS screening of variable loci was done using the ddRAD approach with only 2 species (*C. caretta* and *E. imbricata*), we were able to validate informative nDNA loci with diagnostic alleles/haplotypes for other Chelonioidea taxa, even for species displaying close phylogenetic relationship as *L. olivacea* and *C. caretta*. Considering the number of individuals analyzed and the number of diagnostic sites, we suggest the use of loci 856, 3061, 76958, and 109472 to characterize *Cc* × *Ei* hybrids, loci 421, 3061, and 109472 to characterize *Ei* × *Lo* hybrids, and loci 421, 966, 67959, and 114650 to characterize *Cc* × *Lo* hybrids (Supplementary Table S2). Future genetic studies investigating the hybridization between different species of sea turtles should be able to select more informative loci according to their target species.

According to Vilaça et al. (2012), there are introgressed (at least F2 hybrids) adult females nesting in Bahia (Brazil), and the first interspecific crossing could have occurred at 2 generations ago or a minimum of 40 years. In contrast, our data suggest that only hatchlings (newborns) were confirmed as introgressed hybrids. Considering the age at maturity from 20 to 40 years for *E. imbricata* (Meylan and Donnelly 1999) and from 22 to 29 years for *C. caretta* (Heppel 1998; Casale et al. 2011), we estimate that the minimum time for the first hybridization event was one generation ago (at least 20 years). Since the first hybrid female analyzed in this work was sampled in year 2000 at Bahia, our data suggest that the high-frequency hybridization event in Bahia may have started around 1980. This is supported by Conceição et al. (1990), which in 1989 first recorded hybrid juveniles in the state of Bahia. Bass et al. (1996) also support this hypothesis since they were the first genetic work to report, in 1992, a high incidence of *Cc* × *Ei* hybrid hatchlings (10 of 14 individuals) of females morphologically identified as *E. imbricata* at Praia do Forte, Bahia. This indicates that introgression was likely overestimated by Vilaça et al. (2012), and hybridization may be a more recent phenomenon happening in Brazil. However, another possible hypothesis is that the hybridization may be a recurrent event, and the introgressed hybrids (F2) are much less fertile or inviable, precluding their survival and reproduction.

Studies have reported that the emergence success of hybrids is significantly lower than either hawksbills or loggerheads, although the hatchling production per clutch, breeding and nesting frequency, and hatchling viability of hybrids were similar to parental species (Soares et al. 2017, 2018). However, they only investigated F1 hybrid females and their hatchlings. There is no information about the potential effects of hybridization in other life stages at sea, as survivorship, growth rates, and mating success. Indeed, if all (or the large majority) hybrid adults are F1 hybrids as our results indicate, thus, a most likely explanation is that outbreeding depression (decrease of survival and/or reproductive fitness) may occur mostly in the second and further generations of introgressed individuals. In this situation, the original parental gene combinations can be broken up by recombination in >F1 hybrids, disrupting the coadapted gene complex (Edmands et al. 1999; Goldberg et al. 2005). However, a recent study found immature >F1 hybrids in foraging areas for loggerheads in South-Western Atlantic, which demonstrates that introgressed

individuals may survive the early developmental phases (Brito et al. 2020). Future studies involving nest monitoring, telemetry, diet, and genetics analyses should be performed in order to understand the hybridization outcomes on sea turtle fitness and behavior.

The emergence of high-frequency hybridization cases in Brazil coincides with the period of a great population decline of sea turtles during the XX century. This depletion leads to a reduced chance of potential conspecific encounters, which may be associated with this unique event on the Brazilian coast (Vilaça et al. 2012). Reports of hybridization cases associated with human impact are increasing worldwide for other species (Allendorf et al. 2001; Grabenstein and Taylor 2018). Human activities may lead to a secondary contact between previously isolated populations due to habitat disturbance and environmental changes that increase the hybrids rate (Todesco et al. 2016). Since 1980, sea turtle conservation in Brazil mostly relies on efforts of Projeto TAMAR, a consolidated and successful program aiming at environmental education and monitoring and research of sea turtles. Thereafter, the number of nesting females in monitored beaches has been increasing quickly (Marcovaldi and Chaloupka 2007), but in spite of this greater number of individuals, more recent hybridization events have been reported. A study of 2012 and 2013 nesting seasons showed that the incidence of hybridization in Bahia inferred from hatchlings of *C. caretta* females is 16.66% and for *E. imbricata* females is 8.15% (Soares et al. 2018).

Hybridization in Brazil is a local event with reports of fertile female hybrids in about 300 km of coastline between northern Bahia and Sergipe states. In this work, all female hybrids were originally sampled in rookeries of Bahia and Sergipe beaches, and pelagic individuals were sampled in coastal waters of Ceará, Bahia, Sergipe, and São Paulo states. Other reports of hybrids in Brazil are juveniles from the states of Ceará and Rio Grande do Sul (Cassino Beach), which are 2 important feeding aggregations of *C. caretta* (Proietti et al. 2014b). Further studies focusing on the detailed characterization of hybrids is recommended, mainly in nesting areas worldwide with the overlapping distribution of different sea turtle species.

We confirmed that all the *Cc* × *Ei* F1 hybrids resulted from the crossing between *C. caretta* female and *E. imbricata* male, which indicates a gender bias. This is probably associated with the prevalence of *C. caretta* along the Brazilian coast and the partial overlapping of reproductive season with *E. imbricata* (Vilaça et al. 2012). The beginning of the nesting season for *E. imbricata* overlaps with the nesting peak of *C. caretta* (November and December), when *E. imbricata* males encounter a higher number of *C. caretta* females to mate (Proietti et al. 2014a). Conversely, the encounter between *C. caretta* males and *E. imbricata* females may happen less frequently, since *C. caretta* males leave the mating areas before a large number of *E. imbricata* females arrive at nesting beaches (Vilaça et al. 2012).

Sea turtles present long and complex life cycles and monitoring the consequences of hybridization can be complicated, but it is extremely important to understand their impact on the management of sea turtle populations, particularly for parental species. Particular focus should be directed towards nests and hatchlings of F1 female hybrids to allow monitoring the future consequences of hybridization. We have shown that increasing the resolution of genetic data allows us to better understand this local and atypical phenomenon in Brazil. New detailed genomic approaches should also be able to elucidate the relation between introgression and species-specific adaptive regions of the genome, in relation to lifecycle, foraging habitat, and behavior of hybrids. Thus, further studies should be highly stimulated to expand our comprehension of this particular evolutionary process of potential conservation impact.

Intraspecific Studies of *C. caretta* and *E. imbricata*

Sea turtle genomic structure is quite monotonous, presenting slow cytogenetic and molecular divergence among species (FitzSimmons et al. 1995; Naro-Maciel et al. 2008; Wang et al. 2013). In this study, we were able to identify informative markers for intraspecific analyses that were validated after Sanger re-sequencing for each sea turtle species. The nDNA intraspecific variation found in *C. caretta* and *E. imbricata* analyzed here (Table 2) allowed us to infer important patterns on population structure.

Unlike mtDNA, the nuclear loci used in this study showed that variation within both species is not significantly correlated to the geographic distribution along the Brazilian coast (Figure 3A,C, Supplementary Figure S1). Previous studies using mtDNA data showed significant differences in allelic frequencies between southern and northern Brazilian rookeries for *C. caretta* (Reis et al. 2010a; Shamblin et al. 2014) and *E. imbricata* (Vilaça et al. 2013). For *C. caretta*, 3 genetically distinct clusters based on mtDNA were recognized along the Brazilian coast: northern coast (Bahia and Sergipe), Espírito Santo and Rio de Janeiro (Shamblin et al. 2014). For *E. imbricata*, 2 different mtDNA clusters, although closely related, were reported: Bahia and Rio Grande do Norte (Vilaça et al. 2013).

Assessing the genetic diversity of nuclear markers is important for understanding the contribution of females and males to the population structure of sea turtles, as both sexes can have different reproductive and philopatric behaviors. Lower population structure found in nDNA relative to mtDNA has been previously attributed to male-mediated gene flow (Bowen et al. 2005). Male sea turtles have uncertain philopatry and probably display greater flexibility in their choice of mating areas (FitzSimmons et al. 1997). Similar patterns were found in previous studies using nDNA and are indicative of lack of male philopatry (Bowen et al. 2005; Carreras et al. 2011; Vilaça et al. 2013; Clusa et al. 2018). Thus, the apparently discrepant results for mtDNA and multilocus data could be further explained by male-mediated gene flow between rookeries. Future studies using genome-wide data should validate this hypothesis.

Considering great part of mtDNA haplotype diversity reported in the literature for *C. caretta*, it is possible to distinguish 3 clades with great genetic divergence. They correspond to 2 major lineages—clades I and II—of which the former passed by a more recent split (subclades IA and IB, Figure 4A). Considering multilocus data, neighborNet analysis showed that 9 individuals of *C. caretta* presented greater genetic divergence in relation to Brazilian individuals (highlighted with an ellipse in Figure 4B). These individuals were collected in the southern Brazilian feeding area (ERG) and belong to 3 different mtDNA clades (clades IA, IB, and II). They also appear separately clustered in PCA (Figure 3D) and STRUCTURE analyses (Figure 2 and Supplementary Figure S1). Phylogenetic analysis also resulted in an MCC tree with 2 main lineages, of which one corresponds to a mix of individuals from 3 mtDNA clades (Figure 5).

Phylogeographic studies suggested that the 2 main mtDNA clades I and II of *C. caretta* were isolated by geographic and climatic factors into Atlantic and Indo-Pacific basins during the cooler periods of the Pleistocene (Bowen et al. 1994). As *C. caretta* also tolerates temperate water, migrations via southern Africa, directed by the waters of the Agulhas Current, are possible. The phylogeographic scenario proposed for Shamblin et al. (2014) suggests that mtDNA clade IA had an Indo-Pacific origin, where the earliest diverging lineages of *C. caretta* appear. The earliest colonization was likely from Indo-Pacific lineages invading the Atlantic Ocean. Brazilian haplotypes (CC-A4 and derived ones) seem to be the earliest diverging lineage

within mtDNA clade IB, which was followed by a more recent colonization of the CC-A11.6 precursor from Atlantic to Indian Ocean, as it is closely related to Atlantic lineages. Therefore, transoceanic migration in both directions may be responsible for long-distance gene flow between *C. caretta* populations. Furthermore, current geographic distribution of these lineages presents no phylogenetic concordance, as both lineages are found in both Atlantic-Mediterranean and Indo-Pacific basins (Reis et al. 2010a; Duchene et al. 2012). Despite the small number of samples, this result can suggest a homogenization of *C. caretta* populations at a nuclear level for individuals sharing a common feeding area. This was previously reported for a *C. caretta* population in the southeastern United States and attributed to male-mediated gene flow (Bowen et al. 2005). This behavior should be elucidated using more representative data of the genetic variation through genomic surveys.

For *E. imbricata*, the relationship among previously reported mtDNA haplotypes revealed that there are 7 main clades worldwide (Arantes et al. 2020). Two of them were registered in rookeries from the Atlantic Ocean (LeRoux et al. 2012) and 5 in rookeries from the Indo-Pacific Ocean (Vargas et al. 2016). The neighborNet of nDNA data showed that 5 of 6 individuals of Indo-Pacific mtDNA clades are slightly more distant from individuals that belong to Atlantic mtDNA clades (Figure 4D). The same individuals belong to Indo-Pacific nDNA cluster according to the phylogenetic analyses (Figure 5). They were sampled in the Brazilian feeding aggregations, demonstrating transoceanic migrations for the species.

Despite the separation between *E. imbricata* individuals from Indo-Pacific and Atlantic was not observed in STRUCTURE using nDNA (Supplementary Figure S1), it was slightly observed in PCA (Figure 3D) and neighborNet, and strongly detected in the MCC tree. There is a deep genetic divergence between Indo-Pacific and Atlantic mtDNA lineages of *E. imbricata* dating from the early Pliocene, when the closing of the Isthmus of Panama occurs (Arantes et al. 2020). The geographic pattern of separation between ocean basins found with mtDNA was recovered with nuclear data, except for one individual (R0242). However, R0242 belongs to an early diverging lineage of the Atlantic mtDNA clade, which displays a low Bayesian posterior probability (0.41). It suggests that this individual is deeply related to all other mtDNA Atlantic lineages, thus a likely remnant of the first Indo-Pacific lineages colonizing the Atlantic Ocean.

Eretmochelys imbricata is adapted to tropical waters and although some transoceanic migrations may occur, American and African continents are supposedly important barriers to species migration directing the current distribution of main lineages of sea turtles (Duchene et al. 2012). In contrast, *C. caretta* individuals were more divergent within the Atlantic than between the Atlantic and Indo-Pacific, probably due to a transoceanic gene flow observed in this species more adapted to temperate water. The barriers to gene flow are not the same for all species of sea turtles, likely due to their different ability of dispersion through the oceans and evolutionary responses to environmental changes (Duchene et al. 2012).

Regarding intraspecific phylogenetic analysis, the use of multilocus data resulted in similar topology and divergence times between species when compared to the previous studies that used mtDNA data (Duchene et al. 2012; LeRoux et al. 2012; Shamblin et al. 2014; Vargas et al. 2016). The divergences between main lineages within *E. imbricata* and *C. caretta* were estimated to have occurred about 5.93 mya and 4.29 mya, respectively. It is consistent with the age of formation of the Isthmus of Panama, associated with the deepest phylogenetic split of intraspecific lineages of different species of sea turtles (Naro-Maciel et al. 2008; Duchene et al. 2012). Besides, assessing multilocus data

made possible to evaluate biparental ancestry and accommodate the stochasticity of the coalescent process by combining information from multiple loci distributed throughout the genome, instead of relying only on inferences based on individual tree topologies (Edwards and Beerli 2000; Brito and Edwards 2009).

Concluding Remarks

Next-Generation Sequencing technologies allowed the initial identification of genome-wide polymorphic loci, which were selected for a Sanger sequencing validation step to characterize multilocus datasets useful for inter and/or intraspecific studies. The high-quality multilocus data provided significant interspecific information for the inference of the phylogeny of sea turtles and the characterization of hybrids. Additionally, another multilocus dataset provided relevant intraspecific data for analyses of population structure. The presented results reveal important enhancements in the genetic resolution of the hybridization process and population structure of sea turtles.

Supplementary Material

Supplementary material is available at *Journal of Heredity* online.
Figure S1. STRUCTURE bar plots representing *K* = 2 using correlated allele frequencies model for *Eretmochelys imbricata* and *Caretta caretta* from different Brazilian populations. The *x* axis represents each individual analyzed and the *y* axis represents the estimated admixture proportions related to each population. This graphic was obtained with CLUMPAK.
Figure S2. STRUCTURE bar plots representing *K* = 5 using independent allele frequencies model. The *x* axis represents each individual analyzed and the *y* axis represents the estimated admixture proportions related to each parental species. This graphic was obtained with CLUMPAK. The asterisks (*) show misidentified individuals. Cc: *Caretta caretta*, Ei: *Eretmochelys imbricata*, Lo: *Lepidochelys olivacea*, Cm: *Chelonia mydas*, >F1: introgressed hybrid.
Figure S3. NewHybrids analysis of 13 individuals previously identified as introgressed hybrids. Each vertical bar represents one individual and the *y* axis represents its posterior probability of belonging to different classes: *C. caretta* (Pure 1), *E. imbricata* (Pure 2), F1 hybrid, F2 hybrid, backcross with *C. caretta* (BC1) or backcross with *E. imbricata* (BC2). This graphic was obtained with R package HybridDetective.

Funding

Acknowledgments

Data Accessibility

We have deposited the primary data underlying these analyses as follows:
  - Genbank accession number for DNA sequences: MT23095–MT231002 and MT235978–MT23615.
  - Sampling locations and individual data are available in Supplementary Tables S1 and S3.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001. The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol*. 16:613–622.

Anderson EC. 2008. Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philos Trans R Soc B*. 363:2841–2850.

Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160:1217–1229.

Arantes LS, Vargas SM, Santos FR. 2020. Global phylogeography of the critically endangered hawksbill turtle (*Eretmochelys imbricata*). *Genet Mol Biol*. 43:e20190264.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics*. 141:743–753.

Bass AL, Good DA, Bjorndal KA. 1996. Testing models of female reproductive migratory behaviour and population structure in the Caribbean hawksbill turtle, *Eretmochelys imbricata*, with rntDNA sequences. *Mol Ecol*. 5:321–328.

Bjorndal KA, Bolten AB, Moreira L, Bellini C, Marcovaldi MA. 2006. Population structure and diversity of Brazilian green turtle rookeries based on mitochondrial DNA sequences. *Chelonian Conserv Biol*. 5:262–268.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 10:e1003537.

Bowen BW, Bass AL, Soares L, Toonen RJ. 2005. Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Mol Ecol*. 14:2389–2402.

Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC. 1994. Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. *Evolution*. 48:1820–1828.

Bowen BW, Karl SA. 2007. Population genetics and phylogeography of sea turtles. *Mol Ecol*. 16:4886–4907.

Bowen BW, Nelson WS, Avise JC. 1993. A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. *Proc Natl Acad Sci U S A*. 90:5574–5577.

Briscoe DK, Parker DM, Balazs GH, Kurita M, Saito T, Okamoto H, Rice M, Polovina JJ, Crowder LB. 2016. Active dispersal in loggerhead sea turtles (*Caretta caretta*) during the "lost years." *Proc R Soc B Biol Sci*. 283:20160690.

Brito PH, Edwards SV. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*. 135:439–455.

Brito C, Vilaça ST, Lacerda AL, Maggioni R, Marcovaldi MÂ, Vélez-Rubio G, Proietti MC. 2020. Combined use of mitochondrial and nuclear genetic markers further reveal immature marine turtle hybrids along the South Western Atlantic. *Genet Mol Biol*. 43:e20190098.

Cadena EA, Parham JF. 2015. Oldest known marine turtle? A new protostegid from the Lower Cretaceous of Colombia. *PaleoBios*. 32:1–42.

Carreras C, Pascual M, Cardona L, Marco A, Bellido JJ, Castillo JJ, Tomás J, Raga JA, Sanfélix M, Fernández G, *et al.* 2011. Living together but remaining apart: atlantic and Mediterranean loggerhead sea turtles (*Caretta caretta*) in shared feeding grounds. *J Hered*. 102:666–677.

Casale P, Mazaris A, Freggi D. 2011. Estimation of age at maturity of loggerhead sea turtles *Caretta caretta* in the Mediterranean using length-frequency data. *Endanger Species Res*. 13:123–129.

Conceição MB, Levy JA, Marins LF, Marcovaldi MA. 1990. Electrophoretic characterization of a hybrid between *Eretmochelys imbricata* and *Caretta caretta* (Cheloniidae). *Comp Biochem Physiol Part B*. 97:275–278.

Clusa M, Carreras C, Cardona L, Demetropoulos A, Margaritoulis D, Rees A, Hamza A, Khalil M, Levy Y, Turkozan O, et al. 2018. Philopatry in loggerhead turtles *Caretta caretta*: beyond the gender paradigm. *Mar Ecol Prog Ser*. 588:201–213.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*. 12:499–510.

Dmitriev DA, Rakitov RA. 2008. Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Comput Biol*. 4:e1000113.

Driller M, Vilaca ST, Arantes LS, Carrasco-Valenzuela T, Heeger F, Chevallier D, de Thoisy B, Mazzoni CJ. 2020. Optimization of ddRAD-like data leads to high quality sets of reduced representation single copy orthologs (R2SCOs) in a sea turtle multi-species analysis. *bioRxiv*. 2020.04.03.024331.

Drummond AJ, Bouckaert RR. 2014. *Bayesian evolutionary analysis with BEAST 2*. Cambridge University Press. ISBN: 978-1-107-01965-2.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Thomas P Gilbert M, Morin PA. 2012. Marine turtle mitogenome phylogenetics and evolution. *Mol Phylogenet Evol*. 65:241–250.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 4:359–361.

Eaton DA. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*. 30:1844–1849.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Edmands S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution*. 53:1757–1768.

Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*. 54:1839–1854.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611–2620.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 10:564–567.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 164:1567–1587.

FitzSimmons NN, Limpus CJ, Norman JA, Goldizen AR, Miller JD, Moritz C. 1997. Philopatry of male marine turtles inferred from mitochondrial DNA markers. *Proc Natl Acad Sci U S A*. 94:8912–8917.

FitzSimmons NN, Moritz C, Moore SS. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol Biol Evol*. 12:432–440.

Flot J-F. 2010. Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments: computer program note. *Mol Ecol Resour*. 10:162–166.

Gaos AR, Lewison RL, Jensen MP, Liles MJ, Henriquez S, Chavarria S, Pacheco CM, Valle M, Melero D, Gadea V, et al. 2018. Rookery contributions, movements and conservation needs of hawksbill turtles at foraging grounds in the eastern Pacific Ocean. *Mar Ecol Prog Ser*. 586:203–216.

Goldberg TL, Grant EC, Inendino KR, Kassler TW, Claussen JE, Philipp DP. 2005. Increased infectious disease susceptibility resulting from outbreeding depression. *Conserv Biol*. 19:455–462.

González-Garza BI, Stow A, Sánchez-Teyer LF, Zapata-Pérez O. 2015. Genetic variation, multiple paternity, and measures of reproductive success in the critically endangered hawksbill turtle (*Eretmochelys imbricata*). *Ecol Evol*. 5:5758–5769.

Grabenstein KC, Taylor SA. 2018. Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol Evol*. 33:198–212.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol Appl*. 7:1008–1025.

Hedrick PW. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol*. 22:4606–4618.

Heppel SS. 1998. Application of life-history theory and population model analysis to turtle conservation. *Copeia*. 2:367–375.

Hirayama R. 1998. Oldest known sea turtle. *Nature*. 392:705–708.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23:254–267.

Joly S, Bruneau A. 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from Rosa in North America. *Syst Biol*. 55:623–636.

Joly S, Bryant D, Lockhart PJ. 2015. Flexible methods for estimating genetic distances from single nucleotide polymorphisms. *Methods Ecol Evol*. 6:938–948.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 27:3070–3071.

Karl SA, Bowen BW, Avise JC. 1995. Hybridization among the ancient mariners: characterization of marine turtle hybrids with molecular genetic assays. *J Hered*. 86:262–268.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28:1647–1649.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour*. 15:1179–1191.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33:1870–1874.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 34:772–773.

Lara-Ruiz P, Lopez GG, Santos FR, Soares LS. 2006. Extensive hybridization in hawksbill turtles (*Eretmochelys imbricata*) nesting in Brazil revealed by mtDNA analyses. *Conserv Genet*. 7:773–781.

Leroux RA, Dutton PH, Abreu-Grobois FA, Lagueux CJ, Campbell CL, Delcroix E, Chevalier J, Horrocks JA, Hillis-Starr Z, Troëng S, et al. 2012. Re-examination of population structure and phylogeography of hawksbill turtles in the wider Caribbean using longer mtDNA sequences. *J Hered*. 103:806–820.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25:1451–1452.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. *Annu Rev Genet*. 45:331–355.

Marcovaldi M, Chaloupka M. 2007. Conservation status of the loggerhead sea turtle in Brazil: an encouraging outlook. *Endanger Species Res*. 3:133–143.

Meylan AB, Donnelly M. 1999. Status justification for listing the hawksbill turtle (*Eretmochelys imbricata*) as critically endangered on the 1996 IUCN red list of threatened animals. *Chelonian Conserv Biol*. 3:25.

Molfetti E, Vilaça ST, Georges JY, Plot V, Delcroix E, Le Scao R, Lavergne A, Barrioz S, dos Santos FR, de Thoisy B. 2013. Recent demographic history and present fine-scale structure in the Northwest Atlantic leatherback (*Dermochelys coriacea*) turtle population. *PLoS One*. 8:e58061.

Monzón-Argüello C, Loureiro NS, Delgado C, Marco A, Lopes JM, Gomes MG, Abreu-Grobois FA. 2011. Príncipe island hawksbills: genetic isolation of an eastern Atlantic stock. *J Exp Mar Biol Ecol*. 407:345–354.

Moody RTJ. 1974. The taxonomy and morphology of *Puppigerus camperi* (Gray), an Eocene Sea Turtle from Northern Europe. *Geology*. 25:153–186.

Moore MK, Ball RM. 2002. Multiple paternity in loggerhead turtle (*Caretta caretta*) nests on Melbourne Beach, Florida: a microsatellite analysis. *Mol Ecol*. 11:281–288.

Naro-Maciel E, Le M, FitzSimmons NN, Amato G. 2008. Evolutionary relationships of marine turtles: a molecular phylogeny based on nuclear and mitochondrial genes. *Mol Phylogenet Evol*. 49:659–662.

Nishizawa H, Narazaki T, Fukuoka T, Sato K, Hamabata T, Kinoshita M, Arai N. 2014. Genetic composition of loggerhead turtle feeding aggregations: migration patterns in the North Pacific. *Endanger Species Res*. 24:85–93.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 7:e37135.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.

Proietti MC, Lara-Ruiz P, Reisser JW, da Silva Pinto L, Dellagostin OA, Marins LF. 2009. Green turtles (Chelonia mydas) foraging at Arvoredo Island in Southern Brazil: Genetic characterization and mixed stock analysis through mtDNA control region haplotypes. *Genet Mol Biol*. 32:613–618.

Proietti MC, Reisser J, Marins LF, Marcovaldi MA, Soares LS, Monteiro DS, Wijeratne S, Pattiaratchi C, Secchi ER. 2014a. Hawksbill × loggerhead sea turtle hybrids at Bahia, Brazil: where do their offspring go? *PeerJ*. 2:e255.

Proietti MC, Reisser J, Marins LF, Rodriguez-Zarate C, Marcovaldi MA, Monteiro DS, Pattiaratchi C, Secchi ER. 2014b. Genetic structure and natal origins of immature hawksbill turtles (*Eretmochelys imbricata*) in Brazilian waters. *PLoS One*. 9:e88746.

Putman NF, Mansfield KL. 2015. Direct evidence of swimming demonstrates active dispersal in the sea turtle "lost years". *Curr Biol*. 25:1221–1227.

Reis EC, Soares LS, Vargas SM, Santos FR, Young RJ, Bjorndal KA, Bolten AB, Lôbo-Hajdu G. 2010a. Genetic composition, population structure and phylogeography of the loggerhead sea turtle: colonization hypothesis for the Brazilian rookeries. *Conserv Genet*. 11:1467–1477.

Reis EC, Soares LS, Lôbo-Hajdu G. 2010b. Evidence of olive ridley mitochondrial genome introgression into loggerhead turtle rookeries of Sergipe, Brazil. *Conserv Genet*. 11:1587–1591.

Schlötterer C. 2004. The evolution of molecular markers—just a matter of fashion? *Nat Rev Genet*. 5:63–69.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT, *et al*. 2014. Geographic patterns of genetic variation in a broadly distributed marine vertebrate: new insights into loggerhead turtle stock structure from expanded mitochondrial DNA sequences. *PLoS One*. 9:e85956.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68:978–989.

Stewart KR, Dutton PH. 2011. Paternal genotype reconstruction reveals multiple paternity and sex ratios in a breeding population of leatherback turtles (*Dermochelys coriacea*). *Conserv Genet*. 12:1101–1113.

Soares LS, Bjorndal KA, Bolten AB, Marcovaldi MA, Luz PB, Machado R, Lo R, McDaniel SF, Payton AC, Waltzek TB, Wayne ML. 2018. Effects of hybridization on sea turtle fitness. *Conserv. Genet*. 19:1311–1322.

Soares LS, Bolten AB, Wayne ML, Vilaça ST, Santos FR, Marcovaldi MA, Bjorndal KA. 2017. Comparison of reproductive output of hybrid sea turtles and parental species. *Mar. Biol*. 164:9.

Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, Heredia SM, Hahn MA, Caseys C, Bock DG, *et al*. 2016. Hybridization and extinction. *Evol Appl*. 9:892–908.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3–new capabilities and interfaces. *Nucleic Acids Res*. 40:e115.

Vargas SM, Araújo FC, Monteiro DS, Estima SC, Almeida AP, Soares LS, Santos FR. 2008. Genetic diversity and origin of leatherback turtles (*Dermochelys coriacea*) from the Brazilian coast. *J Hered*. 99:215–220.

Vargas SM, Jensen MP, Ho SY, Mobaraki A, Broderick D, Mortimer JA, Whiting SD, Miller J, Prince RI, Bell IP, *et al*. 2016. Phylogeography, genetic diversity, and management units of hawksbill turtles in the Indo-Pacific. *J Hered*. 107:199–213.

Vilaça ST, Lara-Ruiz P, Marcovaldi MA, Soares LS, Santos FR. 2013. Population origin and historical demography in hawksbill (*Eretmochelys imbricata*) feeding and nesting aggregates from Brazil. *J Exp Mar Biol Ecol*. 446:334–344.

Vilaça ST, Vargas SM, Lara-Ruiz P, Molfetti É, Reis EC, Lôbo-Hajdu G, Soares LS, Santos FR. 2012. Nuclear markers reveal a complex introgression pattern among marine turtle species on the Brazilian coast. *Mol Ecol*. 21:4300–4312.

Wallace BP, DiMatteo AD, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Mortimer JA, Seminoff JA, Amorocho D, Bjorndal KA, *et al*. 2011. Global conservation priorities for marine turtles. *PLoS One*. 6:e24510.

Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, *et al*. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet*. 45:701–706.

Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. 2017. HybridDetective: a workflow and package to facilitate the detection of hybridization using genomic data in r. *Mol Ecol Resour*. 17:e275–e284.

Zhang DX, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol*. 12:563–584.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. 30:614–620.