Dataset Paper

Molecular Data for the Sea Turtle Population in Brazil

Sibelle Torres Vilaça and Fabricio Rodrigues dos Santos

Laboratório de Biodiversidade e Evolução Molecular (LBEM), Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Avenida Antônio Carlos 6627, 31270-010 Belo Horizonte, MG, Brazil

Correspondence should be addressed to Fabricio Rodrigues dos Santos; fsantos@icb.ufmg.br

Received 3 June 2013; Accepted 20 June 2013

Academic Editors: G. Hill and S. Kleinsteuber

Copyright © 2013 S. T. Vilaça and F. R. dos Santos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We report here a dataset comprising nine nuclear markers for the Brazilian population of Cheloniidae turtles: hawksbills (Eretmochelys imbricata), loggerheads (Caretta caretta), olive ridleys (Lepidochelys olivacea), and green turtles (Chelonia mydas). Because hybridization is a common phenomenon between the four Cheloniidae species nesting on the Brazilian coast, we also report molecular markers for the hybrids E. imbricata × C. caretta, C. caretta × L. olivacea, and E. imbricata × L. olivacea and for one hybrid E. imbricata × C. mydas and one between three species C. mydas × E. imbricata × C. caretta. The data was used in previous studies concerning (1) the description of frequent hybrids C. caretta × E. imbricata in Brazil, (2) the report of introgression in some of these hybrids, and (3) population genetics. As a next step for the study of these hybrids and their evolution, genome-wide studies will be performed in the Brazilian population of E. imbricata, C. caretta, and their hybrids.

1. Introduction

From the seven known sea turtle species, five species nest on the Brazilian coast: leatherback (Dermochelys coriacea), green (Chelonia mydas), olive ridley (Lepidochelys olivacea), loggerhead (Caretta caretta), and hawksbill (Eretmochelys imbricata). The Brazilian sea turtle population differs from other worldwide populations because of its high hybrid frequency. Almost 43% of nesting E. imbricata individuals were reported as hybrids in a short stretch of the Brazilian coast (Bahia state), while other sites where E. imbricata nests did not show presence of hybrids [1, 2].

The sea turtles that nest in the Brazilian coast have been shown to form a separate genetic pool from other populations [3–5]. Studies with species nesting in Brazil show that their populations are differentiated from other worldwide populations. Recent telemetry studies corroborate the results from genetic data: C. caretta, E. imbricata, C. mydas, and L. olivacea individuals that nest in Brazilian beaches tend to stay in feeding aggregations within the Brazilian continental shelf [6–9]. On the other hand, Brazilian feeding aggregations are characterized by a mixture of turtles coming from different regions worldwide. Mixed-stock results based on mitochondrial DNA showed that E. imbricata feeding areas receive migrants from Africa, Caribbean, and Pacific Ocean [10]; C. caretta foraging aggregations are characterized by a mixture of Brazilian, Australian, Mediterranean, and northwestern Atlantic turtles [3]; and C. mydas feeding grounds are characterized by the contribution of other Atlantic sites [4]. Regarding L. olivacea, no genetic study dealing with feeding aggregations in Brazil was published so far.

Sea turtle populations nesting in Brazil exhibit a significant genetic differentiation from other turtle populations and are also characterized by their unique high incidence of hybrids. Even though hybrids have been reported in other populations, they were observed only sporadically (Table 1). Moreover, the Brazilian population is singular since more than one type of hybrid is present along the coast. With the exception of D. coriacea, the other four species are capable of hybridizing [1, 2]. Hybrids involving four different species are currently described, and introgression (i.e., backcrossing with one parental species) is not observed in all of them. Generally, most hybrids are F1, with only a small portion being reported as >F1.

The interspecific hybridization is recognized in several studies, and so far five hybrid types were described in Brazilian waters based on morphology and nuclear markers: the frequent hybrid E. imbricata × C. caretta; the less frequent
hybrids *C. caretta* × *L. olivacea*, *E. imbricata* × *L. olivacea*, and *C. mydas* × *L. olivacea*; and even one hybrid between three species *C. mydas* × *E. imbricata* × *C. caretta*. Even though most hybrids found are F1, this hybrid between three species is supposed to be an *E. imbricata* × *C. caretta* F1 that crossed with a *C. mydas* [2]. In a large survey, Viola et al. [2] found that feeding aggregations of *E. imbricata* and *C. caretta* within the Brazilian continental shelf do not exhibit the presence of hybrids, with only two exceptions of one *C. caretta* × *L. olivacea* hybrid being found in São Paulo state and one *C. caretta* × *E. imbricata* being found in the oceanic feeding area of Atol das Rocas.

Here we describe the data from Viola et al. [2]. The dataset presented in this paper refers to the first populational study in sea turtles using nuclear sequences. It focuses on the Brazilian population of *E. imbricata*, *C. caretta*, *L. olivacea*, *C. mydas*, and their hybrids. A total of five nuclear markers were sequenced, and four microsatellites were genotyped in samples that already had a mitochondrial locus (D-loop) typed in previous studies [1, 3]. Detailed information of the allele frequencies in several nuclear loci is obtained from the data.

### 2. Methodology

The DNA of 387 sea turtle samples from the Brazilian coast was sequenced (Figure 1). We sequenced the DNA of four Cheloniidae species that nest in Brazil: 168 samples from *C. caretta*, 121 from *E. imbricata*, 22 from *L. olivacea*, and nine *C. mydas*. We chose to analyse these four species for three main reasons: (i) to construct a detailed database of the allele frequencies in the Brazilian populations, (ii) to establish the typical alleles of each species, and (iii) to use the alleles present in each species to investigate the hybrids present in the Brazilian coast. For these presumably "pure" samples, all individuals had both morphology and mitochondrial DNA (mtDNA) of the respective species. This is a strong indication that these samples belong to "nonhybrid" individuals, since previous studies showed that hybrids had intermediate morphology (or a mix of different morphological characters)

<table>
<thead>
<tr>
<th>Hybrid cross</th>
<th>Site</th>
<th>Analysis</th>
<th>N</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chelonia mydas</em> × <em>Eretmochelys imbricata</em></td>
<td>Western Atlantic (Suriname)</td>
<td>Allozymes</td>
<td>23</td>
<td>[20]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Suriname)</td>
<td>mtDNA, scnDNA</td>
<td>1</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Eastern Pacific (Mexico)</td>
<td>mtDNA, scnDNA, morphology</td>
<td>1</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA, morphology, scnDNA, microsatellites, nuclear sequences</td>
<td>1</td>
<td>[2]</td>
</tr>
<tr>
<td></td>
<td>Western Pacific (Japan)</td>
<td>Morphology</td>
<td>?</td>
<td>[23]</td>
</tr>
<tr>
<td><em>Chelonia mydas</em> × <em>Caretta caretta</em></td>
<td>Western Pacific (Australia)</td>
<td>Morphology</td>
<td>?</td>
<td>C. Limpus, personal communication [22]</td>
</tr>
<tr>
<td></td>
<td>Western Pacific (Japan)</td>
<td>Morphology</td>
<td>?</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA, scnDNA</td>
<td>4</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Canada)</td>
<td>mtDNA, morphology</td>
<td>1</td>
<td>[25]</td>
</tr>
<tr>
<td><em>Caretta caretta</em> × <em>Eretmochelys imbricata</em></td>
<td>Western Pacific (Japan)</td>
<td>Morphology</td>
<td>?</td>
<td>[26]</td>
</tr>
<tr>
<td></td>
<td>Western Pacific (China)</td>
<td>Morphology</td>
<td>1</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Brazil)</td>
<td>Allozymes</td>
<td>1</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (USA)</td>
<td>mtDNA, scnDNA</td>
<td>2</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA, morphology, scnDNA, microsatellites, nuclear sequences</td>
<td>50</td>
<td>[1, 2]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA</td>
<td>10</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (USA)</td>
<td>Morphology</td>
<td>1</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td>Western Pacific (Japan)</td>
<td>Isozymes</td>
<td>?</td>
<td>[31]</td>
</tr>
<tr>
<td><em>Lepidochelys kempii</em> × <em>Caretta caretta</em></td>
<td>Western Atlantic (USA)</td>
<td>mtDNA, scnDNA</td>
<td>1</td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>Western Atlantic (USA)</td>
<td>mtDNA, microsatellites</td>
<td>3</td>
<td>[32]</td>
</tr>
<tr>
<td><em>Lepidochelys olivacea</em> × <em>Chelonia mydas</em></td>
<td>Western Atlantic (Brazil)</td>
<td>Morphology</td>
<td>?</td>
<td>M. Marcovaldi, personal communication [22]</td>
</tr>
<tr>
<td><em>Lepidochelys olivacea</em> × <em>Eretmochelys imbricata</em></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA, morphology, scnDNA, microsatellites, nuclear sequences</td>
<td>2</td>
<td>[1, 2]</td>
</tr>
<tr>
<td><em>Lepidochelys olivacea</em> × <em>Caretta caretta</em></td>
<td>Western Atlantic (Brazil)</td>
<td>mtDNA, morphology, scnDNA, microsatellites, nuclear sequences</td>
<td>14</td>
<td>[2, 3]</td>
</tr>
</tbody>
</table>

1 All turtles belonged to a single hatch.
and mtDNA from different species. These loci were used to describe the genetic diversity within species and to establish the typical (private) alleles for each species. Caution was taken for areas where hybrids had been previously reported; except for the nesting sites in Bahia and Sergipe coastlines, no hybrid was previously registered among nesting or bycatch individuals from the sampling sites. This is particularly important since samples from presumably “pure” individuals from these areas (Bahia and Sergipe) were taken under extra care in establishing private alleles, since they could be hybrid samples.

Of the 387 samples, 66 individuals previously identified as hybrids (morphology of one species and mtDNA from a different one) were analysed with nuclear markers. Those included 50 hybrids of *C. caretta* × *E. imbricata*, two hybrids of *E. imbricata* × *L. olivacea*, and 14 hybrids of *L. olivacea* × *C. caretta* analysed.

Some samples of *C. caretta* × *E. imbricata* hybrids were especially interesting, since they allowed a more detailed view of the hybridization process. Samples of four siblings derived from a single clutch (R0264, R0265, R0267, and R0268) were collected in Praia do Forte, Bahia, and possessed *C. caretta* mitochondria, but the morphology indicated a possible hybridization between *E. imbricata* and *C. mydas*. Another sample used included one hatchling (R0025) of a *C. caretta* × *E. imbricata* hybrid female (R0024). Both samples had mtDNA from *C. caretta*. Besides the four siblings from a single clutch and the hatchling R0025, all other hybrid samples were adult nesting females.

We have also included one bycatch sample (R0384) that was previously classified by morphology as *C. caretta* but identified by mtDNA as a *L. olivacea* × *C. caretta* hybrid from the São Paulo State.

A total of five nuclear markers were sequenced to evaluate the presence of interspecific variation. We used four exons (brain-derived neurotrophic factor (*BDNF*), oocyte maturation factor (*CMOS*), and two recombination activating genes (*RAG1* and *RAG2*)) and one intron (RNA fingerprint protein 35 gene (*R35*)) to identify species-specific alleles and their frequency in hybrids.
PCR reaction mixes of 15 µL included 50 ng of genomic DNA, 1 U of Taq polymerase (Phoenutria), 200 µM of dNTPs, 1X Tris-KCl buffer with 1.5 mM MgCl₂ (Phoenutria), and 1 µM of each primer. PCR enhancers and primer sequences used for each amplified locus are shown in Table 2. The amplification program consisted of 3 min at 94°C, followed by 35 cycles of 40 s at 94°C, 45 s at 45–50°C, 50 s at the annealing temperature of each primer, and a final extension step of 10 min at 72°C. After amplification, PCR products were checked by running in a 0.8% agarose gels and stained with ethidium bromide. These products were cleaned by precipitation using 20% polyethylene glycol and 2.5M NaCl before loading to the sequencing reactions, which were performed using either of the amplification primers. The sequencing reaction was performed with ET DYE Terminator Kit (GE Healthcare) according to the manufacturer's instructions. Then, sequencing products were precipitated and run in the automatic sequencer MegaBACE 1000 (GE Healthcare).

High-quality consensus sequences were obtained using the programs Phred [11], Phrap [12], and Consed 16.0 [13]. The consensus sequences of the autosomal loci were aligned by the Clustal X algorithm implemented in MEGA5 [14] together with the two E. imbricata and C. caretta reference sequences published by Naro-Maciel et al. [15]. Polymorphic sites were identified by visual inspection in Consed or using Polymphed 6.11 [16, 17].

We genotyped four autosomal microsatellites developed for L. olivacea and C. caretta. These loci included OR1 and OR3 [18], and Cc1G02 and Cc1G03 [19]. All genotypes were evaluated to determine species-specific alleles.

Polymerase chain reaction (PCR) mixes of 9 µL included 1 µL of genomic DNA (~40 ng), 1 U of Taq Platinum polymerase (Invitrogen), 200 µM of deoxynucleoside triphosphates, 1X Tris-KCl buffer (Invitrogen), 1.5 mM MgCl₂ (Invitrogen), 1 mM of the forward primer labeled with an m13 tail, 10 mM of the reverse primer, and 10 mM of the m13 primer with fluorescence FAM or HEX. The amplification program consisted of 3 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at annealing temperatures depending on the locus (55°C for OR1, OR2, and OR3 and 60°C for Cc1G02 and Cc1G03), 30 s at 72°C, and a final extension step of 30 min at 72°C. The amplicons were diluted fivefold with Milli-Q water. Genotyping reaction mixes of 10 µL included 2 µL of diluted amplicon, 0.25 µL of ET 550-R (GE Healthcare), and 7.75 µL of Tween20 0.1%. The running conditions followed the manufacturer's recommendations (GE Healthcare) for genotyping in an automated MegaBACE 1000 DNA analysis system. The peaks were analyzed in the Fragment Profiler Program (GE Healthcare) for allele scoring.

Dataset was constructed as follows. From the 387 samples typed for the nine nuclear markers, we selected the samples with a maximum of two missing markers. We did this selection, so the analysis to infer the genetic diversity of the turtle populations, and hybrid inference were not affected by the missing data. The final dataset was composed of 223 samples of the four species, and their hybrids, with at least seven typed loci. The only exception was a hybrid from E. imbricata × L. olivacea and one C. mydas sample, both typed for six loci.

Table 2: Primer sequences used to amplify the nuclear sequences with annealing temperatures and enhancers used in the PCR reaction.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Annealing temperature (°C)</th>
<th>Enhancers</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAG1F</td>
<td>AGTCATCTCTTGCCAGGTC</td>
<td>61.5</td>
<td>BSA IX</td>
</tr>
<tr>
<td>RAG1R</td>
<td>CAGCAGGAAACAAGTTAGGC</td>
<td>68</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>RAG2F</td>
<td>CTGTATCTTCCCCCTTC</td>
<td>55.3</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>RAG2R</td>
<td>GTTGTCACCTGGTAGGCC</td>
<td>68</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>R35F</td>
<td>CAAAGTATCTCCTTTCG</td>
<td>65.5</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>R35R</td>
<td>CAGCCTCTGTATCGAAAGG</td>
<td>68</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>CMOSF</td>
<td>ATTTGTCCTACTACAGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMOSR</td>
<td>ATATGTGCCCCCTGTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDNFF</td>
<td>TCTGGAGAGCCTAAGTGGG</td>
<td>65.5</td>
<td>Gelatin 0.001%</td>
</tr>
<tr>
<td>BDNFR</td>
<td>TAAACCAGCACGAACTC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Dataset Description

The dataset associated with this Dataset Paper consists of 8 items which are described as follows.

Dataset Item 1 (Table). The detailed genetic data obtained. The column popID is an identification number of each population, given by a combination of morphology and mtDNA and, therefore, previous to the analysis with nuclear loci. Number 1 refers to E. imbricata × C. caretta hybrids, number 2 to E. imbricata × L. olivacea hybrids, number 3 to E. imbricata “pure” samples, number 4 to C. caretta “pure” samples, number 5 to L. olivacea × C. caretta hybrids, number 6 to L. olivacea “pure” samples, and number 7 to C. mydas “pure” samples. The column Sample Code is an identification code for each sample. The codes starting with R0XX are deposit codes in the DNA Bank DB-LBEM at the Federal University of Minas Gerais. The column Hybrid/Species is an identification code that summarizes the results obtained with the nuclear data and classifies the sample in a category of hybrid or pure individual. The code EixCc refers to an E. imbricata × C. caretta hybrid; Ei refers to an E. imbricata pure individual; Cc refers to a C. caretta pure individual, EixCcxCm refers to an E. imbricata × C. caretta × C. mydas hybrid, EixCm refers to an E. imbricata × C. mydas hybrid,
EixLo refers to an *E. imbricata* × *L. olivacea* hybrid, LoxCc refers to an *L. olivacea* × *C. caretta* hybrid, Lo refers to an *L. olivacea* pure individual, and Cm refers to a *C. mydas* pure individual. The column Morphology gives the species classification based on morphology. Codes are the same as in the column Hybrid/Species. The column mtDNA refers to the mitochondrial results from the locus D-loop. Codes are the same as in the column Hybrid/Species. The columns 6 to 23 identify the alleles found for each of the loci typed. Columns OR1 Allele 1 and OR1 Allele 2 identify the two alleles for the microsatellite loci OR1. Each column is one allele present in this locus. Columns OR3 Allele 1 and OR3 Allele 2 are the alleles for the microsatellite loci OR3. Columns CCIG02 Allele 1 and CCIG02 Allele 2 are the alleles for the microsatellite loci CCIG02. Columns CCIG03 Allele 1 and CCIG03 Allele 2 are the alleles for the microsatellite loci CCIG03. Columns RAG1 Allele 1 and RAG1 Allele 2 are the alleles for the locus RAG1. Columns CMOS Allele 1 and CMOS Allele 2 are the alleles for the locus CMOS. Columns RAG2 Allele 1 and RAG2 Allele 2 are the alleles for the locus RAG2. Columns R35 Allele 1 and R35 Allele 2 are the alleles for the locus R35. Columns BDNF Allele 1 and BDNF Allele 2 are the alleles for the locus BDNF. No data is shown by a question mark. In the table, the asterisk (*) indicates samples and/or loci with introgression with *E. imbricata*, the double asterisk (**) indicates samples and/or loci with introgression with *C. caretta*, and the triple asterisk (***) indicates samples and/or loci with introgression with *L. olivacea*.

**Dataset Item 2 (Table).** The haplotypes (alleles) found. The column Haplotype refers to the haplotype identification code. The column Gene indicates the locus which the haplotype (allele) was found. A total of five codes are found in this column: BDNF, R35, RAG1, RAG2, and CMOS, which represent the name of the nuclear locus sequenced. The column Species refers to which species the haplotype is typical. The code Lo refers to *L. olivacea*, Cm refers to *C. mydas*, Ei refers to *E. imbricata*, Cc refers to *C. caretta*, and Ei/Cc refers to a haplotype found in both *E. imbricata* and *C. caretta*, and Ei/Lo refers to a haplotype found in *E. imbricata* and *L. olivacea*. The last column, GenBank Accession Number, refers to the number of identification in the GenBank.

**Column 1:** Haplotype  
**Column 2:** Gene  
**Column 3:** Species  
**Column 4:** GenBank Accession Number

**Dataset Item 3 (Table).** The GenBank accession numbers for each sample. No data is shown by a question mark. Each gene is represented by two columns, corresponding to the two alleles found.

**Column 1:** Sample Code  
**Column 2:** RAG1 Allele 1  
**Column 3:** RAG1 Allele 2  
**Column 4:** CMOS Allele 1  
**Column 5:** CMOS Allele 2  
**Column 6:** RAG2 Allele 1  
**Column 7:** RAG2 Allele 2  
**Column 8:** R35 Allele 1  
**Column 9:** R35 Allele 2  
**Column 10:** BDNF Allele 1  
**Column 11:** BDNF Allele 2

**Dataset Item 4 (Nucleotide Sequences).** Sequences with the BDNF exonic alignment. The five aligned sequences are identified as haplotype number followed by the GenBank reference number.

**Dataset Item 5 (Nucleotide Sequences).** Sequences with the CMOS exon alignment. The eleven aligned sequences are identified as haplotype number followed by the GenBank reference number.

**Dataset Item 6 (Nucleotide Sequences).** Sequences with the R35 intron alignment. The thirteen aligned sequences are identified as haplotype number followed by the GenBank reference number.

**Dataset Item 7 (Nucleotide Sequences).** Sequences with the RAG1 exon alignment. The nine aligned sequences are...
identified as haplotype number followed by the GenBank reference number.

**Dataset Item 8 (Nucleotide Sequences).** Sequences with the RAG2 exon alignment. The six aligned sequences are identified as haplotype number followed by the GenBank reference number.

**Dataset Availability**

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at http://dx.doi.org/10.1155/2013/196492. In addition, Dataset Items 1 and 2 associated with this Dataset Paper are available at doi:10.1111/j.1365-294X.2012.05685.x, and Dataset Items 3–8 are available under DRYAD entry doi:10.5061/dryad.j5240.

**Conflict of Interests**

There is no conflict of interests in the access or publication of this dataset.

**References**


Submit your manuscripts at http://www.hindawi.com