[**https://github.com/SandraLudwig/Evolutionary-Insights.git**](https://github.com/SandraLudwig/Evolutionary-Insights.git)

**Supplementary Material of**

**Evolutionary insights on demography of the Loggerhead turtles**

*Sampling*

**Table S1.** Sampling details of *Caretta caretta* populations of the Southwest Atlantic.

| **Nesting area** | **Monitoring stretch (km)** | **Nesting seasons (years)** | **N** | **Reference** |
| --- | --- | --- | --- | --- |
| Rio de Janeiro (RJ) |  | 2005/2006 | 49 | Shamblin et al., (2014) |
| Espírito Santo (ES) |  | 1999-2021 | 236 | this study |
| Bahia (BA) |  | 1999-2020 | 62 | this study |
| Sergipe (SE) |  | 2004-2011 | 35 | this study |

*Laboratory protocols*

We isolated the genomic DNA (gDNA) from the samples using a saline solution (Bruford et al. 1992) and CTAB 2% (Doyle & Doyle 1987) resuspending the DNA pellet in TE Buffer until reaching 50 ng / µL. The gDNA was quantified using NanoDrop ND100 spectrophotometer (Thermo Scientific) and, the quality was check through electrophoresis in 1% agarose gel, Ladder 100 base pairs (bp), and were observed using a UV trans-illuminator (L-PIX Touch 20x20 Loccus©). We amplified the mtDNA control region (D-loop) in both strands using the LCM15382 and H950 primers (Abreu-Grobois et al., 2006) through Polymerase Chain Reactions (PCR’s). Each xx µL PCR including xx ng of gDNA, x x reaction buffer, xxx of MgCl2,xx µM of dNTPs, xx µM of both primers, and xx µL of Taq xxx (marca?). The PCR cycling profile conditions were as follows: initial denaturation at 94ºC for 3 min, followed by 35 cycles of denaturation step of 94ºC for 30 s, annealing at 51°C for 30 s, and extension at 72°C for 40 s, followed by a final extension at 72°C for 10 min. PCR products were sequenced in both directions using *Big Dye Terminator* components (Applied Biosystems) according to the manufacturer's protocols. The genotyping was carried out using 15 microsatellite loci of the nDNA: Cc7G11, Cc1F01, Cc1G02, Cc1G03, CcP7D04, CcP2F11, CcP7C06, CcP8D06, CcP1F09, CcP5C11, CcP1F01, CcP1G03, CcP1B03, CcP5C08, and CcP5H07 (Schuelke 2000; Shamblin et al., 2007; 2009). The microsatellite were splitted in three multiplexes with distinct fluorophosphorus, such as: 1) NED-CcP1G02, FAM-CcP7C04, VIC-CcP7B07, NED-CcP7E11, and FAM-CcP8E07; 2) NED-CcP2H12, FAM-CcP2G10, VIC-CcP1H11, PET-CcP5F01, VIC-CcP7G11, and FAM-CcP7C06; 3) FAM-CcP1F01, NED-CcP1G03, VIC-CcP1B03, FAM-CcP5C08, and PET-CcP5H07. The amplifications were performed in 12.5 µL total volume including 1 ng of gDNA, 1x reaction buffer, xxx of MgCl2,100 µM of dNTPs, 0.2 µM of both primers, 0.2x fluorophorus, and 0.2 µL of Taq xxx (marca?).

*Data analysis*

*Geneland* was chosen because it detects spatial genetic population structure based on the distribution of allelic frequencies (SSR´s dataset) by populations and correlates with geographic distance (UTM coordinates) through an optimal value of K estimation. was set using a correlated alleles model, 500.000 iterations, thinning of 200, a burn‐in of 500. Default settings were used for the maximum rate of the Poisson process, and the maximum number of nuclei in the Poisson–Voronoi tessellation. The most likely value of K was determined by examining the posterior probabilities averaged over multiple runs (10 runs, K = 1–10), and choosing the K value with the highest average posterior probability.

First, we evaluated if there is any genetic barrier between the SWA nesting areas that could explain the great spatial population structure detected (previously by Ludwig et al., in press, MABI, and this study). To reach that, we used Monmonier’s maximum difference algorithm (Monmonier, 1973) using SSR´s dataset through Delaunay triangulation (Brassel & Reif, 1979) that is implemented in *adegenet* package for R (Jombart et al., 2018). The Monmonier and Delaunay triangulation were chosen because have been successfully applied to find boundaries of maximum differences between contiguous polygons of a tessellation, especially to evaluate whether or not there is a genetic and/or geographic barrier and whether or not a correlation has a spatial pattern (Manni et al., 2004; Storfer et al., 2007; Storfer et al., 2010), as detected herein.

Second, we evaluated for an asymmetric migration and gene flow (i.e., migration occurring at a significantly higher rate in one direction than another), in *diveRsity* (*divMigrate* function) for R (Keenan et al. 2013; Sundquist et al., 2016), according to the allele frequency data with a 0.35 filter threshold (e.g., Maas et al. 2018; Ludwig et al., 2021) through the *D* genetic index (Crawford 2010).

Third, we estimated the number of migrants per generation (*Nm*) through a Bayesian Inference using the D-loop dataset in DnaSP under 10,000 and 50,000 recorded Markov chain Monte Carlo (MCMC) steps. Then, in GENECLASS2 (Piry et al., 2004), the last generation of migrants are identified through an assignment test that calculates for each individual of a population the probability of belonging to any other reference population/region or to be a resident of the population/region where it was sampled. This analysis uses the SSR´s dataset and Bayesian estimation of population allele frequencies attributing the Rannala & Mountain, (1997) criterion that calculates the exclusion probabilities through the likelihood method of *L\_home/L\_max\_not\_home* (Paetkau et al., 1995) and Monte Carlo resampling methods (Paetkau et al., 2004), default allelic frequency of 0.01, simulating for 1.000.000 individuals, and a threshold probability (p-value) of 0.05. The migrants were assigned when the probability was ≥ 0.05, when it was ≤0.009 the migrant was considered as coming from an unsampled population, and multiple assignments were allowed.

First, the *Ne* was estimated using SSR´s dataset and performed in NeESTIMATOR v2 (Do et al. 2014), which estimates the allele frequencies differences through the linkage disequilibrium method (LD; Waples 1989) between the SWA nesting areas. This method was chosen because LD assumes that random departures from linkage equilibrium of unlinked loci are inversely proportional with *N*e and takes into account the sample size (Waples and Do 2008), and due to its superior performance in ideal scenarios (Gilbert & Whitlock 2015) when migration rates are low/moderate, with random mating and the sample size is not fixed, the sex is uneven, the population is subdivided, which fits with our study system. We set 0.05 as the lowest allele frequency, Jackknife option as 95% confidence intervals (CIs).

Second, we investigated the past population demography signatures (i.e., expansion, bottleneck, secondary contact) of *C. caretta* of the SWA, using the D-loop dataset, via neutrality tests (Tajima´s D and Fu´s F), which are used to uncover recent demographic events (i.e, population expansion/contraction) and to detect selection in cases where patterns of DNA polymorphism deviate from those predicted by the Wright-Fisher neutral model of evolution (Tajima 1989; Fu and Li 1993; Fu 1996; Fu 1997). Third, a Mismatch Distribution (MD) was estimated the mean pairwise differences (*k*) and Raggedness index (*r*) (Harpending, 1994) from each individual D-loop sequence and tested for significance through 1.000 permutations under the model of sudden expansion (Rogers, 1995).

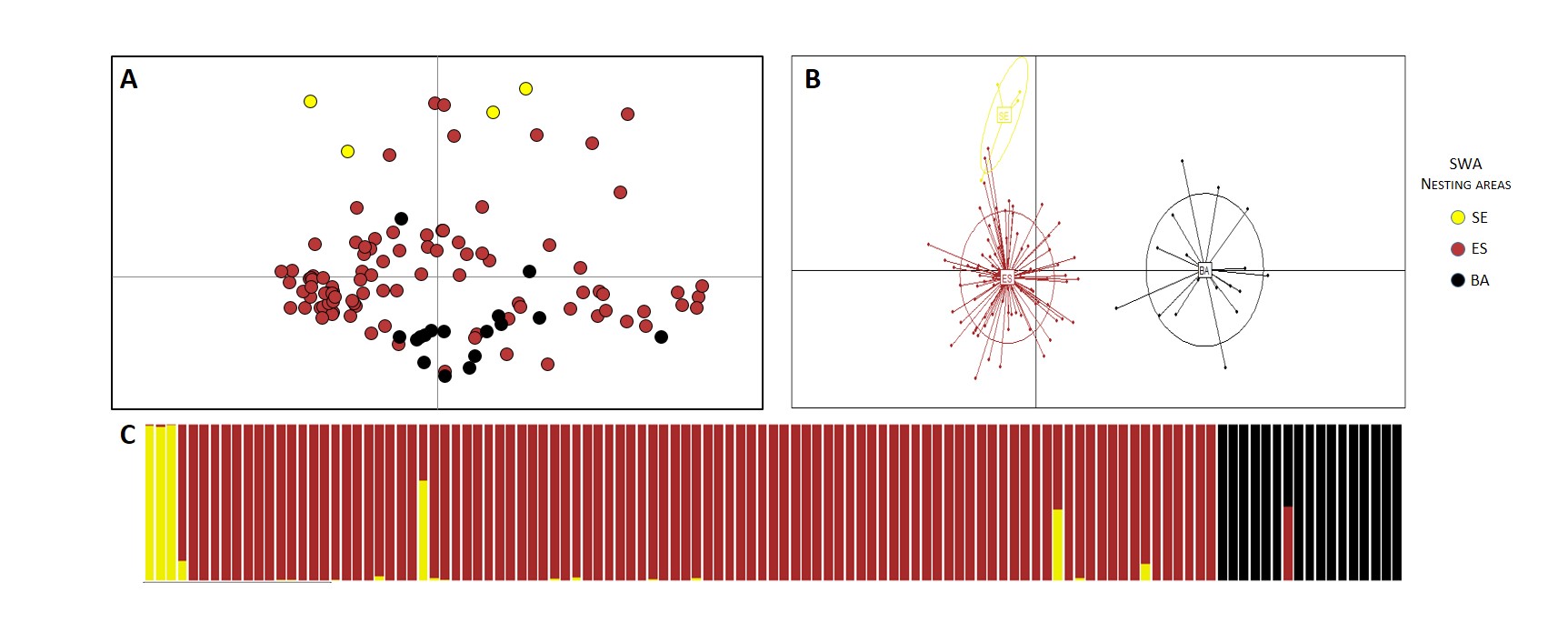
Third, in BOTTLENECK 1.2.02 (Cornuet & Luikart, 1996; Piry et al., 1999) we calculated the expected distribution of alleles frequencies and the significant number of loci with excess *He* based on the stepwise (SMM; Cornuet and Luikart 1996) and the two-phased (TPM; Fu and Chakraborty 1998) mutation models with 95% single-step mutations and 5% multi-step mutations in BOTTLENECK 1.2.02 (Cornuet & Luikart, 1996; Piry et al., 1999). Such analysis detects reductions within the past 2*Ne*–4*Ne* and is expected to show excess heterozygosity in comparison to the number expected under mutation-drift equilibrium. This phenomenon represents the loss of rare alleles and could be assessed by looking at the overall distribution of allele frequency classes (Cornuet & Luikart 1996). The significance of excess heterozygosity was assessed using a one-tailed Wilcoxon sign-rank test, as recommended for fewer than 20 loci (Luikart, 1997), and probability (*P*) for 1,000 simulations.

Forth, was inferred the historical change in the *Ne* over past generations attributing the *Bayesian Skyline Plot* (BSP) method based on the SSR´s dataset in *VarEff* package for R (Nikolic & Chevalet 2014), simulating demographic histories models by steps of constant size for which the posterior probabilities are derived using an approximate likelihood Markov Chain Monte Carlo (MCMC) approach. This analysis simulates past demography (SSR´s dataset) with a coalescent approach, to estimate changes in recent and ancestral *Ne* that rely on motif distance frequencies between alleles to estimate variation in *N*e over time (Global θ = 4*Ne*µ; µ = mutation rate) (Nikolic & Chevalet 2014). Demographic changes were first modeled with MISAT software (Nielsen 1997) using a Two-Phase Mutation model with µ = xxxxx (estimated mutation rate) and 10% of the mutations greater than a single step. Then, we set four independent MCMC runs simulating for 10 million generations sampled at every 1,000 trees, and 25% burn-in. The sexual generation time for *C. caretta* was set to xxx years, as indicated by Vilaça et al. (2021), and we transformed the generation times in time in the past as a pattern. The number of generations ranged from 100 to 1,000 since the origin of each population. For each population, models were run with three separate prior *N*e (minimum, intermediate, and maximum values) through NBAR function that attributes three Global θ as means of effective size being: θ0 of xxxx as *Ne* minimum, θ1 of xxxxx as *N*e intermediate; and θ2 of xxxxx as *N*e maximum. In the last 500 generations, changes to *Ne* were assessed. Inferences were made about recent past expansions and declines based on posterior distributions and their attributes (mean, median, mode, and harmonic values) following Nikolic & Chevalet (2014). The results were displayed graphically, using the *ggplot2* R package (Wickham 2011), for the last 100 years and with a 95% confidence interval (CI).

*Results*

**Table S2.** Deviation from Hardy-Weinberg equilibrium (HWE) in global tests per locus and for all loci and all populations, and the average frequencies of null alleles per locus.

**Figure S1.** Supplementary genetic results on the spatial population structure of *Caretta caretta* of the Southwest Atlantic nesting areas, detected by: (A) Principal Components Analysis; and (B-C) Discriminant Analysis of Principal Components through find.clusters and compoplot functions of adegenet R package.



**Table S3.**