cim.UH
Centro de Investigaciones Marinas
Universidad de La Habana

Revista de Investigaciones Marinas

RESEARCH ARTICLE

Assessing genetic structuring for endangered *Chelonia mydas* (Testudines: Cheloniidae) in southwest Cuba using microsatellites.

Evaluación de la estructuración genética de *Chelonia mydas* (Testudines: Cheloniidae), en peligro de extinción, en el suroeste de Cuba utilizando microsatélites

Luis Javier Madrigal-Roca [1]
Julia Azanza-Ricardo [2*]
Georgina Espinosa-López [1]
Ken Oyama [3]
F. Alberto Abreu-Grobois [4]
Omar Chassin-Noria [5]

[1] Facultad de Biología, Departamento de Bioquímica, Grupo de Genética Aplicada. Calle25, entre I y J, No. 455, Vedado, La Habana, Cuba.

[2] Instituto Superior de Tecnologías y Ciencias Aplicadas (InSTEC). Ave. Salvador Allende, entre Infanta y Rancho Boyeros, No. 1110, Plaza de la Revolución, La Habana Cuba.

[3] Escuela Nacional de Estudios Superiores (ENES) Unidad Morelia, UNAM, Morelia, México. Antigua Carretera a Pátzcuaro 8701, Sin Nombre, Indeco la Huerta, 58190 Mich., México.

[4] Laboratorio de Genética, Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología UNAM, Mazatlán, Sinaloa, México. Cap. Joel Montes Camarena s/n, Cerro del Vigía, 82047 Mazatlán, Sin., México.

[5] Facultad de Biología, CMEB, Universidad Michoacana de San Nicolás de Hidalgo. Morelia Michoacán, México. Gral. Francisco J. Múgica S/N A-1, Felicitas de Río, 58030 Morelia, Mich., México.

Corresponding author:
julia_dragmarino@yahoo.es

Editor:
Erik García-Machado
Laval University.
ULaval Québec, Canadá.

Received: 07.02.2022
Accepted: 09.12.2022

Abstract

Understanding the population genetic structure of the species is essential for determining the possible management units (UM) and their conservation and/or sustainable exploitation with it. *Chelonia mydas* is recognized as an endangered philopatric turtle. This work aims to describe the population structure of the green turtle in southwestern Cuba through traditional analytical approaches and allocation methods. The collections were made between 1998 and 2007 on five beaches in the Cuban southwest. Seven microsatellite *loci* from 149 individuals were amplified and genetic variability parameters were calculated. The population structure was inferred through the use of Wright's F, Analysis of Molecular Variance (AMOVA), and population assignment algorithms based on Bayesian analysis (STRUCTURE) and factorization of sparse non-negative matrices (implemented in R). Most of the *loci* were not in Hardy-Weinberg equilibrium, and several presented linkage disequilibrium. The AMOVA and differentiation statistics suggest the presence of structure at the geographical level analyzed. The highest value of ΔK and the lowest value of cross-entropy were reached for K = 2, a result that suggests that in southwestern Cuba there is the contribution of two ancestral populations of *Chelonia mydas*. Relative migration estimates indicate active genetic exchange between nesting colonies in southwestern Cuba.

Keywords: differentiation, turtle, nesting, assignation, migration.

Resumen

La comprensión de la estructura genética poblacional de las especies es esencial para la determinación de las posibles unidades de manejo (UM) y con ello su conservación y/o explotación sostenible. *Chelonia mydas* es reconocida como una tortuga filopátrida en peligro de extinción. El objetivo de este trabajo es describir la estructura poblacional de

la tortuga verde en el suroccidente de Cuba mediante enfoques analíticos tradicionales y métodos de asignación. Las recolectas se realizaron entre los años 1998 y 2007 en cinco playas del suroccidente cubano. Fueron amplificados siete *loci* de microsatélites de 149 individuos y se calcularon parámetros de variabilidad. La estructura poblacional fue inferida mediante el uso de estadísticos F de Wright, Análisis de Varianza Molecular (AMOVA) y algoritmos de asignación poblacional basados en análisis bayesianos (STRUCTURE) y en factorización de matrices no negativas poco densas (implementado en R). La mayoría de los *loci* no se encontraron en equilibrio de Hardy-Weinberg y varios de estos presentaron desequilibrio de ligamiento. El AMOVA y los estadísticos de diferenciación sugieren presencia de estructura al nivel geográfico analizado. El mayor valor de ΔK y el menor valor de entropía cruzada se alcanzaron para K = 2, resultado que propone que en el suroccidente de Cuba existe el aporte de dos poblaciones ancestrales de *Chelonia mydas*. Los estimados de migración relativa indican un intercambio genético activo entre colonias de anidación del suroccidente cubano.

Palabras clave: diferenciación, tortuga, anidación, asignación, migración.

Introduction

Sea turtles are umbrella species of ocean ecosystems and flag species of conservation (Zhang *et al*., 2021). However, the global population and distribution of these organisms have been markedly reduced due to threats such as overexploitation, bycatch in fishing nets, egg harvesting, marine pollution, habitat loss, and climate change (Seminoff *et al*. al., 2003, Fitak & Johnsen, 2018, Phu *et al*., 2021).

The green sea turtle, *Chelonia mydas* (Linnaeus 1758) is one of six extant species of hard-shelled sea turtles within the family Cheloniidae. Its populations are widely distributed in tropical and subtropical oceans, and it has been listed as "Endangered" by the IUCN since 2006 (Phu *et al*., 2021). The green turtle (*Chelonia mydas*) has a global distribution and is characterized by strong female fidelity to the natal nesting area, moderate male-mediated gene flow, and population overlap in feeding grounds (Álvarez-Varas *et al*., 2021).

Cuban nesting sites are centrally located in the migratory circuits of the Greater Caribbean (Moncada *et al*., 2006), and several reports reflect the abundance of nesting sites in the Cuban Archipelago (Moncada *et al*., 2014; Azanza-Ricardo *et al*., 2018). The Guanahacabibes Peninsula, as well as other coasts and key systems of the southwestern portion of Cuba, constitute nesting areas for sea turtles in the Cuban Archipelago (Azanza-Ricardo *et al*., 2018) of which *C. mydas* is the most abundant species (Moncada *et al*., 2014). For this reason, for more than 20 years various conservation and research activities have been developed in the region (Azanza-Ricardo *et al*., 2019). The vulnerability of sea turtles to environmental changes and, above all, human action, entails monitoring their populations to carry out effective management to prevent their extinction (Poloczanska *et al*., 2009, Azanza-Ricardo *et al*., 2018, Matley *et al*., 2019).

The evaluation of the structure of the populations is of considerable interest because it constitutes the previous step to answering more complex questions related to migration rates and the identification of conservation units. However, for many species, the demarcation of geographic populations is problematic, and, in these cases, clustering methods provide the best solution (Manel *et al*., 2005). Although identifying individual's populations of origin may appear to be a relatively straightforward problem, it requires knowledge of the number and structure of populations, as well as the appropriate characters and statistical tools for assigning individuals to populations (Hansen *et al*., 2001).

Two of the goals of conservation genetics are the identification of unique evolutionary units (EU) as targets for the management and the maintenance of genetic diversity within those units. Microsatellite analysis can contribute to both goals, even more efficiently than other

techniques that have been applied in this field, such as mitochondrial DNA studies (Ashley & Dow, 1994). Recently, Microsatellites are been used for population genomics since microsatellites are more sensitive than genomic datasets to the number of loci and individuals, although less sensitive to the number of populations sampled (Aguirre-Liguori *et al*., 2020). Evolutionary genetic studies are also using microsatellites (Jorde *et al*., 2001). Microsatellites are highly variable, noncoding regions of nuclear DNA consisting of repeats of single sequence motifs, usually with more than ten alleles (Goldstein & Pollock, 1997). Another practical advantage of microsatellites is that only small amounts of DNA are required (Gulcher, 2012), making possible non-lethal sampling and analysis of old archived samples with only small amounts of heavily degraded DNA (Hutchinson *et al*., 1999). However, the greatest advantage of this type of molecular marker is the numerous and novel statistical tests that can be implemented on them, aimed at detecting bottlenecks, migration rates and relationships between individuals (Hansen *et al*., 2001).

Population differentiation is a key process of evolution. Consequently, the measurement and quantification of this phenomenon are of central importance. Accordingly, Wright (1951) founded the so-called *F* statistics, also known as fixation indices. Recently, new indices and approximations have been described as alternatives to the F_{ST} parameter, such as the D index (Ma *et al*., 2015). On the other side, assignment methods, which make use of genetic information to decipher the population membership of individuals or groups of individuals, have been used to study a wide range of evolutionary and ecological processes (Manel *et al*., 2005). Statistics based on individuals for assigning these to populations have caught the attention of the scientific community and, consequently, several methods and software have been explicitly developed to respond to this type of problem (Hansen *et al*., 2001).

All methods of assigning individuals to a population have as a common property that they employ the enormous amount of information available on the multi-*locus* genotypes of individuals. For each individual analyzed, the probability of belonging to each of the possible populations is calculated. Individuals are then assigned to the population from which their multi-*locus* genotype is most likely to be derived. The main difference between techniques consists in the principles according to which the assignments are made (Hansen *et al*., 2001; Manel *et al*., 2005). In general, the power of assignment tests depends on a set of factors: 1) genetic differentiation between populations; 2) the number of populations sampled; 3) degree of polymorphism in the *loci*; 4) the number of *loci* studied and 5) sample sizes (Hansen *et al*., 2001).

In grouping problems, the categories are not predefined and must be constructed from the data. Clustering methods attempt to break down the mix of individuals from different populations by creating groups within which linkage disequilibrium is minimized. These groups can then be considered as gene pools or populations. Thus, clustering methods can simultaneously delineate clusters of individuals based on their multi-*locus* genotypes and assign individuals to the identified groups, typically employing a Monte Carlo Markov chain (MCMC) approach. These analyses are particularly useful when genetic information for potential source populations is not available, the boundaries between populations are unclear, or when some (but perhaps not all) potential sources have been sampled (Manel *et al*., 2005).

Within the framework of the Cuban Archipelago, few studies have been devoted to inferring the population structure of the green turtle. Among them are the one carried out by Ruiz-Urquiola *et al*. (2010) with mitochondrial markers and Azanza-Ricardo (2009), which combines mitochondrial markers and microsatellites. They used methods based on differentiation statistics (*e. g.* F_{ST}), AMOVAs, and Bayesian algorithms (STRUCTURE). This study implement an integrative study based on multiple approaches (classics statistics like

F_{ST}, Bayesian assignment methods, and sparse non-negative matrix factorization algorithms) to describe the population structure of *C. mydas* using microsatellite *loci* in nesting localities of the southwestern region of Cuba.

Materials and methods

Study area

The study area covered the most important nesting rookeries in southwestern Cuba. The biological material was obtained from 149 individuals (N) from Cayo Largo (28), El Guanal on Isla de la Juventud (12), Cayo Real (Cayería de San Felipe; 8) and on the beaches of East (Antonio and Perjuicio; 40), Center (La Barca and El Holandés; 35) and West (Caleta de los Piojos; 26) of the Guanahacabibes Peninsula (Table I). The samplings were made between 1998 and 2007, from the months of May to September (nesting season of the green turtle). The samples were preserved in ethanol (90%) at room temperature.

Extraction and amplification of microsatellite DNA sequences

DNA was extracted from 30-40 mg of muscle tissue, using a modification of the protocol of Hillis *et al.* (1996). The extracted DNA was amplified for seven microsatellite *loci*, using as 5'-3 'primers those proposed by Aggarwal *et al.* (2004) [OR-2, OR-3, and OR-8] and FitzSimmons *et al.* (1995) [Ei8, Cm72, Cm84, and Cc117]. The selected *loci* are moderately variable and useful to discriminate between populations and not individuals.

Microsatellite genotyping

The microsatellite PCR products were mixed with a molecular weight marker (Liz 500-250) and denatured for five minutes at 95 °C with HiDi formamide. The mixture was loaded in an ABI 3730 sequencer (Applied Biosystems). Each genotype was then scored after analyzing the amplification products with the

Table I: Summary of global data from microsatellites analyses in the Cuban southwest green turtle *Chelonia mydas* (N = 149). Except for the number of non-missing observations (**n**), all the other statistics were calculated after the random imputation of data according to the description present in materials and methods.

Locus	Reference	Repeat motif	Size ranges (pb)	n	A	A_e	H_O	H_E	F_{IS}	F_{ST}	F_{IT}	R_{ST}	HW p
Ei8	FitzSimmons *et al.* (1995)	(CA)19	125-175	145 / 145	7	2,59	0,87	0,62	-0,45	0,03	-0,41	0,02	0,00
CM84	FitzSimmons *et al.* (1995)	(CA)15	304-356	136 / 136	22	12,81	0,91	0,93	-0,02	0,04	0,02	0,01	0,00
CM72	FitzSimmons *et al.* (1995)	(CA)33	136-288	141 / 141	27	12,45	0,84	0,92	0,06	0,04	0,10	0,07	0,00
OR2	Aggarwal *et al.* (2004)	(GT)8GCC (GT)5	112-188	146 / 146	25	14,54	0,91	0,93	0,01	0,01	0,03	-0,03	0,01
OR8	Aggarwal *et al.* (2004)	(TC)23	130-174	143 / 141	18	8,33	0,88	0,88	-0,01	0,02	0,01	0,00	0,01
OR3	Aggarwal *et al.* (2004)	(TC)9(AC) 6GC(AC)2	132-184	134 / 132	21	5,66	0,81	0,83	0,00	0,02	0,02	0,00	0,47
Cc117	FitzSimmons *et al.* (1995)	(CA)17	208-256	112 / 112	10	6,28	0,83	0,84	0,00	0,02	0,02	0,00	0,10

n: number of non-missing observations of loci in every chromosome (Chromosome 1 / Chromosome 2). **A:** allelic diversity. **A_e:** effective allelic diversity. **H_O:** Observed heterozygosity. **H_e:** expected heterozygosity. F statistics according to Weir and Cockerham (1984): **F_{IS}**, **F_{ST}** y **F_{IT}**. F statistics according to Slatkin (1995): **R_{ST}**. **HW p**: Exact probability associated to Hardy-Weinberg equilibrium test with 10 000 permutations.

Peakscanner program (Applied Biosystems, 2006). To check the quality of the genotyping, the Micro-Checker 2.2.3 (Van Oosterhout *et al*., 2004) program was used.

Data processing and descriptive analysis of genetic data

Before processing, the missing data were randomly assigned by sampling values with replacement and making the substitutions for each of the regions (that is, the lost data from the West of Guanahacabibes were imputed only with observed data from the West of Guanahacabibes, and so on) *locus* by *locus*. According to Bethlehem (2011), random imputation preserves the distribution of the data, is easy to implement, is computationally efficient, and is also unbiased under the missing completely at random assumption. The premises of Hardy-Weinberg equilibrium were tested with the hw.test function of the R package pegas (Paradis, 2010). For the Monte Carlo procedure 10,000 replicates were used. Linkage disequilibrium was analyzed with the function LD2 from the same package.

Allele richness ($\mathbf{N_A}$), private alleles ($\mathbf{A_P}$), observed heterozygosity ($\mathbf{H_O}$), expected heterozygosity ($\mathbf{H_E}$), and F_{IS} statistics were calculated by regions. In this case, allele richness was calculated with the allelicrichness function of the pegas package, using the extrapolation method proposed by Foulley & Ollivier (2006). Private alleles were inferred with de function private_alleles from the package poppr 2.8.6 (Kamvar *et al*., 2014, 2015) of R. All other measures were obtained with the function divBasic from the diveRsity (Keenan *et al*., 2013) package of R. In the case of F_{IS} statistic, confidence intervals at 95% were calculated also with divBasic through bootstrapping (10,000 permutations).

Likewise, allelic diversity (**A**), effective allelic diversity ($\mathbf{A_E}$), observed heterozygosity ($\mathbf{H_O}$), expected ($\mathbf{H_E}$) heterozygosity, and *F*-statistics were also calculated per *locus*. For this purpose, the functions A (gstudio package), Ae (gstudio R package), H (pegas R package), and Fst (pegas R package) were used, respectively. Additionally, the R_{ST} statistic was determined with the function of the same name, which computes a fixation index for microsatellites that considers the step mutation model and is more appropriate for this type of marker (Paradis, 2020).

As an initial exploratory method, it was verified the existence of differences between the average allele richness by region, by executing the non-parametric test of multiple comparisons of Kruskal-Wallis with the kruskal.test function in R (Hollander & Wolfe, 1973). To determine differences between set of observations, Nunn's *post hoc* test (Nunn, 1964) was executed with the dunn.test function of the dunn.test package, using the correction proposed by Benjamini & Hochberg (1995) and a threshold of 0.05.

Genetic differentiation among nesting locations

Values of the genetic differentiation statistics F_{ST} and G_{ST} were used as criteria to group localities. Both were calculated with the function fastDivPart of the R package diveRsity with 10,000 replicates for Monte Carlo method. Each one of these statistics provides alternative pieces of information not provided by the others, complementing them. For each of the values calculated by pairs of localities, the associated probability value was determined using the pairwiseTest function of the strataG package (Archer *et al*., 2016) of R (10,000 permutations). F_{ST}- family statistics (F_{ST} and G_{ST}), are measures widely used in the literature, and can offer a means to reach comparable results with prior publications. With the values obtained by pair of localities, a heat graph and a dendrogram were constructed with the heatmap.2 function of the R gplots package.

AMOVA and Discriminant Analysis of Principal Components

In virtue of the temporal extension of samplings, the presence of structure among sampling years was also

tested using AMOVA. Molecular analysis of variance was also performed to test the validity of the division of the data by sampling location. For this, the poppr.amova function was used, from the poppr 2.8.6 package of R. The function randtest from the package ade4 of R (Dray & Dufour, 2007, Bougeard & Dray, 2018) was used to test phi-statistics (10,000 permutations were applied). To visualize in a spatial order the distribution of the variance calculated with the AMOVA, discriminant analysis of principal components was executed with the dapc function of the adegenet package (Jombart, 2008, Jombart & Ahmed, 2011) of R. The first 31 principal components (according to the highest score 'a' obtained during the optimization problem) and three discriminant functions were used.

Inference of the individual ancestry coefficients and the Q matrix in STRUCTURE

The STRUCTURE 2.0 program (Pritchard & Wen, 2002) was used to group individual genotypes based on the Bayesian approach and Markov Chains of Monte Carlo (MCMC). The mixing model was considered, which admits mixed individuals by introducing a Q vector to denote the mixing proportions for each individual (Pritchard *et al*., 2000). According to the recommendations of Wang (2017), and in the face of unbalanced sample sizes, the model of uncorrelated allele frequencies was used, and an initial value of alpha smaller than 1 (0.1). The length of the initial run and the number of Markov chains were 5,000 and 10,000, respectively. *A priori* number of populations between 1 and 12 was tested and 20 replications were made for each of these K values. The selection criterion for the best run of the algorithm was based on ΔK (Evanno *et al*., 2005). This value is defined as the second-order rate of change of the probability of K. With the best model, the matrix Q of admixture coefficients was calculated, which were represented by a bar graph.

Inference of the individual ancestry coefficients and the Q matrix with the sparse nonnegative matrix factorization algorithm in R

The coefficients of admixture in R were estimated employing sparse nonnegative matrix factorization algorithms optimized by least squares, which were implemented using the snmf function of the LEA package (Frichot & François, 2014) of R. The code used by the snmf function allows the estimation of the frequencies of homozygotes and heterozygotes and avoids the assumptions of Hardy-Weinberg equilibrium (Frichot *et al*., 2014). This method is fast, precise, and constitutes a non-parametric approach. Unlike STRUCTURE, snmf function does not depend on parametric assumptions and can be an option to use in a larger number of situations (François, 2016). The algorithm was executed for K values between 1 and 12, and 200 repetitions were implemented, each with a maximum of 200 iterations. The 1E-5 value was used as the tolerance threshold, 5% as a percentage of hidden genotypes, and 100 as the regularization parameter (α). To decide which algorithm presented the best results, the predictive criterion introduced by Frichot *et al*. (2014) was used. This is based on cross-entropy and hidden genotype imputation. The best of the models generated was used to calculate the Q matrix. In turn, this Q matrix was represented by a bar graph, in which each bar corresponds to an individual and summarizes the information corresponding to the ancestral proportions present in each one of them.

Congruence levels between assignment methods

A principal component analysis was performed using the fast.prcomp function of the gmodels package in R. The two-dimensional space defined by the first two main components (85% of the variance) was used to project and compare the spatial arrangement of the groups defined in STRUCTURE and R's LEA package. The congruence

rate was defined as the number of common assignments between STRUCTURE and snmf algorithm divided by the total number of individuals. Finally, a summary map was constructed in which the profiles for each of the studied localities were illustrated according to the groups defined by each method. For this, pie charts were constructed with the average admixture coefficients for each of the nesting rookeries.

Gene flow

The function divMigrate from the R package diveRsity was used to estimate the relative migration rates from allele frequency data between the nesting areas from the southwest Cuban territory. The method implemented by this algorithm was described by Sundqvist *et al*. (2016). For this analysis, the measures *Nm*, G_{ST}, and *D* were used. G_{ST} and *D* offer, complementary measures under variable demographic scenarios (Sundqvist *et al*. 2016), which are not available for Cuban populations of *Chelonia mydas*. For this reason, both were employed here. On the other side, *Nm* (*i. e*. effective number of migrants) provide complementary information and could be a more conservative and global measure of relative migration (Alcala *et al*., 2014).

Results

Checking the quality of genotyping with the Micro-Checker program

In the case of *loci* OR8 in East of Guanahacabibes, CM72 in Isla de la Juventud and Cayo Largo, and OR2 in Cayo Largo, an excess of homozygotes was detected, which may indicate the presence of null alleles in the molecular data matrix. For all other cases, *i. e. locus* and nesting locality combinations, which constitute the vast majority (38 of 42; 90%), no evidence of scoring errors due to DNA polymerase slippage during amplification, large alleles marginalized (dominance of small alleles) or null alleles was detected. Consequently, the work matrix was considered suitable for subsequent analysis.

Genetic diversity

The allelic richness by regions presented values from 7.92 in San Felipe to 15.70 in Center of Guanahacabibes. Statistically significant differences were detected between the mean allele richness by region according to Kruskal-Wallis's test (X^2 = 11.627, df = 5, p = 0.04). San Felipe was different from Center and East of Guanahacabibes according to Dunn´s test (Table I). The rest of the nesting locations were statistically similar between them. The nesting area with the highest number of private alleles was East of Guanahacabibes ($\mathbf{A_P}$ = 10), while Isla de la Juventud did not have any.

The expected (H_E) and observed (H_O) heterozygosity varied between locations (Table I). As a general trend, the H_E values ranged between 0.73 and 0.86, while H_O values were between 0.71 and 0.91. The lowest values corresponded to the San Felipe region. In all cases, the observed heterozygosity was higher than expected in the studied localities. In all localities, F_{IS} statistics reach negative values, indicating from low (-0.02 in Isla de la Juventud) to moderate (-0.14 in East of Guanahacabibes) heterozygotes excess. However, based on the confidence intervals, none of the F_{IS} values were statistically significant.

The *loci* OR8, Ei8, CM84, and CM72 were in Hardy-Weinberg disequilibrium (Table II). Likewise, when the samples were divided strictly by localities, it was possible to determine that most of the *loci* in the West of Guanahacabibes, Isla de la de Juventud, and Cayo Largo were far from the Hardy-Weinberg equilibrium. On the other hand, in the case of the central region of the Guanahacabibes peninsula and San Felipe, except for two *loci* (Ei8 and CM72, respectively), all the remaining ones comply with the equilibrium premises. With respect to linkage equilibrium, a notable proportion of pairs of microsatellites (8 of 21) do not comply with the premises. Nevertheless, all *loci* were employed in the analysis because of the snmf method employed is not sensible to them.

Without considering the grouping by nesting localities, the allele richness (**A**, Table II) was between 7 and

Table II: Genetic variability of Cuban southwest green turtle *Chelonia mydas* by localities (N = 149).

Locality	Coordinates	Sampling year	n	r	A_P	H_0	H_E	F_{IS}	F_{IS}_CI	Cat_Dunn_Test
Occident of Guanahacabibes	21.8195; -84.8499	02; 04; 06; 07	20; 4; 1; 1	12,50	3	0,86	0,76	-0,12	[-0.19; -0.05]	ab
Center of Guanahacabibes	21.8733; -84.7326	04; 06; 07	9; 12; 14	15,70	17	0,91	0,85	-0,08	[-0.11; -0.04]	a
East of Guanahacabibes	21.9122; -84.6316	98; 02; 04; 06; 07	1; 3; 6; 23; 7	15,68	19	0,87	0,84	-0,03	[-0.09; 0.02]	a
Cayos de San Felipe	21.9628; -83.5840	99	8	7,92	4	0,70	0,67	-0,03	[-0.16; 0.05]	ab
Isla de la Juventud	21.4847; -82.7251	3	12	10,37	0	0,89	0,82	-0,09	[-0.19; -0.01]	b
Cayo Largo	21.6261; -81.4735	6	28	13,93	4	0,85	0,83	-0,02	[-0.09; 0.03]	ab

Coordinates: For localities with more than one beach sampled, coordinates represent a consensus location of the area, non the beach itself. **Sampling year**: last two digits of every year are reported. **n**: Sample size (by year). **r**: Mean allelic richness. A_P: Private alleles. H_0: Observed heterozigosity. H_E: Expected heterozigosity. F_{IS}: F_{IS} statistic according to Weir & Cockerham (1984). Significant values for a significance level of 0.05 are highlighted with an asterisk. F_{IS}_CI: F_{IS} confidence interval at 95%. **Cat_Dunn_Test**: Categories of similarity between localities depending on the allelic richness and according to the **a posteriori** test of Dunn.

25. The effective allele richness (**Ae**) was between 2.59 and 14.54. For most *loci*, the expected heterozygosity was higher in comparison with the observed heterozygosity. The F statistics calculated for each microsatellite *loci* are also detailed in Table II.

Genetic differentiation among nesting locations

In the case of F_{ST} and G_{ST} statistics, Isla de la Juventud - Center of Guanahacabibes (F_{ST} = 0.007; G_{ST} = 0.003) and Cayo Largo - East of Guanahacabibes (F_{ST} = 0.007; G_{ST} = 0.003) presented the lowest differentiation values. These groups, in turn, form a conglomerate that differs from West of Guanahacabibes and San Felipe regions (Fig. 1). However, in all pairs of localities the F_{ST} and G_{ST} values were statistically different from 0 for at least a significance level of 0.05, suggesting genetic differentiation between organisms from different nesting locations.

The AMOVA executed to verify the presence of genetic structure on a temporal scale was not significant (ϕ = 0.009, σ^2 = -0.08, p = 0.7). By the other side, according to the other AMOVA performed, the grouping by localities (West, Center and East of Guanahacabibes, San Felipe, Isla de la Juventud, and Cayo Largo) only explained 2.6% of the covariance. However, the phi-statistics resulted significative (Fig. 2), in accordance with what was obtained with the differentiation indices.

The results of the discriminant analysis of principal components show that San Felipe is separated from the remaining localities (Fig. 2). The remaining localities converge in the same region of the ordering space defined with the analysis (which comprises 75% of the variance observed in the data). However, there is no overlap between West of Guanahacabibes and the localities of Center of Guanahacabibes and Isla de la Juventud, nor between Isla de la Juventud and Cayo Largo (Fig. 2). A similar result was obtained with the grouping based on the genetic differentiation indices.

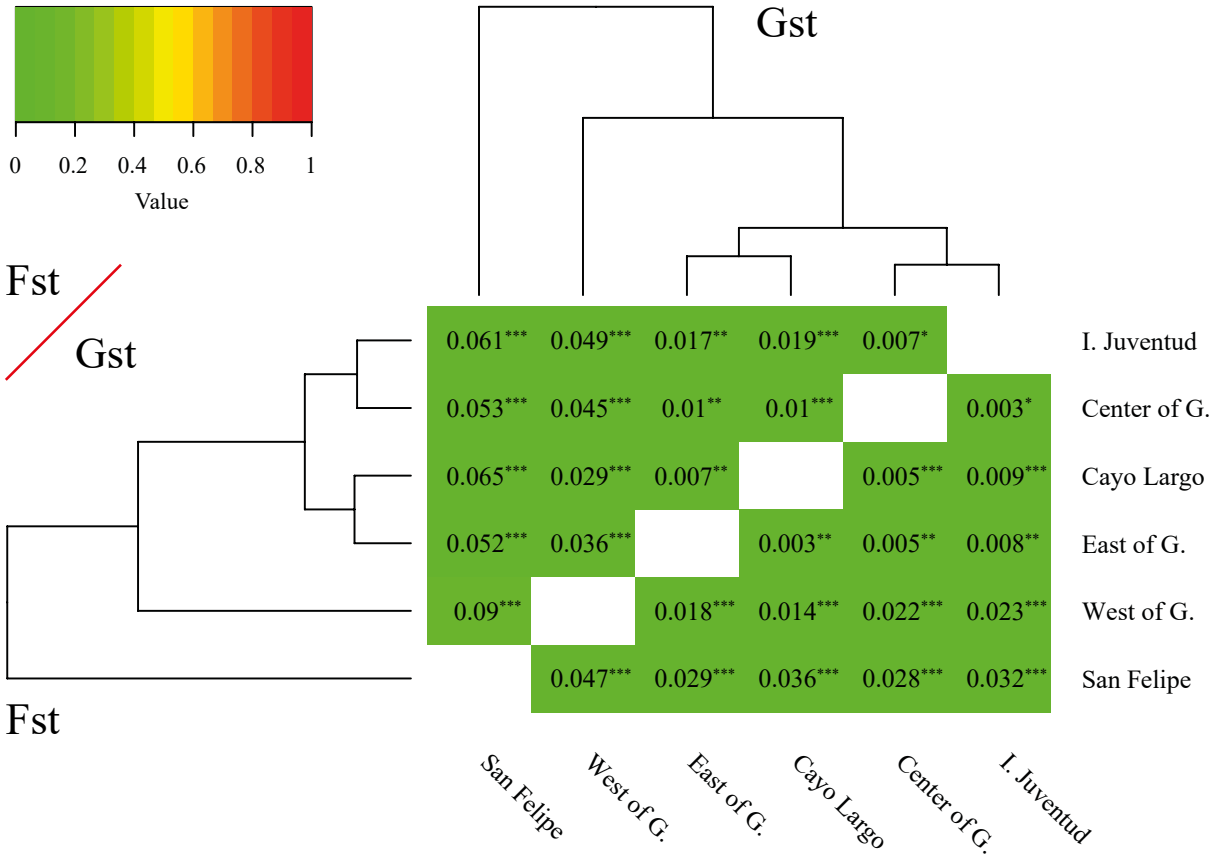

Fig 1. F_{ST}/G_{ST} heatmap among localities after genetic analyses using microsatellites for the Cuban southwest green turtle *Chelonia mydas* (N = 149 individuals). The darker the color, the higher the statistics values. The measures reported comes of the implementation of fastDivPart function of the diveRsity R package (B = 10 000 permutations). Asterisks indicate those values significantly different from 0 based on the 0.05 (*), 0.01 (**) and 0.001 (***) significance levels. Results of a hierarchical cluster analysis under the complete method is represented here as dendrograms. As a measure of dissimilarity was employed the statistics of genetic differentiation.

Population structure inferred by STRUCTURE and snmf algorithm

The STRUCTURE result suggest the highest value of ΔK is achieved for K = 2. The coefficients of addition are represented in Fig. 3A. In the case of R, it was also determined that the number of ancestral populations that best explains the studied data corresponds to K = 2. For this value of K, the lowest value of cross-entropy was reached, a criterion used as a discriminant between the simulations generated. The coefficients of addition are represented in Fig. 3B.

When projecting the groups derived from the analysis in Structure (Fig. 3C) and R (Fig. 3D) in the space defined by the first two main components of the PCA, it was obtained a large overlap of the two areas, with small differences in representation. The calculated congruence rate showed a 77.8% level of fit between both methods (116 consistent classifications out of 149). This high value of similarity between the results can be observed in the presence profiles of the groups identified in the rookeries (pie charts in Fig. 4).

Fig. 4 clearly shows an association between the localities Center of Guanahacabibes, East of

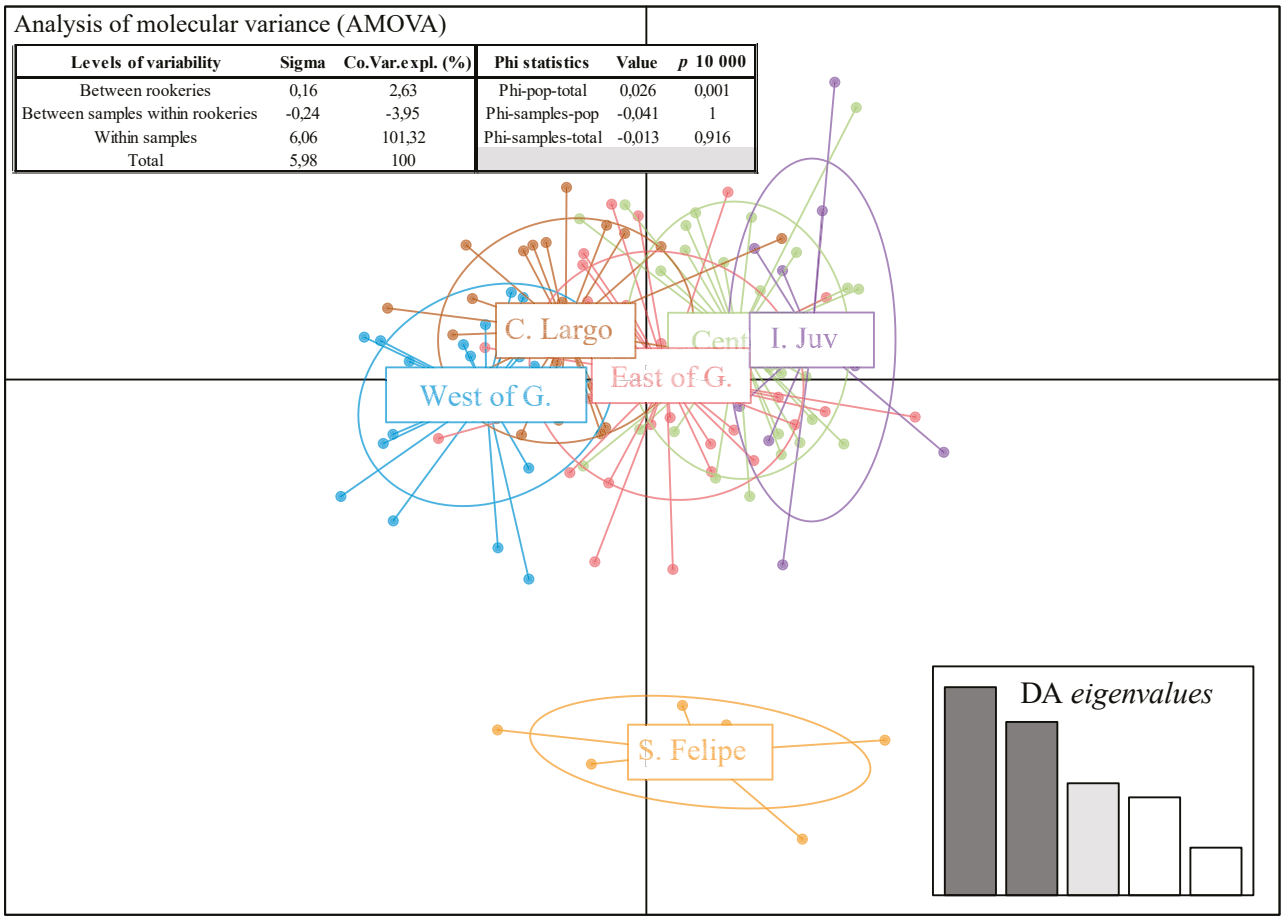

Fig 2. Genetic clustering using Discriminant Principal Components Analysis (DPCA) after genetic analyses using microsatellites for the Cuban southwest green turtle *Chelonia mydas* (N = 149 individuals). Results of AMOVA are printed in the ordination space. For this last analysis, the hierarchical level tested was the nesting areas (level populations or pop in the table). Both, components of variability and phi-statistics are reported.

Guanahacabibes, Isla de la Juventud, and Cayo Largo according to their ancestry profiles. However, San Felipe shows a higher proportion of ancestral group 2, while West of Guanahacabibes presents a composition with a predominance of ancestral group 1.

Gene flow

The patterns of relative migration between nesting localities with the *Nm* statistics (Fig. 5A) show that there is a core of extensive exchange of genetic information among Cayo Largo, East of Guanahacabibes, and Center of Guanahacabibes. Isla de la Juventud and San Felipe seem to be the more isolated nesting beaches of southwest Cuba. Similar results were obtained with the G_{ST} statistic (Fig. 5B). On the other side, according to *D* statistics, the exchanges are greater between West of Guanahacabibes, East of Ganahacabibes, and Cayo Largo (Fig. 5C). A significant asymmetric gene flow was detected for each of the statistics used. Considering *Nm* and G_{ST} (Fig. 5D y 5E), West of Guanahacabibes contributes a greater number of alleles to Cayo Largo in relation to what it receives from the latter locality. On the other hand, according to statistic *D*, the aforementioned pattern is maintained, but it is also evident

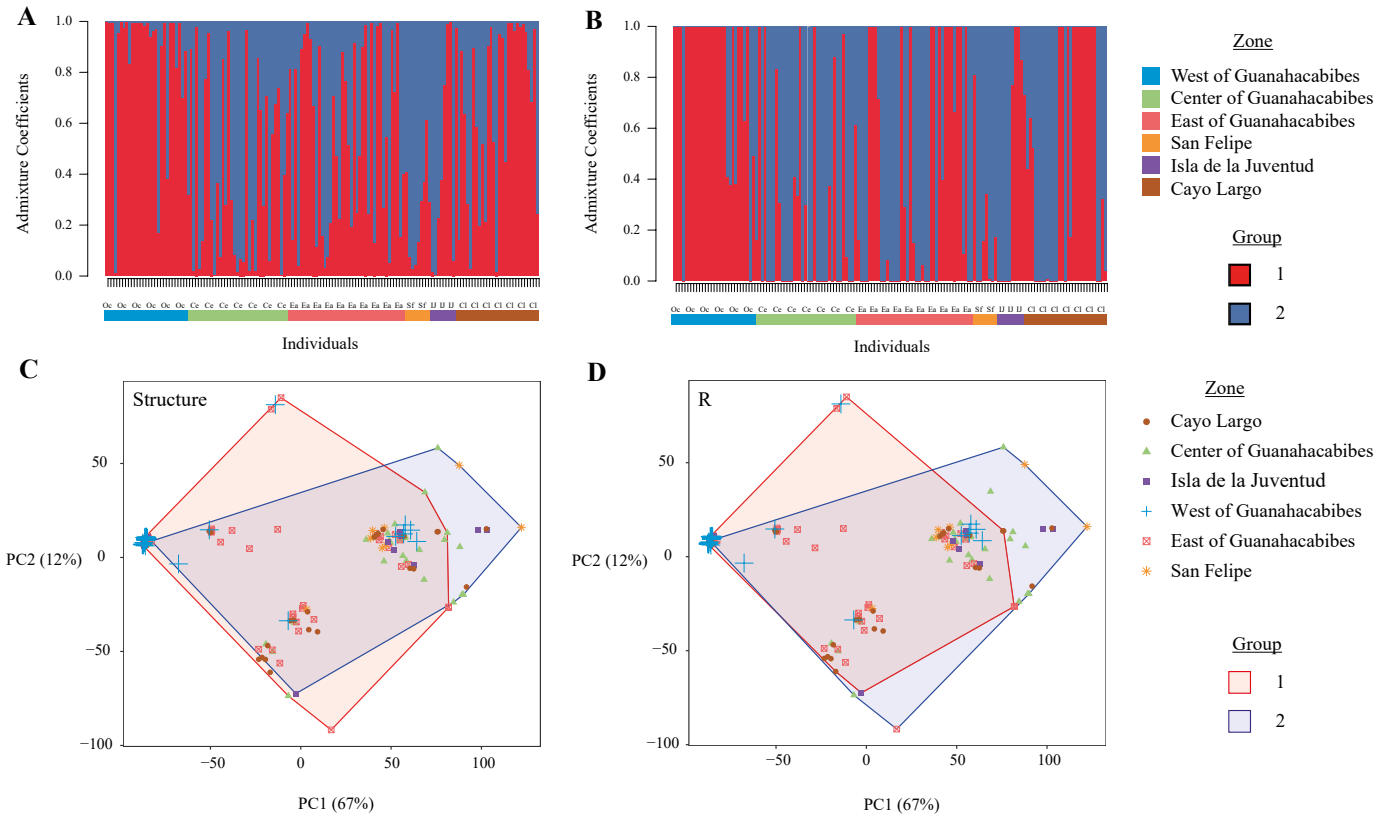

Fig 3. Values of the admixture coefficients calculated using the STRUCTURE software (A) and the R LEA package (B) with a model of two ancestral populations (K = 2) from seven microsatellite loci of *Chelonia mydas* (N = 149) in the region southwestern Cuba. Also is represented the arrangement of individuals in the two-dimensional space defined by the first two components of the principal component analysis performed. The groups refer to the distribution of each individual to the groups defined by the analysis of the population structure of the Structure software (C) and by the analysis LEA package in R (D), from the admixture coefficients.

that the flow from San Felipe to Cayo Largo and East and Center of Guanahacabibes is greater than in the opposite direction (Fig. 5F).

Discussion

In this study, when considering the artificial grouping by nesting regions, all the differentiation statistics used were different from zero (according to *p* values; Fig. 1), which also confirms the presence of a structure based on the hierarchical stratum used. However, based on the absolute values of the differentiation statistics (*F* and *D* family) obtained, which are relatively low as expected for microsatellites, the detected divergence can be considered mild to moderate, especially between Isla de la Juventud, East of Guanahacabibes, Center of Guanahacabibes, and Cayo Largo. Consequently, the studied individuals cannot be considered as members of a panmictic population (free and random mating). According to Roberts *et al.* (2004), a possible explanation for the relatively low values of the F_{ST} statistic despite its statistical significance may be that it is simply an inappropriate statistic with which to estimate population subdivision using microsatellites. This is supported by the fact that a high number of alleles were found at these *loci*. In essence, a high level of heterozygosity within populations can obscure differences between populations, resulting in the underestimation of subdivision.

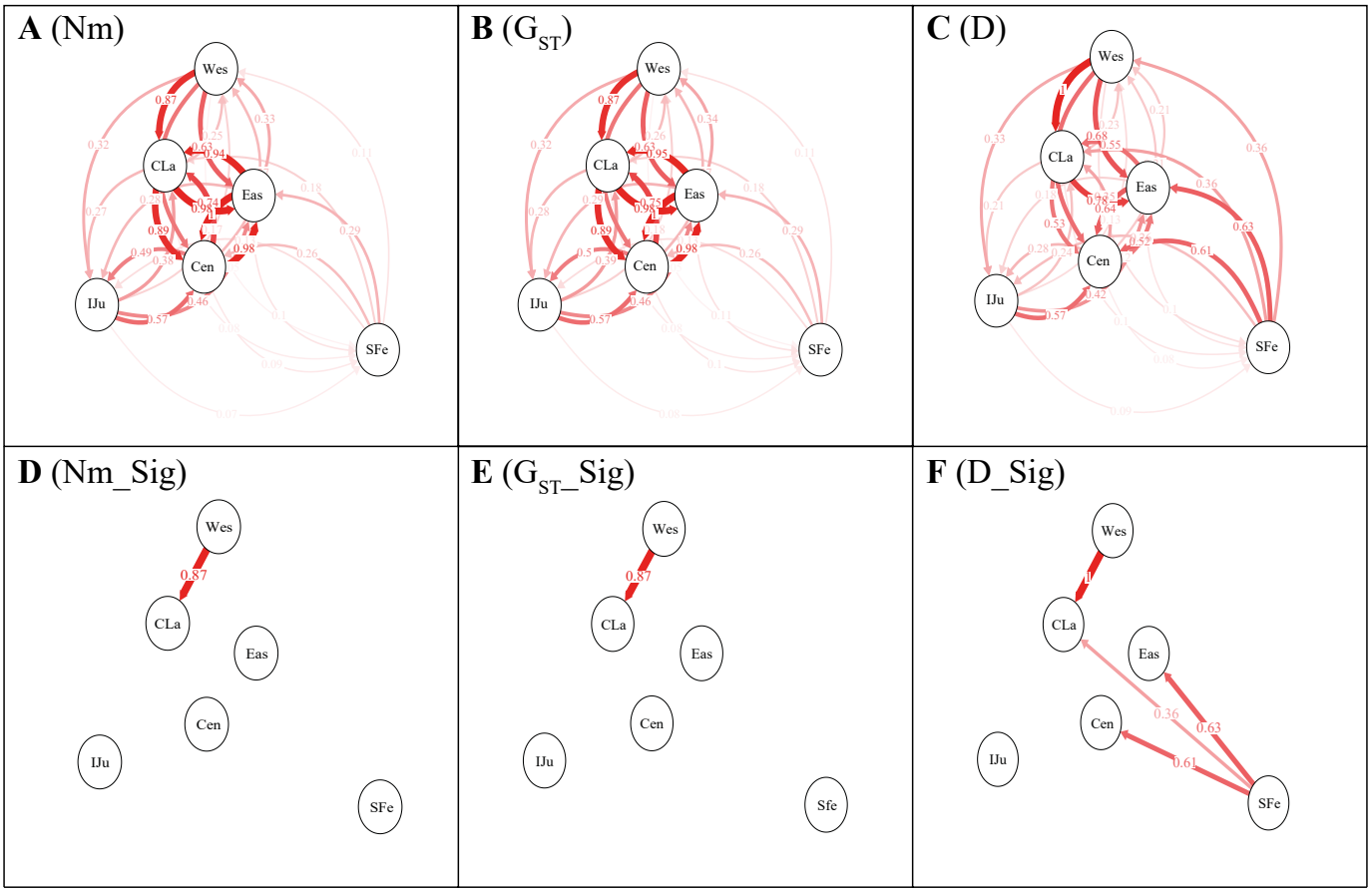

Fig 4. Relative migration measures among populations of *Chelonia mydas* defined by the nesting beaches of southwestern Cuba as estimators of gene flow. The statistics used were N_M (A), G_{ST} (B) and D (C). Significant asymmetric gene flow relationships (D - F) are also reported according to Sundqvist *et al.* (2016). The thicker the line and the darker the color, the greater the magnitude of genetic exchange between localities. Wes: West of Guanahacabibes; Eas: East of Guanahacabibes; Cen: Center of Guanahacabibes; CLa: Cayo Largo; IJu: Isla de la Juventud; SFe: San Felipe.

According to Bowen and Karl (2007), the first population genetic tests on green turtles was carried out by protein electrophoresis, taking nesting beaches as samples. Low genetic diversity was found, a result that became recurrent in subsequent studies with mitochondrial DNA (mtDNA) and nuclear DNA (nDNA), possibly due to low metabolic rates and long generation times. First genetic study of *C. mydas* performed in Cuba, included 11 samples taken in 1998 from the San Felipe keys and 17 from the Guanahacabibes peninsula in 2000. All seven haplotypes found in that study belonged to lineage II. Additionally, three endemic haplotypes were described (Ruiz-Urquiola *et al.*, 2010). No significant genetic structure was found between the Guanahacabibes peninsula and the San Felipe keys, indicating that they represent a single breeding population with high genetic endemism. An AMOVA executed for these data considers the Cuban rookeries as an individual unit, compared to the remaining rookeries of the Greater Caribbean, and places the Cuban nesting turtles as a genetically distinct management unit (Ruiz-Urquiola *et al.*, 2010). After, Azanza (2009) with a larger sample established that considering the genetic differences

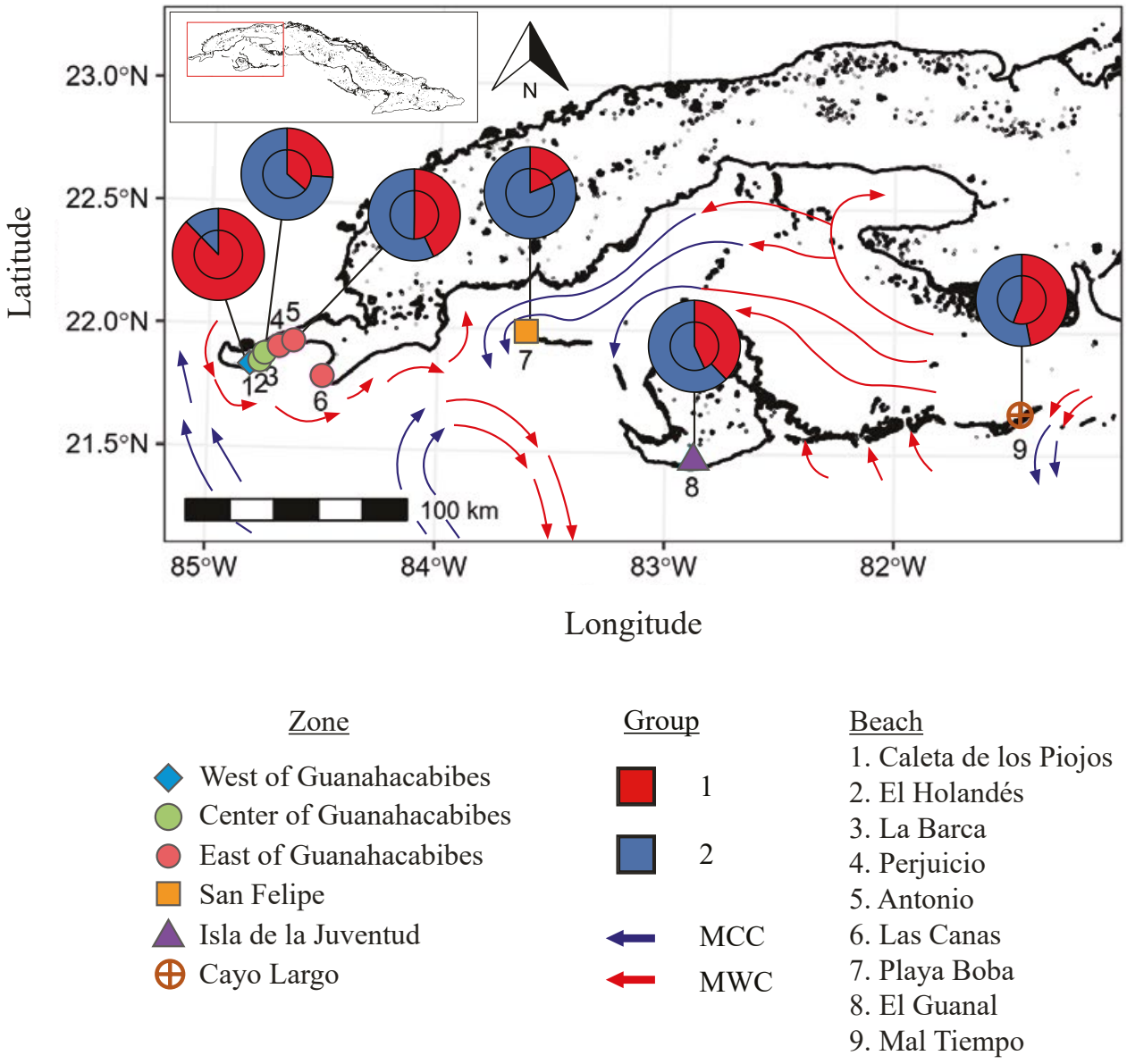

Fig 5. *Chelonia mydas* (N = 149 individuals) nesting locations sampled for microsatellites analyses in Cuban southwestern area (blue diamond: Occident of Guanahacabibes; green circle: Center of Guanahacabibes; light red circle: East of Guanahacabibes; orange square: San Felipe; purple triangle: Isla de la Juventud; brown crossed circumference: Cayo Largo). Beaches are numbered as follows: 1) Caleta de los Piojos; 2) El Holandés; 3) La Barca; 4) Perjuicio; 5) Antonio; 6) Las Canas; 7) Playa Boba; 8) El Guana; 9) Mal Tiempo. Pie charts (blue color as ancestral group 1 and red color as ancestral group 2): Average addition coefficients by regions obtained from the work with the LEA package in R (external pie chart) and Structure (internal pie chart). Red arrow means marine warm currents (MWC) and blue arrows means marine cold currents (MCC).

found based on mtDNA and microsatellite *loci*, Guanahacabibes, San Felipe, Isla de la Juventud, and Cayo Largo nesting areas should be managed as independent evolutionary units.

It is accepted that sea turtles migrate through ocean basins to reach foraging and breeding areas, but they show a significant population structure based on nesting beaches (Ferrera *et al*., 2021). Multiple studies have been carried out on the green turtle to try to elucidate the population subdivision of this globally distributed and threatened species. Thus, using mtDNA markers, Bowen *et al*. (1992) demonstrated matriarchal philopatry by detecting significant population genetic subdivisions between turtles from different nesting grounds. These authors provided evidence in favor of the natal homing hypothesis by finding an extreme example of differentiation between individuals from populations in Suriname and Ascension Island, which share a common foraging area. Bowen and Karl (2007) report that green turtles from Suriname and Ascension Island did not have shared haplotypes (differentiation based on mtDNA). However, results based on five nuclear *loci* suggest that the two populations are indistinguishable, and the corresponding estimates of gene flow suggest current or at least recent exchange. Microsatellites corroborate this same information. Studies based on the mentioned markers suggest gene flow between nesting colonies. Similar results were found by Nishizawa *et al*. (2011).

The genetic divergence is supported in our study by results of the AMOVA as well, which was significant for nesting localities. This result can be contrasted with those obtained from the discriminant analysis of principal components. Except for San Felipe, all the remaining localities coincided in the ordering space. However, in that space, each locality presents positions in which they coincide with certain localities while there is no overlap with others. The previous proposition about the existence of a moderate population structure can also be supported by the fact that, even when the grouping by nesting localities was significant, the percentage of covariance explained in the AMOVA by this hierarchical level does not exceed 3%. A lack of a strong genetic structure with microsatellites has been previously found in marine turtles despite the existence of a strong genetic structure with mtDNA (FitzSimmons *et al*., 1997; Lee, 2008).

The absence of a strong genetic differentiation, based on microsatellite *loci*, may be conditioned by male-mediated gene flow, which is a common event in sea turtles (Roberts *et al*., 2004). In this sense, it is valid to assume that the nuclear gene flow must be greater than the mitochondrial due to the apparent lack of fidelity for the nesting sites of the males. However, there are studies suggesting male philopatry for courtship areas within natal regions (FitzSimmons *et al*. 1997), suggesting that male-mediated gene flow occurs through mating outside of natal courtship areas. These interactions are facilitated by the mixing of different colonies along migratory routes, in foraging areas, or on breeding grounds. Mating of individuals outside of natal areas acts as a buffer against significant differentiation between turtle's rookeries (Ferrera *et al*., 2021).

As can be seen in the bar graphs with the ordered Q coefficients, in the case of snmf algorithm (Fig. 3B), the discrimination is more defined between clusters, with some individuals unequivocally assigned to a single group. In the case of STRUCTURE, the discrimination is more diffuse, with components of both groups for each individual. However, both programs arrive at the same general result with a great consistency rate (≈ 78%): the model of two ancestral populations is the one that best explains the characteristics of the analyzed data. It should be noted, however, that all methods based on cluster analysis involve high uncertainty unless the true populations are strongly divergent (Manel *et al*., 2005). The fact that deviations from the assumptions have been found may be a cause of the slight differences between the results obtained in R and STRUCTURE.

The demarcation of two population groups with these analyses is not absolute, since once the turtles reach reproductive status, these organisms embark on extensive migrations through which the individual populations segregate towards the corresponding natal

sites to lay eggs (a phenomenon known as philopatry). This complicated life cycle process determines that individual populations are difficult to monitor, and their genetic structures and distribution are never fully known (Ruiz-Urquiola *et al*., 2010).

It is difficult to detect the genetic structure of a population from molecular markers alone. According to Ruiz-Urquiola *et al*. (2010), the genetic structure of green turtles is defined by the geographic distribution of critical habitats and dispersal and migration patterns, in combination with the philopatric nature of the species. Azanza-Ricardo (2009) found high levels of nest site fidelity in females which was later corroborated with satellite markers (Ruiz Valdés, 2017). Results obtained from flipper tagging robustly indicate that Cuban sites constitute a focal point for regional populations that are in migration or those that use Cuban habitats as foraging or development areas (Moncada *et al*., 2006).

The method implemented in this paper for the estimation of relative gene flow was an auxiliary mean for testing the genetic structuring of the Cuban green turtle population in southwest Cuba. According to Sundqvist *et al*. (2016), this method is best performed when gene flow is intermediate. In consequence, when existing very low differentiation between populations, the approach fails to detect underlying patterns of migration. However, the intention here was to recreate a plausible scenario that illustrates the subjacent cause of diffuse structure at the geographical level used here.

With its size and central position between the Caribbean Sea and the Gulf of Mexico, Cuba is an essential piece of the biogeography of the region. The island represents a physical barrier that influences the regional marine currents between the Caribbean Sea, the Gulf of Mexico, and the Florida peninsula. The Cuban archipelago is made up of the main island (1,200 km long) surrounded by 4,000 keys and islets. The submerged continental shelf is the largest in the Caribbean insular region and can be further divided into four smaller sections subdivided by regions with narrow shelves. Multiple investigations carried out in the last 24 years on dissimilar Cuban marine organisms have revealed three general patterns of population structure along the Cuban coast: 1) north-south rupture, 2) east-west division in the south and 3) genetic differentiation at a local scale (García-Machado *et al*., 2018).

Three main ocean current systems run along and around the Cuban coast: a western current off southern Cuba, a northeastern current off northwestern Cuba, and a western current off northeastern Cuba. These general circulation patterns are complex and include temporal changes, cyclonic and anticyclonic eddies, nearshore meanders and countercurrents, products of local factors, and current strength. In addition, intense meteorological events, such as hurricanes, promote strong but ephemeral changes in patterns of marine circulation, salinity, and sediment suspension (García-Machado *et al*., 2018). If the influence of the marine currents in the area is analyzed, which intervene in the migratory movements of the turtles, it is possible to justify the mixture of genetic information between zones (due to the sampling of migrant individuals in the localities, which become observations confused) and the profiles generated from the mean values of the addition coefficients by regions (Fig. 4).

Gene flow between populations of *C. mydas* can probably occur between neighboring rookeries located less than 500 km, and rarely between colonies located at a greater distance (Dethmers *et al*., 2006). The fact that the nesting rookeries sampled in the study were located less than 500 km away in all cases thus justifies the exchange of information between localities and the allocation profiles observed in Fig. 5. Also, Bourjea *et al*. (2007) found genetic differences between sites located more than 500 km away but not between nearby areas (< 150 km). According to Shamblin *et al*. (2020), females are capable of dispersing between nesting beaches located less than 15 km apart based on recapture data, while haplotypic evidence suggests discrete

natal homecoming for island groups 70 km apart. If populations are linked by gene flow, it is practically impossible to achieve 100% correct assignments, simply because some individuals in a population are immigrants or descended from immigrants from other populations (Hansen *et al*., 2001).

The primary effects of gene flow are the reduction of genetic differences between populations and the increase of genetic variation within populations. Note how, in Fig. 3, many of the points within each locality exceed the ellipsoid (a measure of intra-population variation). Gene flow can be considered, then, as the cohesive force that holds geographically distant populations together in the context of a few evolutionary units. In the absence of another evolutionary force, gene flow between populations will lead to increased genetic homogeneity. According to the migration islands model (assumed in this study), little gene flow is required for populations to be genetically connected. Even with low levels of gene flow, eventually, all populations involved in the phenomenon will become genetically identical. On the other hand, genetic drift causes the divergence of isolated subpopulations. Consequently, the actual amount of divergence between subpopulations can be understood as the balance between the homogenizing effects of gene flow, which makes subpopulations more similar, and the disruptive effects of genetic drift, which causes differentiation between subpopulations (Allendorf & Luikart, 2007).

Based on the ancestry profiles (Fig. 4) and the asymmetric flow relationships (Fig. 5D-F), it can be considered that both, West of Guanahacabibes and San Felipe constitute sources of genetic information and the remaining localities constitute sinks. Fig. 5 shows how the gene flow is stronger from the source localities to the sink than in the opposite direction. In this sense, the asymmetric flow relationships detected comply with this pattern in all cases and for all the statistics used.

Several colonies in southwestern Cuba are important nesting sites for green turtles (Azanza-Ricardo, 2009). The results derived from the genetic data used in this research suggest the existence of at least three genetically distinct population groups based on the profiles presented in Fig. 4: 1) West of Guanahacabibes, 2) San Felipe and, 3) Center of Guanahacabibes, East of Guanahacabibes, Cayo Largo and Isla de la Juventud. Consequently, it is appropriate from a practical perspective to consider these groups for the management of zones in conservation efforts.

From our results, it can be observed that the levels of genetic diversity of the individuals sampled for each locality are high. This is evidenced by values of observed heterozygosity higher than the expected heterozygosity and low modular and negative sense F_{IS} values (sense suggesting an excess of heterozygotes). The high genetic connectivity between nesting beaches may justify these results. When analyzing the microsatellite loci without considering any clustering, it can be observed that the species-specific markers CM72 and CM84 are among those with the highest allele richness, as expected. Furthermore, in this case, the trend found for each sampling location is reversed and, in most cases, the expected heterozygosity values are higher than expected. Although statistically significant differences were detected between nesting localities based on allele richness, it should not be ignored that the divergence in the allelic richness of San Felipe concerning other localities detected by the non-parametric tests of Kruskal-Wallis and Nunn could be derived from the low number of individuals analyzed corresponding to that site. This fact may also be the cause of the low values of heterozygosity detected for the area. It should also be noted that despite the differences in the collection periods, no association was detected between the genetic structure and the years of sampling.

Also, several microsatellites showed deviations from the Hardy-Weinberg equilibrium by sampling region. Such deviations may be due to selection, genetic imbalance, the presence of null alleles, mutations, or non-random mating (Stephens, 2004). Because microsatellite

loci are not typically coded, they are unlikely to be subject to selective pressures (Li *et al*., 2004). However, high mutation rates can impact the expectations regarding the Hardy-Weinberg equilibrium (Ferrera *et al*., 2021).

Conclusions

The nesting colonies of *Chelonia mydas* in southwestern Cuba were composed of individuals from two ancestral populations. Also, was detected consistent structure within the region suggesting the establishment of, at least, three independent management units: 1) West of Guanahacabibes, 2) San Felipe and, 3) Center of Guanahacabibes, East of Guanahacabibes, Cayo Largo and Isla de la Juventud. Of these three groups, the first two seems to be sources of genetic information, and the third is a sink for it. Therefore, they should be taken into account to implement effective management programs that contribute to the preservation of the genetic diversity and survival of these populations in the Greater Caribbean. The programs used to implement the allocation methods had high levels of adjustment and the values of the addition coefficients observed for the individuals of each of the allocation beaches, as well as the calculated migratory estimators, suggest the existence of gene flow between localities. Likewise, when verifying the performance of different methods and obtaining equivalent results, the choice of the R approximation (snmf function of the lea package) is suggested as the primary technique, since this approach is free of assumptions and adjusts, consequently, to a greater number of practical situations in the research framework.

Acknowledgements

We thank the protected areas administration, and the forest rangers protecting and monitoring the sea turtle nesting areas; the Ministry of Science, Technology, and Environment for the authorization to access natural areas and for their support of the monitoring work; to all workers and volunteers for their effort to protect sea turtles in Cuba. We also thank Laboratorio de Genética of Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología UNAM and Centro de Investigaciones en Ecosistemas in UNAM, Morelia, for facilities and equipment used in our genetic analyses.

Statements

Funding

This study has received funds from the UNAM-UH and TWAS PhD program.

This study has received funding from The Ocean Foundation and Turtle Hospital of Marathon, Florida for its development.

Conflict of interest

The authors have no financial or non-financial conflicts of interest to declare that are relevant to the content of the manuscript.

Ethical behaviour

The authors have followed all applicable international, national, and institutional recommendations related to the use and handling of animals for research.

Permits for sampling and other permits

No permits were required for the conduct of this research.

Author Contributions

Conceptualization, J.A.R., G.E.L. and F.A.A.G.; Methodology, L.J.M.R., J.A.R. and G.E.L.; Software, L.J.M.R., F.A.A.G. and O.C.N.; Validation, G.E.L., K.O., FAAG and O.C.N.; Formal Analysis, L.J.M.R., J.A.R. and O.C.N.; Investigation, L.J.M.R. and J.A.R.; Resources, K.O. amd F.A.A.G.; Data Curation, L.J.M.R. and J.A.R.; Writing – Original Draft Preparation, L.J.M.R., J.A.R. and G.E.L.; Writing – Review & Editing, K.O., F.A.A.G. and O.C.N.; Visualization, L.J.M.R. and J.A.R.; Supervision, K.O., G.E.L. and F.A.A.G.; Funding Acquisition, J.A.R and K.O.

Bibliographic references

Aggarwal, R. K., Velavan, T. P., Undaykumar, D., Hendre, P. S., Shanker, K., Choudhury, B. C., Singh, L. (2004). Development and Characterization of novel microsatellite markers from the olive ridley sea turtle (*Lepidochelys olivacea*). *Mol. Ecol. Not.*, *4*, 77-79.

Aguirre-Liguori, J. A., Luna-Sánchez, J. A., Gasca-Pineda, J., & Eguiarte, L. E. (2020). Evaluation of the minimum sampling design for population genomic and microsatellite studies: An analysis based on wild maize. *Front. Genet.*, *870*.

Alcala, N., Goudet, J., Vuilleumier, S. 2014. On the transition of genetic differentiation from isolation to panmixia: what we can learn from gst and d. *Theor. Popul. Biol. 93*, 75-84.

Allendorf, F. W., Luikart, G. H. (2007). *Conservation and the Genetics of Populations*. Blackwell Publishing, Oxford, UK.

Álvarez-Varas, R., Rojas-Hernández, N., Heidemeyer, M., Riginos, C., Benítez, H. A., Araya-Donoso, R., Reséndiz, E., Lara-Uc, M., Godoy, D. A., Muñoz-Pérez, J. P., Alarcón-Ruales, D. E., Alfaro-Shigeto, J., Ortiz-Álvarez, C., Mangel, J. C., Vianna, J. A., Véliz, D. (2021). Green, yellow or black? Genetic differentiation and adaptation signatures in a highly migratory marine turtle. *Proc. R. Soc. B*, *288*, 20210754. doi: https://doi.org/10.1098/rspb.2021.0754.

Archer, F. I., Adams, P. E., Schneiders, B. B. (2016). strataG: An R package for manipulating, summarizing and analysing population genetic data. *Mol. Ecol. Resour.* doi:10.1111/1755-0998.12559

Ashley, M. V., Dow, B. D. (1994). The use of microsatellite analysis in population biology: Background, methods and potential applications. In B. Schierwater, B. Streit, G. P. Wagner, R. DeSalle (eds.). *Molecular Ecology and Evolution: Approaches and Applications*. Birkhäuser Verlag Basel, Suiza.

Azanza-Ricardo, J. (2009). *Estrategia reproductiva de la tortuga verde, Chelonia mydas, (Testudines, Cheloniidae) y su impacto en la estructura genética de áreas de anidación del occidente del archipiélago cubano*. (Tesis presentada en opción al Grado Científico de Doctor en Ciencias Biológicas), Centro de Investigaciones Marinas, Universidad de La Habana, Cuba.

Azanza-Ricardo, J., Gerhartz Muro, J. L., Martín-Viaña, Y. F., Moncada Gavilán, F., Bretos, F., Medina-Cruz, Y., Nodarse Andreu, G., Pérez Martín, R., García-Alfonso, E. (2018). Marine turtle conservation in Cuba: Achievements and challenges facing persistent and emerging threats. *Bull. Mar. Sci.*, *94* (2), 297-312.

Azanza-Ricardo, J., Martín-Viaña, Y. F., Martínez, Y., Márquez, L., Pérez Alarcón, A., Calderón Peña, R., Ulloa López, D., Pérez Álvarez, P., Gómez Gutiérrez, C., Vásquez Proveyer, L., Varela, R., Garea Moreda, B., Moncada Gavilán, F., Perdigón Cuellar, D., Fernández, J. C. (2019). Avances del proyecto universitario "Prevención de los efectos del cambio climático en especies amenazadas". *Monteverdia, 12* (2), e3273.

Benjamini, Y., Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc*. Series B (Methodological), *57*, 289-300.

Bethlehem, J. (2011). Nonresponse in Surveys. In: Lovric, M. (eds.). *International Encyclopedia of Statistical Science*. Springer-Verlag, Berlin.

Bougeard S., Dray S. (2018). Supervised Multiblock Analysis in R with the ade4 Package. *J. Stat. Softw.*, *86* (1), 1-17. doi: 10.18637/jss.v086.i01 (URL: https://doi.org/10.18637/jss.v086.i01).

Bourjea, J., Lapégue, S., Gagnevin, L., Broderick, D., Mortimer, J. A., Ciccione, S., Roos, D., Taquet, C., Grizel, H. (2007): Phylogeography of the green turtle, *Chelonia mydas*, in the Southwest Indian Ocean. *Mol. Ecol.*, *16* (1), 175-200. doi: http://dx.doi.org/10.1111/j.1365-294X.2006.03122.x.

Bowen, B. W., Karl, S. A. (2007). Population genetics and phylogeography of sea turtles. *Mol. Ecol.*, *16,* 4886–4907.

Bowen, B. W., Meylan, A. B., Ross, J. P., Limpus, C. J., Balazs, G. H., Avise, J. C. (1992). Global Population

Structure and Natural History of the Grenn Turtle (*Chelonia mydas*) in Terms of Matriarchal Phylogeny. *Evolution*, *46* (4), 865-881.

Dethmers, K. E. M., Broderick, D., Moitz, C., FitzSimmons, N. N., Limpus, C. J., Lavery, S., Whiting, S., Guinea, M., Prince, R. I. T., Kennett, R. (2006). The genetic structure of Australasian green turtles (*Chelonia mydas*): exploring the geographical scale of genetic exchange. *Mol. Ecol.*, *15*, 3931-3946. doi: 10.1111/j.1365-294X.2006.03070.x.

Dray, S., Dufour A. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.*, *22* (4), 1-20. doi: 10.18637/jss.v022.i04 (URL: https://doi.org/10.18637/jss.v022.i04).

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*, 241-252.

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. doi: 10.1111/j.1365-294X.2005.02553.x

Ferrera, A., Formia, A., Ciofi, C., Natali, C., Agyekumhene, A., Allman, P. (2021). Genetic structure of olive ridley sea turtles (*Lepidochelys olivacea*) in Ghana, West Africa. *J. Exp. Mar. Biol. Ecol.*, *544*, 151614. DOI: https://doi.org/10.1016/j.jembe.2021.151614.

Fitak, R. R., Johnsen, S. (2018). Green sea turtle (*Chelonia mydas*) population history indicates important demographic changes near the mid-Pleistocene transition. *Mar. Biol.*, *165*, 110. DOI: https://doi.org/10.1007/s00227-018-3366-3.

FitzSimmons, N. N., Moritz, C., Limpus, C. J., Pope, L., Prince, R. (1997). Geographic structure of mitochondrial and nuclear gene polymorphisms in Australian green turtle populations and male-biased gene flow. *Genetics, 147,* 1843–1854.

Fitzsimmons, N. N., Moritz, C., Moore, S. S. (1995). Conservation and dynamics of microsatellite *loci* over 300 million years of marine turtle evolution. *Mol. Biol. Evol.*, *12*, 432–440.

Foulley, J. L., Ollivier, L. (2006). Estimating allelic richness and its diversity. *Livestock Science*, *101*, 150-158.

François, O. (2016). Running Structure-like Population Genetic Analyses with R. *R Tutorials in Population Genetics*. U. Grenoble-Alpes.

Frichot, E., François, O. (2014). LEA: an R package for Landscape and Ecological Association studies. URL: http://membres-timc.imag.fr/Olivier.Francois/lea.html.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., Francois, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *194* (4), 973-983.

García-Machado, E., Ulmo-Díaz, G., Castellanos-Gell, J., Casane, D. (2018). Patterns of population connectivity in marine organisms of Cuba. *Bull. Mar. Sci*, *94* (2), 193-211. doi: https://doi.org/10.5343/bms.2016.1117.

Goldstein, D. B., Pollock, D. D. (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.*, *88*, 335-342.

Gulcher, J. (2012). Microsatellite Markers for Linkage and Association Studies. *Cold. Spring. Harb. Protoc.* DOI: 10.1101/pdb.top068510.

Hansen, M. M., Kenchington, E., Nielsen, E. E. (2001). Assigning individual fish to populations using microsatellite DNA markers. *Fish Fish.*, *2*, 93-112.

Hillis, D., Moritz, C., Baner, B. (1996). *Molecular Systematics*. Sinauer Associates, Inc. Sunderland, Massachusetts, USA.

Hollander, M., Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. John Wiley & Sons. New York, USA.

Hutchinson, W. F., Carvalho, G. R., Rogers, S. I. (1999). A nondestructive technique for the recovery of DNA from dried fish otoliths for subsequent molecular genetic analysis. *Mol. Ecol.*, *8*, 893-894.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*: 1403-1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T., Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi: 10.1093/bioinformatics/btr521.

Jorde, L. B., Watkins, W. S., & Bamshad, M. J. (2001). Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.*, *10*(20), 2199-2207.

Kamvar, Z. N., Brooks, J. C. and Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.*, 6, 208. doi: 10.3389/fgene.2015.00208.

Kamvar, Z. N., Tabima, J. F., Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. doi: 10.7717/peerj.281.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., Prodöhl, P. A. (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors, *Methods. Ecol. Evol*. doi: 10.1111/2041-210X.12067.

Lee, P. (2008). Molecular ecology of marine turtles: New approaches and future directions. *J. Exper. Mar. Biol. Ecol.,* 356, 25–42.

Li, Y. C., Korol, A. B., Fahima, T., Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol., 21*(6), 991-1007.

Ma, L., Ji, Y., Zhang, D. (2015). Statistical measures of genetic differentiation of populations: Rationales, history and current states. *Curr. Zool.*, *61* (5), 886-897.

Manel, S., Gaggiotti, O. E., Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends. Ecol. Evol.*, *20* (3), 136-142.

Matley, J. K., Eanes, S., Nemeth, R. S. and Jobsis, P. D. (2019). Vulnerability of sea turtles and fishes in response to two catastrophic Caribbean hurricanes, Irma and Maria. *Sci. Rep.*, *9*, 14254. https://doi.org/10.1038/s41598-019-50523-3.

Moncada Gavilán, F., Azanza-Ricardo, J., Martín-Viaña, Y.F., Gerhartz Muro, J.L., Nodarse Andreu, G., Cruz, Y. (2014). Programa de Monitoreo de Tortugas Marinas. En *Estado actual de la biodiversidad marino-costera, en la región de los Archipiélagos del Sur de Cuba* [pp. 130-141]. Hernández Ávila (Cp). Centro Nacional de Áreas Protegidas. La Habana. Cuba. Impresos Dominicanos s.r.l.

Moncada, F., Abreu, A., Muhila, A., Bell, C., Troeng, S., Bjorndal, K., Bolten, A., Meylan, A., Zurita, J., Espinosa, G., Nodarse, G., Márquez, R., Foley, A., Ehrhart, L. L. (2006). Movement Patterns of Green Turtles (*Chelonia mydas*) in Cuba and adjacent Caribbean Inferred from Flipper Tag Recaptures. *J. Herpet., 40,* 22-34.

Nishizawa, H., Abe, O., Okuyama, J., Kobayashi, M., Arai, N. (2011). Population genetic structure and implications for natal philopatry of nesting green turtles *Chelonia mydas* in the Yaeyama Islands, Japan. *Endang. Species. Res*, *14*, 141-148. doi: 10.3354/esr00355.

Paradis, E. (2020). *Population Genomics with R*. CRC Press, Taylor & Francis Group, New York, USA.

Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419-420.

Peare, T., Parker, P. G. (1996): Local genetic structure within two rookeries of *Chelonia mydas* (the green turtle). *Heredity*, *77*, 619-628.

Phu, J. L., Pilcher, N. J., Cavada-Blanco, F. (2021). A Survival Blueprint for the Conservation of the Green Sea Turtle *Chelonia mydas* in Sabah, Malaysia. *EDGE of Existence programme*, Zoological Society of London, UK.

Poloczanska, E. S., Limpus, C. J., Hays, G. C. (2009). Vulnerability of Marine Turtles to Climate Change. In: Sims D.W. (eds.). *Advances in Marine Biology* [pp. 151-211], Vol. 56, Burlington: Academic Press. ISBN: 978-0-12-374960-4.

Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multi*locus* genotype data. *Genetics*, *155*, 945-959.

Pritchard, J. K., Wen, W. (2002). *Documentation for structure software: Version 2.* URL: https://web.stanford.edu/group/pritchardlab/software/readme_structure2.pdf.

Roberts, M. A., Schwartz, T. S., Karl, S. A. (2004). Global Population Genetic Structure and Male-Mediated Gene Flow in the Green Sea Turtle (*Chelonia mydas*). Analysis of Microsatellite *Loci*. *Genetics*, *166*, 1857-1870.

Ruiz Valdéz, C. (2017). *Identificación de posibles áreas de alimentación de Chelonia mydas (Testudines, Cheloniidae)*

del archipiélago cubano. (Tesis de Licenciatura). Facultad de Biología. Universidad de La Habana. 58 pp

Ruiz-Urquiola, A., Riverón-Giró, F. B., Pérez-Bermúdez, E., Abreu-Grobois, F., González-Pumariega, M., James-Petric, B. L., Díaz-Fernández, R., Álvarez- Castro, J. M., Jager, M., Azanza-Ricardo, J., Espinosa-López, G. (2010). Population genetic structure of greater caribbean green turtles (*Chelonia mydas*) based on mitochondrial dna sequences, with an emphasis on rookeries from southwestern Cuba. *Rev. Invest. Mar*, *31* (1), 33-52.

Seminoff, J. A., Jones, T. T., Resendiz, A., Nichols, W. J., Chaloupka, M. Y. (2003). Monitoring green turtles (*Chelonia mydas*) at a coastal foraging area in Baja California, Mexico: multiple indices to describe population status. *J. Mar. Biol. Ass*, *U. K.*, *83*, 1355-1362.

Shamblin, B. M., Hart K. M., Martin, K. J., Ceriani, S. A., Bagley, D. A., Mansfield, K. L., Ehrhart, L. M., Campbell, J. N. (2020). Green turtle mitochondrial microsatellites indicate finer-scale natal homing to isolated islands than to continental nesting sites. *Mar. Ecol. Prog. Ser*, *643*, 159-171. https://doi.org/10.3354/meps13348.

Stephens, S. H. (2004). *Genetic analysis of the Kemp's ridley sea turtle (Lepidochelys kempii) and estimates of effective population size*. (Doctoral Thesis). Texas A&M University.

Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P., Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecol. Evol.*, *6* (11), 3461-3475.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes*, *4*, 535-538.

Wang, J. (2017). The computer program Structure for assigning individuals to populations: easy to use but easier to misuse. *Mol. Ecol. Resour.*, *17*, 981-990.

Weir, B., Cockerham, C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358-1370.

Wright, S. (1951). The Genetical Structure of Populations. *Ann. Eugen. 15,* 323–354.

Zhang, T., Lin, L., Gaillard, D., Cheng, F., Chen, H., Li, M., Wu, S., Wang, Z., Shi, H. (2021). Rapid authenticity testing of artificially bred green turtles (*Chelonia mydas*) using microsatellite and mitochondrial DNA markers. *PeerJ*, *9*, e12410. DOI: http://Doi.org/10.7717/peerj.12410.

How to cite this article

Madrigal-Roca, L.J., Azanza-Ricardo, J., Espinosa-López, G., Oyama, K., Abreu-Grobois, F. A., Chassin-Noria, O. (2023). Assessing genetic structuring for endangered *Chelonia mydas* (Testudines: Cheloniidae) in southwest Cuba using microsatellites. *Rev. Invest. Mar.*, *43*(1), 69-89.