ORIGINAL PAPER

The first assessment of the genetic diversity and structure of the endangered West Indian manatee in Cuba

Anmari Alvarez-Aleman[1,2,3] · Margaret E. Hunter[4] · Thomas K. Frazer[1,7] · James A. Powell[3] · Eddy Garcia Alfonso[5] · James D. Austin[1,6]

Abstract

The coastal waters of Cuba are home to a small, endangered population of West Indian manatee, which would benefit from a comprehensive characterization of the population's genetic variation. We conducted the first genetic assessment of Cuban manatees to determine the extent of the population's genetic structure and characterize the neutral genetic diversity among regions within the archipelago. We genotyped 49 manatees at 18 microsatellite loci, a subset of 27 samples on 1703 single nucleotide polymorphisms (SNPs), and sequenced 59 manatees at the mitochondrial control region. The Cuba manatee population had low nuclear (microsatellites $H_E = 0.44$, and SNP $H_E = 0.29$) and mitochondrial genetic diversity ($h = 0.068$ and $\pi = 0.00025$), and displayed moderate departures from random mating (microsatellite $F_{IS} = 0.12$, SNP $F_{IS} = 0.10$). Our results suggest that the western portion of the archipelago undergoes periodic exchange of alleles based on the evidence of shared ancestry and low but significant differentiation. The southeast Guantanamo Bay region and the western portion of the archipelago were more differentiated than southwest and northwest manatees. The genetic distinctiveness observed in the southeast supports its recognition as a demographically independent unit for natural resource management regardless of whether it is due to historical isolation or isolation by distance. Estimates of the regional effective population sizes, with the microsatellite and SNP datasets, were small (all $N_e < 60$). Subsequent analyses using additional samples could better examine how the observed structure is masking simple isolation by distance patterns or whether ecological or biogeographic forces shape genetic patterns.

Keywords Antillean manatee · Population genetics · Conservation · Microsatellites · Single nucleotide polymorphisms (SNPs) · Mitochondrial DNA

Introduction

The West Indian manatee (*Trichechus manatus*) is a charismatic aquatic mammal with broad distribution throughout the West Atlantic, Caribbean, and coastal Central and South America (Lefebvre et al. 2001). Manatees play an essential trophic role as grazers of seagrass beds (Jackson et al. 2001) and are valued as an ecotourism draw in many countries (Sorice et al. 2006). The two subspecies, *T. m. manatus* (Antillean manatee) and *T. m. latirostris* (Florida manatee)

✉ Anmari Alvarez-Aleman
aalvarez@cmaquarium.org

1 School of Natural Resources and Environment, University of Florida, 2035 McCarty hall D, Gainesville, FL 32611, USA

2 Centro de Investigaciones Marinas, Universidad de La Habana, Calle 16 # 114 Entre 1ra y 3ra Plaza, Havana, Cuba

3 Clearwater Marine Aquarium Research Institute, Clearwater Marine Aquarium, Clearwater, FL, USA

4 U.S. Geological Survey, Wetland and Aquatic Research Center, 7920 NW 71st Street, Gainesville, FL 32653, USA

5 Refugio de Fauna Lanzanillo-Pajonal-Fragoso, Empresa Provincial para la Protección de la Flora y la Fauna, Villa Clara, Cuba

6 Department of Wildlife Ecology and Conservation, University of Florida, 110 Newins-Ziegler Hall, Gainesville, FL 32611, USA

7 College of Marine Science, University of South Florida, 140 Seventh Avenue South, KRC 3109, St. Petersburg, FL 33701, USA

have individually been listed as "Endangered" by the International Union for the Conservation of Nature (IUCN) due to over-harvesting and habitat degradation (Deutsch 2008; Self-Sullivan and Mignucci-Giannoni 2008). Under the U.S. Endangered Species Act, the species was recently reclassified from Endangered to "Threatened," primarily based on the large Florida manatee population (Federal Registrar 2017). Data on the status of *T. manatus* is limited across much of its range, apart from areas in the Caribbean possessing relatively large remnant manatee populations (Puerto Rico, Mexico, Belize, and Florida). Of particular interest is the status of this species in Cuba, the largest island of the West Indies where manatees have been under-studied (Quintana-Rizzo and Reynold 2010; Alvarez-Alemán et al. 2018a). Cuba may have valuable population given the large extent of coastal habitat (coastline of approximately 5975 km), which, if suitably managed, could support a large manatee population (Schill et al. 2015). Furthermore, Cuba could play an essential role for the species given its central geographic location within the Caribbean, making the archipelago an evolutionarily important 'steppingstone' across northern Caribbean islands and continental populations.

Historically, manatees occupied much of the coastal habitat of Cuba, with concentrations in regions where numerous river mouths provide access to freshwater (Cuni 1918; Lefebvre et al. 2001; Jiménez-Vázquez 2015). However, manatee numbers in Cuba were assumed to be significantly reduced due to hunting by the start of the twentieth century and were considered perilously low by the 1970s (Lefebvre et al. 2001; Alvarez-Alemán et al. 2018a). Today, reversing the decline in manatee numbers continues to be challenged mainly by poaching and accidental deaths due to net entanglements, but also likely by habitat degradation and boat strikes (Alvarez-Alemán et al. 2018a, 2021). For example, the heavy loss (> 26%) of seagrass beds in the Gulf of Batabanó (Martínez-Daranas et al. 2009), an area of approximately 5580 km² that separates the main island from Isla de la Juventud, has occurred in a region of Cuba where manatees are relatively common (Alvarez-Alemán et al. 2018a, 2021). Furthermore, on the east of the northern and southern coasts, shelf habitats have been altered due to river damming, reducing the volume of freshwater required for seagrasses (Martínez-Daranas and Suárez 2018).

Assessment of the status of manatees in Cuba would benefit from understanding the genetic variation and extent of genetic structuring (Funk et al. 2019). Delineating population structure is an important management tool to protect unique variation within genetically unique groups of species of conservation concern (Moritz 1994; Frankham et al. 2010). Understanding neutral population structure can inform the scale over which dispersal occurs and how genetic drift shapes the neutral genetic changes of small populations (Slatkin 1987). Neutral genetic markers

are also helpful in reconstructing demographic parameters, particularly the effective population size (N_e), which can provide information on the level of variability in a population, and the effectiveness of selection relative to drift (Charlesworth 2009). These factors can be applied to assess future management options, such as planning translocations as a means of ecological restoration or to increase populations suffering from low relative genetic variation (Weeks et al. 2011, Hufbauer et al. 2015). These management options rely on a well-known genetic status to estimate the risks (e.g., inbreeding and outbreeding) associated with selecting the source and recipient populations and locations (Ralls et al. 2018).

Our goal was to add to our knowledge of West Indian manatee genetic variation by filling the knowledge gap in the central portion of the range represented by the archipelago of Cuba. The main objective was to determine the extent of population genetic structuring, characterize the neutral genetic diversity across the archipelago and estimate effective population size. We assessed mitochondrial (mtDNA) and microsatellite markers previously utilized to characterize other manatee populations (Vianna et al. 2006; Hunter et al. 2010, 2012; Nourisson et al. 2011; Tucker et al. 2012). In addition, we generated a single nucleotide polymorphism (SNP) dataset on a subset of manatee samples to provide a larger genome-wide representation across numerous biallelic loci for comparison with the 18 multilocus microsatellites (Väli et al. 2008; Fischer et al. 2017).

Materials and methods

Sample collection and DNA extraction

Manatee tissue samples were obtained from museum specimens (N = 19) and opportunistically from carcasses (N = 37) or manatees rescued from nets (N = 2). In addition, we obtained dermis tissue from 25 manatees sampled at Isla de la Juventud (N = 12) and Guantanamo Bay (N = 13) during capture and health assessment. We preserved tissue samples in SED buffer solution (saturated NaCl; 250 mM EDTA, pH 7.5; 20% DMSO) or 90% ethanol and stored them at room temperature until they could be refrigerated. The collected manatee samples generally clustered into three geographic regions: (1) the northwest coast (NW), primarily in Villa Clara province, (2) the southwest (SW), including Ciénaga de Zapata Peninsula (SW-M) and Isla de la Juventud (SW-IJ), and (3) the southeast (SE) coast of Cuba, primarily Guantánamo Bay (Fig. 1). DNA was extracted using either a standard phenol/chloroform method (Sambrock and Russell 2001) or the QIAGEN DNeasy Blood and Tissue Kit (Valencia, California), quantified using an ND–1000
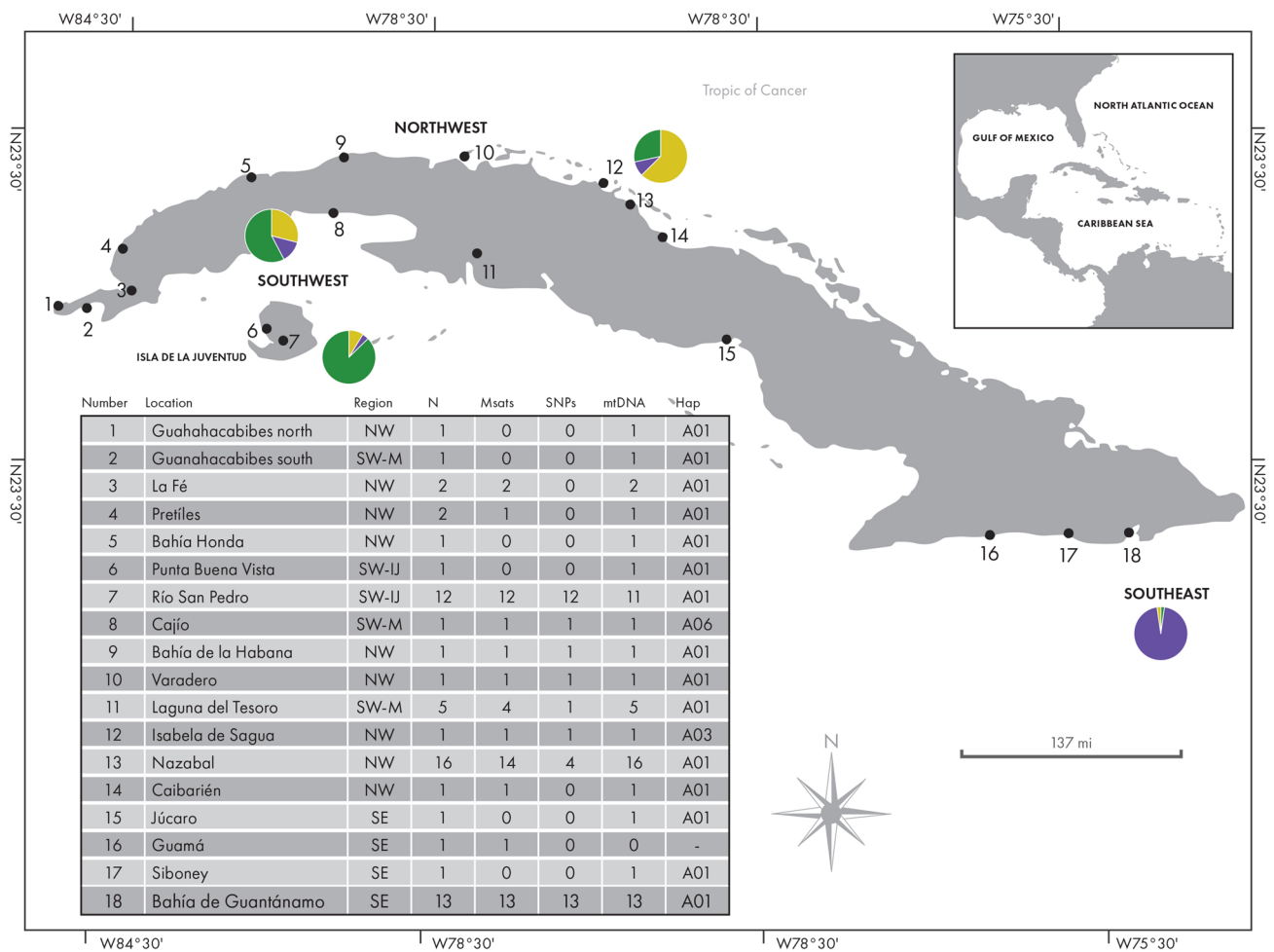
Fig. 1 Distribution of Cuban manatee samples used in this study. The table (inset) includes location number (as denoted on the map), name, region (*NW* Northwest, *SW-M* Southwest mainland, *SW-IJ* Isla de la Juventud, *SE* Southeast), and sample size (N) at that location, and the number of samples that were included in the microsatellite (Msats), SNPs or mitochondrial DNA (mtDNA) dataset. Mitochondrial haplotypes (Hap) are based on reference control region sequences (Vianna et al. 2006). A06 represents a novel haplotype. No haplotype was recovered for location 16. Pie-charts in each sub-region represent the proportion of membership of each of the 3 clusters ($K=3$) in that pre-defined population, obtained with Structure with the 18 loci microsatellite dataset. This proportion was calculated based on 20, 5, 11, 13 individuals from the NW, SW-M, SW-IJ and SE respectively

Number	Location	Region	N	Msats	SNPs	mtDNA	Hap
1	Guahahacabibes north	NW	1	0	0	1	A01
2	Guanahacabibes south	SW-M	1	0	0	1	A01
3	La Fé	NW	2	2	0	2	A01
4	Pretíles	NW	2	1	0	1	A01
5	Bahía Honda	NW	1	0	0	1	A01
6	Punta Buena Vista	SW-IJ	1	0	0	1	A01
7	Río San Pedro	SW-IJ	12	12	12	11	A01
8	Cajío	SW-M	1	1	1	1	A06
9	Bahía de la Habana	NW	1	1	1	1	A01
10	Varadero	NW	1	1	1	1	A01
11	Laguna del Tesoro	SW-M	5	4	1	5	A01
12	Isabela de Sagua	NW	1	1	1	1	A03
13	Nazabal	NW	16	14	4	16	A01
14	Caibarién	NW	1	1	0	1	A01
15	Júcaro	SE	1	0	0	1	A01
16	Guamá	SE	1	1	0	0	-
17	Siboney	SE	1	0	0	1	A01
18	Bahía de Guantánamo	SE	13	13	13	13	A01

spectrophotometer (Nanodrop, Wilmington, DE) and standardized to 20 ng/µL prior to PCR amplification.

Microsatellite genotyping and marker assessment

We used 18 polymorphic microsatellite loci designed for West Indian manatees (García-Rodríguez et al. 2000; Pause et al. 2007) and previously used to assess genetic structure and diversity in many manatee regional populations (Belize, Hunter et al. 2011; Florida, Tucker et al. 2012; Puerto Rico, Hunter et al. 2012; Mexico, Nourisson et al. 2011) (Table S1). We used eight multiplex PCR reactions (Davis, 2014) consisting of a total volume of 13.4 µL containing: 1 µL of 20 ng DNA, 6.7 µL 1×Qiagen Multiplex PCR Master Mix Type-it® Kit (3 mM Mg, 200 uM each dNTP,

HotStarTaq Plus DNA Polymerase) (Qiagen, Valencia, CA), 0.02−0.06 µL forward+reverse primer and brought to volume with ddH$_2$O. The PCR reaction profile was 95 °C for 5 min, 34−40×(95 °C for 30 s, 54−60 °C for 1 min, 72 °C for 1 min), final extension at 72 °C for 10 min, ending with a 4 °C hold. PCR products were electrophoresed on an Applied Biosystems 3730XL automated genetic analyzer with a Genescan Rox500 size standard (Applied Biosystems, Inc., Foster City, CA, USA). Positive and negative controls were included for each PCR plate. Genotypes were scored using GENEMARKER software version 2.7.4 (SoftGenetics LLC, State College, PA, USA). Samples that did not have 100% multilocus amplification success (primarily museum or highly degraded carcass samples) were reamplified a minimum of three times. In order to calculate the error rate,

we re-genotyped the 18 loci at an average of 8 individuals (~15% of samples). The mean error rate per locus (eL) was calculated as the ratio between the number of single-locus genotypes, including at least one allelic mismatch (mL), and the number of replicated single-locus genotypes (nt). We retained genotypes with no more than two missing loci for our final dataset.

Null allele frequencies across all samples were estimated using the null.all function in the R (R Development Core Team 2019) package GenPopReport, version 3.0.4 (Gruber and Adamack 2013). This function bootstraps genotypes to estimate the probability of identifying the observed number of homozygotes of each allele, then a second bootstrap calculates the 95% confidence interval for each following locus (Chakraborty et al. 1994). Null alleles are not significant when confidence intervals overlap zero. Due to the observed genetic structure between east and west samples (see page 15, Population Structure, Results) and evidence of substructuring and admixture among NW and SW manatee genotypes, we also subset the data and re-ran null.all for the western manatees (NW + SW-M + SW-IJ) and the SE separately. We examined the impact of null alleles on analyses that assume HWE by comparing the results of pairwise F_{ST}, STRUCTURE (see below) and the genetic diversity and differentiation, using the full and the reduced datasets with and without the loci displaying potential null alleles, respectively.

We tested for departure from Hardy–Weinberg equilibrium (HWE) using the R package Pegas, version 0.13 (Paradis 2010) with sequential Bonferroni adjustment correction for multiple tests (Rice 1989). We examined HWE for the full Cuban dataset and separately for NW, SW, and SE regions. We tested for linkage disequilibrium (LD) between all pairs of loci (P-value based on 1000 permutations) using Fstat 2.9.4. with sequential Bonferroni adjustments.

Single nucleotide polymorphism genotyping and filtering

We used the whole genome profiling service at Diversity Array Technology (DArT; http://www.diversity arrays. com/). This platform implements sequencing complexity reduced representations applied on next generation sequencing platforms (RadSeq). Complexity reduction was achieved with the PstI and SphI restriction enzymes method, and genotyping followed methods outlined in Grewe et al. (2015). We selected 33 samples with high molecular weight and 260/280 ratios of > 1.6 to search for single nucleotide polymorphisms (SNPs). Sequences were bioinformatically processed using DArTseq analytical pipelines. In the primary pipeline, the FASTQ files were first processed to filter away poor-quality sequences, applying more stringent selection criteria to the barcode region compared to the

rest of the sequence (barcode region: Min Phred pass score 30, Min pass percentage 75, and whole read: Min Phred pass score 10, Min pass percentage 75). In that way, the assignment of the sequences to specific samples carried in the "barcode split" step was reliable (Grewe et al. 2015). Identical sequences were collapsed into "fastqcoll files" and "groomed" using DArT PLD's proprietary algorithm that corrects low-quality bases from singleton tags into correct bases. The "groomed" fastqcoll files were used in the secondary pipeline for DArT PLD's SNP and SilicoDArT (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14). For SNP calling, all tags from all libraries included in the DArTsoft14 analysis are clustered using DArT PL's C++ algorithm at the threshold distance of 3. These effects were followed by parsing the clusters into separate SNP loci using a range of technical parameters, especially the balance of read counts for the allelic pairs. Calling quality was assured by a high average read depth per locus (Grewe et al. 2015).

Using DartR version 1.1.11 (Gruber et al. 2017) (Table S4), we conducted additional quality control filtering steps by filtering out loci with less than 100% reproducibility, generated by reproducing the data independently for 30% of loci. We filtered out loci with call rates < 90%, a value usually arising from failure to call a SNP due to a mutation at one or both of the restriction enzyme recognition sites. In this case, the data are missing owing to low coverage (Gruber et al. 2017). A single SNP per locus was selected, and loci with minor allele frequencies of < 0.05 were omitted (Table S5). Furthermore, we only retained individual genotypes with ≤ 15% missing data. Departure from HWE was tested with the function gl.filter.hwe in DartR version 1.1.11 (Gruber et al. 2017). Finally, we used the OutFLANK method (Whitlock and Lotterhos 2015) to detect putatively non-neutral SNPs for filtration out of the dataset. This method infers the distribution of F_{ST} for loci unlikely to be strongly affected by spatially diversifying selection, using data on a large set of loci with unknown selective properties (Whitlock and Lotterhos 2015).

Population structure

We assessed the extent of genetic structure using a model-based approach implemented in STRUCTURE version 2.3.4 (Pritchard et al. 2000) and using the non-model multivariate approach of Principal Component Analysis (PCA). Both methods were explored for the microsatellite and SNP datasets separately. For STRUCTURE, we ran a 150,000 burn-in followed by 150,000 MCMC chains with 15 replicate runs per assessed genetic cluster (K) using the admixture model and default settings, and we evaluated models of K from 1 to 10. Support for K was evaluated based on the mean likelihood model support (LnP(K)) and the ad hoc ΔK, based

on the rate of change in the LnP(K) between successive K values (Evanno et al. 2005). The ΔK measure is expected to correctly identify the uppermost hierarchical level of genetic structure (Evanno et al. 2005); however, results can be downwardly biased when uneven sampling is involved (Puechmaille 2016). Therefore, to reduce the effect of uneven sampling across regions of Cuba, we used four additional supervised methods, MedMeaK, MaxMeaK, MedMedK, and MaxMedK (Puechmaille 2016), to estimate the number of clusters. This method assigns genotypes to a cluster if the mean (or median) membership coefficient is above a threshold membership coefficient (Puechmaille 2016). We computed all four estimators across thresholds varying from 0.5 to 0.8. Run summaries, ΔK, MedMeaK, MaxMeaK, MedMedK, and MaxMedK, were generated posthoc using STRUCTURE SELECTOR (Li and Liu 2018). STRUCTURE run summaries were generated using CLUMPP (Jakobsson and Rosenberg 2007) and visualized using Distruct (Rosenberg 2004).

STRUCTURE identifies clusters of genotypes by assuming they fit HWE, which could be affected by high frequencies of null alleles. Because PCA is model-free, it does not require assumptions based on HWE or LD (Jombart et al. 2009). PCA was done using the function dudi.pca in the R package ade4R, version 1.7–13 (Dray and Dufour 2007) with weighted vectors and mean centering. We transformed the absolute variance of PC axes to the percentage of total variance represented in the genotype data.

We conducted an Analysis of Molecular Variance (AMOVA) separately on the microsatellite and SNP datasets. We used AMOVA to measure the extent of hierarchical structure based on the results from STRUCTURE and PCA (see Results). Microsatellite (only 18 loci were tested as AMOVA does not rely on HWE assumptions) and SNP variation were examined separately. We nested SW-ML + SW-IJ + NW into one regional group, with SE as a second, for the highest level of variation summarized by AMOVA. The 95% confidence intervals for F-statistics were assessed using 1000 bootstrap permutations, calculated using GenoDive version 3.0 (Meirmans and Tienderen 2004).

Non-spatial estimates of structure (i.e., STRUCTURE and AMOVA) assume an island model (Wright 1943). Because allele frequency differences among our regional samples may be a result of geographic distance (particularly between the SE and other regions), we assess the potential influence of spatial dependence on the microsatellite genetic structure of manatees using spatial PCA (sPCA) (Jombart 2008) implemented in adegenet, version 2.1.2 (Jombart 2015). Spatial PCA summarizes spatial autocorrelation (Moran's I) among allele frequencies with the genetic variance among samples (PCA) to produce composite eigenvalues to detect spatial patterns. Positive eigenvalues represent neighboring genotypes that tend to be highly similar (i.e., positive spatial autocorrelation or global structure), while the negative values represent neighboring genotypes that tend to be highly dissimilar (negative spatial autocorrelation, or local structure). We modified Delaunay triangulation (type 1) for the connection network so those neighboring connections had to follow the coastline or across open water (i.e., removing connections between genotypes sampled on opposite sides of the island) using the edit.nb = TRUE argument in the chooseCN function. We implemented the Monte-Carlo tests to assess the significance of the largest positive (global.rtest) and largest negative (local.rtest) eigenvalues using 9999 permutations. We visualized spatial genetic patterns through the plot.spca function. We subsequently conducted a post hoc sPCA as before but omitted the SE samples, such that the connection network only joined manatees from the NW to the SW genotypes around the western end of the archipelago to determine whether the north–south global structure persisted with the exclusion of the more distant SE genotypes. We did not analyze SNP genotypes due to the small sample size and limited geographic coverage (Fig. 1).

Genetic diversity and differentiation

We quantified variation across the complete microsatellite and SNP datasets and for the three main genetic clusters identified using the statistical metrics discussed above (Meirmans 2015). We estimated expected (H_E), observed heterozygosity (Ho), and the inbreeding coefficient (F_{IS}) for diagnosed clusters and overall. F_{IS} bootstrapping over loci was implemented with 1000 bootstraps using the function boot.ppfis from the R package HierFstat, version 0.04–22 (Goudet 2005). We also calculated the average number of alleles per locus (A) and allelic richness (allelic diversity corrected for differences in sample size, AR) for microsatellites. Diversity measures were calculated using GenoDive and the PopGenReport package (Gruber and Adamack 2013).

The power of these 18 microsatellite loci to detect genetic heterogeneity among at most three regional samples (NW, SW, and SE, see results) was tested using POWSIM vers. 4.0 (Ryman and Palm 2006). POWSIM determines whether we would expect to detect genetic heterogeneity given the empirical allele frequencies under a model of drift alone. We simulated 1000 iterations at multiple effective populations sizes (300, 500, 700, and 1000) and then drew samples reflecting our empirical samples sizes, with F_{ST} calculated at 10, 20, and 100 generations of drift. Fisher's exact tests were used to test for homogeneity. Tests were also conducted at generation 0, which is expected to give false significance (α) at approximately 0.05. The proportion of significant tests over 1000 replicates at each calculated value of F_{ST} was

used to indicate the power of the markers to detect differentiation due to drift. Results were compared against the empirical estimate of global F_{ST} (Nei 1987) generated using GenoDive.

We calculated pairwise fixation (G''_{ST}), which is standardized, allowing for comparisons of fixation generated from microsatellite and SNP datasets (Hedrick 2005). We also calculated differentiation (Jost et al. 2008), which is independent of within-population heterozygosity. Pairwise estimates were calculated between SE and combined NW + SW-M + SW-IJ and three separated groups, SE, NW, and SW (SW-IJ + SW-M) samples based on structure analyses. Calculations were done using GenoDive and included 999 bootstrap replicates to estimate *P*-values.

Effective population size

Recent effective population size (N_e) was estimated using the bias-corrected version of the linkage disequilibrium (LD) method implemented in NeEstimator version 2 (Waples and Do 2008; Do et al. 2014). We evaluated jackknife estimates of 95% confidence intervals, reducing potential bias associated with standard chi-squared confidence intervals (Do et al. 2014). We calculated N_e assuming random mating and evaluated minimum allele frequencies of 0.02. N_e was calculated island-wide using the 18 loci microsatellite and 1703 SNP dataset. The total dataset was analyzed, and subsets of samples were independently evaluated as well, NW (only in the microsatellite dataset), SE, SW (SW-IJ and SW-M), and west samples combined (NW + SW-IJ + SW-M), reflecting different levels of the genetic structure identified (see "Results" section).

Mitochondrial haplotype assessment

We sequenced a portion of the mitochondrial control region (CR) spanning 410 base pairs for 59 manatees using primers CR4 and CR5 (García-Rodríguez et al. 1998) or a combination of these with two sets of shorter primers for problematic samples (CR4short- 5′ CAGATTCCCAACCACATGG 3′ and CR5short-5′ TTAAGGAGCATGATGTGCTTATGC 3′). Individual PCRs consisted of 10 ng of genomic DNA, 1 × Taqreaction buffer, 200 μm dNTPs, 0.25 μm of each primer, and 1 U of Taq polymerase (Titanium) per PCR. The cycling reaction was performed as follows: 94 °C for 1 min, followed by 35 cycles of 95 °C for 45 s, 53 °C for 1 min and 72 °C for 45 s, and a final extension period of 72 °C for 10 min. The annealing temperature was usually 53 °C, but some samples were amplified by adjusting the PCR annealing temperature to 50 °C. PCR products were cleaned with ExoI/SAP at 37 °C for 45 min, followed by 80 °C for 15 min (Applied Biosystems, Foster City, CA, USA). DNA sequences were obtained using Big Dye

terminator sequencing chemistry (Applied Biosystems) in 6.75 μL reactions, containing 0.83 μL Terminator v3.1 Ready Reaction mix, 1.67 μL 5 × reaction buffer, 0.22 μL of primer, 1 μL of PCR product, and 3.92 μL ddH$_2$O. Cleaned sequence products [0.2 μm Sephadex columns (Princeton Separations, Freehold, NJ, USA)] were electrophoresed on a 3130xl Genetic Analyzer (Applied Biosystems). All fragments were bi-directionally sequenced. The sequences were manually confirmed, trimmed, and aligned in Genious 11.1.5.

We evaluated mitochondrial diversity as the total number of haplotypes (*H*), haplotype diversity (*h*), and nucleotide diversity (π) using DnaSP version 5.10 (Librado and Rozas 2009). We also compared our new sequences to existing haplotypes (García-Rodríguez et al. 1998; Vianna et al. 2006; Hunter et al. 2012; Satizábal et al. 2012). Due to the near-zero haplotype diversity, multi-haplotype population genetic analyses were omitted.

Results

Genotyping and marker evaluation

We successfully genotyped 53 manatee samples at 16–18 microsatellite loci. High levels of missing data were primarily associated with museum samples (only 1 of 19 successfully genotyped at > 15 loci) and carcasses (26 of 37 genotyped at > 15 loci). The museum sample successfully genotyped was omitted due to unknown metadata sample collection information. We also omitted three calves collected with mothers to reduce bias in the analyses, resulting in 49 microsatellite genotypes (Peterman et al. 2016). Regionally, 40% were genotyped from the NW (N = 20), 33% from the SW (N = 11 from SW-IJ and N = 5 from SW-M), and 27% from the SE (N = 13) (Fig. 1).

Six out of the 18 microsatellites loci had signals of potential null alleles for the total dataset (TmaSC05, TmaE01, TmaE07, TmaA03, TmaE04, TmaA02) (Table S2). Compared with previous studies utilizing the same set (or subsets) of loci, this frequency of null alleles is relatively high (see Table S1). Assessment of the NW + SW group of samples identified the same six null loci, while evaluation of the SE manatees identified a single locus with possible null alleles (TmaE7). Most signatures of null alleles were restricted to western samples that also reflect some degree of admixture. The geographic pattern of null alleles suggested that these significant tests may reflect demographic processes rather than true null alleles. Furthermore, true null alleles are more common in very large populations with high mutation rates in the microsatellite flanking regions (Chapuis and Estoup 2007). Both conditions are not expected in the Cuba

population relative to others that have been studied (Nourisson et al. 2011; Hunter et al. 2010, 2012; Tucker et al. 2012).

Consistent with previous manatee genetic studies utilizing loci in common, there was no evidence of linkage among microsatellite loci. Estimates of genotyping error were 2% overall, with 14 loci having no detected amplification errors. Four loci each produced a single allelic mismatch across replicated samples ranging from 9 to 13 depending on the locus (Table S2). Each observed genotyping mismatch was associated with DNA extracted from a carcass.

The initial number of SNPs obtained from the DArT pipeline was 8419, and the final number was 1703 polymorphic loci among 30 manatees after filtering (Table S4). All loci were at HWE. Of the 27 manatees retained, after omitting three calves, 15% of samples were derived from carcasses (N = 4) and 85% from live manatees (N = 23). These SNP genotypes represented the NW (N = 4), SW-IJ (N = 12), and the SE (N = 12).

Population structure

Both microsatellite and SNP datasets revealed a west–east clustering pattern at the highest hierarchical level assignment where (1) SE manatees were differentiated from (2) NW, SW-M, and SW-IJ manatees (Fig. 2). For 18 microsatellite loci, the highest LnP(K) was at K = 4 (LnP(K) = -1345.8); however, ΔK identified the greatest rate of change at K = 2 (ΔK = 13.2; Fig. 2). The MedMeaK, MaxMeaK, MedMedK, and MaxMedK methods identified two or three genetic clusters depending on the stringency of the membership coefficient thresholds and the dataset [e.g., 0.5—0.8 (Table S4)]. For K = 2, all western Cuba manatee genotypes reflected a largely uniform second cluster, although some individuals reflected mixed ancestry (K = 2, Fig. 2). At K = 3, the two additional clusters were co-distributed between the NW, SW-M, and SW-IJ genotypes, again providing a picture of differentiation between SE and those to the west, and limited structuring among NW and SW manatees (Fig. 2). In the case of both models (K = 2 and K = 3), the SE manatees were a homogeneous cluster (Fig. 2). After omitting loci with possible null alleles, STRUCTURE identified two clusters based on both LnP(K) (-890.5) and ΔK (50.48; Fig. 2), a result that reinforces the west–east pattern observed with the larger microsatellite dataset.

SNP STRUCTURE results identified the largest LnP(K) at K = 7 (LnP(K) = -40,281.08) although there were minimal differences among most LnP(K) between K = 3–8. ΔK identified K = 2 (ΔK = 197.5) as representing the highest level of structure for the SNP data (Fig. 2). The MedMeaK, MaxMeaK, MedMedK, and MaxMedK methods identified three genetic clusters at the membership coefficient

thresholds of 0.5 and 0.6 and 2 at the more conservative threshold of 0.7 and 0.8 (Table S4).

Results from the PCA mirrored the STRUCTURE analysis in that it separated the SE genotypes from the remainder of the manatee samples examined (Fig. 3). For the 18 microsatellite loci dataset, this east–west separation was identified in the 2nd axis, explaining 10.97% of the total variation. The first Principal Component (PC, 13.43%) did not reflect the variance among regional samples. However, the centroid of the SW-M was centrally located along the 1st PC axis between the NW and SW-IJ (Fig. 3), possibly reflecting a pattern of coastal isolation by distance. The first two PCs of the SNP results explained 17.66% and 8.33% of the total variation. The first PC differentiated the SE manatees from the western samples, and the second PC explained most of the variation between the NW and SW-IJ (Fig. 3). Results from the microsatellite AMOVA found that 77.9% of variation in allele frequencies was observed within individuals ($F_{IT} = 0.221$, 95% CI 0.090–0.337), 12.8% among individuals within regional populations ($F_{IS} = 0.142$, 95% CI 0.022–0.249), and 9.3% among regional populations ($F_{ST} = 0.093$, 95% CI 0.046–0.146). Individual locus results were highly variable, with 6 loci displaying significance among individual (F_{IS}) and 7 loci significant among region (F_{ST}) variance, with only 2 loci being significant at both F-statistics (Table S5). For the SNP AMOVA 79.7% of variation in allele frequencies was observed within individuals ($F_{IT} = 0.203$, 95% CI 0.188–0.217), 6.2% among individuals within regional populations ($F_{IS} = 0.072$, 95% CI 0.060–0.084), and 14.1% among three regional populations ($F_{ST} = 0.141$, 95% CI 0.130–0.151).

Spatial PCA results from 18 microsatellite loci had large eigenvalues at the first two positive λ scores (0.4366 and 0.3622 respectively). Similarly, the eigenvalue decomposition shows that λ_1 and λ_2 explain a large proportion of the spatial autocorrelation and the genetic variance (Figs. S1-S2). Significant spatial autocorrelation was detected for the global test for λ_1 (0.0644, P = 0.0001), but local tests were not significant for λ_{44} (0.0350, P = 0.2411). We subsequently retained λ_1 and λ_2 to examine their contribution to global spatial patterns (Figs. S1-S2). The first eigenvalue depicted a clear east–west pattern reflecting the relatively large genetic distance between Guama + Bahia de Guantanamo (SE) and the rest of the manatee genotypes. Interestingly, the first eigenvalue (λ_1) represented greater spatial autocorrelation but slightly less genetic variance than the second λ_2 (Fig. S1). For λ_2, most of the spatial pattern appears to be explained by variation between the NW and samples representing the Gulf of Batabanó (SW-ML + SW-IJ) and some SE genotypes perhaps reflecting subtle substructure among north and south coasts (Fig. S2). The results of the post hoc sPCA omitting

Fig. 2 Plot of posterior probability of assignment for Cuba manatees based on: **A** 18 microsatellite loci, **B** 12 microsatellite loci, and **C** 1703 SNP loci. Each includes the mean likelihood plot (LnP(*K*)) from 15 replicates per *K*, the Δ*K* plot, and two bar graphs illustrating individual genome assignment to one of K clusters (*K* = 2 or *K* = 3). Geographic areas are NW, Nazabal; SW-M; samples from Isla de la Juventud (SW-IJ); and SE, primarily Guantanamo Bay (see Fig. 1)
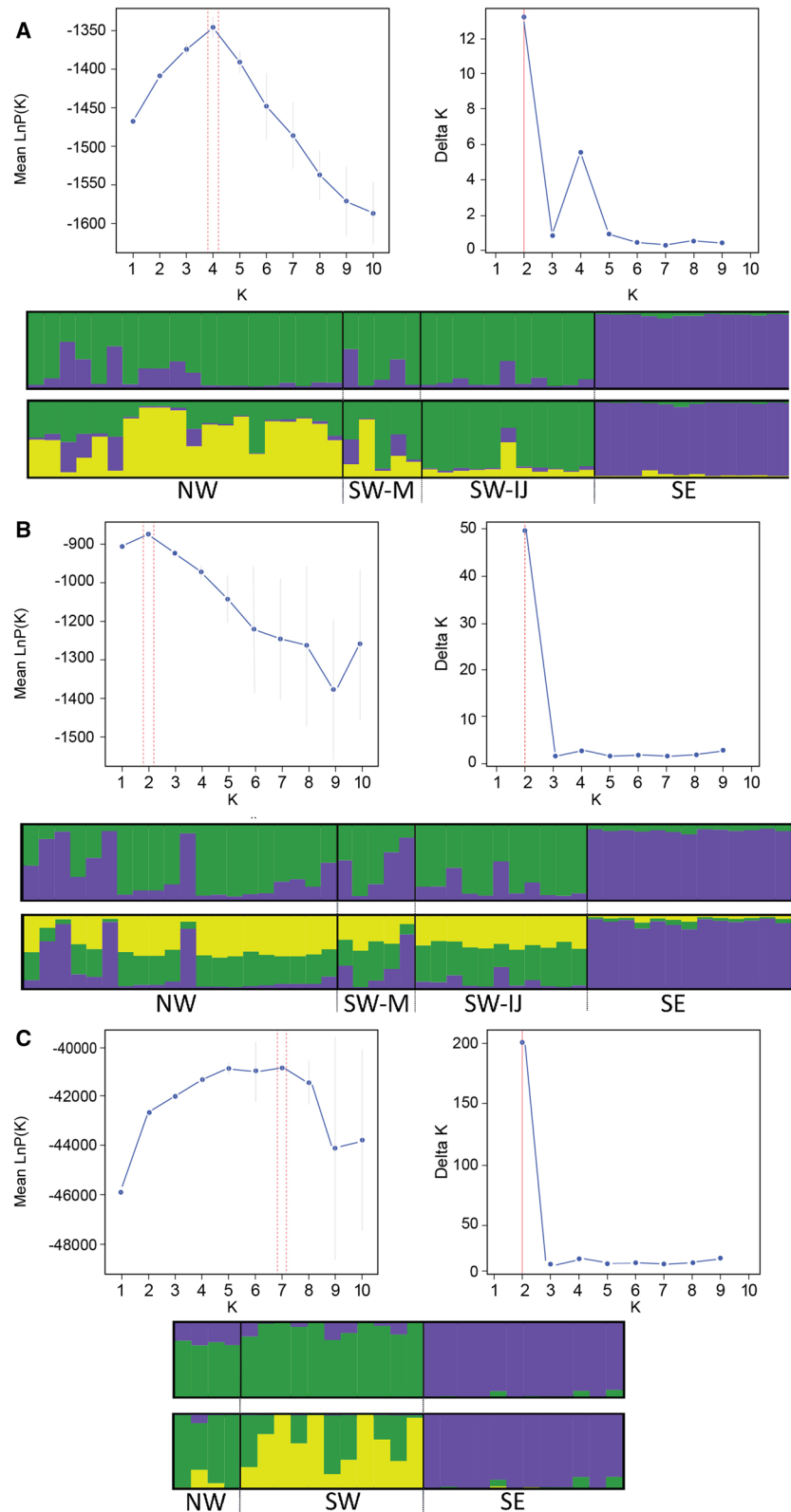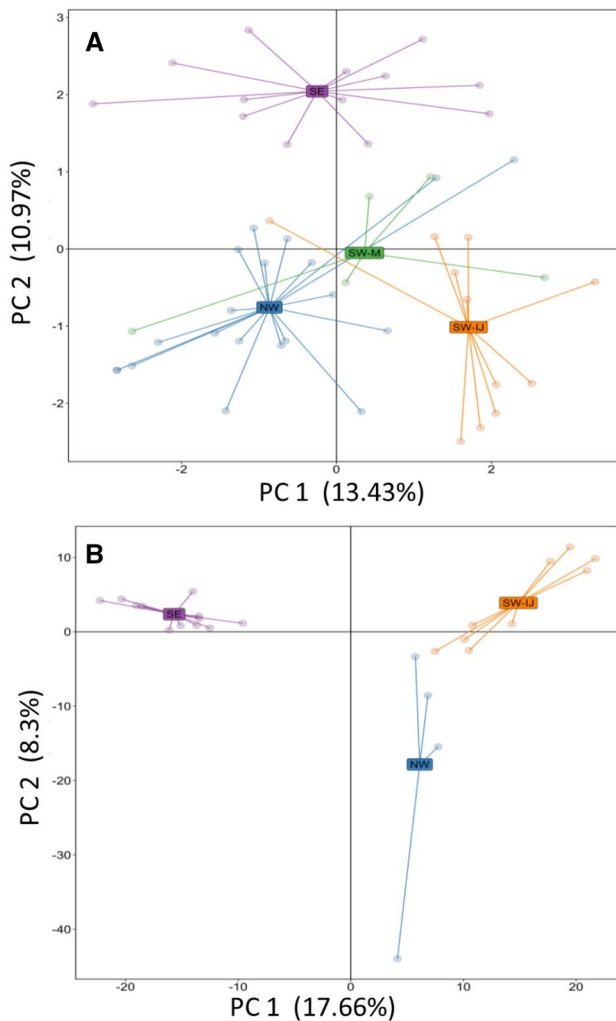
Fig. 3 Principal components analysis (PCA) on **A** 18 microsatellite loci and **B** 1703 SNP loci. *NW* northwest of Cuba, *SW-M* southwest of Cuba-main island, *SW-IJ* southwest of Cuba-Isla de la Juventud, and *SE* southeast of Cuba, mainly samples from Guantanamo Bay (see Fig. 1)

SE samples identified the greatest global pattern of differentiation between the genotypes from the furthest east on the north coast (locations 12–14) and those from the SW-IJ and location 11 on the south coast (Fig. 1) in a pattern that was very similar to what was seen in eigenvector 2 from the full sPCA (Fig. S3).

Genetic diversity and measures of fixation

Microsatellite genetic variation had only minor differences when comparing the 12 and 18 loci microsatellite datasets. For example, the average number of alleles was 3.2 and 3.4, and H_E was 0.443 and 0.460 for 12 and 18 loci, respectively (Table 1). The overall mean estimate of F_{IS} was greater for the 18 locus dataset, but the 95% CI overlapped zero in both

microsatellite datasets (Table 1). Positive (non-zero) F_{IS} was observed in the NW manatees for the full microsatellite and SNP datasets (Table 1).

The power to reject the null hypothesis of no heterogeneity among clusters (i.e., POWSIM) was high (> 0.8) with an N_e of 500 after 20 generations of drift, increasing to 0.99 with a N_e of 300 (Table 2). Under these two effective population sizes, global F_{ST} was expected to be 0.020 and 0.033, respectively. After only 10 generations of drift, power dropped to 0.824 at N_e of 300 ($F_{ST}=0.017$) and 0.480 at N_e of 500 ($F_{ST}=0.010$). Power decreased with increasing N_e under both short periods (10 and 20 generations) of drift (Table 2). Power also increased as the number of generations of drift was assumed. However, our empirical estimate of F_{ST} for the three genetic clusters was 0.099 (95% CI, 0.060–0.139), which exceeded the expected F_{ST} estimates under a pure drift model, indicating that these microsatellite loci have sufficient power to reject genetic homogeneity across Cuba at realistic values of N_e.

Pairwise estimates of fixation and genetic differentiation were significantly greater than zero for the full and reduced microsatellite datasets and SNPs, regardless of the number of genetic clustered assessed. Pairwise estimates of G''_{ST} between two clusters (SE and combine west, NW + SW-M + SW-IJ) was 0.180 ($P=0.001$) at 18 loci and 0.147 ($P=0.001$) for 12 loci. Pairwise D was 0.089 ($P=0.001$) and 0.070 ($P=0.001$) for 18 and 12 loci respectively. Considering three clusters, differentiating SW-M + SW-IJ from the NW, the 18 microsatellite pairwise G''_{ST} ranged from 0.117 (NW-SW) to 0.234 (SE-SW), with slightly lower but comparable values for the 12 loci dataset (Table 3). Pairwise D reflected similar patterns with higher differentiation between SE comparisons and slightly higher values for the 18 loci relative to the 12 loci dataset. Calculations of G''_{ST} based on the SNP data were 0.202 ($P=0.001$) for two clusters (SE vs. west) and ranged from 0.110 to 0.234 for three clusters (Table 3). SNP D was 0.07 ($P=0.001$) for two clusters and ranged from 0.037 to 0.081 for three clusters (Table 3). Overall, both microsatellite loci and SNPs reflected highly similar patterns of fixation and differentiation, with the highest values associated with pairwise comparisons with the SE.

Effective population size

Estimates for the total population N_e were 50.5 (95% CI 25.8–63.7) and 11.8 (95% CI 7.8–19.2) for microsatellites and SNPs, respectively. Despite the small sample sizes, the overall well-defined CIs for N_e could indicate that our estimates are plausible (Waples and Do 2010). In general, estimates of recent N_e were lower and had a smaller 95% CI for

Table 1 Diversity statistics over the 12 and 18 microsatellite loci and 1703 SNP loci examined for manatees from Cuba

Marker	Pop	N	A	AR	P	H_E	H_O	F_{IS}	F_{IS} CI 95%
12 loci	NW	20	2.9	2.2	1	0.418	0.368	0.145	− 0.099–0.163
	SW	16	2.5	2.3	0	0.461	0.425	0.008	− 0.133–0.166
	SE	13	2.1	2.1	0	0.431	0.508	− 0.177	− 0.390–0.002
	Combined	49	3.2	–	–	0.443	0.441	− 0.045	− 0.174–0.007
18 loci	NW	20	3.1	2.3	3	0.469	0.356	0.243	0.102–0.371
	SW	16	3.1	2.5	0	0.492	0.449	0.087	− 0.046–0.225
	SE	13	2.2	2.1	1	0.398	0.419	− 0.052	− 0.210–0.090
	Combined	49	3.4	–	–	0.463	0.410	0.115	− 0.020–0.212
	NW	4	1.719	1.616	8	0.324	0.243	0.248	0.219–0.277
SNPs	SW	11	1.874	1.594	109	0.299	0.290	0.028	− 0.011–0.045
(1703)	SE	12	1.797	1.635	107	0.282	0.280	0.009	− 0.008–0.027
	Combined	27	1.796	–	–	0.305	0.287	0.094	0.080–0.108

N sample size, *A* mean number of alleles per locus, *AR* allelic richness, *P* private alleles, H_E Expected Heterozygosity, H_O Observed Heterozygosity, F_{IS} Inbreeding coefficient and confidence interval, *P* Private alleles

Table 2 POWSIM simulation results for three periods (generations) of genetic drift at four different effective population sizes (N_e) 300 to 1000

N_e	Generations of drift		
	10	20	100
300	0.824 (0.017)	0.993 (0.033)	–
500	0.480 (0.010)	0.896 (0.020)	–
700	0.328 (0.007)	0.720 (0.014)	1.00 (0.049)
1000	0.241 (0.005)	0.492 (0.010)	1.00 (0.049)

Values represent the power to detect drift and the expected F_{ST} (in parentheses) following isolation at 10, 20 or 100 generations

Table 3 Pairwise G''_{ST} (upper matrix) and D_{Jost} (lower matrix bold values) values between three regional genetic clusters: NW, SW, and SE

	NW	SW	SE
Microsatellites			
NW	–	0.103/0.117	0.144/0.191
SW	**0.049/0.060**	–	0.166/0.234
SE	**0.067/0.093**	**0.081/0.120**	–
SNPs			
NW	–	0.110	0.195
SW	**0.037**	–	0.234
SE	**0.068**	**0.081**	–

Microsatellite estimates (top) are calculated for 12 loci (null alleles omitted) and 18 loci (12 loci/18 loci). All pairwise estimates were significantly greater than zero (all $P \le 0.003$)

Table 4 Estimates of effective population size (N_e) for the total dataset and for regional populations. Estimates for SW-M (both datasets) and NW (SNPs) were omitted due to small sample sizes but are included in the Total and West combined estimates

Marker	Independent comparisons	N_e	95% CI-Jackknife
Microsatellites 18 loci			
Total (N = 49)	880	50.5	25.8–63.7
SE (N = 13)	202	6.1	2.5–15.5
NW (N = 20)	647	30.0	9.0–∞
SW (N = 16)	621	37.7	14.9–∞
West combined (N = 36)	841	46.5	19.6–632.6
SNPs 1703 loci			
Total (N = 27)	1445744	11.8	7.8–19.2
SE (N = 12)	921193	10.5	4.2–38.7
SW (N = 11)	1106133	11.0	5.9–53.6
West combined (N = 15)	1267422	19.3	8.9–60.2

Included are the mean N_e estimates, bootstrap 95% credibility intervals, and jackknife estimates

mean estimate but overlapping 95% CIs for microsatellites than SNPs. The combined western estimates (NW + SW) had relatively larger mean values and had a higher upper 95% CI than estimates based on the total population, possibly reflecting the impact of higher admixture evident from the structure analysis.

Mitochondrial control region results

We successfully sequenced 410 bp of the control region for 59 manatee samples resulting in three unique haplotypes. Haplotype frequencies were highly skewed, with 57 sequences identified as haplotype A01 (Genbank

SNPs relative to microsatellite loci (Table 4). The exceptions were the estimates, for the SE region, which had a lower

Accession MZ355766-MZ35574, MZ355776-MZ355794, MZ355796-MZ355824), and a single sample as haplotype A03 (MZ355795) (García-Rodríguez et al. 1998). A third identified haplotype had two mutational differences from haplotype A01 and was designated as a new haplotype, A06 (MZ355775). All haplotypes represent the haplotype cluster I, widespread across the northwestern Caribbean (Belize), Gulf of Mexico and Florida, and Greater Antilles (García-Rodríguez et al. 1998).

Diversity values for the mtDNA sequence data were expectedly low given that there were only three haplotypes with highly skewed frequencies identified in 59 manatees. Overall, haplotype diversity (h) was 0.068 and nucleotide diversity (π) was 0.00025.

Discussion

Cuba is the largest island in the Caribbean, and it has large expanses of coastal habitats with over 600 rivers that could support a considerable number of manatees (Alcolado 2006; Baisre and Arboleya 2006). Historically, manatees likely inhabited much of the coastal habitat of Cuba, although no estimates of population abundance are available. The low levels of haplotype diversity, microsatellite heterozygosity, and allelic variation detected in this study are characteristic of small, isolated populations of manatees that have undergone extended bottlenecks and long-term human pressures (Hunter et al. 2010).

Microsatellite gene diversity in Cuba at 12 loci ($H_E = 0.44$) was comparable to other reported populations in the Caribbean and Gulf of Mexico (e.g., Florida H_O 0.45, H_E 0.48, Tucker et al. 2012; Belize H_O 0.46, H_E 0.45, Hunter et al. 2010; Puerto Rico H_O 0.45, H_E 0.45, Hunter et al. 2012; Mexico H_O 0.44–0.47, H_E 0.41–0.46, Nourisson et al. 2011). The average number of alleles for Cuba at 12 and 18 loci (3.2 and 3.4, respectively) was lower than those obtained for Florida and Puerto Rico (4.8 and 3.9, respectively) (Tucker et al. 2012; Hunter et al. 2012). These previous studies used the same microsatellite loci (or subsets thereof), making diversity comparisons more easily interpretable, especially for heterozygosity. Microsatellite allele size assignment can vary across studies and labs (Davison and Chiba 2003; Amos et al. 2007) and therefore require careful cross-calibration to confirm allele assignment across genotypes. However, comparing heterozygosity estimates should be less sensitive to allele size calling, assuming allelic dropout rates are not highly skewed among studies.

The level of microsatellite genetic diversity was lower relative to what has been reported as average for demographically challenged mammal populations ($H_E = 0.50 \pm 0.027$) (Garner et al. 2005). These were populations affected at least by one or more demographic risk factors such as

population declines, bottlenecks, reduction of population range, and isolation from conspecifics (Garner et al. 2005). Microsatellite diversity was also lower when compared with average estimates from disturbed mammal populations due to hunting ($H_E = 0.60 \pm 0.02$) and habitat fragmentation ($H_E = 0.59 \pm 0.023$) (DiBattista 2008). Most endangered species and populations have lower genetic diversity than non-endangered species with large populations (Frankham 2005). Low diversity values can be associated with an increased vulnerability to environmental, demographic, and stochastic variation and, consequently, an increased probability of extinction (Frankham et al. 2010); hence, it is wise to investigate whether they negatively impact the population. Substantial losses in genetic diversity can occur at the population level when unique genetic diversity is lost, even in large populations (Garner et al. 2005). Inbreeding was detected in the full microsatellite dataset in the NW population only, and may be a near-future problem in other subpopulations, as the abundance of animals appears to be low (Frankham et al. 2010). Close monitoring would be required to detect other signals of inbreeding or inbreeding depression (loss of fitness).

Moderate levels of genetic differentiation between the SE (predominantly Guantanamo Bay) and the rest of Cuba were consistent between the microsatellite and SNP datasets. Whether this genetic differentiation is due to allopatric isolation (presence of a physical barrier) or simply a product off geographic distance remains unknown. Our tests for global population structure due to geographic distance were significant and primarily separated the SE and the remainder of the geographic distribution of samples. However, the extent to which the SE represents a genetically distinct population will require continued geographic sampling to determine where a population separation may occur in the primarily understudied eastern half of the archipelago. In general, parts of the east coast of Cuba might have limited freshwater accessibility due to the construction of dams, but also low-quality seagrass beds resulting from the building of roads and bridges connecting the keys with the mainland (i.e., Archipelago Sabana Camaguey in the northeast) (Alcolado et al. 1998; Martinez-Daranas and Suarez 2018). The environmental deterioration associated with these anthropogenic structures reduces manatees' freshwater and forage habitat (Alcolado et al. 1998).

The dispersal and movement patterns of manatees in Cuba are not well-studied, so the identified population structure cannot be directly associated with limited home ranges or a lack of long-distance movement. Two unpublished studies have been conducted, one in NW-Isla de la Juventud and another in SE-Guantanamo Bay, involving three and eight satellited tracked manatees. The tracked manatees remained within 50 km of the tagging site along the coastline, suggesting relatively short distance movements compared to studied

manatee populations (Reid et al. 2015; Alvarez-Alemán et al. 2018a). However, West Indian manatees have also been documented to travel long distances along coastlines with suitable habitat (Fertl et al. 2005), and mark-recapture observations of manatees moving between Cuba and Florida over open water indicate that more extended distance movements (Alvarez-Alemán et al. 2010, 2018b) could allow for the maintenance of connectivity between the regions of Cuba (Hanski and Mononen 2011). Within the Caribbean manatees, subpopulation differentiation has also been identified in the Puerto Rico and Belize populations within relatively small distances—suggesting limited home ranges and dispersal in some populations or individual manatees (Hunter et al. 2010, 2012). Further studies of behavior and habitat use, complemented with habitat availability across the archipelago, will be required to understand the current genetic structure patterns.

The lack of clear geographic structure among manatees sampled from the NW and SW of Cuba suggests that movement around the island's western point is not restricted. The SW-mainland (i.e., sites 2, 8, 11, Fig. 1) appears to be an admixture of ancestry representing NW and Isla de la Juventud (SW-IJ). However, four of the five manatees sampled in the SW-mainland were obtained from an artificial translocated population in the Laguna del Tesoro (a lagoon in Zapata Swamp disconnected from the ocean), the recipient of translocated manatees in 1990 (Alvarez-Alemán et al. 2018a). At least two of these translocated individuals were moved to the Laguna del Tesoro from the north coast, and the rest came from the southwest.

Mitochondrial control region sequence variation was extremely low, far more than most other populations across the species range (García-Rodríguez et al. 1998; Lima et al. 2021). Only Florida (*T. m. latirostris*) manatees have less haplotype diversity, with only one haplotype (A01) reported across 28 individuals (Vianna et al. 2006). In Cuba, the same haplotype is predominant for many manatees sequenced from Puerto Rico, Dominican Republic, Mexico, and Belize (Vianna et al. 2006; Hunter et al. 2010, 2012). Only two manatees had alternative haplotypes, including A03 and A06, identified here and sequenced previously (Hernandez et al. 2013). All three observed haplotypes are closely related (in the 'A' clade) within "Cluster I" which is ubiquitous in the West Indies (and Florida) and frequently detected along the Caribbean coast of Central America (Vianna et al. 2006; Satizábal et al. 2012). The low haplotype diversity in Cuba manatees could reflect a historical bottleneck or a founder event. The absence of the haplotype B01 in Cuba, particularly in the SE samples, suggests low levels of female dispersal in the Greater Antilles. This haplotype was dominant in Puerto Rico (N = 74) and the Dominican Republic (N = 4) but also reported in Colombia (N = 1) (Vianna et al. 2006; Hunter et al. 2012; Satizábal et al. 2012).

The effective population size estimates for the full Cuba dataset were low, with $N_e \leq 53$ for the microsatellite and SNP datasets. These values were lower than the Florida population estimates ($N_e = 1260$) and the N_e values for the two Florida clusters (West and East) using the same LD method (Tucker et al. 2012). Alarmingly, low estimates were detected for the SE cluster, which perhaps reflects a higher level of isolation when compared with the other genetic clusters. The low values of N_e suggest that the Cuba population could be recovering from a recent bottleneck, historically small population size, or limited gene flow (Lansink et al. 2020). Populations of mammals that have been assumed to have gone through a bottleneck often display low values of N_e, as is the case of the Finnish wolf, Saimaa ringed seal, Canada lynx and Iberian lynx population, all of which had N_e values below 50 (Aspi et al. 2006; Casas-Marce et al. 2013; Valtonen et al. 2014; Prentice et al. 2017).

Several authors have proposed $N_e = 50$ and $N_e = 100$ to be enough for preventing inbreeding depression and retaining the evolutionary potential for fitness in perpetuity in naturally outbreeding diploid species in the short term, respectively (Soulé 1980; Franklin 1980; Jamieson and Allendorf 2012). However, Frankham et al. (2014) criticized these numbers and suggested that these should be raised to at least 100 and 1000 and reevaluated as more data become available. The Convention on Biological Diversity (CBD) post-2020 global diversity framework recommends using the proportion of populations with N_e above 500 as a criterion for maintaining genetic diversity and adaptive capacity, as suggested by Hoban et al. (2020). Applying these principles, the Cuban manatee population appears to be genetically depauperate and requires a conservation strategy to increase population size to improve the genetic variability.

Conclusion

Our results suggest that regular gene flow occurs across the western portion of the archipelago, but there is possibly less connectivity between the SE Guantanamo Bay region and the western portion of the archipelago. The genetic distinction of the SE population supports recognition of it as a demographically independent population for natural resource management (Funk et al. 2012), regardless of whether it is due to historical isolation or isolation by distance. Improvement of genetic diversity and connectivity in the short-term could benefit from a network of national marine protected areas in critical habitats and corridors to secure connections among more isolated regions and reduce manatee mortality. Improving habitat will be necessary, particularly along the eastern portions of the archipelago.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Alcolado PM (2006) Diversidad ecológica. Diversidad, utilidad y estado de conservación de los biotopos marinos. In: Claro R (ed) La Biodiversidad Marina de Cuba. Instituto de Oceanología Ministerio de Ciencia, Tecnología y Medio Ambiente, La Habana, p 39

Alcolado P, Menéndez G, García-Parrado P, Zuniga D, Martinez-Darana B et al (1998) Cayo Coco, Sabana-Camagüey Archipelago, Cuba. In: Kjerfve (ed) CARICOMP—Caribbean Coral Reef, Seagrass and Mangrove Sites. UNESCO, Paris, p 347

Alvarez-Alemán A, Powell JA, Beck C (2010) First report of a Florida manatee (*Trichechus manatus latirostris*) in Cuba. Aquat Mamm 36:148–153

Alvarez-Alemán A, García E, Forneiro Martin-Viana Y, Hernández-González Z, Escalona-Domenech R et al (2018a) Status and conservation of manatees in Cuba: historical observations and recent insights. Bull Mar Sci. https://doi.org/10.5343/bms.2016.1132

Alvarez-Alemán A, Austin JD, Jacoby CA, Frazer TK (2018b) Cuban connection: regional role for Florida's manatees. Front Mar Sci. https://doi.org/10.3389/fmars.2018.00294

Alvarez-Alemán A, Garcia-Alfonso E, Powell JA, Jacoby CA, Austin JD, Frazer TK (2021) Causes of mortality for endangered Antillean manatees in Cuba. Front Mater Sci 8:646021

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10–14

Aspi J, Roininen E, Ruokonen M, Kojola I, Vilà C (2006) Genetic diversity, population structure, effective population size, and demographic history of the Finnish wolf population. Mol Ecol 15:1561–1576. https://doi.org/10.1111/j.1365-294X.2006.02877.x

Baisre JA, Arboleya Z (2006) Going against the flow: effects of river damming in Cuban fisheries. Fish Res 81:283–291. https://doi.org/10.1016/j.fishres.2006.04.019

Casas-Marce M, Soriano L, López-Bao JV, Godoy JA (2013) Genetics at the verge of extinction: insights from the Iberian lynx. Mol Ecol 22:5503–5515. https://doi.org/10.1111/mec.12498

Chakraborty R, Zhong Y, Jin L, Budowle B (1994) Nondetectability of restriction fragments and independence of DNA fragment sizes within and between loci in RFLP typing of DNA. Am J Hum Genet 55:391–401

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. https://doi.org/10.1093/molbev/msl191

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205. https://doi.org/10.1038/nrg2526

Cuni LA (1918) Contribución al estudio de los mamíferos acuáticos observados en las costas de Cuba. Memorias de la Sociedad Cubana de Historia Natural. Felipe Poey 3:83–126

Davis MC (2014) Relatedness, inbreeding and kinship of the endangered Florida manatee (*Trichechus manatus latirostris*). Doctoral Thesis. University of Florida

Davison A, Chiba S (2003) Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. Mol Ecol Notes 3:321–323

Deutsch CJ (2008) *Trichechus manatus* ssp. *latirostris*. The IUCN Red List of Threatened Species 2008: e.T22106A9359881. https://doi.org/10.2305/IUCN.UK.2008.RLTS.T22106A9359881.en. Accessed 17 May 2020

DiBattista JD (2008) Patterns of genetic variation in anthropogenically impacted populations. Conserv Genet 9:141–156. https://doi.org/10.1007/s10592-007-9317-z

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour 14(1):209–214

Dray S, Dufour A (2007) The ade4: package implementing the duality diagram for ecologists. J Stat Softw 22(4):1–20. https://doi.org/10.18637/jss.v022.i04

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

Federal Register (2017) Endangered and threatened wildlife and plants; reclassification of the West Indian manatee from

endangered to threatened. Federal register: the daily journal of the United States, Fed Reg 64, Washington, DC

Fertl D, Schiro AJ, Regan GT, Beck CA, Adimey N, Price-May L et al (2005) Manatee occurrence in the northern Gulf of Mexico, West of Florida. Gulf Caribb Res 17:69–94

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis helleri*. BMC Genomics 18:69. https://doi.org/10.1186/s12864-016-3459-7

Frankham R (2005) Genetics and extinction. Biol Conserv. https://doi.org/10.1016/j.biocon.2005.05.002

Frankham R, Ballou JD, Briscoe D (2010) Introduction to conservation genetics. Cambridge University Press, New York

Frankham R, Corey JA, Bradshaw C, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biol Conserv 170:56–63

Franklin IR (1980) Evolutionary change in small populations. In: Soulé W (ed) Conservation biology: an evolutionary-ecological perspective. Sinauer Associates, Sunderland, pp 135–150

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27(9):489–496. https://doi.org/10.1016/j.tree.2012.05.012

Funk WC, Forester BR, Converse SJ, Darst C, Morey S (2019) Improving conservation policy with genomics: a guide to integrating adaptive potential into US Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet 20:115–134

García-Rodríguez AI, Bowen BW, Domning DP, Mignucci-Giannoni AA, Marmontel M et al (1998) Phylogeography of the West Indian manatee (*Trichechus manatus*): how many populations and how many taxa? Mol Ecol 7:1137–1149

García-Rodríguez AI, Bowen BW, Domning D (2000) Isolation and characterization of microsatellites DNA markers in the Florida manatee (*Trichechus manatus latirostris*) and their application in selected sirenian species. Mol Ecol 9:2161–2163

Garner A, Rachlow JL, Hicks JF (2005) Patterns of genetic diversity and its loss in mammalian populations. Conserv Biol 19:1215–1221

Goudet J (2005) HIERFSTAT, a package for R to computer and test hierarchical F-statistics. Mol Ecol Notes 5:184–186

Grewe PM, Feutry P, Hill PL, Gunasekera RM, Schaefer KM et al (2015) Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. Nature. https://doi.org/10.1038/srep16916

Gruber B, Unmack PJ, Berry OF, Georges A (2017) DARTR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18:691–699

Gruber B, Adamack A (2013) Introduction to PopGenReport using PopGenReport Ver. 1.6.6. User Manual

Hanski I, Mononen T (2011) Eco-evolutionary dynamics of dispersal in spatially heterogeneous environments. Ecol Lett 14:1025–1034. https://doi.org/10.1111/j.1461-0248.2011.01671.x

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Hernández D, Álvarez-Alemán A, Bonde RK, Powell JA, García-Machado E (2013) Diversidad haplotípica en el manatí *Trichechus manatus* en Cuba: Resultados preliminares. Rev De Investig Mar 33:58–61

Hoban S, Bruford M, Jackson JD, Lopes-Fernandes M, Heuertz M (2020) Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Cons. https://doi.org/10.1016/j.biocon.2020.108654

Hufbauer RA, Szucs M, Kasyon E, Youngberg C, Koontz MJ, Richards C, Melbourne BA (2015) Three types of rescues can avert extinction in a changing environment. Proc Natl Acad Sci USA 112(33):10557–10562. https://doi.org/10.1073/pnas.1504732112

Hunter ME, Auil-Gomez N, Tucker KP, Bonde RK, Powell J et al (2010) Low genetic variation and evidence of limited dispersal in the regionally important Belize manatee. Anim 13:592–602

Hunter ME, Mignucci-Giannoni AA, King TL, Tucker KP, Bonde RK (2012) Comprehensive genetic investigation recognizes evolutionary divergence in the Florida (*Trichechus manatus latirostris*) and Puerto Rico (*T.m. manatus*) manatee populations and subtle substructure in Puerto Rico. Conserv Genet 13:1623–1635

Jackson JBC, Kirby MX, Berger WH, Bjorndal KA, Botsford LW et al (2001) Historical overfishing and the recent collapse of coastal ecosystems. Science. https://doi.org/10.1126/science.1059199

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jamieson IG, Allendorf F (2012) How does the 50/500 rule apply to MVPs? Trends Ecol Evol. https://doi.org/10.1016/j.tree.2012.07.001

Jiménez-Vázquez O (2015) Manatíes y delfines en sitios arqueológicos precolombinos de Cuba. Novit Carib 8:30–39

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. Heredity 102:330–334. https://doi.org/10.1038/hdy.2008.130

Jombart T (2015) A tutorial for the spatial Analysis of Principal Components (sPCA) using adegenet 2.0.0. Imperial College London, MRC Centre for Outbreak Analysis and Modelling

Jost L (2008) GST and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

Lansink GM, Esparza-Salas R, Joensuu M, Koskela A, Bujnáková D (2020) Population genetics of the wolverine in Finland: the road to recovery? Cons Gen 21:481–499. https://doi.org/10.1007/s10592-020-01264-8

Lefebvre LW, Marmontel M, Reid JP, Rathbun GB, Domning DP (2001) Status and biogeography of the West Indian manatee. In: Woods CA, Sergile FE (eds) Biogeography of the West Indies, 2nd edn. CRC Press, Boca Raton, pp 425–474

Li YL, Liu JX (2018) Structure selector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour 18:176–177

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

Lima CS, Magalhães RF, Santos FR (2021) Conservation issues using discordant taxonomic and evolutionary units: a case study of the American manatee (*Trichechus manatus*, Sirenia). Wildlife Res. https://doi.org/10.1071/WR20197

Martinez-Daranas B, Cano Mallo M, Clero Alonso L (2009) Los pastos marinos de Cuba: estado de conservación y manejo. *Serie Oceanologica* 5: 24–44. ISSN 2072-800x.

Martínez-Daranas B, Suárez AM (2018) An overview of Cuban seasgrasses. Bull Mar Sci 94:269–282

Meimans PG, Tienderen AH (2004) Genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol 4(4):792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. Mol Ecol 24:3223–3231. https://doi.org/10.1111/mec.13243

Moritz C (1994) Defining "Evolutionarily significant units" for conservation. Tree 9:373–375

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nourisson C, Morales-Vela B, Padilla-Saldivar J, Tucker KP, Clark AN (2011) Evidence of two genetic clusters of manatees with low genetic diversity in Mexico and implication for their conservation. Genetica 139:833–842

Paradis E (2010) Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420

Pause KC, Nourissom C, Clark A, Kellog ME, Bonde RK et al (2007) Polimorphic microsatellite DNA markers for the Florida manatee (*Trichechus manatus latirostris*). Mol Ecol 7:1073–1076

Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. PeerJ 4:e1813

Prentice MB, Bowman J, Khidas K, Koen EL, Row JR et al (2017) Selection and drift influence genetic differentiation of insular Canada lynx (*Lynx canadensis*) on Newfoundland and Cape Breton Island. Ecol Evol 7:3281–3294. https://doi.org/10.1002/ece3.2945

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

Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. Mol Ecol Resour 16:608–627

Quintana-Rizzo E, Reynold III J (2010) Regional Management Plan for the West Indian Manatee (*Trichechus manatus*). CEP Technical Report No. 48. UNEP Caribbean Environment Programme, Kingston, Jamaica.

R Development Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ralls K, Ballou JD, Dudash MR, Eldridge MDB, Fenster CB et al (2018) Call for a paradigm shift in the genetic management of fragmented populations. Conserv Lett 11:e12412. https://doi.org/10.1111/conl.12412

Reid J, Bonde RK, Butler SM, Slone DH (2015) First manatee capture, health assessment and radio tracked movement in Guantanamo Bay, Cuba. Presented at the 21st Biennial Conference on the Biology of Marine Mammals, San Francisco

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol 4(1):137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Mol Ecol 6(3):600–602. https://doi.org/10.1111/j.1471-8286.2006.01378.x

Sambrock J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor

Satizábal P, Mignucci-Giannoni A, Duchene S, Caicedo-Herrera D, Perea-Sicchar CM (2012) Phylogeography and sex-biased dispersal across riverine manatee populations (*Trichechus inunguis* and *Trichechus manatus*) in South America. PLoS ONE. https://doi.org/10.1371/journal.pone.0052468

Schill SR, Raber GT, Roberts JJ, Treml EA, Brenner J, Halpin PN (2015) No reef is an island: integrating coral reef connectivity data in the design of regional scale marine protected area networks. PLoS ONE 10:e0144199

Self-Sullivan C, Mignucci-Giannoni A (2008). *Trichechus manatus* ssp. manatus. The IUCN Red List of Threatened Species 2008: e.T22105A9359161.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236(4803):787–792. https://doi.org/10.1126/science.3576198

Sorice MG, Shaffer GS, Ditton RB (2006) Managing endangered species within the use-preservation paradox: the Florida manatee (*Trichechus manatus latirostris*) as a tourism attraction. Environ Manage 37(1):69–83. https://doi.org/10.1007/s00267-004-0125-7

Soulé ME (1980) Thresholds for survival: maintaining fitness and evolutionary potential. In: Soulé W (ed) Conservation biology: an evolutionary-ecological perspective. Sinauer, Sunderland, pp 151–169

Tucker KP, Hunter ME, Bonde RK, Austin JD, Clark AM et al (2012) Low genetic diversity and minimal population substructure in the endangered Florida manatee: implications for conservation. J Mammal. https://doi.org/10.1644/12-MAMM-A-048.1

Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol 17:3808–3817

Valtonen M, Palo JU, Aspi J, Ruokonen M, Kunnasranta M, Nyman T (2014) Causes and consequences of fne-scale population structure in a critically endangered freshwater seal. BMC Ecol 14:22. https://doi.org/10.1186/1472-6785-14-22

Vianna JA, Bonde RK, Caballero S, Giraldo JP, Lima RP et al (2006) Phylogeography, phylogeny and hybridization in trichechid sirenians: implications for manatee conservation. Mol Ecol. https://doi.org/10.1111/j.1365-294X.2005.02771

Waples RS, Do C (2008) LdNe: a program for estimating effective population size from data on linkage disequilibrium. Mol Ecol Resour 8:753–756

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol Appl 3(3):244–261. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ et al (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evol Appl 4:709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. Amer Naturalist. https://doi.org/10.1086/682949

Wright S (1943) Isolation by distance. Genetics 28:114–138