STARS

2023

Variation in Genetic Structure and Dispersal of Juvenile Green Turtles

Gustavo Stahelin
University of Central Florida

VARIATION IN GENETIC STRUCTURE AND
DISPERSAL OF JUVENILE GREEN TURTLES

by

GUSTAVO DAVID STAHELIN
B.S. Universidade Federal de Santa Catarina, 2006

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Biology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2023

Major Professor: Katherine L. Mansfield

ABSTRACT

Sea turtles are long-lived, globally distributed animals with a complex life-history. Individuals from different populations often share the same foraging areas (mixed stock aggregations). Understanding patterns of dispersal and connectivity between reproductive populations and mixed stock aggregations is fundamental for the development of effective conservation plans. Recently, green sea turtle (*Chelonia mydas*) populations in several reproductive areas have increased, providing an opportunity to evaluate how demographic changes in reproductive areas impact dispersal to, and the composition of, mixed stock aggregations. In this dissertation, I evaluated how dispersal from reproductive populations in the Greater Caribbean to mixed stock aggregations may have changed over time (Chapter 2). I analyzed mitochondrial DNA haplotypes from samples collected from nesting females captured at Melbourne Beach, Florida, USA, and in-water juveniles from two mixed stock aggregations in central Florida (Indian River Lagoon and Trident Basin) over two time periods. Over a 15-year period there were small variations in the composition of the mixed stocks, without a clear relationship to the recent growth in reproductive populations. I developed a modification to the established "many-to-many" mixed stock model to use the distance between source populations and mixed stock aggregations to weight model estimates. In Chapter 3 I created a simulation to understand how sample size and the level of similarity in relation to haplotype frequency between source populations can impact mixed stock model estimates. I determined that a minimum of 150 samples from each mixed stock aggregation is required to accurately estimate contributions from source populations to mixed stock aggregations for most cases using data currently available in the literature. Improving the resolution of the genetic marker used (i.e., increasing the distinction of haplotype frequencies between source

populations) can produce similar results using a smaller number of samples. Finally, in Chapter 4 I evaluated genetic structure of green turtle populations in the Greater Caribbean using a next-generation sequencing approach. I used the same sampling scheme as Chapter 2, with samples from a nesting beach (Melbourne Beach, FL) and two mixed stock aggregations (Indian River Lagoon and Trident Basin, Florida). I identified 4 distinct populations within the samples, and similar to the mtDNA assessment in Chapter 2, the genomic approach also showed small variations in the composition of mixed stock aggregations over a 15-year period. I used a coalescent model to evaluate how these populations diverged from one another, and found strong support for current gene flow among all 4 populations. Results from my analyses reiterate the complexity of sea turtle's dispersal dynamics, and the level of connectivity among populations in the Greater Caribbean. Future studies using mixed stock analysis should consider sample size with more than 150 samples per mixed stock aggregation and the use of more refined genetic markers. Also, genomic assessments of across multiple reproductive aggregations are required for a deeper understanding of other aspects of their ecology.

endless summer months: Doc Ehrhart, Erin Seney, Dean Bagley, Bill Redfoot, JLO, Franko, Janet Hochella, Jim Stevenson, Jen Elliot, and current and past students of the UCF MTRG.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Dispersal is a fundamental part of many species' life cycle. Some species of birds, for instance, have an extended parental care phase, with young individuals dispersing into new areas only after reaching a certain life or developmental-stage (Russell et al., 2004), while in other species, as sea turtles, juveniles may start dispersal after hatching from the eggs with no direct parental care involved (Santos et al., 2016). Dispersal into new areas and the distribution of individuals across a land- or seascape can be influenced by multiple features, and are dependent upon environmental and demographic parameters (Doligez et al., 2003; Engelbrecht et al., 2007; Leach et al., 2016). For migratory species, a combination of environmental factors can contribute to orientation; birds that can adjust their flight path based on wind direction and waves as landscape features (Richardson, 1990) or using magnetoreceptors on their retina (Mouritsen & Hore, 2012). Selection of a new habitat for dispersing individuals is critical for individuals survival, as it can impact food availability and have impacts on reproductive success and fitness (Wiklund & Kaitala, 1995) or predation avoidance (Lundvall et al., 1999). Understanding patterns of species' dispersal is especially important for the conservation and recovery of endangered species.

Post-dispersal habitats for migratory and widely distributed species are often composed of individuals from multiple populations, known as mixed stock aggregations (e.g., seabirds, (Young, 2010); whales, (Brüniche-Olsen et al., 2018; Carroll et al., 2020); sea turtles, (Bowen et al., 1996; Phillips et al., 2022; Stahelin et al., 2022)). Connections between reproductive areas and habitats where individuals disperse to can be evaluated with the use of genetic markers (Pella & Masuda, 2001). Haplotype frequencies vary between populations, and one can calculate probability of a group of individuals sampled at a given location belonging from a specific source population (Pella

& Masuda, 2001). Mixed stock analysis (MSA) is a method used to determine where individuals found in mixed stock aggregations originated from (Anderson et al., 2008; Bolker et al., 2007; Debevec, 2000; Neaves et al., 2005; Pella & Masuda, 2001). Mixed stock models can incorporate the size of each source population as a covariate to more accurately account for sampling error (Okuyama & Bolker, 2005). However, there is a data gap in the literature regarding how variation in population sizes can change the composition of these mixed stock aggregations. This is important because genetic diversity within and among populations contributes to the persistence of wild populations (Frankham, 2005). Gene flow between populations helps reduce inbreeding, which can have deleterious effects on populations and increase the risk of extinction (Frankel & Soulé, 1981). Population genomics is a powerful tool to evaluate species' genetic parameters (Funk et al., 2019; Volk et al., 2021), and samples collected at mixed stock aggregations could be used as a way to evaluate genetic structure and gene flow between reproductive populations.

Appropriate sample size is a fundamental requirement for reliable statistical analysis (Ioannidis, 2005). Variability within datasets and sampling biases can easily skew results and hamper inferences regarding the study's goal (Baker et al., 2009; Jenkins & Quintana-Ascencio, 2020). Although, in some cases, adequate sample sizes might be hard to obtain. Locations with limited or difficult access (see Jenkins & Quintana-Ascencio, 2020), attempting to sample species with low abundances (e.g., endangered species; (Storfer, 1996)), or even lack of funding (Button et al., 2013) can reduce researchers' capacity to collect enough samples. Sample size becomes especially important when trying to differentiate between greatly similar datasets. This is the case for several sea turtle mixed stock aggregations, in which haplotype frequencies between areas is very similar;

mixed stock models may have a reduced capacity to estimate the source populations of sampled individuals (Shamblin et al., 2017).

Sea turtles are globally distributed in tropical, subtropical, and temperate areas (Bowen & Karl, 2007). Extant species of sea turtles started differentiating nearly ~100 million years, with varying hypotheses on how these species colonized the different oceans (Jensen et al., 2019; Phillips et al., 2022). Despite surviving multiple evolutionary pressures over time, all current species are facing some level of extinction risk, mostly caused by human pressures (Fuentes et al., 2016; Jensen et al., 2018; Lewison et al., 2014; Miller et al., 2019; Wallace et al., 2013). Genomic approaches can deepen our understanding of how these species are handling or responding to reduced population sizes or the impacts of recent population growth (Chaloupka et al., 2008; Marcovaldi & Chaloupka, 2007; Mazaris et al., 2017; Seminoff et al., 2015), and provide a clearer picture of effective measurements or conservation strategies for the future.

The green sea turtle (*Chelonia mydas*) is an ideal species to study some of the concerns highlighted above. Green turtles use mixed stock aggregations as foraging and developmental habitats during their entire life cycle (Bjorndal & Bolten, 2008; Phillips et al., 2022; Shamblin et al., 2018; Stahelin et al., 2022; van der Zee et al., 2019). Reproductive populations in many areas are experiencing an increased number of nests, which is an indication of population growth (Chaloupka et al., 2008; Seminoff et al., 2015). Temporal evaluation of genetic parameters at mixed stock aggregations can provide insights into the status of their genetic diversity and genetic structure between populations.

My dissertation aims to address questions regarding the impacts of recent population growth on the composition of green sea turtle mixed stock aggregations, and evaluate the sensitivity and efficacy of commonly used techniques used in mixed stock models for sea turtle studies. For that, in Chapter 2 I conduct a mixed stock analysis using samples collected at an important green turtle rookery in Melbourne Beach, Florida, USA, and two mixed stock aggregations in the east coast of central Florida, USA (Indian River Lagoon, and Trident Submarine Basin) during two sampling periods. I compare contribution estimates from source populations to mixed stock aggregations to recent variations in number of nests laid at reproductive sites. For this goal, I developed a modification to the standard many-to-many mixed stock model by incorporating distance between source populations (rookeries) and mixed stock aggregations to the model. In Chapter 3 I conduct a simulation to evaluate the effect of using finer resolution markers to differentiate nesting populations on mixed stock models, and the impact of sample size on mixed stock models accuracy. In Chapter 4 I use a genomic approach to compare the genetic structure between genetic clusters identified in two mixed stock aggregations in Florida (Indian River Lagoon and Trident Basin) and the reproductive population from Melbourne Beach. I also evaluate how genetic structure for these sites varied in response to population growth and use a coalescent model to understand patterns of gene flow, divergence of current populations in the Greater Caribbean, and effective population sizes. Finally, I make recommendations of best practices for future studies regarding methodological adjustments and statistical approaches that can improve both our ecological understanding of sea turtles biology and future conservation actions derived from such analyses.

CHAPTER 2: INCORPORATING DISTANCE METRICS AND TEMPORAL TRENDS TO REFINE MIXED STOCK ANALYSIS

Publication note

Abstract

The distribution of marine organisms is shaped by geographic distance and oceanographic features like currents. Among migratory species, individuals from multiple populations may share feeding habitats seasonally or across life stages. Here, we introduce a modification for many-to-many mixed stock models to include distance between breeding and foraging sites as an ecological covariate and evaluate how the composition of green turtle, *Chelonia mydas*, juvenile mixed stock aggregations changed in response to population growth over time. Our modified many-to-many model is more informative and generally tightens credible intervals over models that do not incorporate distance. Moreover, we identified a decrease in genetic diversity in a Florida nesting site and two juvenile aggregations. Mixed stock aggregations in central Florida have changed from multiple sources to fewer dominant source populations over the past ~20 years. We demonstrate that shifts in contributions from source populations to mixed stock aggregations are likely associated with nesting population growth. Furthermore, our results highlight the importance of

long-term monitoring and the need for periodical reassessment of reproductive populations and juvenile aggregations. Understanding how mixed stock aggregations change over time and how different life stages are connected is fundamental for the development of successful conservation plans for imperiled species.

Introduction

Dispersal shapes species' distributions and genetic structure; organisms dispersing into new areas may select suitable habitats based on factors such as availability (MacPherson, 1998), temperature (Freitas et al., 2016), resources, and competition (Davoren et al., 2003; Michelot et al., 2017). Among mobile organisms, some have defined home ranges and low dispersal, such as the maned sloth (*Bradypus torquatus*)(Chiarello et al., 2004), while others, such as the saltwater crocodile (*Crocodylus porosus*), are distributed more broadly and travel hundreds of kilometers for food and reproduction (Fukuda et al., 2022). Particularly among migratory organisms, individuals originating from multiple populations may share the same habitat and resources (e.g., fishes (O'Leary et al., 2014), and whales (Brüniche-Olsen et al., 2018; Carroll et al., 2020)). These shared habitats can be occupied by individuals from different populations during a specific time of the year, as in gray whales (*Eschrichtius robustus*) that seasonally share foraging habitats in the Pacific Ocean (Brüniche-Olsen et al., 2018), or during multiple years as in juvenile sea turtle foraging aggregations (Bowen et al., 1996).

In organisms for which different populations share a habitat, mixed stock analysis (MSA) is a useful technique to understand genetic connectivity between source populations and mixed stock aggregations (Paxton et al., 2013). There are different approaches for mixed stock calculations,

most commonly using maximum likelihood (Anderson et al., 2008; Debevec, 2000) or Bayesian inference (Bolker et al., 2007; Neaves et al., 2005; Pella & Masuda, 2001). In addition, there are models designed to use haploid or single-parent inherited markers (e.g., mitochondrial DNA [mtDNA]) (Bolker et al., 2003; Bolker et al., 2007; Okuyama & Bolker, 2005) and nuclear or diploid markers (e.g., microsatellites, allozymes, minisatellites, among others) (Anderson et al., 2008; Debevec, 2000; Neaves et al., 2005). Regardless of the approach, such methods usually compare genetic marker frequencies from mixed stock aggregations to known reproductive populations to estimate contributions to mixed stocks (Pella & Masuda, 2001). For mtDNA and similar markers, Bolker et al. (2007) introduced a model that considers the contributions from all available source populations to multiple mixed stocks simultaneously (e.g., many-to-many models). Many-to-many models allow robust and more realistic estimates of source populations than previous models that only accepted one mixed stock aggregation as a possible destination (e.g., many-to-one models (Bolker et al., 2007; Pella & Masuda, 2001)). Despite the important methodological advancements introduced by Bolker et al. (2007), the many-to-many model introduced in the R package 'mixstock' was not designed to consider site-specific variables such as the distance between source and destination sites. Nishizawa et al. (2013) added a matrix of distances between source populations and mixed stocks as priors to the model as a means to account for distance in a many-to-many framework; however, it is unclear how their distance matrix is incorporated into the model.

Understanding how demographic parameters and dispersal patterns impact mixed stock composition is fundamental for implementing conservation plans for endangered and threatened species, especially if such variables change over time. Under scenarios where individuals are free

and capable of moving in any direction, we would expect large source populations to contribute more individuals to mixed stocks than small populations. However, most organisms disperse between habitats according to biogeographical and resource constraints (Davoren et al., 2003; MacPherson, 1998; Michelot et al., 2017). Also, temporal changes in source population sizes and stochastic events altering dispersal patterns can hamper our ability to characterize source populations for mixed stocks (Pella & Masuda, 2001). A mixed stock model implemented by Okuyama and Bolker (2005) weights model contribution estimates based on the size of each source population. Mixed stock models often require robust sampling from all sites under evaluation to obtain reliable estimates (Bolker et al., 2003), leading to combined datasets from previously published studies, regardless of the time period when samples were collected (Bowen et al., 1996; Carroll et al., 2020; Naro-Maciel, Gaughran, et al., 2014; O'Leary et al., 2014; Proietti et al., 2012; Shamblin et al., 2017; van der Zee et al., 2019). A concern with such an approach is that source populations are not necessarily constant over time (Chaloupka et al., 2008; Seminoff et al., 2015), just as haplotype frequencies in mixed stock aggregations might fluctuate (Bjorndal & Bolten, 2008). Furthermore, the distance traveled by individuals from source populations to foraging aggregations is, in general, an important variable impacting dispersal (Roland et al., 2000; Vanschoenwinkel et al., 2007), and is often not considered in mixed stock assessments (Bowen et al., 1996; Brüniche-Olsen et al., 2018; Carroll et al., 2020; O'Leary et al., 2014; Paxton et al., 2013; Shamblin et al., 2012) (but see (Naro-Maciel et al., 2014; Nishizawa et al., 2013; Proietti et al., 2012)). Therefore, there is a need to evaluate the impact of temporal changes in reproductive population demographics on mixed stock aggregations and to develop models that can better account for the distance between breeding and mixed stock aggregations into many-to-many MSAs.

The green turtle (*Chelonia mydas*) is an ideal organism to evaluate how demographic variations and distance between source populations (rookeries) impact mixed stock aggregations. First, green turtle foraging aggregations are typically composed of individuals from multiple populations (Bjorndal & Bolten, 2008; Bolker et al., 2007; Bowen et al., 1996; Shamblin et al., 2012, 2017; van der Zee et al., 2019). Second, in the past two decades, the number of nests in several green turtle rookeries in the Greater Caribbean and the western North Atlantic have increased at different rates (Chaloupka et al., 2008; Seminoff et al., 2015). Recently, van der Zee et al. (van der Zee et al., 2019) suggested that changes in contributions observed in a juvenile mixed stock in Bonaire could be associated with a variation in the size of the source nesting populations. Third, green turtles leave nesting beaches as hatchlings and swim away from the coast to offshore habitats where they reside for a number of years (Witherington et al., 2012). Even though oceanic-stage green turtles are not complete passive drifters and may actively swim and orient (Putman & Mansfield, 2015), there is substantial evidence suggesting marine turtle juvenile dispersal is also influenced by oceanographic currents, especially during the first few years of their life cycle (Mansfield et al., 2021; Putman et al., 2020; Putman & Naro-Maciel, 2013) (but see (Naro-Maciel et al., 2017)). Therefore, we can approximate the distance traveled by individuals between rookeries and mixed stock areas by following the main marine currents connecting the different areas. Lastly, the east coast of central Florida, USA, hosts one of the largest nesting aggregations for green turtles in the western North Atlantic (Chaloupka et al., 2008) and several mixed stock aggregations (Ehrhart et al., 2007; Redfoot & Ehrhart, 2013).

Here, we evaluate the impact of temporal changes in rookery sizes and in green turtle mixed stock aggregations in the Greater Caribbean and western North Atlantic while accounting for distance

9

traveled between rookeries and mixed stock aggregations. To achieve these goals, we (*i*) modify the many-to-many mixed stock model to weight estimates based on the distance between rookeries and mixed stocks, and use the modified model to (*ii*) evaluate how variations in rookery sizes impacted MSA estimates over a two-decade period, and (*iii*) assess how MSA estimates change in response to variation in mixed stock haplotype frequencies and rookery size over the same period.

Methods

Study site and data collection

Adding to the available data on haplotype frequencies for rookeries and mixed stocks (Supplementary tables S1-S3), we collected data from nesting female green turtles in the Brevard County portion of the Archie Carr National Wildlife Refuge, in Melbourne Beach, Florida, USA (28.04° N, 80.55° W to 27.87° N, 80.45° W – hereafter referred to as "MB") (Ehrhart et al., 2014). We sampled juveniles at two mixed stock foraging sites: the Indian River Lagoon about two kilometers south of the Sebastian Inlet (27.82° N, 80.43° W – "IRL"), and at Trident Basin at Port Canaveral (28.42° N, 80.59° W – "TRID"), both on the east coast of central Florida, USA (Ehrhart et al., 2007; Redfoot & Ehrhart, 2013). All specimens used in this study were collected in accordance with animal care and use protocols approved by the Institutional Animal Care and Usage Committee at the University of Central Florida (IACUC 2020-04, 2020-18, 2020-138, and their predecessors). Skin and blood samples collection were conducted under permits MTP-231, NMFS 19508, and their predecessors.

We defined two sampling periods, "old" and "new", within the rookery and mixed stock samples: for the rookery, samples collected before 2000 = MB_{old}, and samples collected in 2016 to 2018 =

MB$_{new}$. At the mixed stock sites, samples from 2003 to 2005 = IRL$_{old}$ and TRID$_{old}$, while samples from 2016 to 2018 = IRL$_{new}$ and TRID$_{new}$. Nesting female samples were assigned to a sampling period based on their first encounter, while juvenile mixed stock samples were assigned to a sampling period if any of the capture dates occurred during the years examined in this study. We recorded the standard straight carapace length (SCL) from the nuchal notch to the tip of the longest pygal scute when possible (Bolten, 1999). We extracted DNA from either skin or blood samples. Skin samples were collected using a 4-mm biopsy punch and stored in 95% ethanol at room temperature. Blood samples were collected from the dorsal cervical sinus into vials with sodium or lithium heparin, centrifuged to separate plasma, and red blood cells were frozen at -20 °C. For most of the blood samples collected from nesting females before the year 2000, a subset of the whole blood was also stored at room temperature in lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 10 mM NaCl, 1% SDS, pH 8.0) using a 1:10 ratio of blood to buffer (Bagley, 2003).

Laboratory analyses

We extracted genomic DNA using either a Qiagen DNeasy blood and tissue kit following the manufacturer's protocol or a Serapure Bead method with adaptations (Faircloth & Glenn, 2016; Rohland & Reich, 2012). We used primers LCM15382 (Abreu-Grobois et al., 2006) and CM16437 (Shamblin et al., 2012) to amplify an 829 bp fragment of the mitochondrial control region (mtDNA). We used 20 µL polymerase chain reactions with final concentrations of 20 mM Tris HCl pH 8.4, 50 mM KCl, 0.25 mM of each dNTP, 1.5 mM MgCl$_2$, 0.5 µM of each primer, 1 unit of *Taq* DNA polymerase, approximately 10 ng of genomic DNA, and water. We set up thermal cyclers to the following conditions: 95°C for 5 min; 40 cycles of 95°C for 30 s, 57°C for 30 s, 72°C for 80 s; and a final extension at 72°C for 10 min; holding at 10°C. Samples with haplotype

CM-A1.1 were screened at one additional locus using primers CM12751F and CM13064R following PCR and sequencing conditions described in Shamblin et al. (2017). We purified all PCR reactions using Exonuclease I (EN0581) and FastAP (EF0651) following the manufacturer's protocol. Samples were sequenced in both directions through Sanger sequencing.

Data analyses

We edited, assembled, and aligned mtDNA sequences to reference haplotypes (829 bp) available from the Archie Carr Center for Sea Turtle Research database (https://accstr.ufl.edu/resources/mtdna-sequences/) using Geneious R8 (Kearse et al., 2012). We created a median-joining haplotype network using PopART v1.7 (Leigh & Bryant, 2015), and calculated pairwise fixation indexes (F_{ST}), nucleotide (π) and haplotype (h) diversities using Arlequin v3.5.2.2 (Excoffier & Lischer, 2010). We used F_{ST} thresholds proposed by Wright (Wright, 1978) to assess population differentiation. We compared genetic variation over time for each sampling site via analysis of molecular variance (AMOVA) with 10,000 permutations in Arlequin.

Mixed Stock Analysis

We modified the many-to-many mixed stock model, originally implemented in the R package 'mixstock' (Bolker et al., 2007). The original model available in the 'mixstock' package accepts a covariate to weight estimates obtained from haplotype frequency data by the relative size of each rookery (Bolker et al., 2007; Okuyama & Bolker, 2005). Although rookery size is an important factor influencing contributions from rookeries to mixed stocks, the current model does not accept site-specific factors such as distance between rookery and mixed stock location, or main marine

12

currents in between. To date, researchers need to input a matrix of values as priors into the many-to-many model in order to add the effect of distance on estimates (Nishizawa et al., 2013). Our assumption is that rookeries might have greater contributions to relatively closer mixed stocks than to distant ones. Similarly, dispersal from rookeries to some mixed stock aggregations can be facilitated by oceanographic conditions. Even though juveniles are capable of orienting and actively swimming in marine currents (Putman & Mansfield, 2015), there is a greater chance for individuals to disperse to areas closer to where currents initially lead them than to other locations. Following this rationale, we modified the many-to-many model to weight the expected contributions by the scaled inverse distance (*P*) between each source population and mixed stock pair. The model we introduce here is:

$$Estimate \sim SourceContribution * SourceSize * P \qquad (1)$$

where *SourceContribution* is the estimated contribution from each rookery based on haplotype frequencies, and *SourceSize* is the estimated size of each source population. The modified model differs from Bolker et al. (Bolker et al., 2007) only by the scaled inverse distance (*P*) matrix. The code and rationale for the base model with *SourceContribution* and *SourceSize* are described in Bolker et al. (2007) and Okuyama and Bolker (2005). See Supplementary Document S1 for details on our modifications. Here, we populated the matrix with values derived from the estimated inverse distances between each rookery and mixed stock by measuring the length of probable paths between sites using available marine currents as vectors for transport between rookeries and mixed stock aggregations (Figure 2.1; Supplementary Tables S4 and S5). Scaled estimates of effective inverse distance (*P*) between each pair of mixed stock (*i*) and rookery (j) were calculated by

13

$$P_{ij} = \frac{1}{D_{ij}} \bigg/ \sum_{j=1}^{n} \frac{1}{D_{ij}} \qquad (2)$$

where D is the estimated distance from the rookery j to the mixed stock i. Given that a variety of factors may influence the direction and intensity of marine currents (Hays, 2017), we considered two different scenarios (Scenarios 1 and 2) in which individuals may take different paths to move between sites (Figure 2.1 – our discussion focuses only on Scenario 1. See Supplementary Table S4 for distance scenario 2). Our goal was to introduce a tool to improve future mixed stock analysis, not necessarily to define dispersal patterns for green turtles.
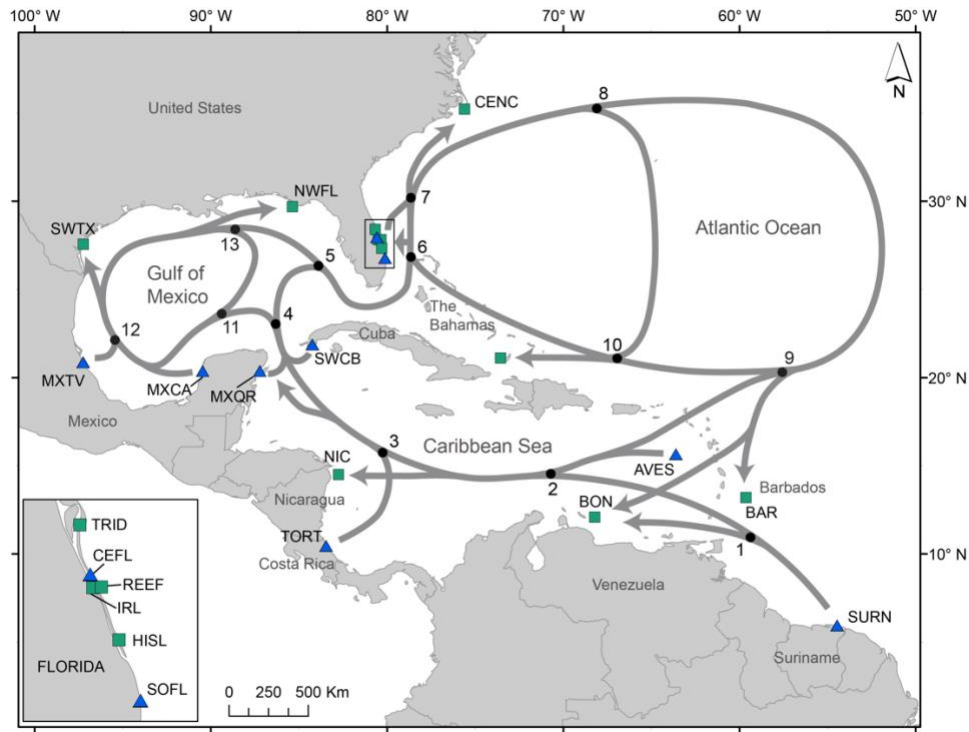
Figure 2.1: Probable routes used by juvenile green turtles to estimate the relative distances between rookeries and foraging areas; general current direction indicated by gray arrows. Blue triangles are rookeries, and green squares are mixed stock aggregation areas. Samples from MB are part of the CEFL rookery. Rookeries: AVES = Aves Island, Venezuela; SURN = Matapica and Galibi, Suriname; TORT = Tortuguero, Costa Rica; MXQR = Quintana Roo, Mexico; MXCA = Campeche and Yucatán, Mexico; MXTV = Tamaulipas and Veracruz, Mexico; SWCB = Guanahacabibes Peninsula, Cuba; SOFL = South Florida, United States; CEFL = Central Florida, United States. Mixed stocks: BAR = multiple areas, Barbados; BON = Lac Bay, Bonaire; NIC = Northeast Nicaragua, Nicaragua; BAH = Southern Bahamas, The Bahamas; SWTX = Southwest Texas, United States; NWFL = Northwest Florida, United States; HISL = Hutchinson Island, United States; RSBI = Reef at Sebastian Inlet, United States; IRL = Indian River Lagoon, United States; TRID = Trident Basin, United States; CENC = Central North Carolina, United States. Map generated in ArcMap 10.8.1 (https://www.esri.com), and arrows and labels were added in Adobe Illustrator 24.3 (https://www.adobe.com).

We used a short fragment of the mtDNA (491 bp) for our MSAs (Supplementary Tables S1-S3), which is contained within the longer (829 bp) fragment. We searched the literature for haplotype frequencies in other mixed stocks and rookeries within the western North Atlantic and Greater Caribbean (Figure 2.1, Supplementary Tables S1-S3). We removed from our final dataset haplotypes found in mixed stock aggregations but not yet described in rookeries (Bolker et al.,

2007), and considered only rookeries in the northwest Atlantic and the Greater Caribbean to reduce noise from unlikely contributors (Engstrom et al., 2002). Though mixed stock data published by van der Zee et al. (2019) uses a timeframe slightly different than the one from IRL and TRID, we included their data in our dataset evaluating variations in haplotype frequencies to assess possible variations in other sites as well.

For rookery size we used a three-year average of the number of nests laid (Table 2.1), based on the best available data we had access to. We estimated rookery size for two time periods: historical (~late 1990s) and recent (early 2010s). Finally, we included only rookeries in the western North Atlantic and Greater Caribbean for which data on the annual number of nests were available for both time periods considered (Table 2.1).

Table 2.1: Number of green turtle nests (source size) by rookery used in mixed stock models, relative size (proportion) in relationship to other rookeries, and variation of relative size between the two time periods (historical and recent). See Figure 2.1 for site abbreviations.

| Rookeries | Source size | | Proportion | | |
	Historical (period)	Recent (period)	Historical	Recent	Variation
CEFL	1,353 (1997-99)[a]	10,129 (2012-14)[a]	1.36%	5.00%	↑ 3.64%
SOFL	827 (1997-99)[a]	6,549 (2012-14)[a]	0.83%	3.23%	↑ 2.40%
MXQR	1,039 (1999-2001)[b]	11,907 (2012-14)[b]	1.05%	5.88%	↑ 4.83%
MXCA	636 (1999-2001)[52]	11,281 (2012-14)[52]	0.64%	5.57%	↑ 4.93%
MXTV	528 (1999-2001)[c]	10,713 (2012-14)[c]	0.53%	5.29%	↑ 4.76%
TORT	86,667 (1997-99)[d]	129,060 (2012-14)[d]	87.30%	63.72 %	↓ -23.58%
SWCB	159 (1997-99)[e]	242 (2010-12)[e]	0.16%	0.12%	↓ -0.04%
SURN	6,562 (1987-89)[f]	19,646 (2008-10)[f]	6.61%	9.70%	↑ 3.09%
AVES	1,500 (1990s)[g]	3,000 (2010s)[g]	1.51%	1.48%	↓ -0.03%

[a](Florida Fish and Wildlife Conservation Commission - Fish and Wildlife Research Institute, 2021); [b](Cuevas Flores et al., 2019); [c](Pineda & Rocha, 2016); [d](Varela et al., 2015); [e](Azanza Ricardo et al., 2013); [f](Seminoff et al., 2015); [g](Nalovic et al., 2020).

We ground-truthed our model to ensure that the modified model results were consistent with the original model in the 'mixstock' package when incorporating a matrix of ones (Supplementary Document S2). We also used a simulated dataset to compare the estimates from our modified model to the approach used by Nishizawa et al. (2013) and determine if results were similar. We compared estimates from the original many-to-many model (MSA$_1$) to estimates from our modified model including a distance matrix (MSA$_2$) to demonstrate how inclusion of a new covariate can impact MSA estimates. For the models described below (MSA$_3$-MSA$_6$) we

populated the distance matrix with values from distance scenario 1 (Supplementary Table S4). To evaluate how changes in rookery sizes impacted MSA estimates, we combined all available data for each mixed stock into a single dataset (Supplementary Table S2) and created one model for each period: MSA_3 – "historical" source size, and MSA_4 – "recent" source size. Finally, to assess how contributions from rookeries changed over time based on mixed stocks haplotype frequencies and rookery sizes, we also built two models: MSA_5 – "old" sampling period and "historical" source size, and MSA_6 – "new" sampling period and "recent" source size. We considered samples from mixed stock BON (van der Zee et al., 2019) to be from comparable timeframes (2006-07 and 2015-16) to IRL and TRID. Therefore, we added haplotype frequencies from BON to our "old" and "new" sampling periods in MSA_5 and MSA_6, and used the same haplotypic data from MSA_3/MSA_4 for all other mixed stocks (Supplementary Table S3). Even though the IRL, TRID, and BON are the only mixed stocks with data to answer our last goal, we included all other mixed stocks in models MSA_5 and MSA_6 to ensure estimates were more accurate. We used 3 chains for each model with a random starting point. We adjusted the number of iterations and burn-in period for models (Supplementary Table S6) to ensure chain convergence by checking the Gelman-Rubin shrink factor (<1.08). To determine evidence of changes between models, we compared estimates by subtracting the posterior distributions for each estimated parameter between two models (e.g., MSA_3 vs MSA_4). The resulting distribution was used in the Test of Practical Equivalence implemented in the R package 'bayestestR' (Makowski et al., 2019) against a null hypothesis (CI = 0.89, range = -0.05 to 0.05). In short, the Test of Practical Equivalence evaluates what proportion of the credible interval of the resulting distribution (i.e., 89% CI) that falls inside the range defined as the null (i.e., -0.05 to 0.05) (Makowski et al., 2019). We chose a credible interval of 89% based on small posterior distributions sample size ($<10,000$) (Kruschke, 2015; Makowski et al., 2019).

Finally, we used linear regressions to test if the mean estimates from models MSA_3-MSA_6 were correlated to the distance between rookeries and mixed stocks.

Results

Biometrics

We sampled a total of 200 turtles among the three locations and two time periods (Table 2.2). The mean SCL of first capture for nesting females in MB_{old} was 100.9 cm (SD 5.1, range 93.4 – 114.1) and for MB_{new} was 97.8 cm (4.7, 90.5 – 108.6). For mixed stock samples, the mean size of first capture at IRL_{old} was 46.8 cm (10.8, 29.5 – 68.6) and for IRL_{new} 48.1 cm (8.4, 32.4 – 66.7), while $TRID_{old}$ was 29.5 cm (2.9, 23.4 – 39.2) and $TRID_{new}$ was 30.9 cm (4.3, 23.7 – 43).

Table 2.2: Haplotypes identified using the 829 bp mitochondrial DNA fragment. MBold (nesting females before 2000), MBnew (nesting females between 2016-2018); IRLold and TRIDold (foraging sites between 2003-2005); IRLnew and TRIDnew (foraging sites between 2016-2018). Grayed rows indicate sequencing of a diagnostic fragment to distinguish between variants of haplotype CM-A1.1. IRL = Indian River Lagoon; Trident = Trident Submarine Basin. *2 samples failed to amplify the diagnostic sequence for CM-A1.1.

Haplotype	Adults		Juveniles			
	MB_{old}	MB_{new}	IRL_{old}	IRL_{new}	$TRID_{old}$	$TRID_{new}$
CM-A1.1	11	9*	7	11	13	24
CM-A1.1.1	1	1	6	8	6	21
CM-A1.1.2	10	6	1	3	7	3
CM-A1.2	5	1	4		2	1
CM-A1.4					1	1
CM-A2.1				1		
CM-A3.1	9	11	13	20	17	11
CM-A5.1			1	2	1	
CM-A8.1			2	1		
CM-A13.1			3			
CM-A16.1			3			
CM-A18.1				1	1	1
CM-A18.2					2	
CM-A26.1				1	2	1
CM-A27.1				1		
CM-A28.1			1		2	
CM-A48.3	2					
Total	**27**	**21**	**34**	**38**	**41**	**39**

Population structure

We identified four different haplotypes in MB (Table 2.2, Supplementary Fig. S1), including two samples with CM-A48.3 in MB_{old}. This is the first time CM-A48.3 has been identified at a nesting site. The short-fragment version of this variant (CM-A48) had previously only been found in Cuba (Ruiz-Urquiola et al., 2010). We identified haplotypes CM-A27.1 and CM-A28.1 for the first time

in a juvenile foraging site on the east coast of Florida. Haplotypes CM-A3.1 and CM-A1.1 were the most frequent both in adult (41.7% and 41.7%) and in-water samples (45.8% and 25.0% in IRL, and 35% and 46.3% at TRID) for both "old" and "new" sampling periods (Table 2.2, Supplementary Fig. S1).

Results from AMOVAs to determine if genetic diversity changed over time per site indicate that most variation was observed within populations and not over time for all sites. However, we did find that among population variation was greater at the MB site, indicating greater change-over-time than found at the in-water sites (Supplementary Table S7). Haplotype (h) and nucleotide (π) diversities decreased for all sites over time. In MB, h decreased from 0.709 (SD = 0.047) before year 2000 to 0.567 (0.056) in 2016-2018, and π decreased from 2.252×10^{-3} (1.482×10^{-3}) to 0.757×10^{-3} (0.693×10^{-3}). For mixed stock aggregations, in the IRL h varied from 0.8 (0.049) in 2003-2005 to 0.65 (0.063) in 2016-2018 and π from 3.526×10^{-3} (2.112×10^{-3}) to 2.899×10^{-3} (1.794×10^{-3}), while in TRID h went from 0.734 (0.05) to 0.553 (0.068) and π from 2.451×10^{-3} (1.566×10^{-3}) to 1.13×10^{-3} (0.885×10^{-3}). The TRID mixed stock saw the highest reduction in haplotype diversity (from 0.732 to 0.553). Despite variation in haplotype and nucleotide diversities, pairwise F_{ST} comparisons did not indicate significant variation within sites over time (Table 2.3). The only differences in F_{ST} were between IRL_{old} and both MB sampling periods, between IRL_{new} and MB_{old}, and between $TRID_{new}$ and both IRL sampling periods. We found no evidence of structuring between sampling timeframes for each site. For rookery data, we also grouped samples with previous studies for the models evaluating changes in haplotype frequencies (MB is part of the CEFL rookery - Supplementary Table S1).

Table 2.3: Pairwise distance between sampling sites. Cells in bold indicate moderate genetic differentiation ($0.05 < F_{ST} < 0.15$)(Wright, 1978). Cells below diagonal show pairwise F_{ST} values. Values above diagonal show estimated p-values obtained from bootstrapping after correcting for multiple comparisons.

	MB$_{old}$	MB$_{new}$	IRL$_{old}$	IRL$_{new}$	TRID$_{old}$	TRID$_{new}$
MB$_{old}$	-	0.710	0.075	0.084	0.710	0.710
MB$_{new}$	0.046	-	0.495	1.000	1.000	0.710
IRL$_{old}$	**0.064**	**0.056**	-	0.930	0.710	0.091
IRL$_{new}$	**0.063**	0.004	0.013	-	1.000	0.132
TRID$_{old}$	0.022	-0.007	0.027	-0.004	-	0.930
TRID$_{new}$	0.033	0.040	**0.077**	0.047	0.011	-

Distance matrix

Comparing the estimates obtained using our modified model and the approach used by Nishizawa et al. (2013) we found that our modified model consistently provided estimates with narrower credible intervals than adding distances as priors (Supplementary Fig. S2). Regarding the inclusion of a distance matrix to a many-to-many approach, estimates from rookery SWCB to the TRID mixed stock and from MXQR, SURN, and AVES to both mixed stocks remained essentially the same between MSA$_1$ and MSA$_2$ (Figure 2.2). However, adding the distance matrix in MSA$_2$ made the credible intervals wider from CEFL, MXCA, MXTV, TORT, and SWCB to the IRL mixed stock, and from MXCA and TORT to the TRID estimates. In contrast, credible intervals were narrower from SOFL to the IRL mixed stock, and from CEFL, SOFL, and MXTV to the IRL. We found a weak relationship between values populated into the distance matrix and rookeries contribution estimates for models MSA$_4$-MSA$_6$ (Supplementary Fig. S3).
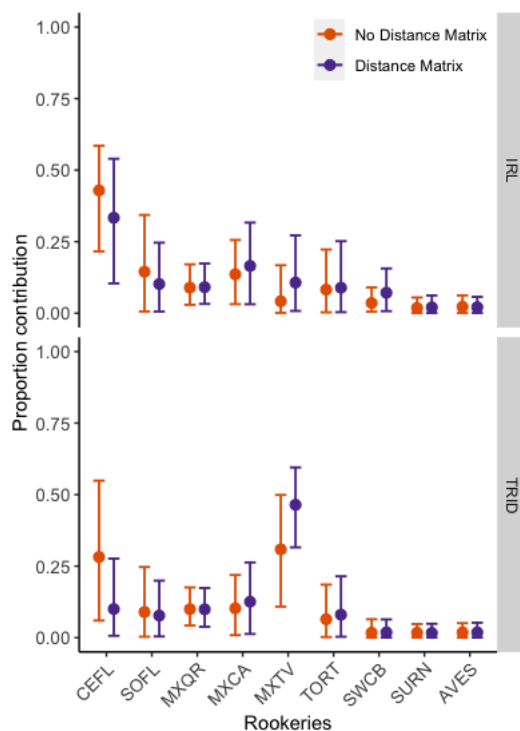
Figure 2.2: Impact of incorporating a distance matrix into many-to-many mixed stock models. Solid points represent mean estimates and vertical bars 95% credible intervals. No Distance Matrix (MSA$_1$) = standard many-to-many model from package 'mixstock'; Distance Matrix (MSA$_2$) = modified many-to-many model with a distance matrix. See Figure 2.1 for site abbreviations.

Effect of rookery size on mixed stocks

All rookeries showed an increase in the average number of nests per season from historical to recent time periods (Table 2.1). However, given the different rates of increase, the relative contribution from each rookery (the number of nests divided by the total number of nests in all rookeries for each time period) changed over time. For AVES and SWCB, their relative proportion remained virtually unchanged (~1.5% and ~0.14% respectively). On the other hand, TORT had the largest increase in absolute numbers (over 42,000 annually), but its proportion decreased from 87.30% to 63.72%. For CEFL, SOFL, MWQR, MXCA, MXTV, and SURN there was an increase in their relative proportion at similar rates (3.09 – 4.93%).

We found little evidence of changes in contributions to mixed stocks as a response to changes in rookery sizes alone (MSA_3 vs MSA_4 – Figure 2.3). Some mixed stocks analyzed have a single main contributing source population: MXTV is the main contributor to TRID and SWTX, SURN is the main contributor to BAR, and TORT appears as the main contributor to BON, HISL, BAH, and NIC. Contributions to IRL, CENC, RSBI, and NWFL appear to come from multiple sources without a clear single origin. Considering the rookery-centric estimates (Supplementary Fig. S4), individuals from most rookeries disperse similarly among the mixed stocks analyzed (overall mean estimate = 8.33%, SD = 8.25%). Individuals from TORT disperse mainly to NIC, followed by other mixed stock(s) not present in this analysis (UNK). Main destinations for individuals originating from the SURN rookery were BAR, NIC, and UNK. For both TORT and SURN, there is great uncertainty regarding the main mixed stock destinations. Complete results for models using distance scenario 2 are available in Supplementary Tables S8-S9.
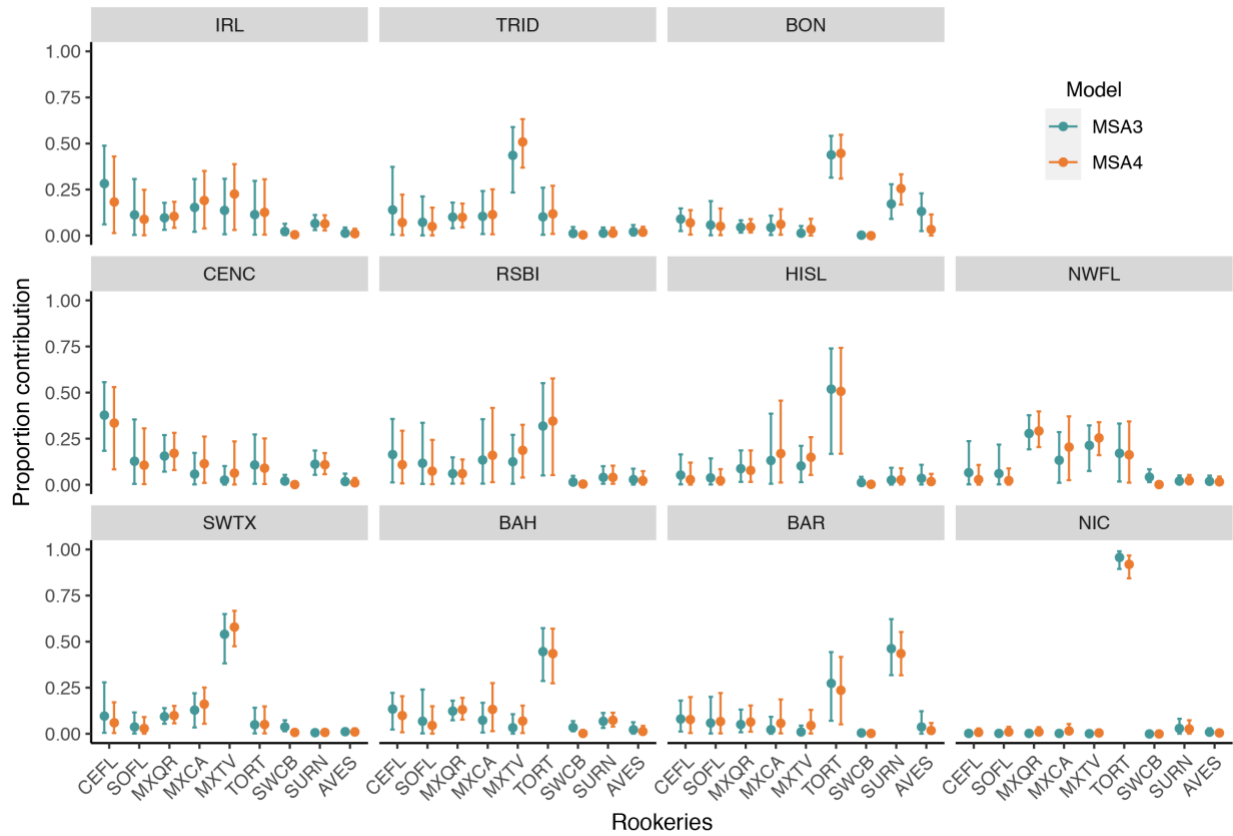
Figure 2.3: Mixed stock-centric estimates comparing different source sizes. Filled dots represent the mean estimate, and vertical bars 95% credible intervals. MSA_3 = "historical" source size. MSA_4 = "recent" source size. * = rookeries with evidence of a difference in contribution in response to source size variation. See Figure 2.1 for site abbreviations.

Haplotype and source size variation

For this objective, we present only the results for mixed stocks with data available for the two sampling periods and the corresponding rookery sizes: IRL, TRID, and BON (MSA_5 and MSA_6, Figure 2.4 – Supplementary Tables S3 and S10 for complete results). The impact of changes in source size on the broader mixed stock estimates was established in the previous section (models MSA_3 and MSA_4). For TRID, we found evidence of an increase in the proportion of individuals from MXTV, and for BON there was a decrease in contributions from SURN. Also, for the IRL mixed stock, recent years have narrower credible intervals and lower mean estimates for rookeries

25

CEFL, SOFL, MXQR, SWCB, SURN, and AVES, while wider credible intervals were observed for MXCA and TORT. For TRID, we observed the same pattern of narrower credible intervals for all rookeries except for MXTV. Finally, for BON, recent years appear with tighter credible intervals for SURN and AVES, while for CEFL, SOFL, MXQR, MXCA, MXTV, and TORT we see wider credible intervals (Figure 2.4). Source-centric estimates for MSA_5 and MSA_6 indicate no changes in destination of individuals from all rookeries over time (Supplementary Fig. S5).
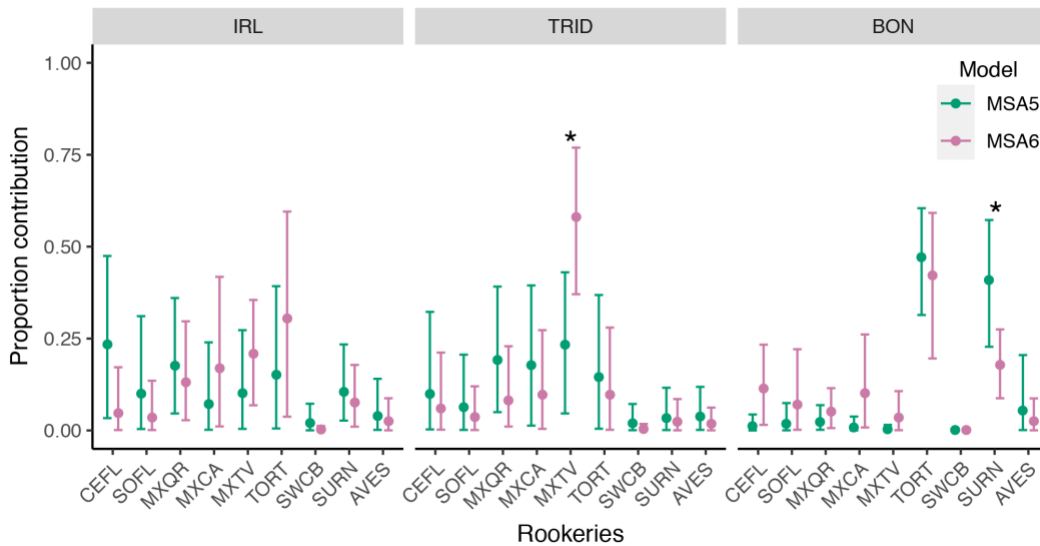
Figure 2.4: Mixed stock-centric estimates with different sampling events and varying source sizes. Circles represent the mean estimate, and vertical bars 95% credible intervals. MSA_5 = "old" mixed stock sampling period and "historical" source size. MSA_6 = "new" mixed stock sampling period and "recent" source size. * = rookeries with evidence of a difference in contribution in response to source size variation. See Figure 2.1 for site abbreviations.

Discussion

Our study introduces an important advancement for mixed stock analysis: a more informative and ecologically meaningful model incorporating a matrix of site to site-specific weighted inverse distances. We demonstrate how demographic variations in source populations and temporal

changes in haplotype frequencies in mixed stocks aggregations can impact MSA estimates. Also, we show how understanding dispersal patterns and connectivity between sites is crucial for management of migratory organisms. Our analyses indicate how the stock composition of juvenile aggregations of green turtles in east central Florida have changed over a 13-year period, simultaneous to a population growth on several source nesting populations. Furthermore, we clearly demonstrate the importance of long-term monitoring and periodic reassessment of both breeding and juvenile aggregations.

For the juvenile IRL and TRID mixed stocks, the mean SCL from individuals sampled in our study was comparable to the mean sizes previously reported for both sites (Long et al., 2021; Redfoot & Ehrhart, 2013). For the rookery MB, the mean size and observed reduction in mean SCL among nesting females are consistent with a trend recently described for this populations (Phillips et al., 2021). The pairwise comparison between sites (Table 2.3) corroborates our decision to treat IRL and TRID as separate mixed stocks. Also, F_{ST} indicates a greater genetic differentiation between MB and the IRL mixed stock, suggesting it is not mostly composed of individuals from MB despite geographical proximity, supporting our assumption that the distance between MB and IRL is much greater than what a straight line between these sites suggests (Supplementary Tables S4-S5). Several green sea turtle nesting sites in the western North Atlantic and Greater Caribbean have increased in both estimated abundance and number of nests laid (Seminoff et al., 2015), including the MB nesting aggregation (Chaloupka et al., 2008). Female sea turtles are known for reproductive natal philopatric behavior (Bowen et al., 1996). Given that the genetic marker we used (mtDNA) is both haplotypic and maternally inherited, a reduction in haplotype diversity in reproductive populations would be expected given the reduced effective size associated with

mtDNA, especially for historically bottlenecked populations. Regarding in-water aggregations, the observed reductions in h and π could be a consequence of changes in the main contributors to each mixed stock (Figure 2.4), with a general homogenization of the genetic pool. Similarly, a recent study on a mixed stock in Lac Bay, Bonaire, found a reduction in nucleotide diversity but no clear change in haplotype diversity over 9 years (van der Zee et al., 2019). However, van der Zee et al. (2019) amplified only the short mtDNA fragment, which could reduce their ability to detect variations. Regardless, results from our analysis indicate a predominance of a single contributor in BON in recent years instead of two from the "old" sampling period (Figure 2.4), supporting the van der Zee et al. (2019) hypothesis of changes in contributors over time. Even though it is unlikely to be observed on all sites simultaneously, after splitting our dataset for IRL and TRID into two sampling periods, we cannot discard the possibility that these reductions are due to small sample sizes. The reduction in sample size for IRL, TRID, and BON mixed stocks in MSA_5/MSA_6 compared to MSA_3/MSA_4 could help explain the increased uncertainty around the estimates (Figure 2.4). Additionally, we acknowledge that our results are a snapshot in time and encompass less than one generation-time for this species; undetected complex ecological processes might be underway (Bjorndal et al., 2005). Future studies with a larger sample size from a single mixed stock and time period could try to address this concern using a resampling approach (e.g., jackknife-based method) to identify how sampling might affect MSA estimates.

We identified variations on the width of credible intervals between our modified model and the original model in the 'mixstock' package (MSA_1 vs MSA_2). Even though we did not specifically test possible causes for variation in credible intervals after the inclusion of the distance matrix, we suspect it could be related to values in the distance matrix that do not match estimates from

28

haplotype frequencies (e.g., haplotype frequencies indicate small contribution from one rookery while the value in the distance matrix suggest higher contribution from the same source). Mixed stocks and/or rookeries with small sample sizes could be more impacted by such variations.

We found no clear evidence of changes in contributors to mixed stocks when considering variation in rookery size alone (MSA_3 and MSA_4; Figure 2.3). Previous studies report little or no difference in estimates when comparing models with rookery size versus models with an uninformative covariate (i.e., equal value to all rookeries) while using a many-to-one framework (Proietti et al., 2012; Shamblin et al., 2017). This is not an unexpected result as MSA estimates are mostly derived from genetic markers (Pella & Masuda, 2001), and the weight provided by covariates might not be enough to change estimates. However, we found evidence of variations in contributions when the haplotypic variation was considered along with rookery size variation (MSA_5 and MSA_6; Figure 2.4). Though, we did not test a model with varying haplotype frequencies and constant rookery size, as rookery sizes did change over time this would be an unrealistic scenario and we could not tease these changes apart. Therefore, we cannot determine if the observed fluctuation in haplotype frequencies (and rookery contribution estimates) was caused by changes in the influx of individuals from source populations to mixed stocks or by variation in source population sizes because both possibilities are intrinsically dependent on one another.

The main contributors to mixed stocks from models MSA_3 and MSA_4 were partially different from previous analyses in the Atlantic Ocean and Greater Caribbean (Bass et al., 1998, 2006; Bass & Witzell, 2000; Bolker et al., 2007; Foley et al., 2007; Lahanas et al., 1998; Luke et al., 2004; Monzón-Argüello et al., 2010; Naro-Maciel et al., 2012; van der Zee et al., 2019), which could be

explained by substantial differences between our dataset and the different datasets and models used by previous studies. However, results from MSA_5 and MSA_6 corroborate findings from studies that identified fluctuations in contributions over time in response to changes in haplotype frequencies in mixed stocks (Bjorndal & Bolten, 2008; van der Zee et al., 2019). An assumption of mixed stock models is that all source populations are represented and adequately sampled (Pella & Masuda, 2001) – an assumption that will rarely be met. Engstrom et al. (2002) suggest not including unlikely contributors to mixed stock models to reduce noise, a decision we also made. However, researchers may have different thresholds to define an unlikely contributor, therefore, this decision becomes arbitrary. Comparing estimates among studies is difficult as new areas are added and more samples are sequenced. Furthermore, our modified model uses effective distance to weight estimates; this adds an extra layer of differentiation among studies, making direct study comparisons even harder. Regardless of agreement (or lack of agreement) between our results and previous studies, we believe that future assessments can improve biological meaning if mixed stocks and rookeries are periodically reassessed for haplotype frequencies.

An increase in juvenile abundance following reproductive population growth and increased number of nests laid is a reasonable expectation. This expectation depends on the fitness of reproductive individuals, the hatching success of the nests laid, and survival and recruitment rates for juveniles. However, Bjorndal et al. (2005) found no correlation between increased number of nests at Tortuguero, Costa Rica, the putative main stock of origin for the mixed stock, and variations in the abundance of green turtles at Union Creek, The Bahamas. One hypothesis was that Union Creek was near carrying capacity for green turtles, and abundance would remain stable over time (Bjorndal et al., 2005). Our models corroborate their findings, showing TORT as the

main contributor (Figure 2.3) despite the reduction in TORT's size in relationship to the other source populations in the region (Table 2.1). The stability of contributions to BAH could be an indication that the carrying capacity hypothesis is still a valid option for Union Creek. Similarly, Long et al. (2021) attributed a decrease in green turtle abundance in the IRL mixed stock between 2001 and 2018 to a general decrease in habitat quality, despite the increased abundance in rookeries. It is possible that juvenile abundance increased in other mixed stock aggregations and that the observations in the IRL and BAH mixed stocks (Bjorndal et al., 2005; Long et al., 2021) are isolated cases. However, a study with green turtles from MB identified a decrease in nesting females' mean size and size at maturity over the past decades, especially after the late 1990s (Phillips et al., 2021). One of the explanations for a decrease in nesting female body size is reduced juvenile mass growth rate (Bjorndal et al., 2013), which, ultimately, could lead to overall reduced reproductive fitness in rookeries. At least for leatherback sea turtles (*Dermochelys coriacea*), reproductive fitness can be impacted by maternal health parameters (Perrault et al., 2012). Interestingly, these data are supported by our genetic analyses that found little change between in-water sites over time, but greater change among time periods for the nesting beach site.

Mixed stock analysis using either the number of nests or the number of nesting females as a proxy for source size should correct estimates by emergence success (total number of hatchlings emerged divided by the total number of hatched eggs in a clutch). Emergence success can vary among seasons, rookeries, species, and can be affected by maternal health and environmental factors (Montero et al., 2018; Perrault et al., 2012). To ensure future mixed stock analyses benefit from more informative rookery sizes, we urge researchers to report the number of nests, hatching success, and emerging success, as well as other basic reproductive parameters from nesting

populations. We second the call by Shamblin et al. (2014) for broad use of longer fragments of mtDNA in reassessments of rookeries that have been only evaluated using the short fragment, and especially, that new studies refrain from sequencing the short fragments only. The development of new diagnostic markers using whole mitogenomic sequences (Shamblin et al., 2012), or a combination of mtDNA with other markers (e.g., nuclear microsattelites), to increase discrimination between rookeries is essential for our understanding of sea turtle evolution and dispersal patterns. Future population and species assessments will benefit from better and more refined genetic information.

Understanding dispersal and connectivity among habitats and across life stages is fundamental for species' conservation. The main feature introduced by our modified model is the capacity for researchers to more easily consider variables that are specific to each pair of source populations and mixed stocks in a many-to-many framework. Prior to our modified model, studies incorporating particle dispersal probabilities or distance between sites often weighted MSA estimates using a many-to-one model framework because the probabilities from a source will differ to each mixed stock, and estimates from multiple mixed stock models need to be combined for a regional overview (Naro-Maciel et al., 2014; Proietti et al., 2012; Putman et al., 2020; Shamblin et al., 2017). Many-to-many models provide estimates with narrower credible intervals than many-to-one models when analyzing the same dataset (Anderson et al., 2013; Bolker et al., 2007; Monzón-Argüello et al., 2010). Our modified model usually provided narrower credible intervals than the approach introduced by Nishizawa et al. (2013) on a many-to-many framework. Site-specific probability matrices that incorporate complex variables such as particle dispersal model

32

estimates will enable researchers to consider multiple cohorts, variation within and among seasons, and multiple variables that can impact oceanographic currents (Putman & Naro-Maciel, 2013).

Our modified many-to-many mixed stock model can incorporate new variables to make models more informative. More importantly, by incorporating distance between rookeries and mixed stocks, or particle dispersal probabilities, models we can better account for unlikely source populations, allowing more realistic estimates of rookery contributions to mixed stocks for robust ocean basin analyses. The short-fragment mtDNA markers used for MSA lack the resolution needed to differentiate between several rookeries (Shamblin et al., 2012). As mixed stock model estimates are mainly derived from haplotype frequencies (Bolker et al., 2007; Pella & Masuda, 2001), under scenarios where the genetic marker used is unable to differentiate populations, covariates can help improve model accuracy. The source code and example script for incorporating the site-specific matrix is available in Supplementary Document S1, and we encourage others to use this approach to incorporate distances, transport probabilities, or any other metric that scales the contributions from each rookery to each mixed stock. Contribution estimates from such models will be more ecologically meaningful and more accurate. Further, we highlight the importance of long-term monitoring and periodic reassessment of mixed stock aggregations regarding stocks of origin, abundance, health status, and other population parameters. We also emphasize the importance of periodical reassessment of haplotype frequencies at rookeries, as well as basic demographic and reproductive parameters. For migratory endangered species such as sea turtles, broad analyses considering multiple rookeries within or among ocean basins with more informative estimates are critical for understanding dispersal, connectivity, and evolution.

Understanding how the composition of mixed stock aggregations shift over time is fundamental for the development of successful conservation plans for endangered and threatened species.

Acknowledgments

Data availability

Supplementary Tables S1-S10, Supplementary Figures S1-S5, and Supplementary Documents S1-S2 for Chapter 2 are available at: https://doi.org/10.1038/s41598-022-24279-2

CHAPTER 3: THE EFFECT OF SAMPLE SIZE, SOURCE POPULATION DIFFERENTIATION, AND SMALL CONTRIBUTORS ON MIXED STOCK MODELS ESTIMATES

Introduction

Sample size is a fundamental aspect of any scientific study. Statistical analyses often may have misleading or incorrect interpretation due to inadequate sampling (Button et al., 2013; Ioannidis, 2005). This is a serious problem for study reproducibility, particularly in studies focusing on endangered species (Shaw et al., 2021) or within the biological sciences (see review in Jenkins & Quintana-Ascencio (2020)). A minimum of 25 samples has been suggested for accurate regression analyses with some variance (Jenkins & Quintana-Ascencio, 2020), but datasets for other types of analysis lack a general guideline of minimum number of samples required for reliable and ecologically meaningful results.

Small sample size occurs in studies for various reasons, ranging from difficulty of access to areas (e.g., tan isolated island (Martins et al., 2008)), difficult logistics (e.g., maintenance of optimal conditions for blood samples collected in the field (Ryser-Degiorgis, 2013), or even lack of funding to process more samples. Highly migratory species, or long-lived species with complex life histories, pose additional challenges to meeting minimum sample sizes. For example, sea turtles are widely distributed and occupy different habitats during different life stages, where data for early life stages are almost non-existent due to logistical challenges to access and locate individuals (Mansfield et al., 2014; Mansfield & Putman, 2013; Phillips et al., 2021; Shamblin et al., 2018). Reasonable or adequate sample sizes might be hard to achieve and may remain a distant goal for many researchers (Witmer, 2005). Large sample sizes can be even more difficult to obtain

for endangered species, where the number or density of individuals available for sampling is likely small, and regulatory agencies may impose limits on the number of samples that can be collect for research (e.g., Endangered Species Act in the United States and similar legislation in other countries).

Despite challenges above, species distribution is vital to ecology and biogeography because it impacts fitness and population biology (Chuine, 2010). To understand distribution patterns for migratory and widely distributed species such as sea turtles, researchers have used satellite telemetry (Godley et al., 2002; Mansfield et al., 2014; Marcovaldi et al., 2010; Phillips et al., 2021; Seminoff et al., 2002; Stoneburner, 1982), stable isotopes (Ceriani et al., 2012; Seminoff et al., 2012; Zbinden et al., 2011), or genetics (Bolker et al., 2007; Naro-Maciel, et al., 2014; Phillips et al., 2022; Shamblin et al., 2017; Shamblin, et al., 2015; Stahelin et al., 2022). Satellite telemetry data acquisition and tag cost are expensive, limiting the number of individuals that can be tracked per study (Ceriani et al., 2012; Roberts et al., 2018). Stable isotopes can be used in combination with satellite telemetry to first characterize areas used by individuals satellite tracked and then use the pre-defined isotopic signatures for non-tracked individuals using large sample sizes at a relatively low cost (Ceriani et al., 2012; Seminoff et al., 2012; Zbinden et al., 2011). For genetics, different markers can be used for source population assignment (e.g., mtDNA, microsatellites, autosomal single nucleotide polymorphisms). Statistical methods compare the frequency of markers found in a group of individuals to estimate the source populations where those individuals came from (Bolker et al., 2007; Pella & Masuda, 2001). Similarly to stable isotopes, most molecular markers can be used in large-scale at relatively low cost.

Mixed stock analyses (MSA) are regularly used to evaluate source populations of a group of individuals captured at mixed stock aggregations. For sea turtles, mixed stock analyses are often conducted using a wide number of samples per study site (Figure 3.1). In Chapter 2, I raised the possibility that uncertainties in some model estimates could be due to reduced sample sizes in sampled locations (Stahelin et al., 2022). This concern arises mostly due to the overlap in haplotype frequencies among some sea turtle rookeries (Supplementary Table S1). Bolker et al. (2003) described how the Markov Chain Monte Carlo (MCMC) applied to mixed stock models is effective handling rare haplotypes and sampling error. Another concern is that mixed stock analysis depends on a minimal discrimination between source populations through haplotype frequencies to determine the distribution of individuals (Pella & Masuda, 2001). Even though the Bayesian approach implemented in the 'mixstock' R package is designed to account for such imperfections in haplotype frequencies between source populations, molecular markers currently used may not provide enough resolution for accurate results (Shamblin et al., 2017).

Aiming to address the concerns defined above regarding possible parameters influencing the accuracy of mixed stock model estimates, in this chapter I simulate (i) the effect of sample sizes on mixed stock estimates, (ii) the impact of higher resolution markers could have on model estimates, and (iii) determine how small contributors affect mixed stock model estimates and our understanding about individuals' dispersal from breeding populations to mixed stock aggregations.

Figure 3.1: Examples of sample sizes used in sea turtle mixed stock analyses found in the literature. Samples were collected either from adult females on a nesting beach (Rookery), or in-water (Mixed stock) sampling sites. Sources: Anderson et al., 2013; Bagley, 2003; Bass et al., 1998, 2006; Bjorndal et al., 2005; Bjorndal & Bolten, 2008; Costa Jordao et al., 2017; Dutton et al., 2014; Formia et al., 2006, 2007; Jensen et al., 2016; Lahanas et al., 1998; Laurent et al., 1998; Luke et al., 2004; Maffucci et al., 2006; Millán-Aguilar, 2009; Monzón-Argüello et al., 2010; Naro-Maciel et al., 2012; Prosdocimi et al., 2012; Read et al., 2015; Ruiz-Urquiola et al., 2010; Shamblin, Bagley, et al., 2015; Shamblin et al., 2012, 2017; Stahelin et al., 2022

Methods

Simulating datasets

Using the statistical software R (R Core Team, 2016), I simulated three different scenarios of haplotype distribution between source populations (rookeries) (Figure 3.2), with three source populations in each scenario. The first scenario (Figure 3.2a), hereafter called the '*distinct*' scenario, contains populations with a low level of shared haplotypes. The second scenario, hereafter called the '*similar*' scenario (Figure 3.2b), simulates populations with a greater similarity in haplotype frequencies between populations. The third scenario, or the '*shared*' scenario (Figure 3.2c), contains populations with an extreme case where most haplotypes are shared among all populations. Comparison of these scenarios evaluates how higher resolution markers that better

discriminate between source populations can impact mixed stock model estimates (i.e., the *similar* scenario assumes a high-resolution marker, whereas the *shared* scenario is based on relatively low resolution). The *shared* scenario simulates haplotypes frequencies resembling the resolution of the 400 bp mtDNA marker found at green sea turtle rookeries in the Atlantic Ocean (e.g., east central Florida, USA, Tamaulipas, Mexico, and Tortuguero, Costa Rica (Supplementary Table S1); or Trindade Island, Brazil, Rocas Atoll, Brazil, and Ascension Island, UK (Costa Jordao et al., 2017)). The *similar* scenario possibly represents resolution using 800 bp fragments of mtDNA, while the *distinct* scenario represents a deeper resolution from the 800 bp fragment. I used a maximum of 19 haplotypes for these simulations to incorporate low frequency haplotypes and to closely simulate the number of haplotypes shared by many rookeries (Supplementary Table S1).
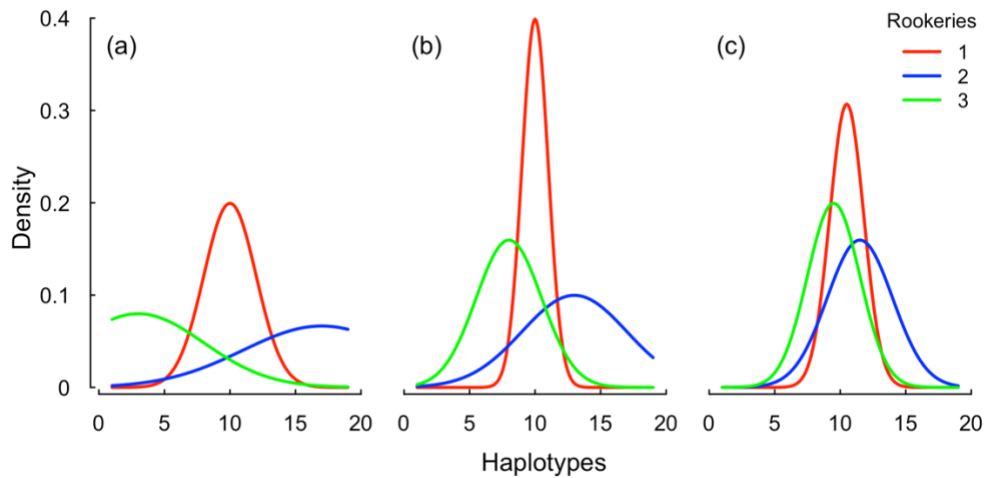
Figure 3.2: Density plots of the haplotype frequencies for the three simulated scenarios: (a) *distinct*, (b) *similar*, and (c) *shared*. Solid lines represent the density of haplotypes occurring in each rookery.

The following datasets were simulated for each of the haplotype frequency scenarios described above. I first created three rookeries based on haplotype frequencies for the corresponding scenario

(Figure 3.2a, b, or c). Then, to simulate a research study collecting samples from these source populations, I randomly sampled without replacement (i.e., to avoid resampling of the same 'individual') from each rookery. Next, I randomly sampled individuals with replacement from the three rookeries generated in the first step to simulate dispersal and capture of individuals in order to create four mixed stock aggregations. During this step, I recorded the proportion of individuals originating from each rookery and moving to each mixed stock to compare the model estimates to the true contribution from each rookery. Finally, I simulated studies collecting individuals at each mixed stock aggregations by randomly sampling individuals without replacement.

The population sizes for rookeries were set to be 500, 1,000, 10,000, or 50,000, while mixed stock aggregations had either 500, 1,000, or 5,000 individuals (Table 3.1). Categorical population sizes reduce the number of simulations but cover a wide range. For each combination of rookery and mixed stock sizes (e.g., rookery size = 500 and mixed stock size = 1,000 - Table 3.1), I sampled a random number of individuals from the corresponding rookeries and mixed stocks. Sample sizes for each sampling event ranged from 25 to 200 for the *distinct* and *similar* scenarios, and from 25 to 300 for the *shared* scenario. Values for the minimum number of samples < 25 often crashed the mixed stock model because all haplotypes were from only one source population (i.e., it was not a mixed stock). Preliminary results indicated that larger sample sizes would be needed for the *shared* scenario to achieve a similar accuracy as the *distinct* and *similar* scenarios. Therefore, I simulated a larger number of samples for the *shared* scenario only. The maximum number of samples was set to 200 (300 for the *shared* scenario) to match a reasonable number of samples in real-world studies (Figure 3.1). Lastly, I repeated the sampling step for each combination of rookery and mixed stock sizes 1,000 times for the *distinct* and *similar* scenarios, and from 1,500 times for the

shared scenario. In total, I evaluated 12,000 models for each of the *distinct* and *similar* scenarios, and 18,000 models for the *shared* scenario.

Table 3.1: Summary table of the parameters used for all simulated datasets.

Population size		Sample size		Number of simulations	
Rookeries	Mixed stocks	*Distinct* and *Similar*	*Shared*	*Distinct* and *Similar*	*Shared*
500	500	25-200	25-300	1,000	1,500
1,000	500	25-200	25-300	1,000	1,500
10,000	500	25-200	25-300	1,000	1,500
50,000	500	25-200	25-300	1,000	1,500
500	1,000	25-200	25-300	1,000	1,500
1,000	1,000	25-200	25-300	1,000	1,500
10,000	1,000	25-200	25-300	1,000	1,500
50,000	1,000	25-200	25-300	1,000	1,500
500	5,000	25-200	25-300	1,000	1,500
1,000	5,000	25-200	25-300	1,000	1,500
10,000	5,000	25-200	25-300	1,000	1,500
50,000	5,000	25-200	25-300	1,000	1,500

Finally, I used the many-to-many mixed stock model from the mixstock package (Bolker et al., 2007) in software R (R Core Team, 2016) to analyze the haplotype frequencies obtained during each simulation described above. See Chapter 2 for a broader review of mixed stock analysis. I clarify that in Chapter 2 I introduced the possibility to consider distance between sites in a many-to-many framework, but simulations presented in this Chapter do not consider distance. Each model used three chains with random starting points and 20,000 iterations (10,000 iterations burn-in). I used the Gelman-Rubin shrink factor < 1.2 from the coda package in R (Gelman & Rubin, 1992; Plummer et al., 2006) to evaluate model convergence.

Data analysis

I used three metrics to evaluate the impact of sample sizes on accuracy of mixed stock model estimates. The first metric was credible interval width (CIW) around model estimates. I calculated the CIW by subtracting the value of the 2.5% credible interval from the value of the 97.5% credible interval (i.e., CI from 35% - 50% = CIW with 15 percentage points). The second metric was the absolute difference between the true contribution from each rookery and mixed stock pair and the mean estimate from the model (hereafter called "absolute model estimate error"). For example, if the true contribution of each rookery to a mixed stock aggregation was 25%-25%-50%, and the model estimate was 30%-30%-40%, then the absolute model estimate error would be 5%-5%-10%. Finally, the third metric was the probability of the true contribution from each rookery to fall inside the credible interval estimated by the model. Related to the absolute model estimate error, I also evaluated the difference between the true contribution from each rookery and mixed stock pair and the mean estimate from the model ("general model estimate error") to determine if mixed stock models were consistently under- or overestimating the parameters across different sample sizes. Note that the only difference between the absolute model estimate error variable and the general model estimate error used to evaluate under- or overestimation is that the former uses absolute numbers. For all metrics, I evaluated how estimates varied for small contributing sources (i.e., source populations with < 5% of true contribution to a mixed stock aggregation).

To test variations in credible intervals I fitted generalized linear models (GLM). For the absolute model estimate error dataset, I used the function *fitDist()* from the *gamlss* package (Rigby & Stasinopoulos, 2005) to fit a series of relevant parametric distributions to select an appropriate distribution using the Akaike Information Criterion (AIC). Following model selection with

fitDist(), I used a beta regression model from the *betareg* package (Cribari-Neto & Zeileis, 2010).

To evaluate the probability of true contribution to be contained inside the credible interval (pCI), I used logistic GLM models with a binomial distribution. I tested a combination of independent variables for both GLMs and beta regression models: rookery sample size, mixed stock aggregation sample size, scaled rookery population size, scaled mixed stock population size, ratio of sampled individuals in rookeries in relationship to rookery population size, ratio of sampled individuals in mixed stock aggregations in relationship to mixed stock aggregation population size, and quadratic terms for rookery and mixed stock sample sizes. I used the AIC for model selection for models described above (Bolker, 2014). To evaluate the general model estimate error I fit a GLM with a gaussian distribution. As the relationship between the model estimate error and other variables was already established during the absolute model estimate error analysis, for the general model estimate error I only evaluated how model estimates varied across both source populations and mixed stock aggregations sample sizes. Lastly, aiming to quantify the level of similarity between haplotype frequency at rookeries (i.e., how difficult it is for models to distinguish between the source of individuals) I calculated the amount of overlap of haplotype frequencies between the different rookeries within each scenario using the *overlapping* package. All statistical analyses were conducted in software R using RStudio (R Core Team, 2016; RStudio Team, 2020).

Results

Credible interval width

The model with the lowest AIC for the three different haplotype distribution scenarios contained the quadratic interaction between sample size for both rookeries and mixed stock aggregations

(Table 3.2 – CredibleIntervalWidth ~ MixedStockSampleSize + MixedStockSampleSize^2).

Credible interval width decreased with an increase in mixed stock aggregations sample sizes for all scenarios, stabilizing the CIW around 15-20 % when mixed stock aggregation samples for *distinct* and *similar* scenarios are > 100 (Figure 3.3, Figure 3.4, Appendix Figure A.1, and Appendix Figure A.2) and > 150 for the *shared* scenario (Figure 3.5, Appendix Figure A.3). Even though the interaction between sample size in rookeries and mixed stock aggregations is important (as indicated by model selection), mixed stock aggregation sample size is a stronger predictor of credible interval width than rookery sample size for all scenarios (Figure 3.3, Figure 3.4, and Figure 3.5).

Table 3.2: Model selection table detailing the four highest ranked models for each scenario used to evaluate the relationship between credible interval width (CIW) and the different variables used in the simulations. MS_SSize = Mixed stock aggregation sample size, SP_SSize = Source population sample size, MS_PopSize = Mixed stock aggregation scaled population size, SP_PopSize = Source population scaled population size, dAICc = Delta AICc, df = degrees of freedom.

Model	dAICc	df	Weight	Scenario
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Distinct*
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	15837.4	7	<0.001	*Distinct*
CIW ~ MS_SSize * SP_SSize	15857.7	5	<0.001	*Distinct*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	15865.2	8	<0.001	*Distinct*
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Similar*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	20298	8	<0.001	*Similar*
CIW ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	20303	6	<0.001	*Similar*
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	20305	7	<0.001	*Similar*
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Shared*
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	33839	7	<0.001	*Shared*
CIW ~ MS_SSize * SP_SSize	33846	5	<0.001	*Shared*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	33890	8	<0.001	*Shared*
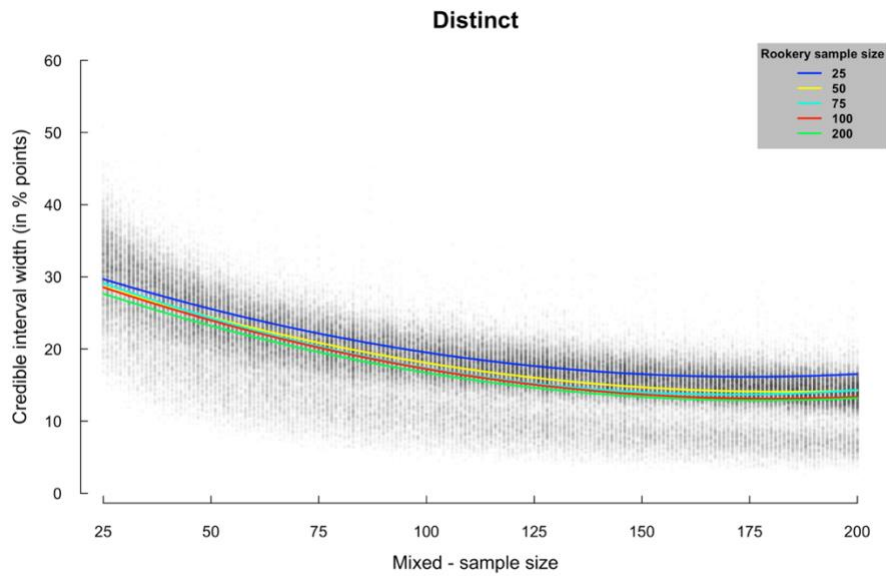
Figure 3.3: *Distinct* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample size. Colors represent fit lines from a generalized linear model with a gaussian distribution for different source population sample sizes.
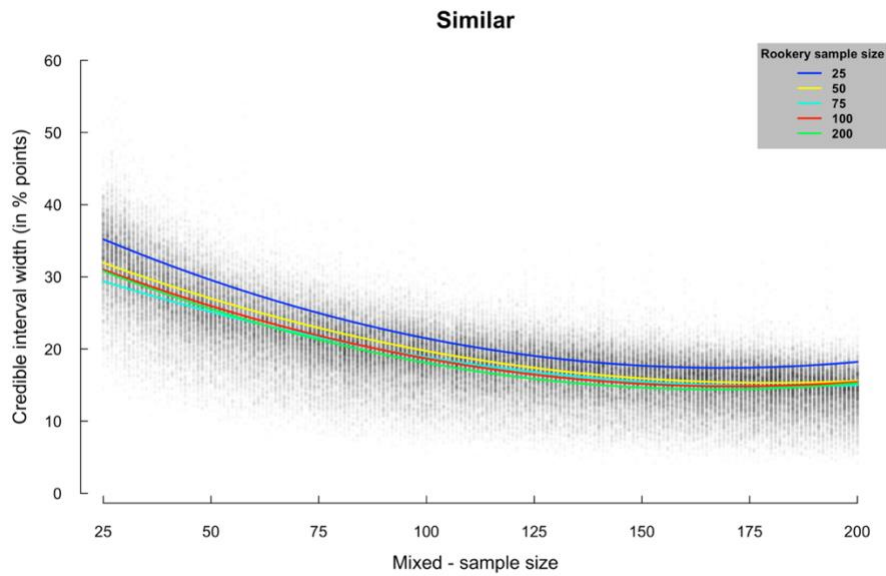
Figure 3.4: *Similar* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample size. Colors represent fit lines from a generalized linear model with a gaussian distribution for different source population sample sizes.

45
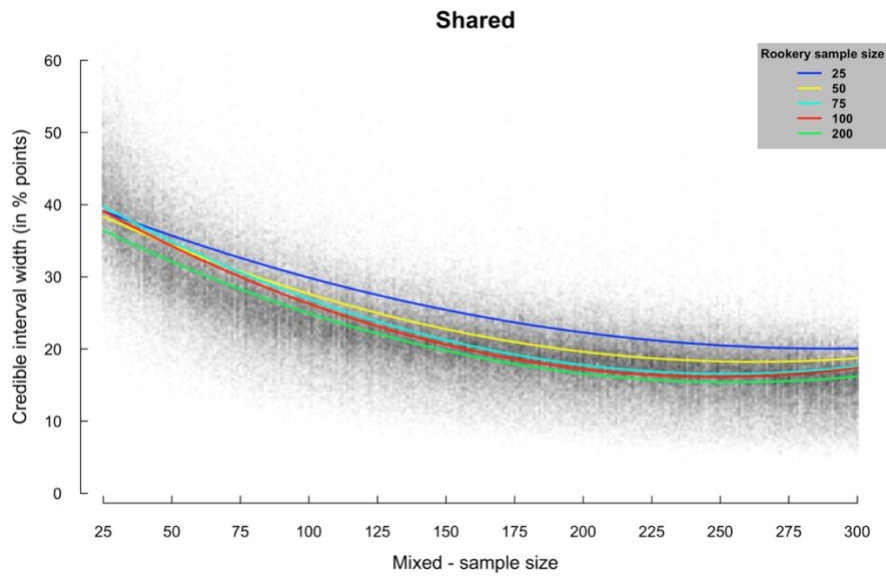
Figure 3.5: *Shared* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample. Colors represent fit lines from a generalized linear model with a gaussian distribution for different source population sample sizes.

The fit lines for all haplotype distribution scenarios have similar shapes (Figure 3.6). with a small difference in CIW between the *distinct* and *similar* scenarios across the entire range of sample sizes. As observed in Figure 3.5, the *shared* scenario contains greater uncertainty at small sample sizes, and requires more samples to achieve a CIW similar to the *distinct* and *similar* scenarios (Figure 3.6). *Similar* and *distinct* curves inflect most strongly around 100 samples, but the *shared* curve inflects most strongly close to 200 samples.
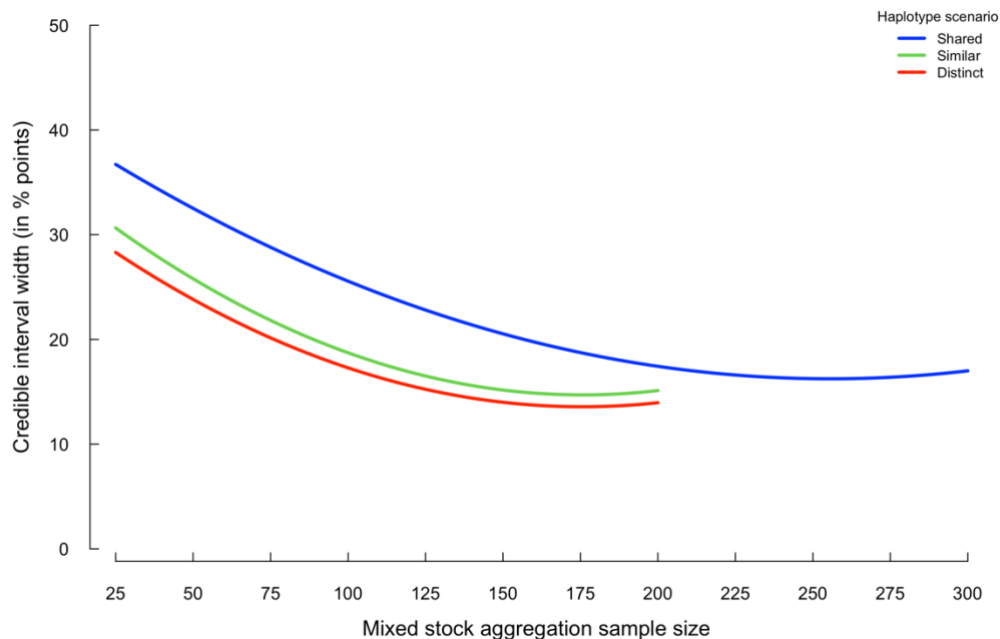
Figure 3.6: Fit lines of credible interval width (in percentage points) as a function of mixed stock aggregation sample size for the different haplotype distribution scenarios. Colors represent fit lines from a generalized linear model with a gaussian distribution for different haplotypic scenarios.

Scenarios with small contributors to mixed stock aggregations (i.e., true contribution $< 5\%$) consistently have lower CIW than source populations with $>10\%$ of true contribution, regardless of the scenario (Figure 3.7). A similar pattern is observed for the *distinct* and *similar* scenarios; however, the trends converge at small sample sizes for the *shared* scenario (Figure 3.7). The fit line for contributors with $> 10\ \%$ of true contribution to mixed stock aggregations is consistent with the general lines observed previously (Figure 3.3 - Figure 3.5). In all scenarios, the curve inflects most strongly around 100 mixed stock aggregation samples and continues to decline more slowly with more samples.
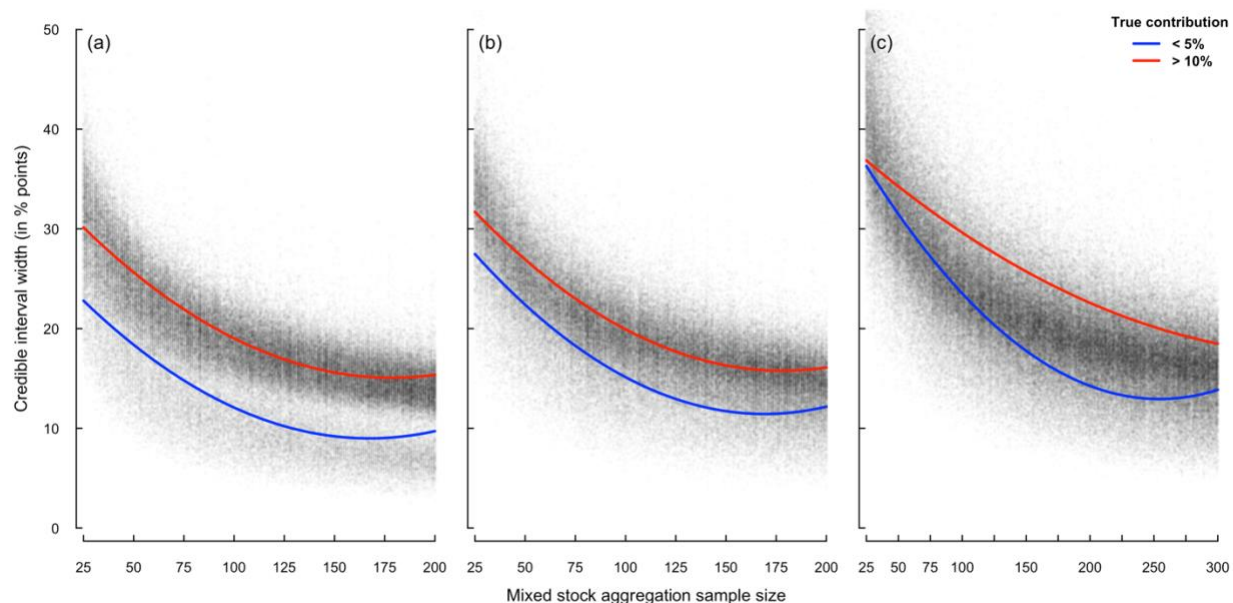
Figure 3.7: Scatter plot of credible interval width (in percentage points) as a function of mixed stock aggregation sample sizes. (a) *distinct* scenario, (b) *similar* scenario, and (c) *shared* scenario. Solid lines represent fit line for models with a subset of the dataset: blue lines represent model estimates with true contribution < 5%; red lines represent model estimates with true contribution > 10%.

Absolute model estimate error

The model with the highest support from AIC for the absolute model estimate error is a quadratic model of mixed stock aggregation sample size (AbsoluteModelEstimateError ~ MixedStockSampleSize + MixedStockSampleSize^2 - Table 3.3). Absolute model estimate error has a negative relationship with mixed stock aggregation sample size, with a similar pattern for all haplotype distribution scenarios (Figure 3.8, Figure 3.9, Figure 3.10, and Appendix Figure A.4, Figure A.5, and Figure A.6). Rookery sample size has no effect on absolute model error estimate alone, and is dependent on the interaction with mixed stock aggregation sample size for a significant change on estimates (Figure 3.8, Figure 3.9, and Figure 3.10).

Table 3.3: Model selection table detailing the four highest ranked models for each scenario used to evaluate the relationship between absolute model estimate error (AMEE) and the different variables used in the simulations. MS_SSize = Mixed stock aggregation sample size, SP_SSize = Source population sample size, MS_PopSize = Mixed stock aggregation scaled population size, SP_PopSize = Source population scaled population size, dAICc = Delta AICc, df = degrees of freedom.

Model	dAICc	df	Weight	Scenario
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Distinct*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	507	7	<0.001	*Distinct*
AMEE ~ MS_SSize * SP_SSize	513	5	<0.001	*Distinct*
AMEE ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	525	8	<0.001	*Distinct*
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Similar*
AMEE ~ MS_SSize * SP_SSize	535	5	<0.001	*Similar*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	536	7	<0.001	*Similar*
AMEE ~ MS_SSize + SP_SSize	542	4	<0.001	*Similar*
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Shared*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	859	7	<0.001	*Shared*
AMEE ~ MS_SSize * SP_SSize	880	5	<0.001	*Shared*
AMEE ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	909	8	<0.001	*Shared*

Figure 3.8: *Distinct* haplotype scenario scatter plot of absolute model estimate error (in percentage points) as a function of mixed stock aggregations sample. Colors represent fit lines from a regression model with a beta distribution for different source population sample sizes.
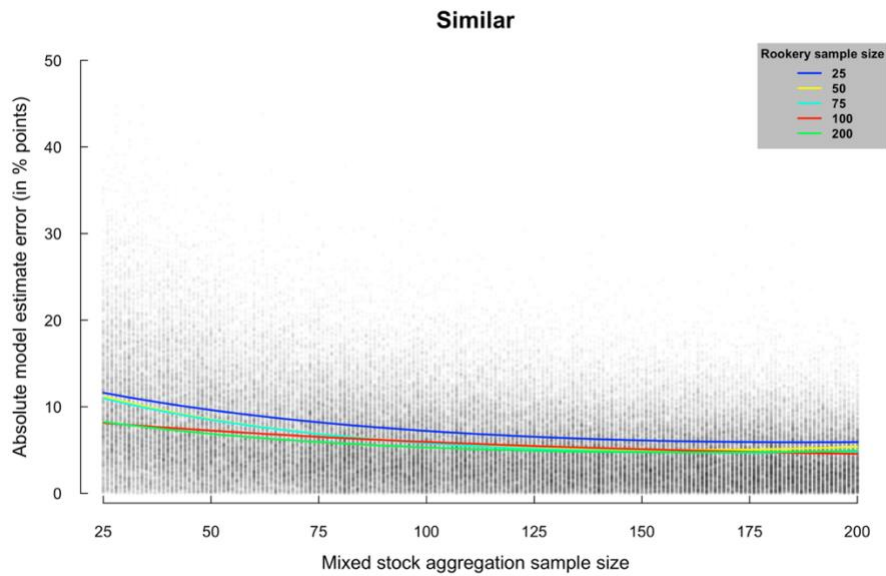
Figure 3.9: *Similar* haplotype scenario scatter plot of absolute model estimate error (in percentage points) as a function of mixed stock aggregations sample. Colors represent fit lines from a regression model with a beta distribution for different source population sample sizes.
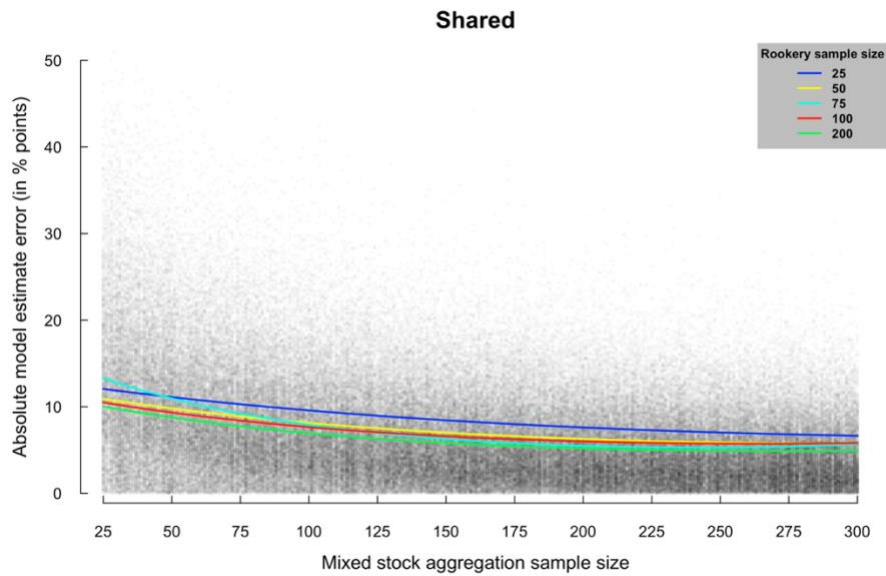
Figure 3.10: *Shared* haplotype scenario scatter plot of absolute model estimate error (in percentage points) as a function of mixed stock aggregations sample size. Colors represent fit lines from a regression model with a beta distribution for different source population sample sizes.

The difference between fit lines for the different haplotype distribution scenarios is not as clear as for the credible interval width (Figure 3.11). The fit lines for the *distinct* and *similar* scenarios are almost indistinguishable, and both lines reach a plateau of accuracy around 5 percentage points when mixed stock aggregation sample size is > 100 samples. A comparable accuracy for the *shared* scenario is reached when mixed stock aggregation sample size is > 200 samples (Figure 3.11).
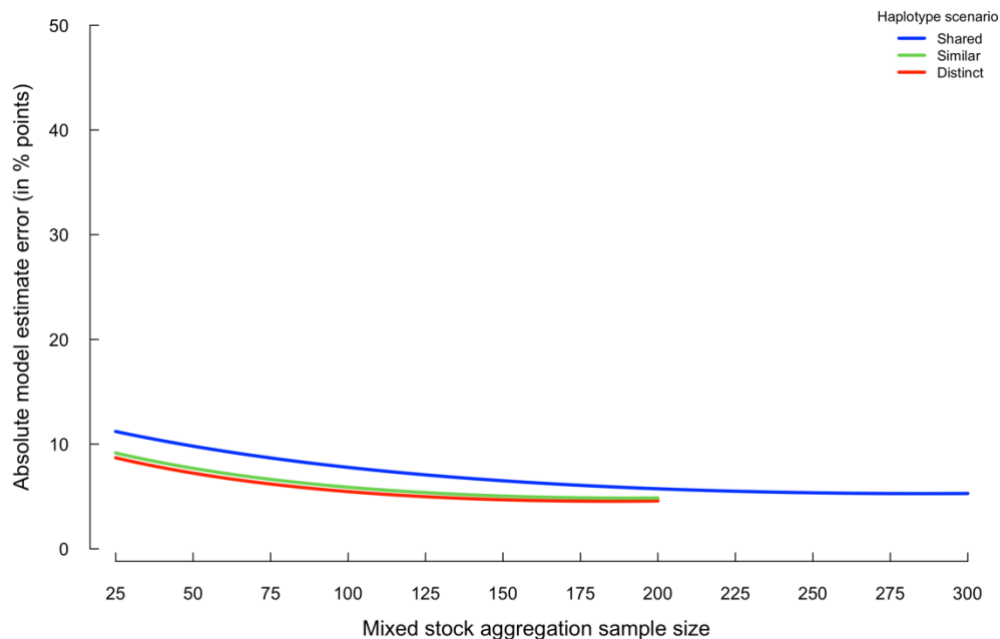
Figure 3.11: Fit lines of absolute model estimate error (in percentage points) as a function of mixed stock aggregation sample size for the different haplotype distribution scenarios. Colors represent fit lines from a regression model with a beta distribution for different haplotypic scenarios.

Even though the scatter plots do not indicate the presence of specific unexplained clusters of data points as in the previous section (Figure 3.3), I evaluated the effect of small contributors (true contribution from a given rookery to a mixed stock aggregation is $< 5\%$) on absolute model estimate error (Figure 3.12), following the same approach conducted for the CIW (used in Figure 3.8 – Figure 3.10). For the *distinct* scenario (Figure 3.12a), absolute model estimate error is similar for both groups when mixed stock aggregation sample sizes are small (< 50), absolute estimate error decreases at a faster rate for the small contributing rookeries (blue line), and is consistently lower for sample sizes > 50. For the *similar* and *shared* scenarios (Figure 3.12b and Figure 3.12c, respectively), absolute model estimate error is greater for the small contributors than the larger contributors when mixed stock aggregation sample size is small, and the two groups have very similar estimate errors when sample size is > 75 for the *similar* scenario and > 100 for the *shared*

52

scenario. In general, the relationship between blue and red fit lines is alike for all scenarios, the main difference is that the *shared* and *similar* scenarios are shifted upwards towards greater uncertainty for small sample sizes. Absolute model estimate error fit lines for rookeries with > 10% of true contribution (red lines) for all scenarios follow the same pattern observed in Figure 3.11.
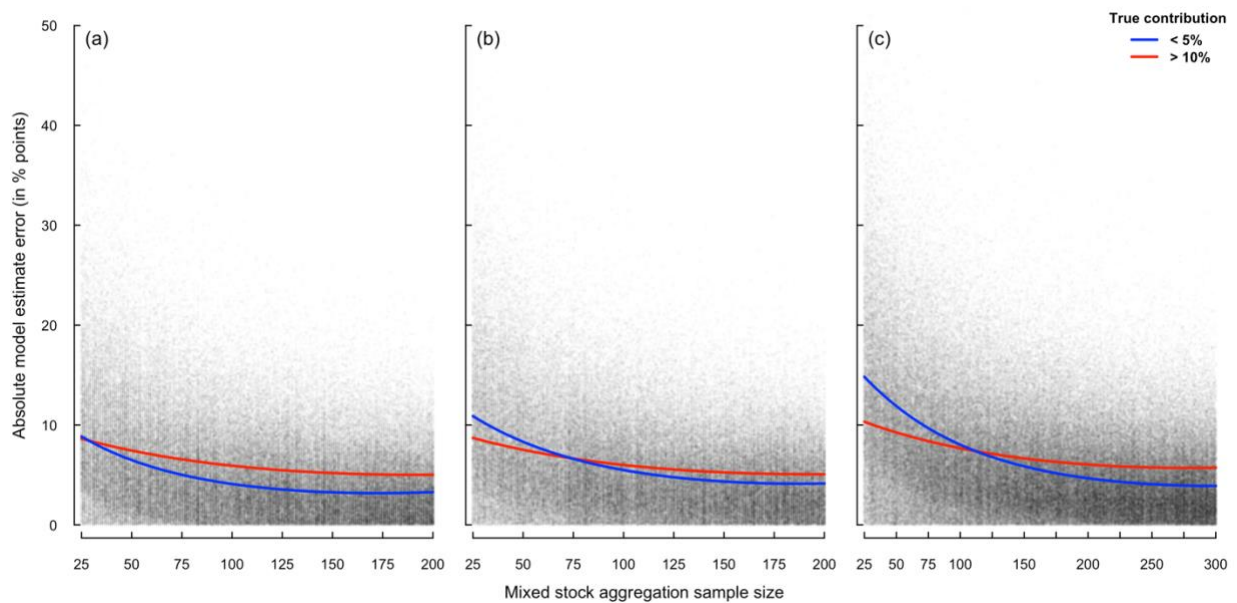
Figure 3.12: Scatter plot of absolute model estimate error (in percentage points) as a function of mixed stock aggregation sample sizes. (a) *distinct* scenario, (b) *similar* scenario, and (c) *shared* scenario. Solid lines represent fit line for models with a subset of the dataset: blue lines represent model estimates with true contribution < 5%; red lines represent model estimates with true contribution > 10%.

General model estimate error

Results indicate that model estimates are overall not biased by sample sizes for all three scenarios tested (Figure 3.13, Figure 3.14, and Figure 3.15). However, contributions from rookeries to mixed stock aggregations tend to be overestimated for small contributors and are underestimated for

contributors with > 10% of true contribution (Figure 3.16). Overestimation for the small

contributors is smaller for the *distinct* scenario and greater for the *shared* scenario (Figure 3.16).
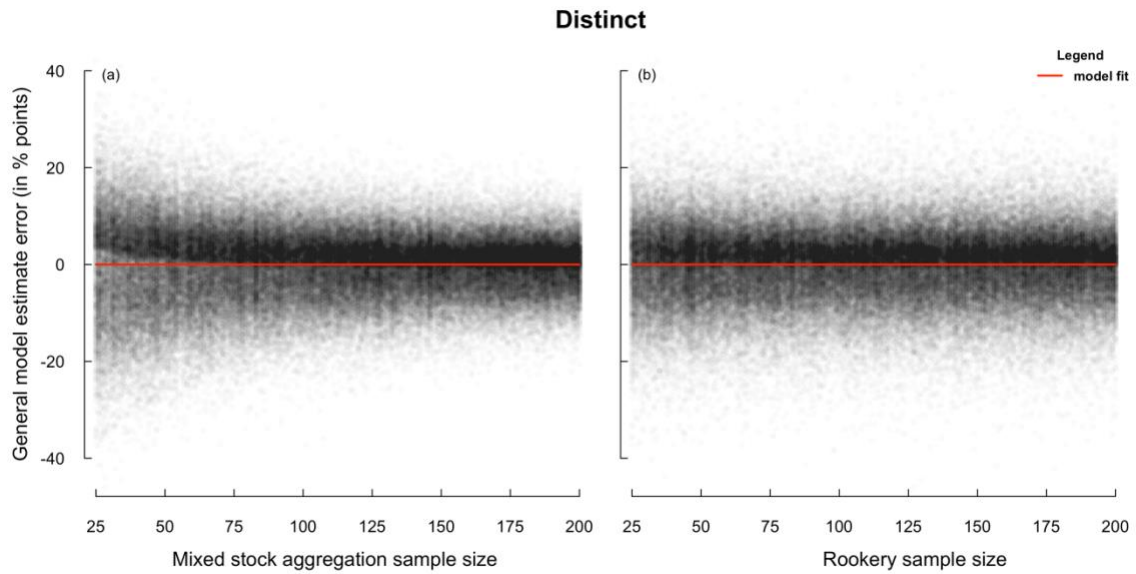
Figure 3.13: *Distinct* haplotype scenario scatter plot of general model estimate error (in percentage points) as a function of (a) mixed stock aggregation sample sizes and (b) rookery sample size. Solid lines represent fit line for a linear regression model.
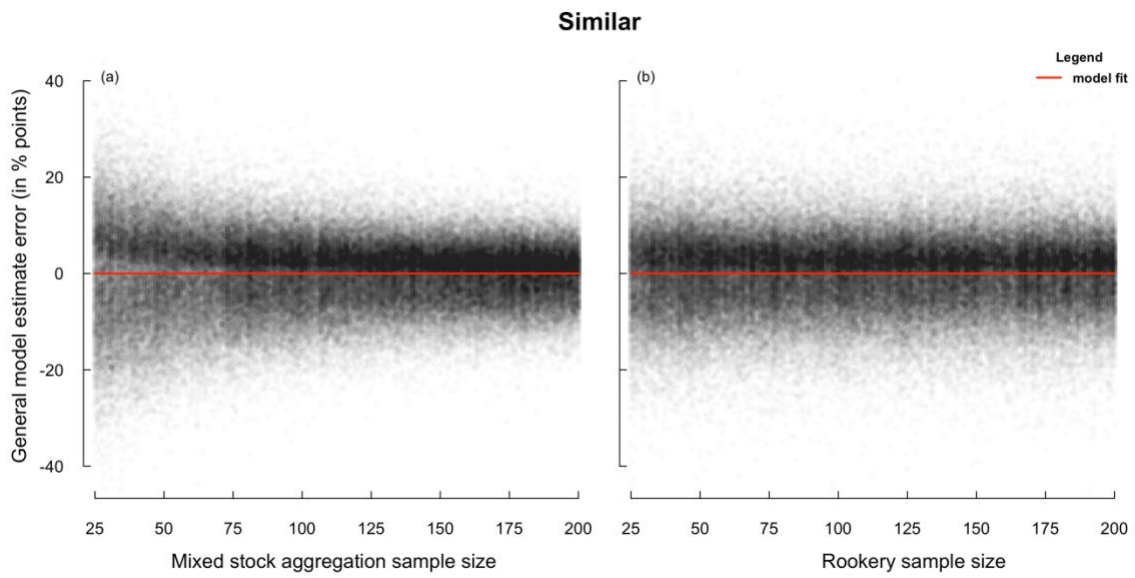
Figure 3.14: *Similar* haplotype scenario scatter plot of general model estimate error (in percentage points) as a function of (a) mixed stock aggregation sample sizes and (b) rookery sample size. Solid lines represent fit line for a linear regression model.
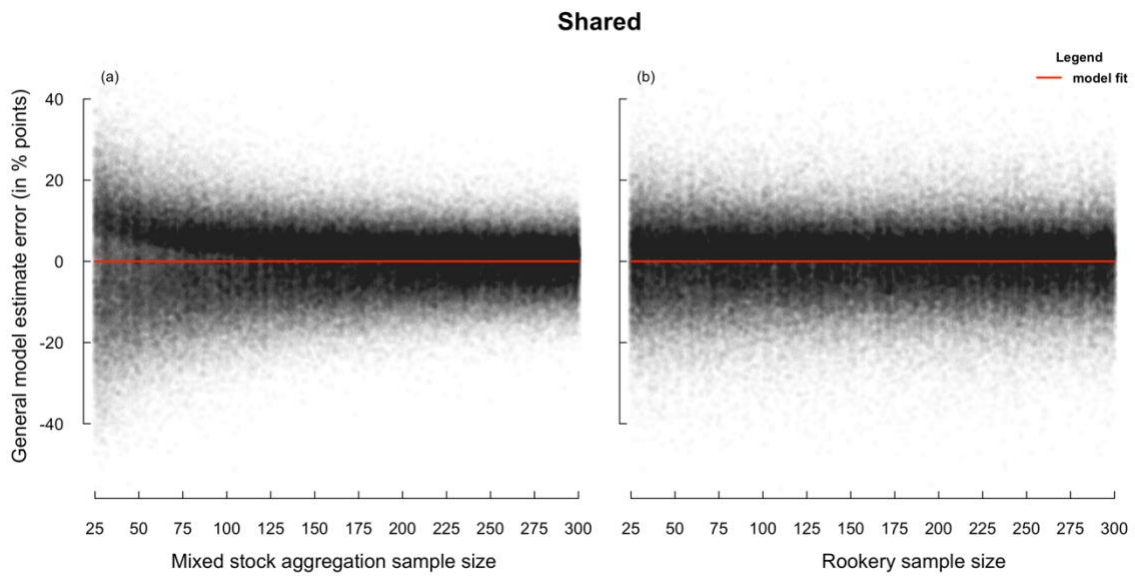
Figure 3.15: *Shared* haplotype scenario scatter plot of general model estimate error (in percentage points) as a function of (a) mixed stock aggregation sample sizes and (b) rookery sample size. Solid lines represent fit line for a linear regression model.
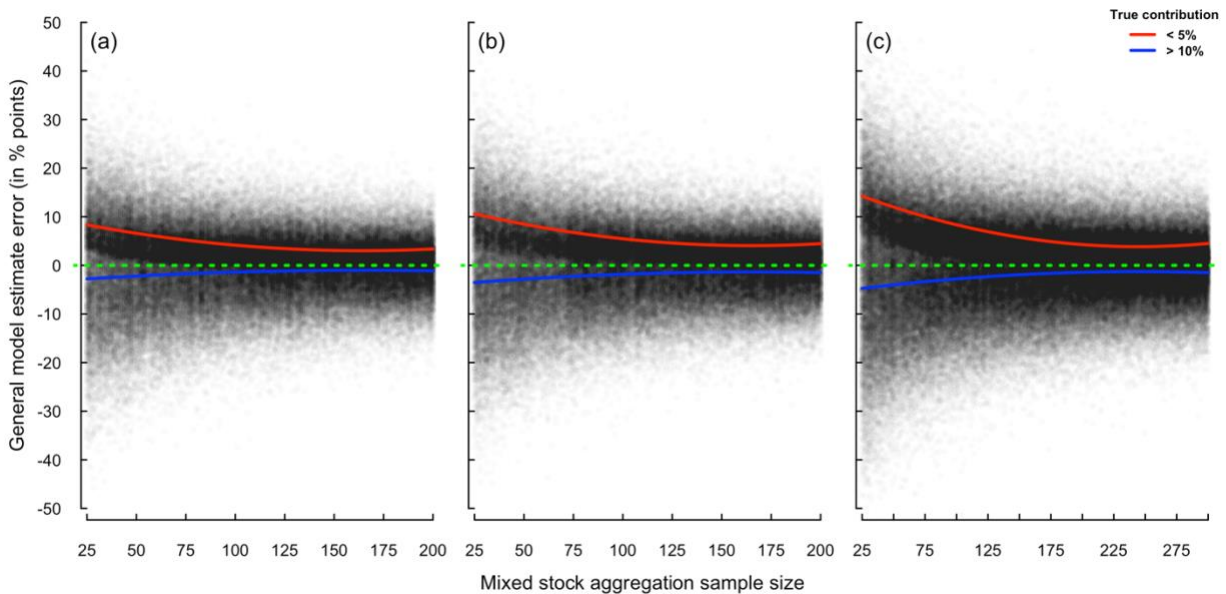
Figure 3.16: Scatter plot of general model estimate error (in percentage points) as a function of mixed stock aggregation sample sizes and true contribution from rookeries. Red lines = small contributors with true contribution < 5%; blue lines = contributors with true contribution > 10%; green dotted line represent the zero value for reference. (a) *distinct*, (b) *similar*, and (c) *shared* scenario.

Probability of true contribution fitting inside the credible interval

The logistic regression models indicate an increased accuracy of models with larger sample sizes (Figure 3.17). For all scenarios the probability of true contributions to fall inside the credible interval range is ~60% for small sample sizes, increasing to ~70% with increased mixed stock aggregation sampling. Rookery sample size has a small effect on model accuracy (Figure 3.17).
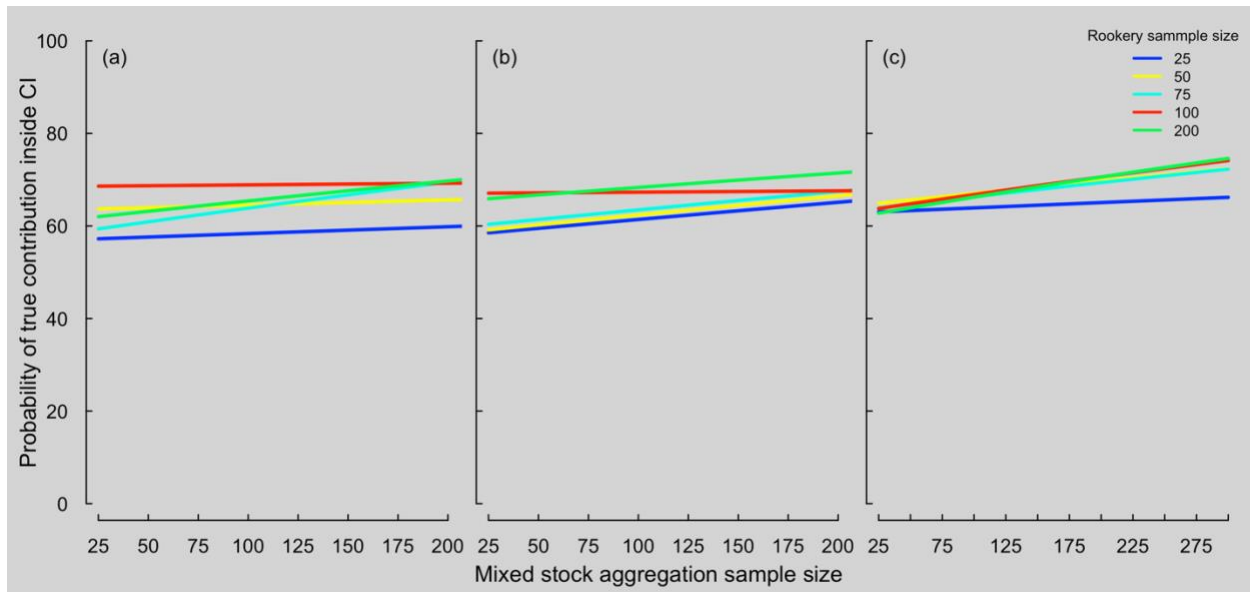
Figure 3.17: Logistic regression models of the probability of the true contribution from source populations to be contained inside the credible intervals as a function of mixed stock aggregation sample size. (a) *distinct*, (b) *similar*, and (c) *shared* scenario. Solid lines represent different rookery sample sizes.

Evaluating the impact of small contributors to the estimates revealed that model accuracy is poor for very small contributors (Figure 3.18). Estimates from small contributors are mostly outside of the credible intervals for small sample sizes in all scenarios (Figure 3.18a-c). For the entire range of sample sizes evaluated, probability of true contribution being contained inside the credible intervals never goes above 60% for the small contributors. For rookeries with > 10% of true contribution to mixed stocks, probability remains around 80% throughout the entire range of simulated sample sizes. Finally, for the true contributions from rookeries to mixed stock aggregations that were not contained inside the credible intervals, I calculated the absolute difference between the true contribution to the edge of the credible intervals (Figure 3.19). We can see that the most points that fell outside of the credible intervals by less than 5 percentage points.
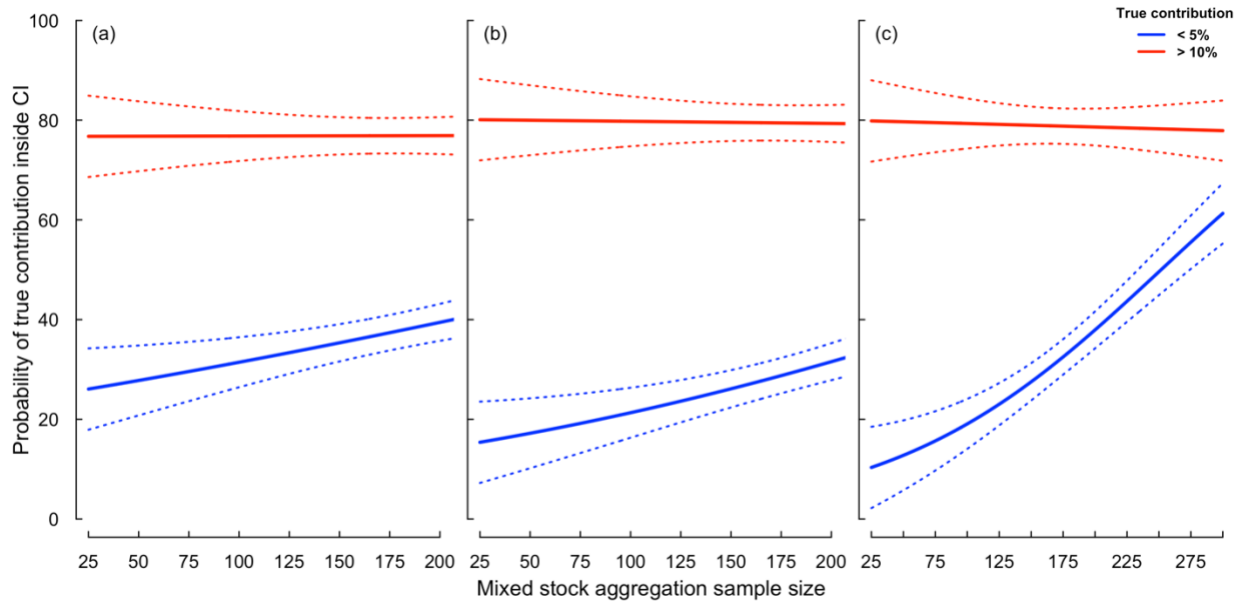
Figure 3.18: Logistic regression models of the probability of the true contribution from source populations to be contained inside the credible intervals as a function of mixed stock aggregation sample size. (a) *distinct*, (b) *similar*, and (c) *shared* scenario. Solid lines represent fit line for models with a subset of the dataset: blue lines represent model estimates with true contribution < 5%; red lines represent model estimates with true contribution > 10%.
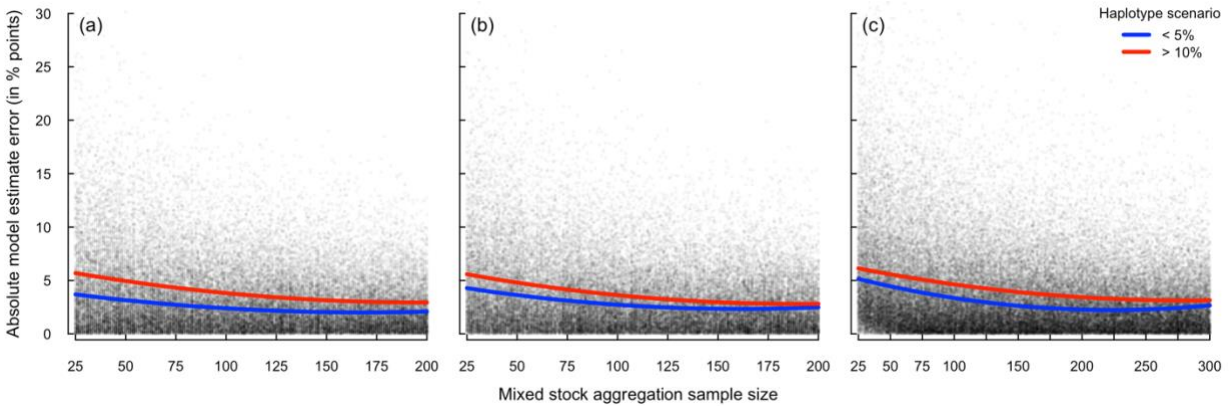
Figure 3.19: Scatter plot of the absolute difference between true contribution from rookeries to mixed stock aggregations and the end of the range of credible intervals as a function of mixed stock aggregation sample sizes for model estimates that the true contribution from rookeries to mixed stock aggregations were not contained inside the credible intervals. (a) *distinct*, (b) *similar*, and (c) *shared* scenario. Blue lines represent model estimates with true contribution < 5%; red lines represent model estimates with true contribution > 10%.

Haplotype frequency overlap

The amount of overlap between rookeries varied considerably between the scenarios (Table 3.4 and Figure 3.2). The *distinct* scenario had an average of 17.7% of overlap, the *similar* scenario had an average of 27.2%, while the *shared* scenario had an average of 45.1%.

Table 3.4: Proportion of overlap between haplotype frequencies between the different rookery haplotype distribution scenario considered in the simulations.

Scenario	Pop1-Pop2	Pop1-Pop3	Pop2-Pop3
Distinct	20.58%	21.01%	11.66%
Similar	25.42%	30.44%	25.78%
Shared	46.14%	48.46%	40.76%

Discussion

Results of this chapter demonstrate the effectiveness of mixed stock models to understand dispersal and composition of mixed stock aggregation sites, even at convoluted haplotype frequency scenarios. There are some caveats to using these models, and here I discuss how using more refined molecular markers and larger sample sizes can influence estimates, model accuracy, and ecological inferences.

No previous studies evaluated in depth the impact of sample size on mixed stock analysis using mtDNA markers. The lack of standardized guidelines might be one of the reasons for a wide range of sample sizes used by researchers (Figure 3.1). For instance, in Chapter 2, I used 72 samples from the Indian River Lagoon and 80 samples from Trident Basin for the mixed stock models even though the number of samples I had access to was much larger and I could have used > 150 samples

for each sampling site and period. The decision on the number of samples used was based on a number that seemed reasonable at the time based on the published literature and funding available for Sanger sequencing. Based on the results presented here, researchers need to be aware of the possible bias in their results driven by the sample size used.

Here, I demonstrate that, except for small contributors, true contributions from rookeries to mixed stock aggregations are often contained inside the CIs, and absolute model estimate error and CIW have a negative relationship with mixed stock aggregation sample size. Accurate mixed stock analysis using mtDNA markers can be achieved near 100 samples for *distinct* and *similar* scenarios and 150-175 samples for the *shared* scenario. Around 50-60 samples from rookeries under any scenario regardless of the size of the reproductive population can also provide enough data for mixed stock models. This is an important finding as researchers could focus on generating high quality data (e.g., sequence longer mtDNA fragments and associated microsatellites) from a reasonable number of reproductive individuals and increase sample size from mixed stock aggregations. Results presented here give a clear direction that future studies should prioritize larger samples sizes in mixed stock aggregations to improve model accuracy. It is important to acknowledge that, (i) the number of samples for mixed stock aggregations suggested here might be extremely hard to achieve in some locations or for specific sea turtle life stages, due the logistical challenges to access the habitats where animals live or to even encounter the animals within these habitats (Mansfield et al., 2014; Phillips et al., 2022; Witherington et al., 2012), and (ii) models used here are more complex than linear regressions, which can certainly impact the effect that sample size has on the response variable. Therefore, despite the suggestion of minimum

sample sizes presented here, larger datasets should be encouraged (Ioannidis, 2005; Jenkins & Quintana-Ascencio, 2020).

I call attention to the fact that, even for large sample sizes and scenarios with a clearer distinction between rookeries (i.e., *distinct* scenario), the true contribution from rookeries to mixed stock aggregations was contained inside the 95% credible interval around 80% of the times (Figure 3.18). Figure 3.19 shows how close to the credible interval most of the points were, though it is unclear why or which variable could be the cause of this issue. A possible solution for this inaccuracy could be to use 99% credible intervals instead of 95% to ensure most of those true contribution points to be included inside the credible intervals in future models.

I also demonstrate the importance of using higher resolution markers for mixed stock analysis, especially in the context of sea turtle assessments. Comparing results between the *shared* and *similar* scenarios indicate that the resolution of the genetic marker used can help reduce the number of samples required for accurate estimates. I highlight that the proportion of overlap between haplotype frequencies used in this Chapter (Table 3.4) is merely a measurement of similarity between rookeries. Other researchers can use metrics of overlap to compare how convoluted haplotype frequencies relate to the scenarios simulated here. Shamblin et al., (2017) used a longer mtDNA fragment and developed extra mitochondrial microsatellites to refine differentiation between rookeries in Florida and Mexico for green turtles captured in the Gulf of Mexico. Several sea turtles rookeries have only been assessed for the short (417 bp) mtDNA fragment, while many mixed stock aggregations have generated data using long mtDNA fragments (e.g., Hancock et al., 2019; Naro-Maciel, Gaughran, et al., 2014; Patrício et al., 2017; Phillips et al., 2022; Stahelin et

al., 2022). This inconsistency between shorter fragments for rookeries and longer fragments for mixed stock aggregations hampers studies' capacity for broader and accurate estimates of dispersal between areas. As a result, researchers normally must decide to either a) consider more rookeries as possible sources at the expense of using a poorer molecular marker, or b) evaluate fewer rookeries with better markers (Phillips et al., 2022; Shamblin et al., 2017; Stahelin et al., 2022). The reason for such mismatch might be that most sea turtle research is conducted by separate groups with their own specific short-term goals, questions to answer, conservation issues to address, and often a limited budget for molecular analyses. Even though mtDNA seems like the logical choice of molecular marker for sea turtles given their philopatric behavior, other markers could be used for mixed stock analysis with potentially greater resolution between areas. Mixed stock analysis in groups other than sea turtles tend to use either multiple loci (e.g., 82 autosomal loci in gray whales; Brüniche-Olsen et al., (2018)), or large sample sizes and multiple loci (e.g., 180 SNP loci across 9991 individuals from 116 populations of steelhead trout; Hess et al., (2016)). The use of other molecular markers for mixed stock analysis could more easily increase differentiation between rookeries and improve model estimates and refine designation of Distinct Population Segments or Regional Management Units for the species.

An interesting finding from these simulations is the small difference between absolute model estimate error for the *shared* and *distinct* and *similar* scenarios (Figure 3.11). Even though there is greater uncertainty around estimates for the *shared* scenario, as seen in CIW (Figure 3.6), the absolute model estimate error becomes consistently smaller and reaches similar accuracy levels as the other scenarios with increased sample sizes (> 150-175 samples). Moreover, for rookeries with > 10% of true contribution to mixed stock aggregations, the probability of true contribution falling

inside the credible intervals is consistently high around 80% (Figure 3.18). These results indicate the robustness of the many-to-many approach, even under extreme conditions as the *shared* scenario.

The bias introduced by small contributors to the analysis is observed in Figure 3.7, with a clear separation of point estimates from the remaining of the results. One hypothesis for a lower credible interval for small contributors is the proximity of the estimates to zero. Credible intervals for these estimates are necessarily bounded by zero, leading to tighter credible intervals. If this hypothesis is correct, the same pattern would be observed on estimates for rookeries with true contribution > 95%, but the simulation code constrained the proportion of the contributions from rookeries to range from ~2% to ~75%. However, the absolute model estimate error associated with small contributors (Figure 3.12) indicates that the mean estimate is farther from the true contribution for small contributors especially in the *similar* and *shared* scenarios (Figure 3.12b and c), suggesting that this hypothesis might not be correct Figure 3.18.

Another explanation for the bias introduced by small contributors would be the presence of haplotypes in mixed stock aggregation samples that are commonly found in small contributing rookeries (i.e., a sampling bias). Mixed stock haplotype frequencies could be skewed towards common haplotypes found in the small contributing rookery especially when the sample size is small, simply due to chance. Mixed stock models appear to often overestimate rookery contribution from small contributors likely by assuming that the mixed stock sample size is truly representative of the mixed stock aggregation pool of haplotypes. Bolker et al., (2003) built into the mixed stock model framework ways to account for biased sampling, especially the ones related

to occurrence of rare haplotypes, but data presented here might be an indication that other cases of bias also affect results. This hypothesis seems to better explain the pattern observed for credible interval width, absolute model estimate error, and probability of contribution inside credible intervals observed.

Importantly, as contributions to mixed stock aggregations must add to 100%, the bias introduced by small contributors has, inevitably, a cascade effect on all other estimates. This is clearly observed on Figure 3.15 and Figure 3.16: there is a consistent overestimation of contributions from small contributors, which is compensated by underestimating the contributions other source populations, resulting in the overall null trend seen on Figure 3.15. Three solutions exist to reduce bias introduced by small contributors to mixed stocks: (a) incorporate distance between mixed stock aggregations and source populations into many-to-many mixed stock models (per Stahelin et al., 2022); (b) increase mixed stock aggregation sample sizes to reduce the risk of biased sampling (Figure 3.7 and Figure 3.12); and/or (c) remove small contributors/source populations from analyses all together (Engstrom et al., 2002; Jensen et al., 2020). In most cases, it might be difficult or almost impossible to identify *a priori* small contributors. For instance, in Chapter 2 I decided to remove all rookeries from the south Atlantic and West Africa to reduce possible noise introduced by unlikely contributors, even though the model introduced in Chapter 2 considering distance between sites could help reduce such noise. It is important to note that Bolker et al., (2003) evaluated how well Markov Chain Monte Carlo (MCMC) estimation can handle rare haplotypes from source populations, but the case of small contributors simulated here seem to escape the robust approach implemented in the many-to-many mixed stock analysis. The reason for this could be due to the occurrence of common haplotypes from the small contributing population in the

mixed stock aggregation sample (biased sampling), which is different than having rare haplotypes; a condition that can be handled by the MCMC approach (Bolker et al., 2003).

Simulations presented in this chapter only consider scenarios with 19 haplotypes. Mixed stock models treat haplotypes that occur in the same areas as a single haplotype (e.g., if haplotypes A, B, and C are only found in rookery X and mixed stocks Y and Z are treated as haplotype D – see Bolker et al., (2007)). Even though more haplotypes available could theoretically mean increased capacity to distinguish rookeries, as markers providing the same information are combined, the number of markers used by the model can be much lower. Therefore, a different number of haplotypes available does not necessarily mean that the simulation results would drastically change. More importantly, sea turtle rookeries can be considered less complex from a haplotype frequency standpoint than the mixed stock aggregations, as philopatric behavior creates similarities among individuals from the same reproductive area. For this reason, it makes sense for fewer samples to be required from rookeries to achieve accurate estimates.

A possible caveat is the impact of the number of source populations on estimates, as all simulations presented here had three source populations. Larger sample sizes could be required for models with more source populations or with more complex scenarios than the ones simulated here. For instance, there might be a minimum number of samples required based on the number of possible source populations. Considering the *shared* scenario as an example: what if we added a fourth source population that overlapped only a small proportion of haplotypes with the three other convoluted populations? Results from the *distinct* and *similar* scenarios indicate that the mixed stock model is robust enough to effectively identify contributions from this 4th population. Still,

65

future studies should address this question to determine the impact of sample size required if more source populations are added to the dataset.

Conservation implications and recommendations

Understanding the level of error involved in estimates based on the sample size used for analysis is of critical importance for researchers and managers to design future experiments. Proper conservation planning and management plans of imperiled species rely on accurate statistical models. Therefore, it is of maximum importance for studies of dispersal and connectivity between populations across ocean basins to provide reliable and accurate estimate to managers.

Based on results here, I would strongly recommend the sea turtle research community to work together on reassessing the genetic haplotypes frequencies of rookeries using the higher resolution mtDNA markers available for each species. This action alone would allow future studies to move away from the *shared* scenario simulated here towards the *similar* or *distinct* ones, improving quality of estimates and interpretation of results without the need of prohibitive sample sizes. Second, to reduce costs and allow larger number of samples to be analyzed, I would recommend different groups to either work together in laboratory steps or use larger sample sizes especially for mixed stock aggregations. Third, researchers should consider using the 99% credible intervals instead of 95% credible interval to ensure the true contribution from rookeries to mixed stock aggregations is contained within the credibility interval. Fifth, the most important recommendation from this simulation, is for future studies to use large sample sizes especially in mixed stock aggregations. Sample size above 150 samples from each mixed stock aggregation evaluated should be the goal for future sea turtle mixed stock studies using mtDNA markers. Model accuracy and,

consequently, ecological inferences about the dynamics between rookeries and mixed stock aggregations can be greatly improved with larger number of samples. Finally, I would recommend that governmental and international agencies coordinating funds for sea turtle research to support the use of the most advanced molecular methodologies available to date, seeking to improve our knowledge regarding population structure and dispersal patterns.

CHAPTER 4: GENETIC STRUCTURE AND DEMOGRAPHIC INFERENCE OF THE GREEN SEA TURTLE FROM THE EAST COAST OF CENTRAL FLORIDA

Introduction

Population genomics is a powerful tool to inform conservation actions and management plans for species of interest (Funk et al., 2019). Imperiled species that have exhibited population declines face an increased risk of inbreeding depression and reduced genetic diversity. This loss of diversity in contemporary time can lead to an increased risk of extinction. For example, a reduction in population size led to increased inbreeding for the Glanville fritillary butterfly (*Melitaea cinxia*), which increased the extinction risk for this species (Saccheri et al., 1998). Similarly, the southern dunlins (*Calidris alpina schinzii*) fitness was impacted by reduced genetic diversity (Blomqvist et al., 2010). Evolutionary processes and threats from human development can trap species in an extinction vortex that can be caused by a drastic reduction or fragmentation of populations, low genetic diversity, genetic drift, and decreased population fitness (Gilpin & Soulé, 1986). Though incipient, conservation planning and monitoring using genomic data can greatly improve our understanding of long-term and complex ecological processes for wild populations (Bernos et al., 2020; Forester et al., 2022; Funk et al., 2019; Nielsen et al., 2020).

Maintenance of genetic diversity is fundamental for species' survival and reduction of extinction risk in wild populations (Frankham, 2005). Populations may experience a decrease in genetic diversity through geographic isolation and/or bottleneck events (Hoelzel et al., 2002). For instance, the northern elephant seal population had a decrease in genetic diversity and fitness following a strong bottleneck (Hoelzel et al., 2002). Prolonged population isolation could lead to speciation,

unless there is sufficient gene flow between one or more populations (Ferrière et al., 2004). Gene flow between populations is a crucial factor in both preventing reduction in population size and loss of genetic diversity over time. Long-term monitoring of genetic parameters and assessment of historical demographic variations in wild populations is important for the conservation of imperiled species.

Sea turtles are a group of species with a philopatric reproductive behavior, in which individuals, after dispersal and developmental stages, return to their natal populations for reproduction (Bowen et al., 1992). Despite female strong reproductive site-fidelity, previous studies indicate low to moderate levels of male-mediated gene flow between reproductive areas (Karl et al., 1992; Roberts et al., 2004; Roden et al., 2013). Foraging habitats for sea turtles, unlike reproductive areas, are composed of individuals from multiple genetically distinct populations (Bjorndal & Bolten, 2008; Bowen et al., 1996; Phillips et al., 2022; Shamblin et al., 2012; Stahelin et al., 2022; van der Zee et al., 2019). Even though sea turtles spend most of their lives in the ocean, most research and conservation actions are beach-centered (Bjorndal, 1999). As a result, management strategies often rely on data from reproductive population to delineate management units or distinct population segments (Seminoff et al., 2015; Wallace et al., 2010).

A recent analysis indicate that the green sea turtle (*Chelonia mydas*) mixed stock aggregations in east coast of central Florida, United States, are composed of individuals from multiple nesting sites throughout the Greater Caribbean (Stahelin et al., 2022). The rookeries identified by Stahelin et al., (2022) as the main contributors to east central Florida mixed stocks are part of the North Atlantic Distinct Population Segments (DPSs), which are going through considerable demographic

growth during the past few decades (Chaloupka et al., 2008; Stahelin et al., 2022; Wallace et al., 2010). Population genomics in mixed stock aggregations can enhance our understanding regarding genetic differentiation between populations within the Northwest Atlantic RMU and future management strategies for this species. The goals of this Chapter are to (i) determine if the number of genetic clusters identified using genomics is similar to the number of source populations Stahelin et al., (2022) identified using mtDNA, (ii) determine how the predominance of genetic clusters varied over time in each of the sampling sites, and (iii) estimate historical demographic parameters for each of the genetic clusters.

Methods

Study site and data collection

I used samples collected from the main reproductive site for green turtles in Florida: at the Brevard County portion of the Archie Carr National Wildlife Refuge, in Melbourne Beach (MB), Florida, USA (Ehrhart et al., 2014). In addition, I also used samples from two mixed stock aggregations in the east coast of Central Florida: the Indian River Lagoon (IRL) site south of the Sebastian Inlet (Ehrhart et al., 2007), and the Trident Submarine Basin (TRID) site at Port Canaveral (Redfoot & Ehrhart, 2013). Biological samples were collected under permits NMFS 19508, MTP-231, and their predecessors. All specimens used in this study were collected in accordance with animal care and use protocols approved by the Institutional Animal Care and Usage Committee at the University of Central Florida (IACUC 2020-04, 2020-18, 2020-138, and their predecessors).

The sampling scheme is the same used in Stahelin et al., (2022): two sampling periods for each site. In short, samples collected from MB (rookery) were grouped based on the first year of capture

of the turtles: from 1985 to 1999 (old) and from 2016 to 2018 (new). For in-water (mixed stock aggregations) data, samples from IRL and TRID were grouped to each sampling period if any of the (re)capture dates occurred from 2003 to 2005 (old) or from 2016 to 2018 (new). I used 60 samples from nesting females (n = 30 per sampling period) and 240 samples from juveniles (n = 60 per sampling period per site). Animals were captured, measured, and tagged following standardized methods for these sites (Ehrhart et al., 2007; Phillips et al., 2021; Redfoot & Ehrhart, 2013). To ensure individual identification, each turtle was tagged with an Inconel tag in each front flipper and, since 1999, with a passive integrated transponder tag. All animals had their straight carapace length (SCL) from the nuchal notch to the tip of the longest pygal scute (Bolten, 1999). Skin and blood samples were collected and stored as described in Stahelin et al., (2022).

DNA extraction and SNP generation

I extracted genomic DNA using a Qiagen DNeasy blood and tissue kit following the manufacturer's protocol or a Serapure Bead method (Rohland & Reich, 2012) with adaptations (Faircloth & Glenn, 2016). We used a double-digest restriction enzyme protocol (Peterson et al., 2012) to generate single nucleotide polymorphisms (SNPs) with SbfI-HF (NEB R3642) and Sau3AI (NEB R0169). I size-selected fragments ranging from 550 to 900 bp using a Pippin Prep (Sage Science) and sequenced pooled libraries of 144 individuals in two lanes of paired-end 150 bp reads in Illumina HiSeq 2500 or HiSeq 4000.

I assessed the quality of RADseq sequencing files using FastQC (Andrews, 2010). To demultiplex individuals, I used 'process_radtags' in *Stacks* v2.41 (Catchen et al. 2013) and aligned sequences to a reference genome (Bentley et al., 2022) using Bowtie 2 (Langmead & Salzberg, 2012) and

SAMtools (Li et al., 2009). I executed the pipeline in software *Stacks* using the *ref_map.pl* wrapper and used *Populations* to keep only one random SNP per loci (Catchen et al. 2013). I removed SNPs with missing genotype rate data > 10% (-geno 0.1), individuals with a missing rate of genotype data > 10% (-mind 0.1), and SNPs with minor allele frequency < 1% (-maf 0.01) using PLINK v1.9 (Purcell et al., 2007). Using the list of filtered SNPs (whitelist) and samples, I generated a new population map and used the whitelist of loci for calculations of population-wide statistics in *Populations* (Catchen et al. 2013).

Data analyses

To identify the number of genetic clusters, I grouped all individuals from MB as one population and all in-water samples as a second population. I used the software ADMIXTURE v1.3 (Alexander et al., 2009), with a maximum likelihood approach to estimate ancestries to identify possible genetic clusters (K = 1-8; the maximum number of K was selected using 2 + the 6 likely sources identified by MSA in Chapter 2) within all samples. I evaluated 2 extra Ks in case evaluation using genomic data could be able to identify additional clusters than the ones suggested by mtDNA data. I used three different clustering algorithms to evaluate consistency in results using different approaches. The first method used was a combination of a Discriminant Analysis of Principal Components (DAPC) and Principal Component Analysis (PCA) implemented in the R package *Adegenet* v2.1.8 (Jombart, 2008; R Core Team, 2016) to detect potential genetic clusters within all samples. For DAPC and PCA, all missing genotype data were completed with the mean allele frequencies. I adjusted the final number of principal components using the a-score and used two discriminant functions. Secondly, I used software STRUCTURE v2.3.4 (Pritchard et al., 2000), a Bayesian-based algorithm with 100,000 iterations of burn-in followed by 500,000

iterations for the posterior distribution. To select the best K for the STRUCTURE results, I used the Evanno method in CLUMPAK, which evaluates variations on the log probability between sequential pairs of K (ΔK)(Evanno et al., 2005; Kopelman et al., 2015). Lastly, I used software ADMIXTURE, which uses a maximum likelihood approach to estimate ancestries for sampled individuals to determine clusters and cross-validation to select the most probable value of K (number of populations/clusters) (Alexander et al., 2009). I included all principal components available (n = 231) and defined the maximum number of clusters as 8 (same used for ADMIXTURE and STRUCTURE).

I used a Fisher's exact test to evaluate the variation in the proportion of genetic clusters found in each sampling period, and in each sampling period for each site (i.e., if the proportion of individuals from Cluster X in site Y changed between sampling periods). Fisher's exact test was run using function *prop.test()* implemented in software R (R Core Team, 2016).

To estimate historical demographic parameters and divergence times between genetic clusters I created coalescent models using software FASTSIMCOAL v2.7 (Excoffier et al., 2021). I created a series of models simulating different arrangements of divergence between populations, the presence of bottleneck(s) in each of the clusters, different arrangements of gene flow, varying population sizes over time, and models with and without a recent population growth or decline. Each model contained 1,000,000 simulations (-n) to get parameter estimates, used the minor allele frequency (-m) to compute the site frequency spectrum (SFS), and mutation rate of 1.2×10^{-8} substitutions per site per generation previously used in green turtle analysis (Fitak & Johnsen, 2018). The minimum observed SFS (-C) was set to 10, and used maximum composite likelihood

for parameter estimation (-M). As FASTSIMCOAL uses an approximation for the parameters, I ran each model at least 100 times, and used the model with the highest likelihood for comparison between models. Finally, I used a non-parametric bootstrap (block-bootstrap) to estimate confidence intervals. Block bootstrap is done by splitting the SNP data into n equally sized blocks, randomly picking blocks n blocks to recreate a SNP datafile, and running the selected model 100 times to obtain the maximum likelihood. The process of creating a random SNP datafile and running through the selected model 100 times was repeated 50 times. Finally, the estimates of the 50 models with the highest maximum likelihood were combined and confidence intervals were calculated using the *boot* package in R (Canty, 2002).

Results

Out of the 231 individuals used for genomic analyses, the mean SCL at first capture in MB_{old} was 101.6 cm (SD 5.2 cm, range 95.6 – 114.1 cm, n = 16), and in MB_{new} was 97.7 cm (4.9 cm, 88.8 – 109.2 cm, n = 26). For in-water samples, the mean SCL in IRL_{old} was 45.3 cm (9.7 cm, 30.2 – 68.6 cm, n = 48) and in IRL_{new} was 48.4 cm (8.34 cm, 32.4 – 66.7 cm, n = 56), while in $TRID_{old}$ it was 29.6 cm (2.9 cm, 23.4 – 39.2 cm, n = 55) and in $TRID_{new}$ was 31 cm (5.2 cm, 23.7 – 54.9 cm, n = 58).

I trimmed sequences to 135 bp to ensure consistent data quality for base calling across samples. A total of 300 samples were initially processed and sequenced with the ddRAD protocol, and 231 samples passed all data quality checks and filtering steps. The mean depth of coverage was 48.4x (SD = 35.6x, range 4.1x – 170.4x). I recovered a total of 14,731 SNPs that were used for downstream analyses.

Population structure

Pairwise F_{ST} between sampling sites and time-period indicates low genetic differentiation among all sites, ranging from 0.0052 to 0.0114 (Table 4.1). Using a PCA to evaluate possible differences between cluster of individuals regarding their site of capture or sampling period indicated that there was a major overlap between all sites and sampling periods (Figure 4.1).

Table 4.1: Pairwise F_{ST} values between sites and time-periods.

	MB_{new}	IRL_{old}	IRL_{new}	$TRID_{old}$	$TRID_{new}$
MB_{old}	0.0145	0.0101	0.0099	0.0114	0.0103
MB_{new}		0.0076	0.0078	0.0084	0.0082
IRL_{old}			0.0054	0.0059	0.0056
IRL_{new}				0.0055	0.0052
$TRID_{old}$					0.0056
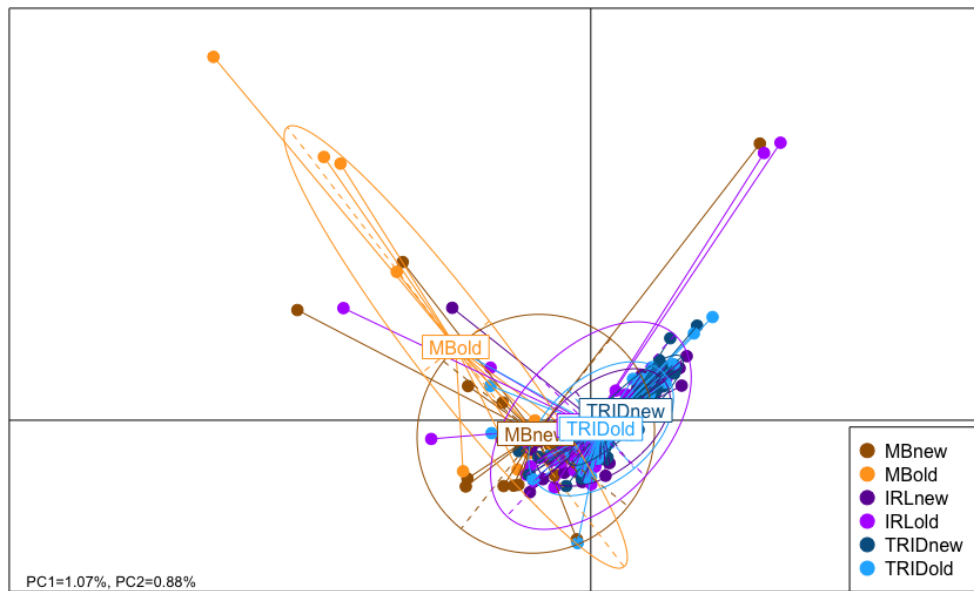
Figure 4.1: Principal component analysis (PCA) with genetic clusters assigned based on sampling site and time-period.

75

I then identified clusters of individuals (function *find.clusters()* from the *Adegenet* package in R), where individuals were grouped into four different clusters (Figure 4.2). Cluster 2 is the most differentiated one, while Clusters 1 and 3 are the most similar (Figure 4.2, Table 4.2). Comparing the sampling site and period to the clusters inferred by DAPC, 63.6% of samples from MB_{old}, fell into Cluster 3, while the remaining 36.4% grouped with Cluster 4 (Table 4.3). For MB_{new}, also had most samples in Cluster 3 (88.5%), 7.7% of individuals in Cluster 4, and only 1 sample (3.8%) grouped with Cluster 2. For in-water samples, IRL_{old} is the only site with individuals in all 4 clusters, with 74.5% of samples falling into Cluster 3, followed by 19.2% in Cluster 1. IRL_{new} had a similar composition with 62.8% of samples in Cluster 3 and 35.3% in Cluster 1, but IRL_{new} has no sample in Cluster 2. Samples from TRID only fell into Clusters 3 and 1, but there was a predominance of individuals from Cluster 3 in $TRID_{old}$ with 69.8%, while Cluster 1 was more abundant in $TRID_{new}$ with 58.5% of samples (Table 4.3).
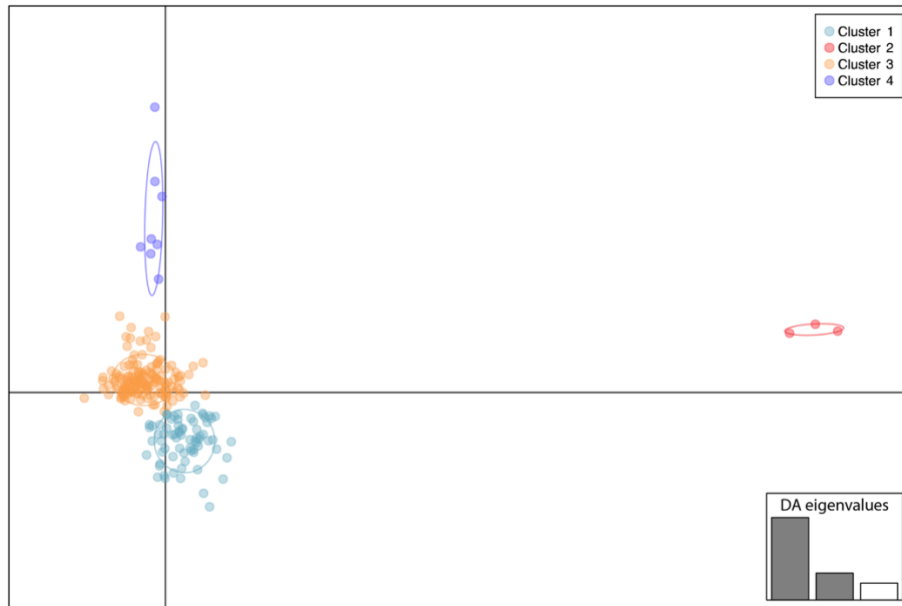
Figure 4.2: Discriminant analysis of principal component using clusters determined by *de* novo approach within samples from all sampling sites and time-periods.

Table 4.2: Pairwise F_{ST} values between inferred *de novo* genetic clusters.

	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0.0126	0.0045	0.0127
Cluster 2		0.0069	0.0858
Cluster 3			0.0058

Table 4.3: Number of samples from each sampling site and time-period versus inferred genetics clusters from *de novo* assignment.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
MB_{old}			7	4
MB_{new}		1	23	2
IRL_{old}	9	2	35	1
IRL_{new}	18		32	1
$TRID_{old}$	13		30	
$TRID_{new}$	31		22	
Total	**71**	**3**	**149**	**8**

Using software STRUCTURE to identify the same clusters as detected using DAPC, I was able to identify only 3 clusters with the Evanno method for best K: Cluster 2 was grouped with Cluster 1 (Figure 4.3). All three clusters identified have some level of admixture with one or both of the other clusters. Following the same rationale and using software ADMIXTURE to identify clusters using clusters from DAPC as populations indicated that samples are mostly admixed (i.e., K = 1 - Figure 4.4) based on cross validation error.

K=1

K=2

K=3
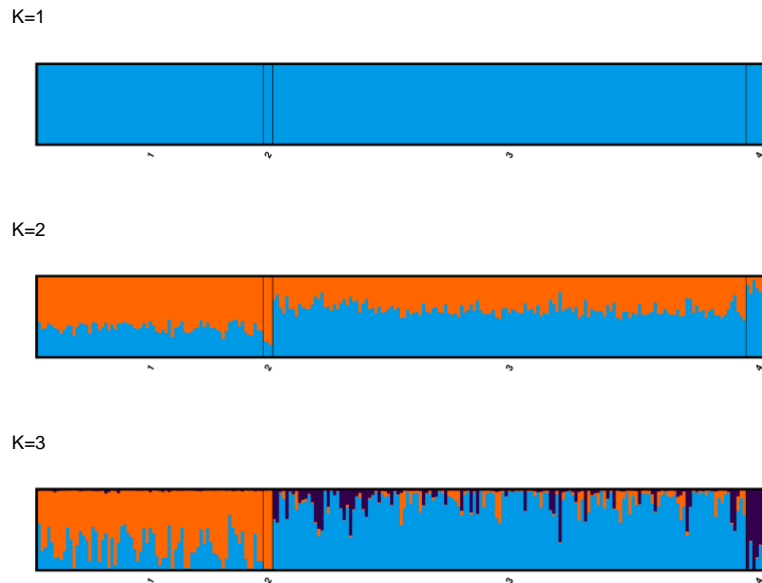
Figure 4.3: Result of the STRUCTURE analysis using *de novo* clusters from the Discriminant Analysis of Principal Component as populations. Values of K indicate the number of populations in each plot. The Evanno-method for K selection indicates K = 3.

K=1

K=2

K=3
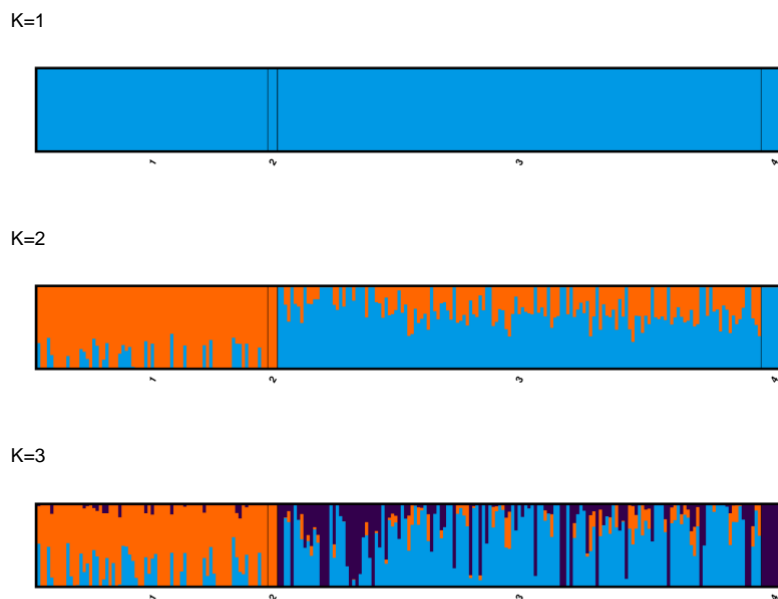
Figure 4.4: Result of the ADMIXTURE analysis using *de novo* clusters from the Discriminant Analysis of Principal Component as populations. Values of K indicate the number of populations in each plot. Cross-validation error indicate K = 1.

Proportion of clusters over time

The proportion of individuals from each genetic cluster is summarized in Table 4.4, along with the results of the Fisher's exact test. I found that Cluster 1 significantly increased in the proportion when all sample were combined, from 21.8% to 37.7% (Table 4.4). Also, Cluster 1 significantly increased in the proportion of individuals from 30.2% to 58.5% while Cluster 3 exhibited a significant decrease in its prevalence at TRID from 69.8% to 41.5%. Variation for all other clusters and sites were not significant.

Table 4.4: Proportion of the clusters identified by the Discriminant Analysis of Principal Components in each sampling period. p-value represents the value obtained from a Fisher's exact test. * indicate significance for $\alpha = 0.05$. ALL = samples from all sampling sites were combined into a single dataset; MB = samples from the Melbourne Beach rookery; IRL = samples from the Indian River Lagoon; TRID = samples from the Trident submarine basin.

| | Sampling period | | | |
Clusters	OLD	NEW	p-value	Site
Cluster 1	21.78%	37.69%	0.01405*	
Cluster 2	1.98%	0.77%	0.8254	
Cluster 3	71.29%	59.23%	0.07824	ALL
Cluster 4	4.95%	2.31%	0.4672	
Cluster 1	0.00%	0.00%	-	
Cluster 2	0.00%	3.85%	1	
Cluster 3	63.64%	88.46%	0.1926	MB
Cluster 4	36.36%	7.69%	0.09399	
Cluster 1	19.15%	35.29%	0.1185	
Cluster 2	4.26%	0.00%	0.4393	
Cluster 3	74.47%	62.75%	0.3033	IRL
Cluster 4	2.13%	1.96%	1	
Cluster 1	30.23%	58.49%	0.01055*	
Cluster 2	0.00%	0.00%	-	
Cluster 3	69.77%	41.51%	0.01055*	TRID
Cluster 4	0.00%	0.00%	-	

Coalescent models

The model with the highest support from maximum likelihood is summarized in Figure 4.5 and Appendix Table B.3. Using the generation time for the Western Atlantic ocean of 38.5 years as a reference (Seminoff, 2004), the split between clusters 1 and 2 (TDI1) happened 31.9 (CI 23.1 – 36.3) Mya, while the split between clusters 1 and 3 (TDI2) was 7.8 (5.2 – 10.0) Mya, and between clusters 1 and 4 (TDI3) was 2.9 (2.3 – 3.9) Mya. Cluster 2 had an increase in population size following TDI1 and has remained with a relatively constant effective population size (Ne) of 526,413 (352,534 – 619,161) individuals since then. Effective population size for cluster 1, on the other hand, grew much larger than cluster 2 following TDI1, reaching 5.6 million individuals (4.2e+6 – 6.8e+6). Cluster 1 had a major reduction in Ne during TDI2 to 1.8 million individuals (1.3e+6 – 2.2e+6), when cluster 3 diverged from cluster 1. Cluster 3 remains with a constant Ne of 506,422 (430,889 – 583,649) since TDI2, while cluster 1 had another reduction in Ne during TDI3, when cluster 4 diverged from cluster 1 in a bottleneck event. Since TDI3, both clusters 1 and 4 remain with a constant Ne of 550,410 (408,638 – 660,872) and 536,892 (388,712 – 623,722) individuals respectively. Following TDI1 there was no gene flow occurring between clusters 1 and 2. After TDI2 there was some gene flow between clusters 1 and 3. Currently, since TDI3 there is gene flow occurring between all 4 clusters at different levels (Figure 4.5 and Appendix Table B.3).
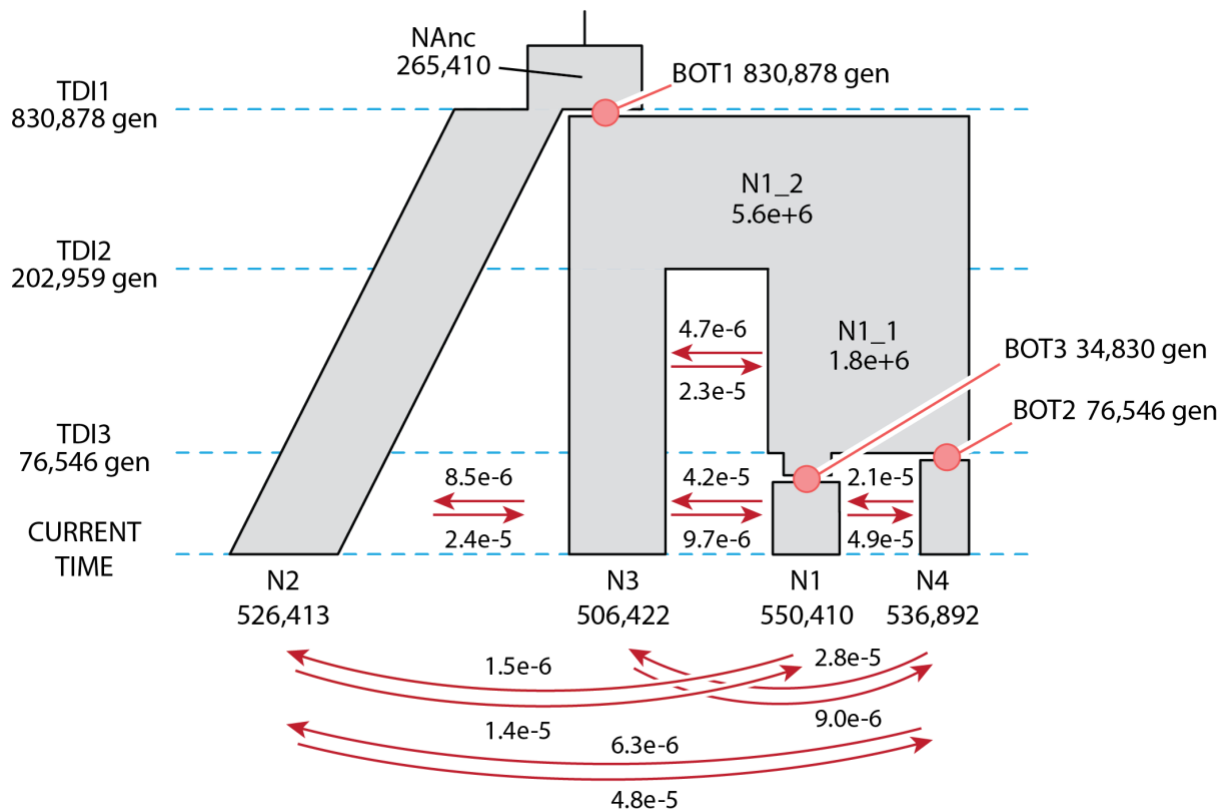
Figure 4.5: Diagram indicating the demographic model for the four genetic clusters identified in all samples. N1-N4 represent the estimate of effective population size for each of the clusters from Figure 4.2. Blue dotted lines indicate the estimate divergence events for each of the clusters. NAnc = Effective population size of ancestral population. TDI = Time of divergence (1 – 3) in generations between clusters. BOT1 – BOT3 indicate the time in generation when a bottleneck happened. Red arrows represent gene flow between the different clusters in number of genes per generation. Gene flow between TDI2 and TDI3 happened only between clusters 1 and 3, while it is happening among all 4 clusters between TDI3 and current time. Column widths generally represent population size, but it is not to scale. Generation times between divergence periods are not to scale.

Discussion

In this chapter, I demonstrated that green sea turtle genetic structure for the Greater Caribbean is composed of 3 major genetic clusters, and I evaluated how dispersal of green sea turtles in the Atlantic Ocean might have happened. I also found evidence of changes in composition of mixed stock aggregations in recent years.

Assuming that the genetic structure for nuclear markers is similar to the one found in mtDNA haplotypes, we would expect to detect a similar number of genetic clusters with both approaches. However, the number of clusters inferred by the *de novo* clustering algorithm (n = 4) was lower than the number of possible source populations for IRL and TRID identified in Chapter 2 (n = 6 - Stahelin et al., (2022)). A lower number of clusters for green turtle populations in the Atlantic Ocean has also been suggested when using microsatellites and mtDNA data (Naro-Maciel, et al., 2014). Fine-scale genetic differentiation of rookeries using longer fragments is useful for designation of regional management units (RMUs - (Shamblin et al., 2012)), but might be insufficient for evaluation of broad-scale connectivity and dispersal (Lovette et al., 2004).

I identified a shift in the composition of TRID aggregation in recent years, currently having a majority of individuals from Cluster 1 instead of Cluster 3 (Table 4.4). A similar pattern was reported in Chapter 2 (Stahelin et al., 2022), where the Tamaulipas rookery, Mexico, became the major contributor to the TRID mixed stock aggregation (Figure 2.4). This could be a strong indication that Cluster 1 includes the Mexican rookeries, while Cluster 3 includes the central Florida ones (Table 4.4). Analysis of mtDNA haplotypes indicate a great similarity between rookeries in Florida, USA, and Mexico, which is also observed in the DAPC and PCA results between Clusters 1 and 3 (Figure 4.2). Sea turtles at the TRID site are smaller (and younger) than the ones found in the IRL, with almost no overlap in size classes between these areas (Ehrhart et al., 2007; Redfoot & Ehrhart, 2013). Size classes at TRID indicate that this is the type of habitat that juvenile green turtles first occupy after recruitment from offshore. Most of the early dispersal stage green turtles captured in the Gulf of Mexico originated from rookeries along the Mexican coast (Phillips et al., 2022; Shamblin et al., 2018, 2023). A number of these individuals leave the

Gulf of Mexico into the western North Atlantic ocean towards the Sargasso sea (Mansfield et al., 2021; Putman & Mansfield, 2015). A change in genetic composition observed at TRID might be a few years in advance from being observed in other locations, like the IRL. Assuming Cluster 1 indeed belongs to the Tamaulipas rookery, this can be a strong indication of a broader shift in juvenile genetic composition following recent growth in nesting populations observed in the region, including the Tamaulipas rookery (Chaloupka et al., 2008; Pineda & Rocha, 2016; Stahelin et al., 2022).

Evaluating the mtDNA haplotype composition of individuals from Cluster 2 revealed that all three individuals have haplotype CM-A8.1, which is the dominant variant in the South Atlantic and West African rookeries (Bjorndal et al., 2006; Encalada et al., 1996; Formia et al., 2006, 2007). The occurrence of haplotype CM-A8.1 in central Florida samples could be explained by the dispersal of individuals that initially followed the Atlantic South Equatorial current (Figure 4.6), which flows west in the South Atlantic and splits into two currents at the northeast region of Brazil: the south branch turns into the Brazil current flowing south along the Brazilian coast, and the north branch becomes the North Brazil current that flows towards the Caribbean (Talley et al., 2011). The North Brazil current could then facilitate the dispersal of individuals into the Caribbean, leading to small proportions of animals from the South Atlantic region at the east central Florida coast (Mansfield et al., 2017; Naro-Maciel et al., 2017; Proietti et al., 2014). Alternatively, a sea turtle farm in the Cayman Islands established a captive breeding program in 1968 with individuals from multiple rookeries (Ascension Island, Costa Rica, Guyana, and Suriname - Barbanti et al., (2019)). Annually, the captive breeding program releases the excess individuals from their stocks into the ocean (Barbanti et al., 2019). Still, only a fraction of the individuals have been genotyped

to determine the genetic composition of the individuals released in the wild. Haplotype CM-A8 is found in > 83% of females reproducing in Ascension Island (Formia et al., 2007) and ~2% of females from Suriname (Bjorndal et al., 2005), two of the source populations for the farm's breeding stocks. For instance, haplotype CM-A8 was previously found in one nesting female found at Hutchinson Island, Florida (Shamblin, et al., 2015), and one nesting female from the MB rookery was reported earlier in this chapter. It is possible, therefore, that both in-water juveniles and nesting females found in Florida with haplotype CM-A8 originated from the captive breeding program in the Cayman Islands and their occurrence in Florida sites might have been facilitate by marine currents.
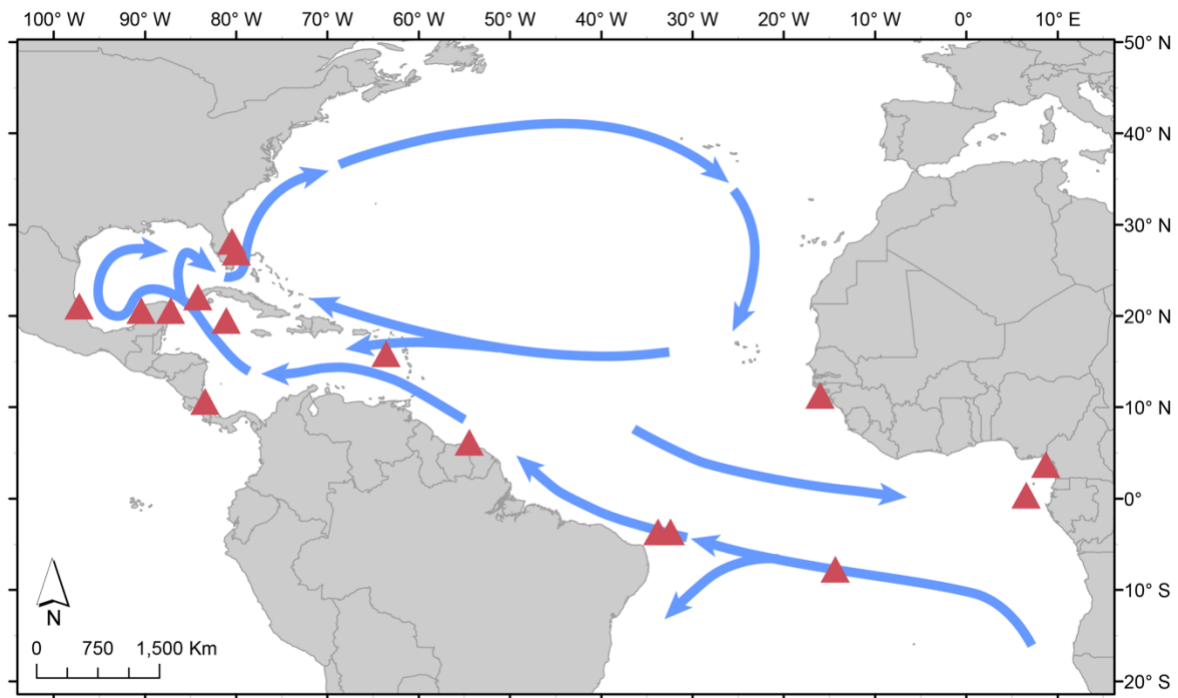
Figure 4.6: Marine currents in the Atlantic Ocean with potential to impact dispersal of sea turtles in the Greater Caribbean. Red triangles represent main green turtle rookeries. Blue arrows represent marine currents. Marine currents adapted from Talley et al., (2011).

Gene flow between all identified clusters is well supported by the coalescent model (Figure 4.5), and has been previously identified for other populations (FitzSimmons et al., 1997; Karl et al., 1992; Naro-Maciel et al., 2014; Roberts et al., 2004; van der Zee et al., 2021). Migration could help explain low levels of differentiation observed here and the occurrence of individuals from multiple clusters within the MB rookery (Table 4.4). However, the presence of Cluster 4 in the MB rookery puts into question the cluster assignment made so far. One hypothesis could be that Cluster 4 population encompasses Tortuguero in Costa Rica, and maybe other rookeries from east Caribbean (Aves Island in Venezuela, and Galibi in Suriname). It could be that a small group of individuals from Cluster 4 migrated to the peninsular Florida recently (during the population decline observed in the past century and could not be detected by coalescent models). Male-mediated gene flow between populations has been discussed and previously identified (FitzSimmons, et al., 1997; Karl et al., 1992; Roberts et al., 2004), which could certainly be the case in my models, but the presence of individuals from multiple clusters in the MB samples raises the possibility for female migration as well. Therefore, it also plausible that Cluster 1 is in fact from Tortuguero and east Caribbean and re-populated the Peninsular Florida region in recent years. Both of these hypotheses could be supported by migration and well-established dispersal corridor between rookeries in the Northwest Atlantic and Caribbean. A better understanding of which, or if any, of the hypotheses detailed here is correct can be addressed by future studies using genomic analysis from multiple Caribbean rookeries.

Data concerns and considerations

Divergence times between the genetic clusters and estimates of effective population sizes for each cluster/population identified here differs from previous estimates for the species by nearly an order

of magnitude (see Fitak & Johnsen, 2018; Jensen et al., 2019; Naro-Maciel, et al., 2014; van der Zee et al., 2021; Vilaça et al., 2021). My estimates of time of divergence are much larger (i.e., events happened longer ago), and effective population sizes are much larger than previous studies. It is unclear the reason for such mismatch between results, as priors used and multiple data quality steps are similar. For instance, here I used the mutation rate of 1.2×10^{-8}, the same used by Fitak & Johnsen, (2018), while Vilaça et al., (2021) used 7.9×10^{-9}. Interestingly enough, van der Zee et al., (2021) tested the effect of mutation rate (between 1.2×10^{-8} and 7.9×10^{-9}) and their results suggest that the difference in mutation rate is not sufficient to explain the variation observed here.

One possibility might be due to issues I ran into during data quality and filtering steps of SNP data. Samples from the "old" time-period were removed from the dataset at a much higher rate than those from the "new" time-period. It is unclear why the sequencing quality of these samples was much reduced, but it could be associated with the storage method before DNA extraction (Ballare et al., 2019). Skin samples were preserved in ethanol at room temperature, and red blood cell samples were either frozen at -20ºC or stored at room temperature on lysis buffer (see Methods in Chapter 2 - (Stahelin et al., 2022)). However, as these samples were stored for 20+ years, fluctuations in storage conditions over the years may have contributed to a general decrease in DNA quality. Overall, samples from the "old" time-periods had a lower depth of sequencing, leading to a higher rate of missing calls across loci.

Another possible bias could be related to loci and sample filtering steps. I found that switching the order in which filters while keeping the same cutoff values for the filters (i.e., individuals with missing genotype data [--mind], minor allele frequency [--maf], and SNPs with missing

genotyping rate [--geno]), changed considerably the number of individuals and SNPs filtered. The order in which these filters are applied is inconsistent between, for instance, the PLINK software manual (Purcell et al., 2007) and previously published papers (Agudelo et al., 2022; Mastrochirico-Filho et al., 2021), without a clear explanation for that choice in all cases. Users may have been applying the filters in multiple orders for not knowing of impacts on the final dataset. Applying filters for individuals with high rate of genotype data missing will generate a dataset with a higher lower number of individuals with many SNPs shared between samples. On the other hand, filtering for loci with high data missing first, will generate a dataset with more individuals and fewer SNPs. This trade off seems obvious, but I have not found a study evaluating the impact of these choices on genetic structure of the resulting dataset.

Conclusions

The use of genomic tools should be the goal and standard practice in near future sea turtle assessments, as is the case for other organisms. A recent publication of reference genomes for most sea turtle species (Bentley et al., 2022) opened the possibility for a variety of studies regarding evolution, ecology, disease, and others. Next-generation sequencing has already been used recently to refine population structure and aspects of sea turtles' ecology (Prakash et al., 2022; van der Zee et al., 2021; Vilaça et al., 2023; Vilaça et al., 2021), and certainly many more studies are coming out to improve our knowledge about their ecology. Conservation plans can benefit greatly from conservation genomics instead of conservation genetic studies, greatly improving the best-data-available practice for managers (Nielsen et al., 2020).

Gene flow between populations in the Caribbean is highly supported by the coalescent models, with varying rates between them. This finding adds to the amount of data indicating that green sea turtle reproductive aggregations are interconnected, which is important for conservation planning. Continuous gene flow between regions and reproductive areas is critical for sea turtles' persistence over time, improving population's phenotypic plasticity to face an ever-changing environment. Finally, studies with long-term in-water data as this one should be encouraged and receive greater support from managing and funding agencies around the world. Understanding of dynamics in these mixed stock aggregations can certainly benefit sea turtle conservation plans.

CHAPTER 5: CONCLUSIONS

This is the first study to evaluate changes in contributions to green sea turtle mixed stock aggregations in response to reproductive population growth (Chapter 2). In addition, I modified the standardly used many-to-many mixed stock model to incorporate distance between source populations (rookeries) and mixed stock aggregations. This approach weights rookery contribution estimates based on distance between rookeries and mixed-stock aggregations. Also, to date, this is the only study to evaluate in detail the impact of sample size and other metrics in mixed stock models estimates (Chapter 3). Finally, I used next-generation sequencing to understand dispersal patterns and assess connectivity between populations from the north Atlantic and Greater Caribbean (Chapter 4).

Despite recent variations in many green turtle rookeries, mixed stock aggregations in the east coast of Central Florida have remained relatively constant in terms of their stock contribution over the past ~20 years (Chapter 2). The only site with a noticeable change was in the Trident Submarine Basin (TRID) mixed stock, where the Tamaulipas rookery in Mexico is the main contributor to juveniles in this area in recent years (2016 - 2018). Interestingly, a similar result was observed using the genomic data, with Cluster 1 becoming the most common source of individuals in TRID with nearly 60% of recent contributions (Chapter 4). The dispersal of juveniles from Mexican and other rookeries in the Caribbean matches the assumptions made in Chapter 2 when we included a distance matrix to weight mixed stock models estimates based on how far an individual would need to travel between their natal rookery and a mixed stock aggregation.

Mixed stock analysis is commonly used in sea turtle studies as a way to understand relationships between reproductive populations and in-water aggregations (Gaos et al., 2020; Medeiros et al., 2019; Phillips et al., 2022; Shamblin et al., 2017; Stahelin et al., 2022). Besides helping address basic ecological questions, these models help conservation managers delineate conservation plans and identify protective actions within and between countries. Chapter 3 demonstrates how mixed stock models can be impacted by using biased sampling due to small sample sizes. Results from this dissertation allow me to make a few recommendations for future studies: (i) I strongly recommend incorporating distance or any other site-specific metric between rookeries and mixed stock aggregations (e.g., particle dispersal model probabilities) in mixed stock analyses; the use of the model introduced in Chapter 2 with a distance matrix can reduce credible intervals and make results more ecologically meaningful. I recommend the use of both (ii) higher resolution mtDNA markers (e.g., longer haplotypes) and (iii) larger sample sizes (> 150 samples) at mixed stock aggregations in order to improve model accuracy and reduce credible intervals (Chapter 3). Ideally, these three recommendations would be implemented simultaneously, but implementation of any or a combination of these in future studies will certainly improve model estimates of contribution from rookeries to mixed stock aggregations. Interpretation of mixed stock models has a higher chance of being ecologically incorrect if these recommendations are not followed, with unintended consequences for the conservation and recovery of these species.

Gene flow between populations is important for adaptation to evolutionary pressures, but also for persistence of populations over time (Crispo, 2008). The coalescent model detailed in Chapter 4 offers support for gene flow between all populations identified using the *de novo* Discriminant Analysis of Principal Components. This result agrees with previous assessments that indicated

different levels of gene flow between sea turtle populations (FitzSimmons, et al., 1997; Karl et al., 1992; Roberts et al., 2004). More importantly, it provides insightful understanding that geographical reproductive areas could be shared by more than one genetic cluster (Table 4.3and Table 4.4 – Chapter 4). Use of genomic assessments and coalescent models in future studies is recommended as these tools can greatly improve our understanding of ecological aspects sea turtles distribution and their evolutionary history (Sovic et al., 2016, 2019).

Finally, additional in-water sea turtle sampling studies should be encouraged and supported by managers and funding agencies. Sea turtles are, in general, late-maturing individuals (Casale et al., 2011; Girondot et al., 2021; Seminoff, 2004; Turner Tomaszewicz et al., 2022), meaning that changes in foraging aggregations may take several years to start reflecting in rookeries. Constant monitoring of mixed stock aggregations can reduce the response time to a new threat and more effectively protect these species. Long-term datasets with enough samples to evaluate changes in populational parameters over time (like the ones used in this study) should, ideally, be maintained or established in several locations around the world to enhance our understanding of sea turtle dispersal and genetic connectivity. More importantly, researchers should aim to use larger datasets and genomic data to answer multiple ecological questions, pushing the bar of the 'best available data' forward for management decisions. Partnerships between conservation programs, managers, and funding agencies should be further encouraged, to improve knowledge in under-represented life stages, species, and areas for more robust and ecologically meaningful analyses.

APPENDIX A:
APPENDIX FIGURES
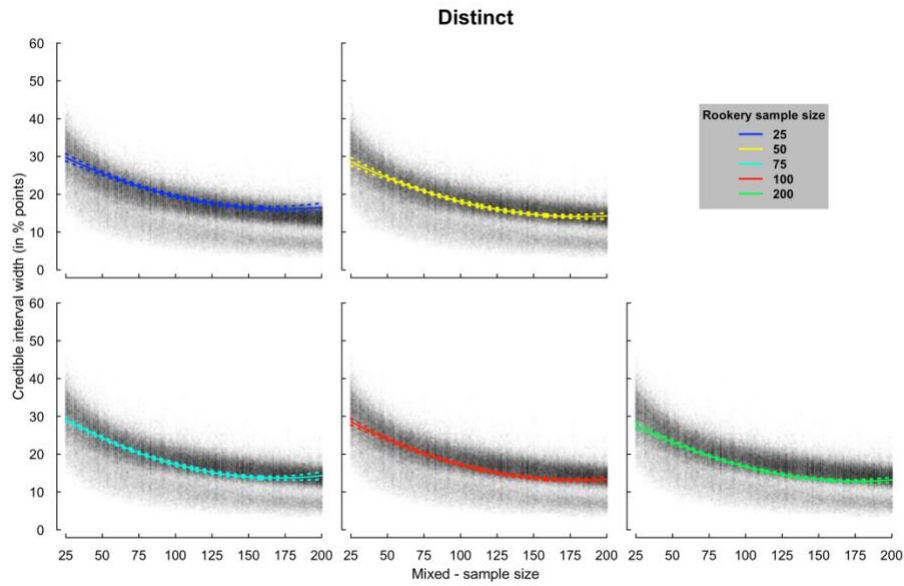
Figure A.1: *Distinct* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.
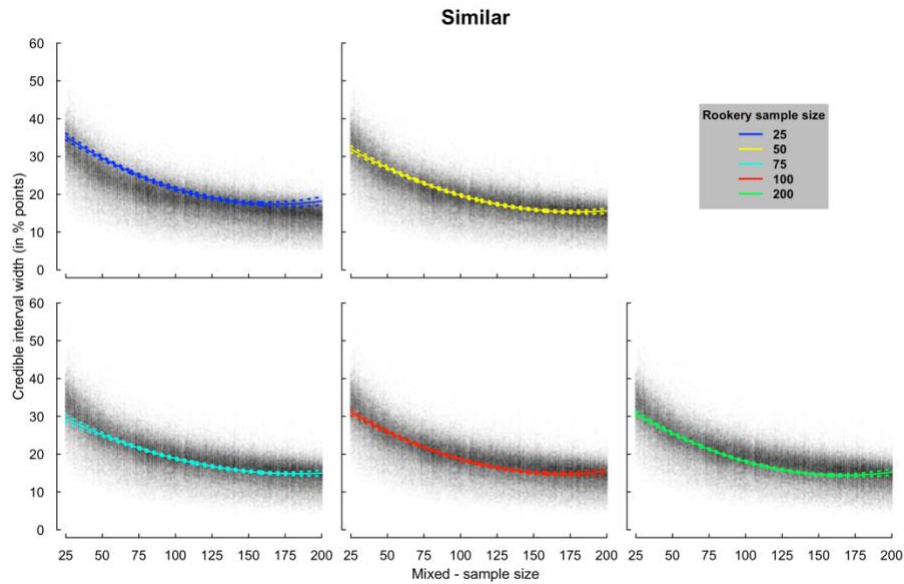
Figure A.2: *Similar* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.
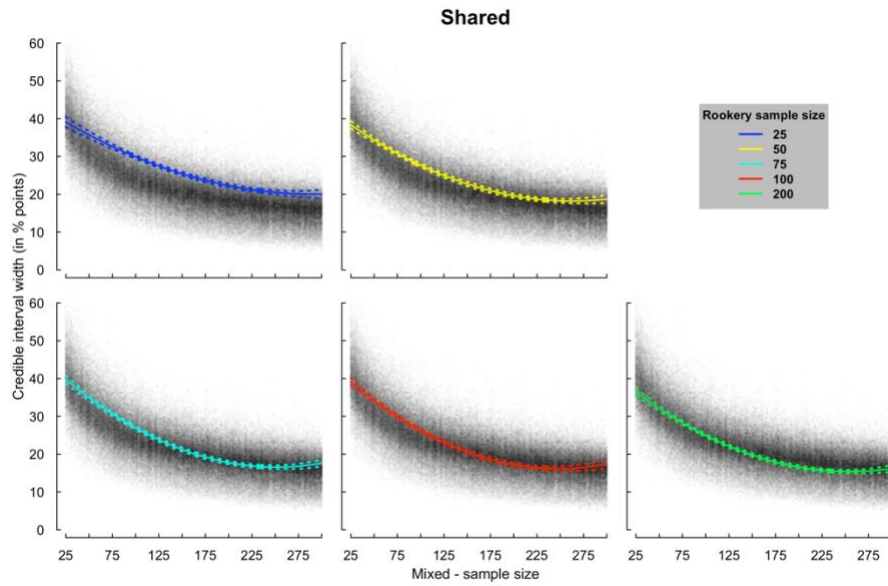
Figure A.3: *Shared* haplotype scenario scatter plot of credible interval width (in percentage points) of mixed stock model estimates as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.
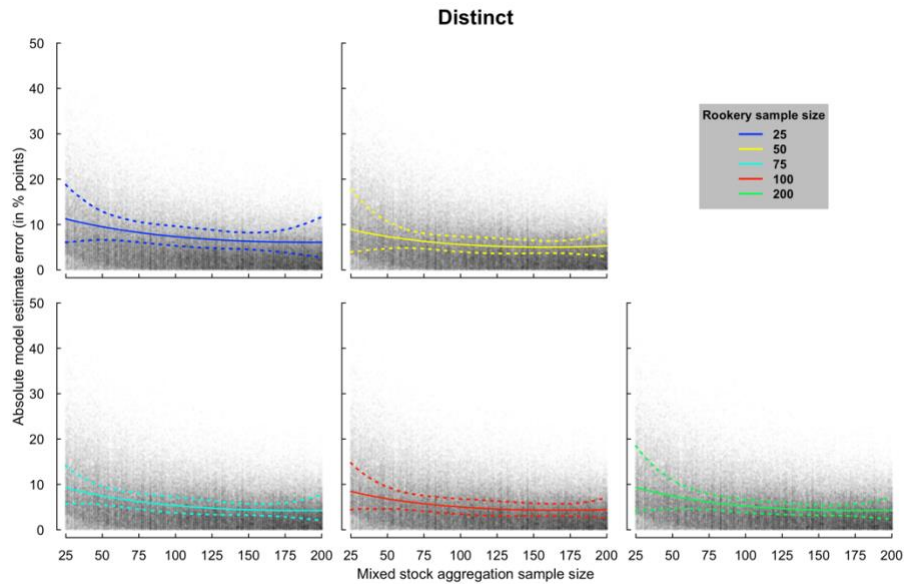
Figure A.4: *Distinct* haplotype scenario scatter plot of absolute model estimate error (in percentage points) as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.
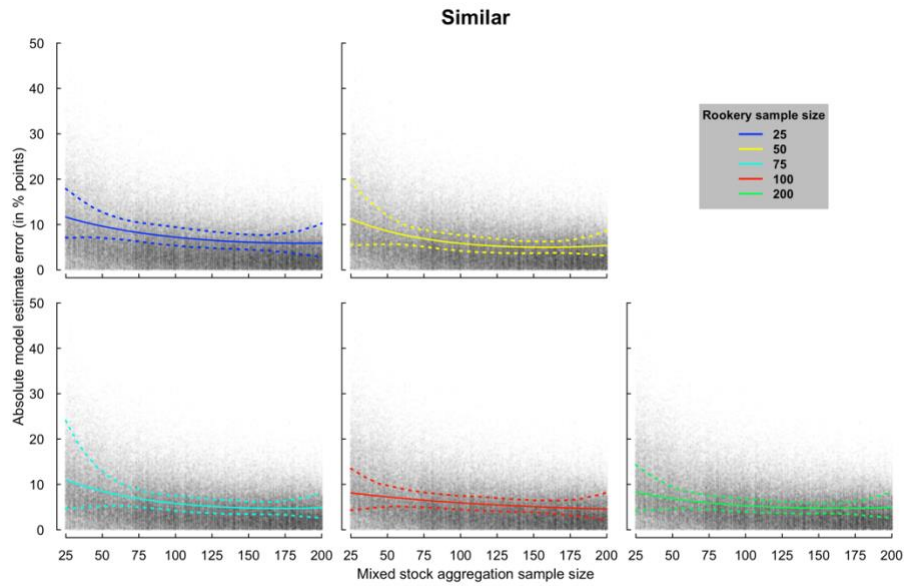
Figure A.5: *Similar* haplotype scenario scatter plot of model absolute estimate error (in percentage points) as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.
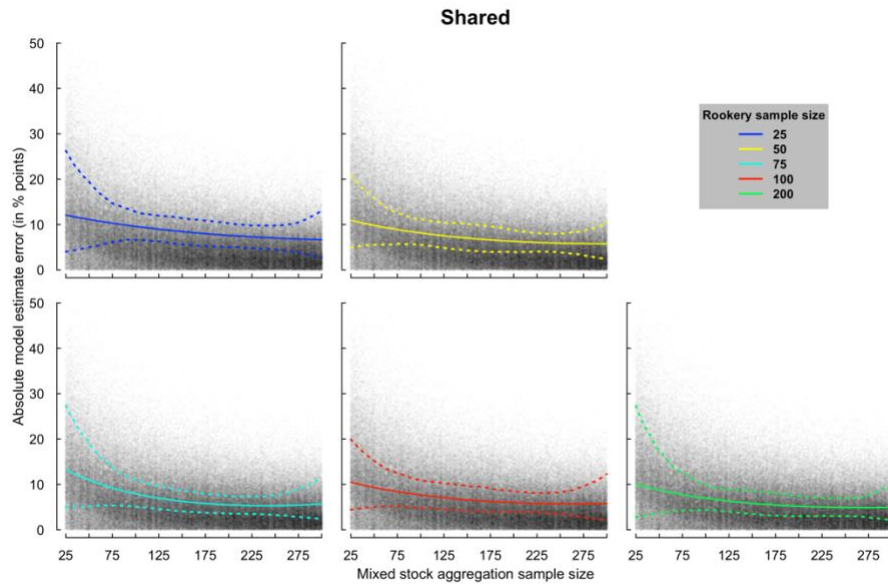
Figure A.6: *Shared* haplotype scenario scatter plot of model absolute estimate error (in percentage points) as a function of mixed stock aggregations sample size. Colors represent fit lines from different rookery sample sizes. Dashed lines represent 95% confidence intervals.

APPENDIX B:
APPENDIX TABLES

Table B.1: Model selection table detailing all models for each scenario used to evaluate the relationship between credible interval width (CIW) and the different variables used in the simulations. MS_SSize = Mixed stock aggregation sample size, SP_SSize = Source population sample size, MS_PopSize = Mixed stock aggregation scaled population size, SP_PopSize = Source population scaled population size, MS_PropPSize = Ratio of sampled individuals in mixed stock aggregation in relationship to mixed stock aggregation population size, SP_PropPSize = Ratio of sampled individuals in source population in relationship to the source population population size. dAICc = Delta AICc, df = degrees of freedom.

Model	dAICc	df	Weight	Scenario
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Distinct*
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	15837.4	7	<0.001	*Distinct*
CIW ~ MS_SSize * SP_SSize	15857.7	5	<0.001	*Distinct*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	15865.2	8	<0.001	*Distinct*
CIW ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	15867.5	6	<0.001	*Distinct*
CIW ~ MS_SSize + SP_SSize	15887.3	4	<0.001	*Distinct*
CIW ~ MS_SSize	17822.7	3	<0.001	*Distinct*
CIW ~ MS_SSize + MS_PopSize	17824.7	4	<0.001	*Distinct*
CIW ~ MS_PropPSize + SP_PropPSize	90893.7	4	<0.001	*Distinct*
CIW ~ MS_PropPSize * SP_PropPSize	90894.7	5	<0.001	*Distinct*
CIW ~ MS_PropPSize	90984.6	3	<0.001	*Distinct*
CIW ~ SP_SSize + SP_SSize^2	108991.8	4	<0.001	*Distinct*
CIW ~ SP_SSize + SP_PopSize	109099.8	4	<0.001	*Distinct*
CIW ~ SP_SSize	109100	3	<0.001	*Distinct*
CIW ~ SP_PropPSize	109999.1	3	<0.001	*Distinct*
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Similar*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	20298	8	<0.001	*Similar*
CIW ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	20303	6	<0.001	*Similar*
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	20305	7	<0.001	*Similar*
CIW ~ MS_SSize + SP_SSize	20327	4	<0.001	*Similar*
CIW ~ MS_SSize * SP_SSize	20328	5	<0.001	*Similar*
CIW ~ MS_SSize	24052	3	<0.001	*Similar*
CIW ~ MS_SSize + MS_PopSize	24053	4	<0.001	*Similar*
CIW ~ MS_PropPSize + SP_PropPSize	114587	4	<0.001	*Similar*
CIW ~ MS_PropPSize * SP_PropPSize	114588	5	<0.001	*Similar*
CIW ~ MS_PropPSize	114666	3	<0.001	*Similar*
CIW ~ SP_SSize + SP_SSize^2	134360	4	<0.001	*Similar*
CIW ~ SP_SSize + SP_PopSize	134678	4	<0.001	*Similar*
CIW ~ SP_SSize	134698	3	<0.001	*Similar*
CIW ~ SP_PropPSize	136837	3	<0.001	*Similar*
CIW ~ MS_SSize + MS_SSize^2	0	4	1	*Shared*

Model	dAICc	df	Weight	Scenario
CIW ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	33839	7	<0.001	*Shared*
CIW ~ MS_SSize * SP_SSize	33846	5	<0.001	*Shared*
CIW ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	33890	8	<0.001	*Shared*
CIW ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	33917	6	<0.001	*Shared*
CIW ~ MS_SSize + SP_SSize	33924	4	<0.001	*Shared*
CIW ~ MS_SSize	42943	3	<0.001	*Shared*
CIW ~ MS_SSize + MS_PopSize	42945	4	<0.001	*Shared*
CIW ~ MS_PropPSize + SP_PropPSize	176671	4	<0.001	*Shared*
CIW ~ MS_PropPSize * SP_PropPSize	176673	5	<0.001	*Shared*
CIW ~ MS_PropPSize	177362	3	<0.001	*Shared*
CIW ~ SP_SSize + SP_SSize^2	211119	4	<0.001	*Shared*
CIW ~ SP_SSize	212024	3	<0.001	*Shared*
CIW ~ SP_SSize + SP_PopSize	212025	4	<0.001	*Shared*
CIW ~ SP_PropPSize	215200	3	<0.001	*Shared*

Table B.2: Model selection table detailing all models for each scenario used to evaluate the relationship between absolute model estimate error (AMEE) and the different variables used in the simulations. MS_SSize = Mixed stock aggregation sample size, SP_SSize = Source population sample size, MS_PopSize = Mixed stock aggregation scaled population size, SP_PopSize = Source population scaled population size, MS_PropPSize = Ratio of sampled individuals in mixed stock aggregation in relationship to mixed stock aggregation population size, SP_PropPSize = Ratio of sampled individuals in source population in relationship to the source population population size. dAICc = Delta AICc, df = degrees of freedom.

Model	dAICc	df	Weight	Scenario
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Distinct*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	507	7	<0.001	*Distinct*
AMEE ~ MS_SSize * SP_SSize	513	5	<0.001	*Distinct*
AMEE ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	525	8	<0.001	*Distinct*
AMEE ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	526	6	<0.001	*Distinct*
AMEE ~ MS_SSize + SP_SSize	532	4	<0.001	*Distinct*
AMEE ~ MS_SSize + MS_PopSize	1014	4	<0.001	*Distinct*
AMEE ~ MS_SSize	1021	3	<0.001	*Distinct*
AMEE ~ MS_PropPSize + SP_PropPSize	8170	4	<0.001	*Distinct*
AMEE ~ MS_PropPSize * SP_PropPSize	8171	5	<0.001	*Distinct*
AMEE ~ MS_PropPSize	8237	3	<0.001	*Distinct*
AMEE ~ SP_SSize + SP_SSize^2	10296	4	<0.001	*Distinct*
AMEE ~ SP_SSize	10361	3	<0.001	*Distinct*

Model	dAICc	df	Weight	Scenario
AMEE ~ SP_SSize + SP_PopSize	10363	4	<0.001	*Distinct*
AMEE ~ SP_PropPSize	10770	3	<0.001	*Distinct*
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Similar*
AMEE ~ MS_SSize * SP_SSize	535	5	<0.001	*Similar*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	536	7	<0.001	*Similar*
AMEE ~ MS_SSize + SP_SSize	542	4	<0.001	*Similar*
AMEE ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	542	8	<0.001	*Similar*
AMEE ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	544	6	<0.001	*Similar*
AMEE ~ MS_SSize	882	3	<0.001	*Similar*
AMEE ~ MS_SSize + MS_PopSize	883	4	<0.001	*Similar*
AMEE ~ MS_PropPSize + SP_PropPSize	8320	4	<0.001	*Similar*
AMEE ~ MS_PropPSize * SP_PropPSize	8322	5	<0.001	*Similar*
AMEE ~ MS_PropPSize	8342	3	<0.001	*Similar*
AMEE ~ SP_SSize + SP_SSize^2	10524	4	<0.001	*Similar*
AMEE ~ SP_SSize + SP_PopSize	10595	4	<0.001	*Similar*
AMEE ~ SP_SSize	10596	3	<0.001	*Similar*
AMEE ~ SP_PropPSize	10988	3	<0.001	*Similar*
AMEE ~ MS_SSize + MS_SSize^2	0	4	1	*Shared*
AMEE ~ MS_SSize * SP_SSize + MS_PopSize + SP_PopSize	859	7	<0.001	*Shared*
AMEE ~ MS_SSize * SP_SSize	880	5	<0.001	*Shared*
AMEE ~ MS_SSize * MS_PopSize + SP_SSize * SP_PopSize	909	8	<0.001	*Shared*
AMEE ~ MS_SSize + SP_SSize + MS_PopSize + SP_PopSize	911	6	<0.001	*Shared*
AMEE ~ MS_SSize + SP_SSize	932	4	<0.001	*Shared*
AMEE ~ MS_SSize + MS_PopSize	1814	4	<0.001	*Shared*
AMEE ~ MS_SSize	1830	3	<0.001	*Shared*
AMEE ~ MS_PropPSize + SP_PropPSize	16573	4	<0.001	*Shared*
AMEE ~ MS_PropPSize * SP_PropPSize	16575	5	<0.001	*Shared*
AMEE ~ MS_PropPSize	16735	3	<0.001	*Shared*
AMEE ~ SP_SSize + SP_SSize^2	21634	4	<0.001	*Shared*
AMEE ~ SP_SSize + SP_PopSize	21806	4	<0.001	*Shared*
AMEE ~ SP_SSize	21809	3	<0.001	*Shared*
AMEE ~ SP_PropPSize	22490	3	<0.001	*Shared*

Table B.3: Model estimates and 95% confidence intervals from coalescent demographic model. Parameter names follow the labels used in Figure 4.5. MIGXY refers to the probability of one individual migrating from population X to population Y.

Parameter	Mean	2.50%	97.50%	Type of estimate
N1	550,410	408,638	660,872	Ne
N1_1	1,815,202	1,347,651	2,179,495	Ne
N1_2	5,644,881	4,190,899	6,777,754	Ne
N2	526,413	352,534	619,161	Ne
NAnc	265,410	177,742	312,172	Ne
N3	506,422	430,889	583,649	Ne
N4	536,892	388,712	623,722	Ne
BOT1	830,878	602,310	943,898	Time in generations
BOT2	76,546	59,973	100,239	Time in generations
BOT3	34,830	26,969	53,599	Time in generations
TDI1	830,878	602,310	943,898	Time in generations
TDI2	202,959	134,516	259,354	Time in generations
TDI3	76,546	59,973	100,239	Time in generations
MIG12 (Current)	1.49E-06	8.20E-07	1.03E-05	Gene flow per generation
MIG13 (Current)	4.23E-05	8.63E-06	7.55E-05	Gene flow per generation
MIG14 (Current)	4.86E-05	1.23E-05	8.54E-05	Gene flow per generation
MIG21 (Current)	1.40E-05	6.25E-06	3.33E-05	Gene flow per generation
MIG23 (Current)	2.44E-05	6.66E-06	1.19E-04	Gene flow per generation
MIG24 (Current)	4.81E-05	4.62E-06	8.18E-05	Gene flow per generation
MIG31 (Current)	9.69E-06	2.04E-06	2.97E-05	Gene flow per generation
MIG32 (Current)	8.50E-06	1.73E-06	1.31E-05	Gene flow per generation
MIG34 (Current)	9.02E-06	3.42E-06	4.34E-05	Gene flow per generation
MIG41 (Current)	2.05E-05	8.20E-06	5.93E-05	Gene flow per generation
MIG42 (Current)	6.31E-06	2.20E-06	1.29E-05	Gene flow per generation
MIG43 (Current)	2.80E-05	9.92E-06	5.77E-05	Gene flow per generation
MIG13 (TDI2-TDI3)	4.65E-06	1.64E-06	8.23E-06	Gene flow per generation
MIG31 (TDI2-TDI3)	2.28E-05	4.60E-06	5.58E-05	Gene flow per generation

REFERENCES

Abreu-Grobois, F. A., Horrocks, J. A., Formia, A., Dutton, P. H., LeRoux, R. A., Vélez-Zuazo, X., Soares, L. S., & Meylan, P. A. (2006). New mtDNA Dloop primers which work for a variety of marine turtle species may increase the resolution of mixed stock analyses. *Proceedings of the 26th Annual Symposium on Sea Turtle Biology*, 179.

Agudelo, J. F. G., Mastrochirico-Filho, V. A., Borges, C. H. de S., Ariede, R. B., Lira, L. V. G., Neto, R. R. de O., Freitas, M. V., Sucerquia, G. A. L., Vera, M., Berrocal, M. H. M., & Hashimoto, D. T. (2022). Genomic selection signatures in farmed Colossoma macropomum from tropical and subtropical regions in South America. *Evolutionary Applications*, *15*(4), 679–693. https://doi.org/10.1111/eva.13351

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Anderson, E. C., Waples, R. S., & Kalinowski, S. T. (2008). An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*, 1475–1486.

Anderson, J. D., Shaver, D. J., & Karel, W. J. (2013). Genetic Diversity and Natal Origins of Green Turtles (*Chelonia mydas*) in the Western Gulf of Mexico. *Journal of Herpetology*, *47*(2), 251–257. https://doi.org/10.1670/12-031

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Azanza Ricardo, J., Ibarra Martín, M. E., González Sansón, G., Abreu Grobois, F. A., Eckert, K. L., Espinosa López, G., & Oyama, K. (2013). Nesting ecology of *Chelonia mydas* (Testudines: Cheloniidae) on the Guanahacabibes Peninsula, Cuba. *Revista de Biologia Tropical*, *61*(4), 1935–1945. https://doi.org/10.15517/rbt.v61i4.12869

Bagley, D. A. (2003). *Characterizing juvenile green turtles, (*Chelonia mydas*), from three east central Florida developmental habitats*. University of Central Florida.

Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., & Coleman, C. I. (2009). Understanding heterogeneity in meta-analysis: the role of meta-regression. *International Journal of Clinical Practice*, *63*(10), 1426–1434. https://doi.org/10.1111/j.1742-1241.2009.02168.x

Ballare, K. M., Pope, N. S., Castilla, A. R., Cusser, S., Metz, R. P., & Jha, S. (2019). Utilizing field collected insects for next generation sequencing: Effects of sampling, storage, and DNA extraction methods. *Ecology and Evolution*, *9*(24), 13690–13705. https://doi.org/10.1002/ece3.5756

Barbanti, A., Martin, C., Blumenthal, J. M., Boyle, J., Broderick, A. C., Collyer, L., Ebanks-Petrie, G., Godley, B. J., Mustin, W., Ordóñez, V., Pascual, M., & Carreras, C. (2019). How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Molecular Ecology*, *28*(7), 1637–1651. https://doi.org/10.1111/mec.15017

Bass, A. L., Epperly, S. P., & Braun-McNeill, J. (2006). Green turtle (*Chelonia mydas*) foraging and nesting aggregations in the Caribbean and Atlantic: Impact of currents and behavior on dispersal. *Journal of Heredity*, *97*(4), 346–354. https://doi.org/10.1093/jhered/esl004

Bass, A. L., Lagueux, C. J., & Bowen, B. W. (1998). Origin of green turtles, *Chelonia mydas*, at "Sleeping Rocks" off the Northeast coast of Nicaragua. *Copeia*, *1998*(4), 1064. https://doi.org/10.2307/1447360

Bass, A. L., & Witzell, W. N. (2000). Demographic composition of immature green turtles (*Chelonia mydas*) from the East Central Florida Coast: Evidence from mtDNA markers. *Herpetologica*, *56*(3), 357–367.

Bentley, B. P., Carrasco-Valenzuela, T., Ramos, E. K. S., Pawar, H., Arantes, L. S., Alexander, A., Banerjee, S. M., Masterson, P., Kuhlwilm, M., Pippel, M., Mountcastle, J., Haase, B., Silva, M. U., Formenti, G., Howe, K., Chow, W., Tracey, A., Sims, Y., Pelan, S., … Komoroske, L. M. (2022). Differential sensory and immune gene evolution in sea turtles with contrasting demographic and life histories. *BioRxiv*, 2022.01.10.475373. https://doi.org/https://doi.org/10.1101/2022.01.10.475373

Bernos, T. A., Jeffries, K. M., & Mandrak, N. E. (2020). Linking genomics and fish conservation decision making: a review. *Reviews in Fish Biology and Fisheries*, *30*(4), 587–604. https://doi.org/10.1007/s11160-020-09618-8

Bjorndal, K. A., & Bolten, A. B. (2008). Annual variation in source contributions to a mixed stock: Implications for quantifying connectivity. *Molecular Ecology*, *17*(9), 2185–2193. https://doi.org/10.1111/j.1365-294X.2008.03752.x

Bjorndal, K. A., Bolten, A. B., & Chaloupka, M. Y. (2005). *Evaluating trends in abundance of immature green turtles,* Chelonia mydas*, in the Greater Caribbean*. *15*(1), 304–314.

Bjorndal, K. A., Bolten, A. B., Moreira, L., Bellini, C., & Marcovaldi, M. Â. (2006). Population structure and diversity of Brazilian green turtle rookeries Based on mitochondrial DNA sequences. *Chelonian Conservation and Biology*, *5*(2), 262–268. https://doi.org/10.2744/1071-8443(2006)5[262:psadob]2.0.co;2

Bjorndal, K. A., Bolten, A. B., & Troëng, S. (2005). Population structure and genetic diversity in green turtles nesting at Tortuguero, Costa Rica, based on mitochondrial DNA control region sequences. *Marine Biology*, *147*(6), 1449–1457. https://doi.org/10.1007/s00227-005-0045-y

Bjorndal, K. A., Parsons, J., Mustin, W., & Bolten, A. B. (2013). Threshold to maturity in a long-lived reptile: Interactions of age, size, and growth. *Marine Biology*, *160*(3), 607–616.

https://doi.org/10.1007/s00227-012-2116-1

Bjorndal, K. a. (1999). Priorities for Research in Foraging Habitats. In K. L. Eckert, K. A. Bjorndal, F. A. Abreu-Grobois, & M. Donnelly (Eds.), *Research and Management Techniques for the Conservation of Sea Turtles* (Issue 4, pp. 1–3).

Blomqvist, D., Pauliny, A., Larsson, M., & Flodin, L.-Å. (2010). Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. *BMC Evolutionary Biology*, *10*(1), 33. https://doi.org/10.1186/1471-2148-10-33

Bolker, B. (2014). *Bbmle: Tools for general maximum likelihood estimation* (R package version 1.0.25).

Bolker, B. M., Okuyama, T., Bjorndal, K. A., & Bolten, A. B. (2007). Incorporating multiple mixed stocks in mixed stock analysis: "Many-to-many" analyses. *Molecular Ecology*, *16*(4), 685–695. https://doi.org/10.1111/j.1365-294X.2006.03161.x

Bolker, B., Okuyama, T., Bjorndal, K. A., & Bolten, A. B. (2003). Sea turtle stock estimation using genetic markers: Accounting for sampling error of rare genotypes. *Ecological Applications*, *13*(3), 763–775. https://doi.org/10.1890/1051-0761(2003)013[0763:STSEUG]2.0.CO;2

Bolten, A. B. (1999). Techniques for Measuring Sea Turtles. In K. L. Eckert, K. A. Bjorndal, F. A. Abreu-Grobois, & M. Donnelly (Eds.), *Research and Management Techniques for the Conservation of Sea Turtles.* (Issue 4, pp. 1–5).

Bowen, A. B. W., Bass, A. L., Diez, C. E., Dam, R. Van, Bolten, A., Bjorndal, A., Miyamoto, M. M., Ferl, R. J., Applications, S. E., & May, N. (1996). Origin of hawksbill turtles in a Caribbean feeding area as indicated by genetic markers. *Ecological Applications*, *6*(2), 566–572.

Bowen, B. W., & Karl, S. A. (2007). Population genetics and phylogeography of sea turtles. *Molecular Ecology*, *16*(23), 4886–4907. https://doi.org/10.1111/j.1365-294X.2007.03542.x

Bowen, B. W., Meylan, B., Ross, J. P., Limpus, C. J., Balazs, G. H., & Avise, J. C. (1992). Global population structure and natural history of the green turtle (*Chelonia mydas*) in terms of matriarchal phylogeny. *Evolution*, *46*(4), 865–881. https://doi.org/10.1111/j.1558-5646.1992.tb00605.x

Brüniche-Olsen, A., Urban, R. J., Vertyankin, V. V., Godard-Codding, C. A. J., Bickham, J. W., & DeWoody, J. A. (2018). Genetic data reveal mixed-stock aggregations of gray whales in the North Pacific Ocean. *Biology Letters*, *14*(10), 1–4. https://doi.org/10.1098/rsbl.2018.0399

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Canty, A. J. (2002). Resampling methods in R: the boot package. *The Newsletter of the R Project*, *2*(3), 2–7.

Carroll, E. L., Ott, P. H., McMillan, L. F., Vernazzani, B. G., Neveceralova, P., Vermeulen, E., Gaggiotti, O. E., Andriolo, A., Scott Baker, C., Bamford, C., Best, P., Cabrera, E., Calderan, S., Chirife, A., Fewster, R. M., Flores, P. A. C., Frasier, T., Freitas, T. R. O., Groch, K., … Jackson, J. A. (2020). Genetic diversity and connectivity of southern right whales (*Eubalaena australis*) found in the Brazil and Chile-Peru wintering grounds and the South Georgia (Islas Georgias Del Sur) feeding ground. *Journal of Heredity*, *111*(3), 263–276. https://doi.org/10.1093/JHERED/ESAA010

Casale, P., Mazaris, A. D., & Freggi, D. (2011). Estimation of age at maturity of loggerhead sea turtles *Caretta caretta* in the Mediterranean using length-frequency data. *Endangered Species Research*, *13*(2), 123–129. https://doi.org/10.3354/esr00319

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Ceriani, S. A., Roth, J. D., Evans, D. R., Weishampel, J. F., & Ehrhart, L. M. (2012). Inferring Foraging Areas of Nesting Loggerhead Turtles Using Satellite Telemetry and Stable Isotopes. *PLoS ONE*, *7*(9), e45335. https://doi.org/10.1371/journal.pone.0045335

Chaloupka, M., Bjorndal, K. A., Balazs, G. H., Bolten, A. B., Ehrhart, L. M., Limpus, C. J., Suganuma, H., Troëng, S., & Yamaguchi, M. (2008). Encouraging outlook for recovery of a once severely exploited marine megaherbivore. *Global Ecology and Biogeography*, *17*(2), 297–304. https://doi.org/10.1111/j.1466-8238.2007.00367.x

Chiarello, A. G., Chivers, D. J., Bassi, C., MacIel, M. A. F., Moreira, L. S., & Bazzalo, M. (2004). A translocation experiment for the conservation of maned sloths, *Bradypus torquatus* (Xenarthra, Bradypodidae). *Biological Conservation*, *118*(4), 421–430. https://doi.org/10.1016/j.biocon.2003.09.019

Chuine, I. (2010). Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1555), 3149–3160. https://doi.org/10.1098/rstb.2010.0142

Costa Jordao, J., Bondioli, A. C. V., Almeida-Toledo, L. F. de, Bilo, K., Berzins, R., Le Maho, Y., Chevallier, D., & de Thoisy, B. (2017). Mixed-stock analysis in green turtles *Chelonia mydas*: mtDNA decipher current connections among west Atlantic populations. *Mitochondrial DNA Part A*, *28*(2), 197–207. https://doi.org/10.3109/19401736.2015.1115843

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, *34*(2). https://doi.org/10.18637/jss.v034.i02

Crispo, E. (2008). Modifying effects of phenotypic plasticity on interactions among natural

selection, adaptation and gene flow. *Journal of Evolutionary Biology*, *21*(6), 1460–1469. https://doi.org/10.1111/j.1420-9101.2008.01592.x

Cuevas Flores, E. A., Guzmán Hernández, V., Guerra Santos, J. J., & Rivas Hernández, G. A. (2019). *El uso del conocimiento de las tortugas marinas como herramienta para la restauración de sus poblaciones y hábitats asociados*. Universidad Autónoma del Carmen.

Davoren, G. K., Montevecchi, W. A., & Anderson, J. T. (2003). Distributional patterns of a marine bird and its prey: Habitat selection based on prey and conspecific behaviour. *Marine Ecology Progress Series*, *256*, 229–242. https://doi.org/10.3354/meps256229

Debevec, E. M. (2000). SPAM (version 3.2): statistics program for analyzing mixtures. *Journal of Heredity*, *91*(6), 509–511. https://doi.org/10.1093/jhered/91.6.509

Doligez, B., Cadet, C., Danchin, E., & Boulinier, T. (2003). When to use public information for breeding habitat selection? The role of environmental predictability and density dependence. *Animal Behaviour*, *66*(5), 973–988. https://doi.org/10.1006/anbe.2002.2270

Dutton, P. H., Jensen, M. P., Frutchey, K., Frey, A., Lacasella, E., Balazs, G. H., Cruce, J., Tagarino, A., Farman, R., & Tatarata, M. (2014). Genetic stock structure of green turtle (*Chelonia mydas*) nesting populations across the pacific Islands. *Pacific Science*, *68*(4), 451–464. https://doi.org/10.2984/68.4.1

Ehrhart, L. M., Redfoot, W. E., & Bagley, D. A. (2007). Marine turtles of the central region of the Indian River Lagoon system, Florida. *Florida Scientist*, *70*(4), 415–434.

Ehrhart, L., Redfoot, W., Bagley, D., & Mansfield, K. (2014). Long-term trends in loggerhead (*Caretta caretta*) nesting and reproductive success at an important western Atlantic rookery. *Chelonian Conservation and Biology*, *13*(2), 173–181. https://doi.org/10.2744/ccb-1100.1

Encalada, S. E., Lahanas, P. N., Bjorndal, K. A., Bolten, A. B., Miyamoto, M. M., & Bowen, B. W. (1996). Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: A mitochondrial DNA control region sequence assessment. *Molecular Ecology*, *5*(4), 473–483. https://doi.org/10.1111/j.1365-294X.1996.tb00340.x

Engelbrecht, B. M. J., Comita, L. S., Condit, R., Kursar, T. A., Tyree, M. T., Turner, B. L., & Hubbell, S. P. (2007). Drought sensitivity shapes species distribution patterns in tropical forests. *Nature*, *447*(7140), 80–82. https://doi.org/10.1038/nature05747

Engstrom, T. N., Meylan, P. A., & Meylan, A. B. (2002). Origin of juvenile loggerhead turtles (*Caretta caretta*) in a tropical developmental habitat in Caribbean Panamá. *Animal Conservation*, *5*(2), 125–133. https://doi.org/10.1017/S1367943002002184

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). fastsimcoal2 : demographic inference under complex evolutionary scenarios. *Bioinformatics*, *37*(24), 4882–4885. https://doi.org/10.1093/bioinformatics/btab468

Faircloth, B., & Glenn, T. (2016). *Preparation of an AMPure XP Substitute (AKA Serapure)*. https://doi.org/10.6079/J9MW2F26

Ferrière, R., Dieckmann, U., & Couvet, D. (2004). *Evolutionary conservation biology* (1st ed.). Cambridge University Press.

Fitak, R. R., & Johnsen, S. (2018). Green sea turtle (*Chelonia mydas*) population history indicates important demographic changes near the mid-Pleistocene transition. *Marine Biology*, *165*(7), 1–7. 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*(4), 1843–1854. https://doi.org/10.1093/genetics/147.4.1843

Florida Fish and Wildlife Conservation Commission - Fish and Wildlife Research Institute, F. W. C. F. W. R. I. (2021). *Index Nesting Beach Survey (INBS)*.

Foley, A. M., Singel, K. E., Dutton, P. H., Summers, T. M., Redlow, A. E., & Lessman, J. (2007). Characteristics of a green turtle (*Chelonia mydas*) assemblage in northwestern Florida determined during a hypothermic stunning event. *Gulf of Mexico Science*, *25*(2), 131–143. https://doi.org/10.18785/goms.2502.04

Forester, B. R., Murphy, M., Mellison, C., Petersen, J., Pilliod, D. S., Van Horne, R., Harvey, J., & Funk, W. C. (2022). Genomics-informed delineation of conservation units in a desert amphibian. *Molecular Ecology*, *31*(20), 5249–5269. https://doi.org/10.1111/mec.16660

Formia, A., Broderick, A., Glen, F., Godley, B., Hays, G., & Bruford, M. (2007). Genetic composition of the Ascension Island green turtle rookery based on mitochondrial DNA: implications for sampling and diversity. *Endangered Species Research*, *3*, 145–158. https://doi.org/10.3354/esr003145

Formia, A., Godley, B. J., Dontaine, J. F., & Bruford, M. W. (2006). Mitochondrial DNA diversity and phylogeography of endangered green turtle (*Chelonia mydas*) populations in Africa. *Conservation Genetics*, *7*(3), 353–369. https://doi.org/10.1007/s10592-005-9047-z

Frankel, O. H., & Soulé, M. E. (1981). *Conservation and evolution*. Cambridge University Press.

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140.

https://doi.org/10.1016/j.biocon.2005.05.002

Freitas, C., Olsen, E. M., Knutsen, H., Albretsen, J., & Moland, E. (2016). Temperature-associated habitat selection in a cold-water marine fish. *Journal of Animal Ecology*, *85*(3), 628–637. https://doi.org/10.1111/1365-2656.12458

Fuentes, M. M. P. B., Gredzens, C., Bateman, B. L., Boettcher, R., Ceriani, S. A., Godfrey, M. H., Helmers, D., Ingram, D. K., Kamrowski, R. L., Pate, M., Pressey, R. L., & Radeloff, V. C. (2016). Conservation hotspots for marine turtle nesting in the United States based on coastal development. *Ecological Applications*, *26*(8), 2708–2719. https://doi.org/10.1002/eap.1386

Fukuda, Y., Moritz, C., Jang, N., Webb, G., Campbell, H., Christian, K., Lindner, G., & Banks, S. (2022). Environmental resistance and habitat quality influence dispersal of the saltwater crocodile. *Molecular Ecology*, *31*(4), 1076–1092. https://doi.org/10.1111/mec.16310

Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Morey, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation Genetics*, *20*(1), 115–134. https://doi.org/10.1007/s10592-018-1096-1

Gaos, A. R., LaCasella, E. L., Kurpita, L., Balazs, G., Hargrove, S., King, C., Bernard, H., Jones, T. T., & Dutton, P. H. (2020). Hawaiian hawksbills: a distinct and isolated nesting colony in the Central North Pacific Ocean revealed by mitochondrial DNA. *Conservation Genetics*, *21*(4), 771–783. https://doi.org/10.1007/s10592-020-01287-1

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Gilpin, M. E., & Soulé, M. E. (1986). Minimum viable populations: processes of species extinction. In M. E. Soulé (Ed.), *Conservation Biology: The Science of Scarcity and Diversity* (pp. 19–34). Sinauer Associates.

Girondot, M., Mourrain, B., Chevallier, D., & Godfrey, M. H. (2021). Maturity of a giant: age and size reaction norm for sexual maturity for Atlantic leatherback turtles. *Marine Ecology*, *42*(5), 1–11. https://doi.org/10.1111/maec.12631

Godley, B. J., Richardson, S., Broderick, A. C., Coyne, M. S., Glen, F., & Hays, G. C. (2002). Long-term satellite telemetry of the movements and habitat utilisation by green turtles in the Mediterranean. *Ecography*, *25*(3), 352–362. https://doi.org/10.1034/j.1600-0587.2002.250312.x

Hancock, J. M., Vieira, S., Taraveira, L., Santos, A., Schmitt, V., Semedo, A., Patrício, A. R., Ferrand, N., Gonçalves, H., & Sequeira, F. (2019). Genetic characterization of green turtles (*Chelonia mydas*) from São Tomé and Príncipe: Insights on species recruitment and dispersal in the Gulf of Guinea. *Journal of Experimental Marine Biology and Ecology*, *518*(June), 151181. https://doi.org/10.1016/j.jembe.2019.151181

Hays, G. C. (2017). Ocean currents and marine life. *Current Biology*, *27*(11), R470–R473. https://doi.org/10.1016/j.cub.2017.01.044

Hess, J. E., Ackerman, M. W., Fryer, J. K., Hasselman, D. J., Steele, C. A., Stephenson, J. J., Whiteaker, J. M., & Narum, S. R. (2016). Differential adult migration-timing and stock-specific abundance of steelhead in mixed stock assemblages. *ICES Journal of Marine Science: Journal Du Conseil*, *73*(10), 2606–2615. https://doi.org/10.1093/icesjms/fsw138

Hoelzel, A. R., Fleischer, R. C., Campagna, C., Le Boeuf, B. J., & Alvord, G. (2002). Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology*, *15*(4), 567–575. https://doi.org/10.1046/j.1420-9101.2002.00419.x

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jenkins, D. G., & Quintana-Ascencio, P. F. (2020). A solution to minimum sample size for regressions. *PLoS ONE*, *15*(2), 1–15. https://doi.org/10.1371/journal.pone.0229345

Jensen, M. P., Allen, C. D., Eguchi, T., Bell, I. P., LaCasella, E. L., Hilton, W. A., Hof, C. A. M., & Dutton, P. H. (2018). Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the World. *Current Biology*, *28*(1), 154-159.e4. https://doi.org/10.1016/j.cub.2017.11.057

Jensen, M. P., Bell, I., Limpus, C. J., Hamann, M., Ambar, S., Whap, T., David, C., & FitzSimmons, N. N. (2016). Spatial and temporal genetic variation among size classes of green turtles (*Chelonia mydas*) provides information on oceanic dispersal and population dynamics. *Marine Ecology Progress Series*, *543*, 241–256. https://doi.org/10.3354/meps11521

Jensen, M. P., Dalleau, M., Gaspar, P., Lalire, M., Jean, C., Ciccione, S., Mortimer, J. A., Quillard, M., Taquet, C., Wamukota, A., Leroux, G., & Bourjea, J. (2020). Seascape genetics and the spatial ecology of juvenile green turtles. *Genes*, *11*(3). https://doi.org/10.3390/genes11030278

Jensen, M. P., FitzSimmons, N. N., Bourjea, J., Hamabata, T., Reece, J., & Dutton, P. H. (2019). The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). *Journal of Biogeography*, *46*(5), 860–870. https://doi.org/10.1111/jbi.13483

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Karl, S. A., Bowen, B. W., & Avise, J. C. (1992). Global population genetic structure and male-mediated gene flow in the green turtle (Chelonia mydas): RFLP analyses of anonymous nuclear loci. *Genetics*, *131*(1), 163–173.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper,

A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kruschke, J. K. (2015). *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press. https://doi.org/10.1016/B978-0-12-405888-0.09999-2

Lahanas, P. N., Bjorndal, K. A., Bolten, A. B., Encalada, S. E., Miyamoto, M. M., Valverde, R. A., & Bowen, B. W. (1998). Genetic composition of a green turtle (*Chelonia mydas*) feeding ground population: Evidence for multiple origins. *Marine Biology*, *130*(3), 345–352. https://doi.org/10.1007/s002270050254

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laurent, L., Casale, P., Bradai, M. N., Godley, B. J., Gerosa, G., Broderick, A. C., Schroth, W., Schierwater, B., Levy, A. M., Freggi, D., Abd El-Mawla, E. M., Hadoud, D. A., Gomati, H. E., Domingo, M., Hadjichristophorou, M., Kornaraky, L., Demirayak, F., & Gautier, C. (1998). Molecular resolution of marine turtle stock composition in fishery bycatch: A case study in the Mediterranean. *Molecular Ecology*, *7*(11), 1529–1542. https://doi.org/10.1046/j.1365-294x.1998.00471.x

Leach, K., Montgomery, W. I., & Reid, N. (2016). Modelling the influence of biotic factors on species distribution patterns. *Ecological Modelling*, *337*, 96–106. https://doi.org/10.1016/j.ecolmodel.2016.06.008

Leigh, J. W., & Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lewison, R. L., Crowder, L. B., Wallace, B. P., Moore, J. E., Cox, T., Zydelis, R., McDonald, S., DiMatteo, A., Dunn, D. C., Kot, C. Y., Bjorkland, R., Kelez, S., Soykan, C., Stewart, K. R., Sims, M., Boustany, A., Read, A. J., Halpin, P., Nichols, W. J., & Safina, C. (2014). Global patterns of marine mammal, seabird, and sea turtle bycatch reveal taxa-specific and cumulative megafauna hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), 5271–5276. https://doi.org/10.1073/pnas.1318960111

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Long, C. A., Chabot, R. M., El-Khazen, M. N., Kelley, J. R., Mollet-Saint Benoît, C., & Mansfield, K. L. (2021). Incongruent long-term trends of a marine consumer and primary producers in a habitat affected by nutrient pollution. *Ecosphere*, *12*(6). https://doi.org/10.1002/ecs2.3553

Lovette, I. J., Clegg, S. M., & Smith, T. B. (2004). Limited utility of mtDNA markers for determining connectivity among breeding and overwintering locations in three neotropical migrant birds. *Conservation Biology*, *18*(1), 156–166. https://doi.org/10.1111/j.1523-1739.2004.00239.x

Luke, K., Horrocks, J. A., LeRoux, R. A., & Dutton, P. H. (2004). Origins of green turtle (*Chelonia mydas*) feeding aggregations around Barbados, West Indies. *Marine Biology*, *144*(4), 799–805. https://doi.org/10.1007/s00227-003-1241-2

Lundvall, D., Svanbäck, R., Persson, L., & Byström, P. (1999). Size-dependent predation in piscivores: Interactions between predator foraging and prey avoidance abilities. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(7), 1285–1292. https://doi.org/10.1139/f99-058

MacPherson, E. (1998). Ontogenetic shifts in habitat use and aggregation in juvenile sparid fishes. *Journal of Experimental Marine Biology and Ecology*, *220*(1), 127–150. https://doi.org/10.1016/S0022-0981(97)00086-5

Maffucci, F., Kooistra, W. H. C. F., & Bentivegna, F. (2006). Natal origin of loggerhead turtles, Caretta caretta, in the neritic habitat off the Italian coasts, Central Mediterranean. *Biological Conservation*, *127*(2), 183–189. https://doi.org/10.1016/j.biocon.2005.08.009

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Mansfield, K. L., Mendilaharsu, M. L., Putman, N. F., dei Marcovaldi, M. A. G., Sacco, A. E., Lopez, G., Pires, T., & Swimmer, Y. (2017). First satellite tracks of South Atlantic sea turtle 'lost years': seasonal variation in trans-equatorial movement. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1868), 20171730. https://doi.org/10.1098/rspb.2017.1730

Mansfield, K. L., & Putman, N. F. (2013). Oceanic habits and habitats. In *Biology of Sea Turtles* (3rd ed., pp. 189–211). CRC Press.

Mansfield, K. L., Wyneken, J., & Luo, J. (2021). First Atlantic satellite tracks of "lost years" green turtles support the importance of the Sargasso Sea as a sea turtle nursery. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1950). https://doi.org/10.1098/rspb.2021.0057

Mansfield, K. L., Wyneken, J., Porter, W. P., Luo, J., & B, P. R. S. (2014). First satellite tracks of neonate sea turtles redefine the ' lost years ' oceanic niche First satellite tracks of neonate sea turtles redefine the ' lost years ' oceanic nicheError: JavaScript error found: CSL error:

Exception: TypeError: "undefined" is n. *Proceedings of the Royal Society*, *281*(March), 1–9. https://doi.org/10.1098/rspb.2013.3039

Marcovaldi, M. A., Lopez, G. G., Soares, L. S., Lima, E. H. S. M., Thomé, J. C. A., & Almeida, A. P. (2010). Satellite-tracking of female loggerhead turtles highlights fidelity behavior in northeastern Brazil. *Endangered Species Research*, *12*(3), 263–272. https://doi.org/10.3354/esr00308

Marcovaldi, M., & Chaloupka, M. (2007). Conservation status of the loggerhead sea turtle in Brazil: an encouraging outlook. *Endangered Species Research*, *3*, 133–143. https://doi.org/10.3354/esr003133

Martins, M., Sawaya, R. J., & Marques, O. A. V. (2008). A first estimate of the population size of the critically endangered lancehead, *Bothrops insularis*. *South American Journal of Herpetology*, *3*(2), 168–174. https://doi.org/10.2994/1808-9798

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Borges, C. H. S., Lira, L. V. G., Mendes, N. J., Agudelo, J. F. G., Cáceres, P., Berrocal, M. H. M., Sucerquia, G. A. L., Porto-Foresti, F., Yáñez, J. M., & Hashimoto, D. T. (2021). Development of a multi-species SNP array for serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*. *Scientific Reports*, *11*(1), 19289. https://doi.org/10.1038/s41598-021-98885-x

Mazaris, A. D., Schofield, G., Gkazinou, C., Almpanidou, V., & Hays, G. C. (2017). Global sea turtle conservation successes. *Science Advances*, *3*(9). https://doi.org/10.1126/sciadv.1600730

Medeiros, L., Monteiro, D. S., Botta, S., Proietti, M. C., & Secchi, E. R. (2019). Origin and foraging ecology of male loggerhead sea turtles from southern Brazil revealed by genetic and stable isotope analysis. *Marine Biology*, *166*(6), 1–14. https://doi.org/10.1007/s00227-019-3524-2

Michelot, C., Pinaud, D., Fortin, M., Maes, P., Callard, B., Leicher, M., & Barbraud, C. (2017). Seasonal variation in coastal marine habitat use by the European shag: Insights from fine scale habitat selection modeling and diet. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *141*(April), 224–236. https://doi.org/10.1016/j.dsr2.2017.04.001

Millán-Aguilar, O. (2009). *Estructura genética poblacional de la tortuga verde,* Chelonia mydas, *en el golfo de México determinada por análisis de sequencias del ADN mitocondrial*. Universidad Nacional Autónoma de México, Mazatlán.

Miller, E. A., McClenachan, L., Uni, Y., Phocas, G., Hagemann, M. E., & Van Houtan, K. S. (2019). The historical development of complex global trafficking networks for marine wildlife. *Science Advances*, *5*(3). https://doi.org/10.1126/sciadv.aav5948

Montero, N., dei Marcovaldi, M. A. G., Lopez–Mendilaharsu, M., Santos, A. S., Santos, A. J. B., & Fuentes, M. M. P. B. (2018). Warmer and wetter conditions will reduce offspring

production of hawksbill turtles in Brazil under climate change. *PLoS ONE*, *13*(11), 1–16. https://doi.org/10.1371/journal.pone.0204188

Monzón-Argüello, C., López-Jurado, L. F., Rico, C., Marco, A., López, P., Hays, G. C., & Lee, P. L. M. (2010). Evidence from genetic and Lagrangian drifter data for transatlantic transport of small juvenile green turtles. *Journal of Biogeography*, *37*(9), 1752–1766. https://doi.org/10.1111/j.1365-2699.2010.02326.x

Mouritsen, H., & Hore, P. J. (2012). The magnetic retina: Light-dependent and trigeminal magnetoreception in migratory birds. *Current Opinion in Neurobiology*, *22*(2), 343–352. https://doi.org/10.1016/j.conb.2012.01.005

Nalovic, M. A., Ceriani, S. A., Fuentes, M. M. P. B., Pfaller, J. B., Wildermann, N. E., & Cuevas, E. (2020). *Sea turtles in the North Atlantic & wider Caribbean Region*. IUCN-SSC Marine Turtle Specialist Group.

Naro-Maciel, E., Bondioli, A. C. V., Martin, M., De Pádua Almeida, A., Baptistotte, C., Bellini, C., Marcovaldi, M. Â., Santos, A. J. B., & Amato, G. (2012). The interplay of homing and dispersal in green turtles: A focus on the southwestern atlantic. *Journal of Heredity*, *103*(6), 792–805. https://doi.org/10.1093/jhered/ess068

Naro-Maciel, E., Gaughran, S. J., Putman, N. F., Amato, G., Arengo, F., Dutton, P. H., McFadden, K. W., Vintinner, E. C., & Sterling, E. J. (2014). Predicting connectivity of green turtles at Palmyra Atoll, central Pacific: A focus on mtDNA and dispersal modelling. *Journal of the Royal Society Interface*, *11*(93). https://doi.org/10.1098/rsif.2013.0888

Naro-Maciel, E., Hart, K. M., Cruciata, R., & Putman, N. F. (2017). DNA and dispersal models highlight constrained connectivity in a migratory marine megavertebrate. *Ecography*, *40*(5), 586–597. https://doi.org/10.1111/ecog.02056

Naro-Maciel, E., Reid, B. N., Alter, S. E., Amato, G., Bjorndal, K. A., Bolten, A. B., Martin, M., Nairn, C. J., Shamblin, B., & Pineda-Catalan, O. (2014). From refugia to rookeries: Phylogeography of Atlantic green turtles. *Journal of Experimental Marine Biology and Ecology*, *461*, 306–316. https://doi.org/10.1016/j.jembe.2014.08.020

Neaves, P. I., Wallace, C. G., Candy, J. R., & Beacham, T. D. (2005). *CBayes: Computer program for mixed stock analysis of allelic data. Free program distributed by the authors over the internet.* (3.0).

Nielsen, E. S., Beger, M., Henriques, R., & von der Heyden, S. (2020). A comparison of genetic and genomic approaches to represent evolutionary potential in conservation planning. *Biological Conservation*, *251*(August), 108770. https://doi.org/10.1016/j.biocon.2020.108770

Nishizawa, H., Naito, Y., Suganuma, H., Abe, O., Okuyama, J., Hirate, K., Tanaka, S., Inoguchi, E., Narushima, K., Kobayashi, K., Ishii, H., Tanizaki, S., Kobayashi, M., Goto, A., & Arai,

N. (2013). Composition of green turtle feeding aggregations along the Japanese archipelago: Implications for changes in composition with current flow. *Marine Biology*, *160*(10), 2671–2685. https://doi.org/10.1007/s00227-013-2261-1

O'Leary, S. J., Dunton, K. J., King, T. L., Frisk, M. G., & Chapman, D. D. (2014). Genetic diversity and effective size of Atlantic sturgeon, *Acipenser oxyrhinchus oxyrhinchus* river spawning populations estimated from the microsatellite genotypes of marine-captured juveniles. *Conservation Genetics*, *15*(5), 1173–1181. https://doi.org/10.1007/s10592-014-0609-9

Okuyama, T., & Bolker, B. M. (2005). Combining genetic and ecological data to estimate sea turtle origins. *Ecological Applications*, *15*(1), 315–325. https://doi.org/10.1890/03-5063

Patrício, A. R., Formia, A., Barbosa, C., Broderick, A. C., Bruford, M., Carreras, C., Catry, P., Ciofi, C., Regalla, A., & Godley, B. J. (2017). Dispersal of green turtles from Africa's largest rookery assessed through genetic markers. *Marine Ecology Progress Series*, *569*, 215–225. https://doi.org/10.3354/meps12078

Paxton, K. L., Yau, M., Moore, F. R., & Irwin, D. E. (2013). Differential migratory timing of western populations of Wilson's Warbler (*Cardellina pusilla*) revealed by mitochondrial DNA and stable isotopes. *The Auk*, *130*(4), 689–698. https://doi.org/10.1525/auk.2013.13107

Pella, J., & Masuda, M. (2001). Bayesian methods for analysis of stock mixtures from genetic characters. *Fishery Bulletin*, *99*(1), 151–167.

Perrault, J. R., Miller, D. L., Eads, E., Johnson, C., Merrill, A., Thompson, L. J., & Wyneken, J. (2012). Maternal health status correlates with nest success of leatherback sea turtles (*Dermochelys coriacea*) from Florida. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0031841

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Phillips, K. F., Martin, K. R., Stahelin, G. D., Savage, A. E., & Mansfield, K. L. (2022). Genetic variation among sea turtle life stages and species suggests connectivity among ocean basins. *Ecology and Evolution*, *12*(11), 1–26. https://doi.org/10.1002/ece3.9426

Phillips, K. F., Stahelin, G. D., Chabot, R. M., & Mansfield, K. L. (2021). Long-term trends in marine turtle size at maturity at an important Atlantic rookery. *Ecosphere*, *12*(7). https://doi.org/10.1002/ecs2.3631

Pineda, O. G., & Rocha, A. R. B. (2016). *Las tortugas marinas en México: logros y perspectivas para su conservación* (1st ed.). CONANP.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11. http://oro.open.ac.uk/22547/

Prakash, S. S., Lal, M. M., Dutton, P. H., Rico, C., & Piovano, S. (2022). Kinship genomics approach to study mating systems in a depleted sea turtle rookery. *Regional Studies in Marine Science*, *51*, 102174. https://doi.org/10.1016/j.rsma.2022.102174

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Proietti, M. C., Reisser, J. W., Kinas, P. G., Kerr, R., Monteiro, D. S., Marins, L. F., & Secchi, E. R. (2012). Green turtle *Chelonia mydas* mixed stocks in the western South Atlantic, as revealed by mtDNA haplotypes and drifter trajectories. *Marine Ecology Progress Series*, *447*(1), 195–209. https://doi.org/10.3354/meps09477

Proietti, Maira C., Reisser, J., Marins, L. F., Marcovaldi, M. A., Soares, L. S., Monteiro, D. S., Wijeratne, S., Pattiaratchi, C., & Secchi, E. R. (2014). Hawksbill × loggerhead sea turtle hybrids at Bahia, Brazil: Where do their offspring go? *PeerJ*, *2014*(1), 1–14. https://doi.org/10.7717/peerj.255

Prosdocimi, L., González Carman, V., Albareda, D. A., & Remis, M. I. (2012). Genetic composition of green turtle feeding grounds in coastal waters of Argentina based on mitochondrial DNA. *Journal of Experimental Marine Biology and Ecology*, *412*, 37–45. https://doi.org/10.1016/j.jembe.2011.10.015

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool aet for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Putman, N. F., & Mansfield, K. L. (2015). Direct evidence of swimming demonstrates active dispersal in the sea turtle "lost years." *Current Biology*, *25*(9), 1221–1227. https://doi.org/10.1016/j.cub.2015.03.014

Putman, N. F., & Naro-Maciel, E. (2013). Finding the "lost years" in green turtles: Insights from ocean circulation models and genetic analysis. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1768). https://doi.org/10.1098/rspb.2013.1468

Putman, N. F., Seney, E. E., Verley, P., Shaver, D. J., López-Castro, M. C., Cook, M., Guzmán, V., Brost, B., Ceriani, S. A., Mirón, R. de J. G. D., Peña, L. J., Tzeek, M., Valverde, R. A., Cantón, C. C. G., Howell, L., Ravell Ley, J. A., Tumlin, M. C., Teas, W. G., Caillouet, C. W., … Mansfield, K. L. (2020). Predicted distributions and abundances of the sea turtle 'lost years' in the western North Atlantic Ocean. *Ecography*, *43*(4), 506–517. https://doi.org/10.1111/ecog.04929

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*.

Redfoot, W., & Ehrhart, L. (2013). Trends in size class distribution, recaptures, and abundance of

juvenile green turtles (*Chelonia mydas*) utilizing a rock riprap lined embayment at Port Canaveral, Florida, USA, as developmental habitat. *Chelonian Conservation and Biology*, *12*(2), 252–261. https://doi.org/10.2744/CCB-0952.1

Richardson, W. J. (1990). Wind and orientation of migrating birds: a review. *Experimentia*, *46*, 416–425. https://doi.org/10.1007/978-3-0348-7208-9_11

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Roberts, A., Silverman, E., & Gifford, S. (2018). Sample size considerations for satellite telemetry and animal distributions. *Journal of Wildlife Management*, *82*(7), 1536–1544. https://doi.org/10.1002/jwmg.21504

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*(4), 1857–1870. https://doi.org/10.1534/genetics.166.4.1857

Roden, S. E., Morin, P. A., Frey, A., Balazs, G. H., Zarate, P., Cheng, I. J., & Dutton, P. H. (2013). Green turtle population structure in the Pacific: New insights from single nucleotide polymorphisms and microsatellites. *Endangered Species Research*, *20*(3), 227–234. https://doi.org/10.3354/esr00500

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Roland, J., Keyghobadi, N., & Fownes, S. (2000). Alpine *Parnassius* butterfly dispersal: Effects of landscape and population size. *Ecology*, *81*(6), 1642–1653. https://doi.org/10.2307/177313

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

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

Russell, E. M., Yom-Tov, Y., & Geffen, E. (2004). Extended parental care and delayed dispersal: northern, tropical, and southern passerines compared. *Behavioral Ecology*, *15*(5), 831–838. https://doi.org/10.1093/beheco/arh088

Ryser-Degiorgis, M. P. (2013). Wildlife health investigations: Needs, challenges and recommendations. *BMC Veterinary Research*, *9*. https://doi.org/10.1186/1746-6148-9-223

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., & Hanski, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, *392*(6675), 491–494. https://doi.org/10.1038/33136

Santos, R. G., Pinheiro, H. T., Martins, A. S., Riul, P., Bruno, S. C., Janzen, F. J., & Ioannou, C. C. (2016). The anti-predator role of within-nest emergence synchrony in sea turtle hatchlings. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1834), 20160697. https://doi.org/10.1098/rspb.2016.0697

Seminoff, J. A. (2004). Chelonia mydas *The IUCN Red List of Threatened Species.* https://doi.org/https://doi.org/10.2305/IUCN.UK.2004.RLTS. T4615A11037468.en

Seminoff, J. A., Allen, C. D., Balazs, G. H., Dutton, P. H., Eguchi, T., Haas, H. L., Hargrove, S. A., Jensen, M., Klemm, D. L., Lauritsen, A. M., Macpherson, S. L., Opay, P., Possardt, E. E., Pultz, S., Seney, E. E., Van Houtan, K. S., & Waples, R. S. (2015). *Status review of the green turtle (*Chelonia mydas*) under the endangered species act*. NOAA Technical Memorandum, NOAA- NMFS-SWFSC.

Seminoff, J. A., Benson, S. R., Arthur, K. E., Eguchi, T., Dutton, P. H., Tapilatu, R. F., & Popp, B. N. (2012). Stable Isotope Tracking of Endangered Sea Turtles: Validation with Satellite Telemetry and δ15N Analysis of Amino Acids. *PLoS ONE*, *7*(5), e37403. https://doi.org/10.1371/journal.pone.0037403

Seminoff, J. A., Resendiz, A., & Nichols, W. J. (2002). Home range of green turtles Chelonia mydas at a coastal foraging area in the Gulf of California, Mexico. *Marine Ecology Progress Series*, *242*, 253–265. https://doi.org/10.3354/meps242253

Shamblin, B. M., Bagley, D. A., Ehrhart, L. M., Desjardin, N. A., Martin, R. E., Hart, K. M., Naro-Maciel, E., Rusenko, K., Stiner, J. C., Sobel, D., Johnson, C., Wilmers, T. J., Wright, L. J., & Nairn, C. J. (2015). Genetic structure of Florida green turtle rookeries as indicated by mitochondrial DNA control region sequences. *Conservation Genetics*, *16*(3), 673–685. https://doi.org/10.1007/s10592-014-0692-y

Shamblin, B. M., Bjorndal, K. A., Bolten, A. B., Hillis-Starr, Z. M., Lundgren, I., Naro-Maciel, E., & Nairn, C. J. (2012). Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology*, *21*(10), 2330–2340. https://doi.org/10.1111/j.1365-294X.2012.05530.x

Shamblin, B. M., Bolten, A. B., Abreu-Grobois, F. A., Bjorndal, K. A., Cardona, L., Carreras, C., Clusa, M., Monzón-Argüello, C., Nairn, C. J., Nielsen, J. T., Nel, R., Soares, L. S., Stewart, K. R., Vilaça, S. T., Türkozan, O., Yilmaz, C., & Dutton, P. H. (2014). Geographic patterns of genetic variation in a broadly distributed marine vertebrate: New insights into loggerhead turtle stock structure from expanded mitochondrial DNA sequences. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0085956

Shamblin, B. M., Dutton, P. H., Bjorndal, K. A., Bolten, A. B., Naro-Maciel, E., Santos, A. J. B.,

Bellini, C., Baptistotte, C., Marcovaldi, M. Â., & Nairn, C. J. (2015). Deeper mitochondrial sequencing reveals cryptic diversity and structure in Brazilian green turtle rookeries. *Chelonian Conservation and Biology*, *14*(2), 167–172. https://doi.org/10.2744/CCB-1152.1

Shamblin, B. M., Dutton, P. H., Shaver, D. J., Bagley, D. A., Putman, N. F., Mansfield, K. L., Ehrhart, L. M., Peña, L. J., & Nairn, C. J. (2017). Mexican origins for the Texas green turtle foraging aggregation: A cautionary tale of incomplete baselines and poor marker resolution. *Journal of Experimental Marine Biology and Ecology*, *488*, 111–120. https://doi.org/10.1016/j.jembe.2016.11.009

Shamblin, B. M., Hart, K. M., Lamont, M. M., Shaver, D. J., Dutton, P. H., LaCasella, E. L., & Nairn, C. J. (2023). United States Gulf of Mexico Waters provide important nursery habitat for Mexico's Green turtle nesting populations. *Frontiers in Marine Science*, *9*(January), 1–14. https://doi.org/10.3389/fmars.2022.1035834

Shamblin, B. M., Witherington, B. E., Hirama, S., Hardy, R. F., & Nairn, C. J. (2018). Mixed stock analyses indicate population-scale connectivity effects of active dispersal by surface-pelagic green turtles. *Marine Ecology Progress Series*, *601*, 215–226. https://doi.org/10.3354/meps12693

Shaw, R. C., Greggor, A. L., & Plotnik, J. M. (2021). The Challenges of Replicating Research on Endangered Species. *Animal Behavior and Cognition*, *8*(2), 240–246. https://doi.org/10.26451/abc.08.02.10.2021

Sovic, M. G., Fries, A. C., & Gibbs, H. L. (2016). Origin of a cryptic lineage in a threatened reptile through isolation and historical hybridization. *Heredity*, *117*(5), 358–366. https://doi.org/10.1038/hdy.2016.56

Sovic, M. G., Fries, A., Martin, S. A., & Lisle Gibbs, H. (2019). Genetic signatures of small effective population sizes and demographic declines in an endangered rattlesnake, Sistrurus catenatus. *Evolutionary Applications*, *12*(4), 664–678. https://doi.org/10.1111/eva.12731

Stahelin, G. D., Hoffman, E. A., Quintana-Ascencio, P. F., Reusche, M., & Mansfield, K. L. (2022). Incorporating distance metrics and temporal trends to refine mixed stock analysis. *Scientific Reports*, *12*(1), 20569. https://doi.org/10.1038/s41598-022-24279-2

Stoneburner, D. L. (1982). Satellite telemetry of loggerhead sea turtle movement in the Georgia Bight. *Copeia*, *1982*(2), 400. https://doi.org/10.2307/1444621

Storfer, A. (1996). Quantitative genetics: A promising approach for the assessment of genetic variation in endangered species. *Trends in Ecology and Evolution*, *11*(8), 343–348. https://doi.org/10.1016/0169-5347(96)20051-5

Talley, L. D., Pickard, G. L., Emery, W. J., & Swift, J. H. (2011). *Descriptive physical oceanography: an introduction* (6th ed.). Academic press.

Turner Tomaszewicz, C. N., Avens, L., Seminoff, J. A., Limpus, C. J., FitzSimmons, N. N.,

Guinea, M. L., Pendoley, K. L., Whittock, P. A., Vitenbergs, A., Whiting, S. D., & Tucker, A. D. (2022). Age-specific growth and maturity estimates for the flatback sea turtle (*Natator depressus*) by skeletochronology. *PLOS ONE*, *17*(7), e0271048. https://doi.org/10.1371/journal.pone.0271048

van der Zee, J. P., Christianen, M. J. A. A., Bérubé, M., Nava, M., Schut, K., Humber, F., Alfaro-Núñez, A., Becking, L. E., & Palsbøll, P. J. (2021). The population genomic structure of green turtles (*Chelonia mydas*) suggests a warm-water corridor for tropical marine fauna between the Atlantic and Indian oceans during the last interglacial. *Heredity*, *127*(June), 510–521. https://doi.org/10.1038/s41437-021-00475-0

van der Zee, J. P., Christianen, M. J. A., Nava, M., Velez-Zuazo, X., Hao, W., Bérubé, M., van Lavieren, H., Hiwat, M., Berzins, R., Chevalier, J., Chevallier, D., Lankester, M. C., Bjorndal, K. A., Bolten, A. B., Becking, L. E., & Palsbøll, P. J. (2019). Population recovery changes population composition at a major southern Caribbean juvenile developmental habitat for the green turtle, *Chelonia mydas*. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-50753-5

Vanschoenwinkel, B., De Vries, C., Seaman, M., & Brendonck, L. (2007). The role of metacommunity processes in shaping invertebrate rock pool communities along a dispersal gradient. *Oikos*, *116*(8), 1255–1266. https://doi.org/10.1111/j.2007.0030-1299.15860.x

Varela, R. G., Quílez, G. Z., & Harrison, E. (2015). *Report on the 2014 green turtle program at Tortuguero, Costa Rica*.

Vilaça, S. T., Maroso, F., Lara, P., de Thoisy, B., Chevallier, D., Arantes, L. S., Santos, F. R., Bertorelle, G., & Mazzoni, C. J. (2023). Evidence of backcross inviability and mitochondrial DNA paternal leakage in sea turtle hybrids. *Molecular Ecology*, *32*(3), 628–643. https://doi.org/10.1111/mec.16773

Vilaça, S. T., Piccinno, R., Rota-Stabelli, O., Gabrielli, M., Benazzo, A., Matschiner, M., Soares, L. S., Bolten, A. B., Bjorndal, K. A., & Bertorelle, G. (2021). Divergence and hybridization in sea turtles: Inferences from genome data show evidence of ancient gene flow between species. *Molecular Ecology*, *30*(23), 6178–6192. https://doi.org/10.1111/mec.16113

Volk, D. R., Konvalina, J. D., Floeter, S. R., Ferreira, C. E. L. L., & Hoffman, E. A. (2021). Going against the flow: Barriers to gene flow impact patterns of connectivity in cryptic coral reef gobies throughout the western Atlantic. *Journal of Biogeography*, *48*(2), 427–439. https://doi.org/10.1111/jbi.14010

Wallace, B. P., DiMatteo, A. D., Hurley, B. J., Finkbeiner, E. M., Bolten, A. B., Chaloupka, M. Y., Hutchinson, B. J., Alberto Abreu-Grobois, F., Amorocho, D., Bjorndal, K. A., Bourjea, J., Bowen, B. W., Dueñas, R. B., Casale, P., Choudhury, B. C., Costa, A., Dutton, P. H., Fallabrino, A., Girard, A., … Mast, R. B. (2010). Regional management units for marine turtles: A novel framework for prioritizing conservation and research across multiple scales. *PLoS ONE*, *5*(12), 1–11. https://doi.org/10.1371/journal.pone.0015465

Wallace, B. P., Kot, C. Y., DiMatteo, A. D., Lee, T., Crowder, L. B., & Lewison, R. L. (2013). Impacts of fisheries bycatch on marine turtle populations worldwide: toward conservation and research priorities. *Ecosphere*, *4*(3), art40. https://doi.org/10.1890/ES12-00388.1

Wiklund, C., & Kaitala, A. (1995). Sexual selection for large male size in a polyandrous butterfly: The effect of body size on male versus female reproductive success in *Pieris napi*. *Behavioral Ecology*, *6*(1), 6–13. https://doi.org/10.1093/beheco/6.1.6

Witherington, B., Hirama, S., & Hardy, R. (2012). Young sea turtles of the pelagic Sargassum-dominated drift community: Habitat use, population density, and threats. *Marine Ecology Progress Series*, *463*, 1–22. https://doi.org/10.3354/meps09970

Witmer, G. W. (2005). Wildlife population monitoring: some practical considerations. *Wildlife Research*, *32*(3), 259. https://doi.org/10.1071/WR04003

Wright, S. (1978). *Evolution and the genetics of populations, volume 4: variability within and among natural populations.* University of Chicago Press.

Young, L. C. (2010). Inferring colonization history and dispersal patterns of a long-lived seabird by combining genetic and empirical data. *Journal of Zoology*, *281*(4), 232–240. https://doi.org/10.1111/j.1469-7998.2010.00698.x

Zbinden, J. A., Bearhop, S., Bradshaw, P., Gill, B., Margaritoulis, D., Newton, J., & Godley, B. J. (2011). Migratory dichotomy and associated phenotypic variation in marine turtles revealed by satellite tracking and stable isotope analysis. *Marine Ecology Progress Series*, *421*, 291–302. https://doi.org/10.3354/meps08871