UNIVERSITAT DE BARCELONA

Sea turtle conservation: genetics and genomics for a better management

Anna Barbanti

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento 4.0. España de Creative Commons.***

This doctoral thesis is licensed under the ***Creative Commons Attribution 4.0. Spain License.***

SEA TURTLE CONSERVATION:

GENETICS AND GENOMICS FOR A BETTER MANAGEMENT

ANNA BARBANTI

UNIVERSITAT DE BARCELONA

Doctoral Thesis

University of Barcelona

Title:

Sea turtle conservation: genetics and genomics for a better management

Thesis presented by
Anna Barbanti

To apply for the title of
Doctor of the University of Barcelona
Genetics Program
Department of Genetics, Microbiology and Statistics
Evolutionary Genetics Group

Marta Pascual Berniola

Supervisor/Tutor

Carlos Carreras Huergo

Supervisor

Anna Barbanti

Barcelona, July 2020

The sea is everything.
It covers seven-tenths of the terrestrial globe.
Its breath is pure and life-giving. It is an immense
desert place where man is never lonely […].
The ocean is the vast reservoir of Nature.

Twenty Thousand Leagues under the Sea
Jules Verne

ACKNOWLEDGEMENTS

First of all, I would like to thank my family and in particular my parents. They are the main reason why I am the person I am today, they always supported me and helped me with my decisions. Without my parents I would have never been able to learn as much as I did, to study in so many different countries and to do all this completely stress free. In the same way, I am grateful to my brother and sister who also have always supported my choices, and as we grow older have become more and more a fundamental presence in my life.

Second, I would like to thank my supervisors, Marta Pascual and Carlos Carreras, or as I would call them many times, my academic parents. You taught me literary everything I know about scientific research. You took me in when I didn't know the first thing about genetics (apart from Mendel) and had very little idea of what science was really all about. You have trained me to be the most efficient I could possibly be, taught me not only science but also time and workload management. You taught me to take on responsibilities and to be independent. And most importantly, you were always there for me, to help, to support, to listen, to explain all the things I didn't understand, making me feel involved and safe in our environment. All this always made me feel grateful, happy and most of lucky for the choice I made four years ago of coming back to your lab in Barcelona. The problem now is that you set my standard very high, and I don't think I will be able to settle for anything less than what I had in the group of Evolutionary Genetics of Barcelona.

I would like to thank Luca Bargelloni and Rafaela Franch for hosting me at the University of Padova and teaching me about new generation sequencing, fortunately in Italian, which made everything much easier for me.
I also want to thank Cynthia Riginos for hosting me at the University of Queensland in Brisbane (Australia) teaching me new population genetic analyses and how to look at things from a different perspective. I also want to thank Cynthia and her crew for welcoming me in their group and making me feel part of it, and for helping me during the global issues caused by the pandemic.

I want to thank Roser for being the best friend one could ever want, always being there with me and for me since I first moved to Barcelona six years ago. We shared scientific troubles, rugby tournaments at 35°C, that little house with the weird door and much more. You have been a key element for my mental sanity ever since I met you and I love you like a sister.

I thank the crew from the Evolutionary Genetics lab! I want to mention Victor (VOS) for all your help in the lab since I was a newbie, the infinite discussions about non-important things and the post-lunch cat videos. Victor (VOM), Cris, Silvia and Carla, for all the fun we had together at the beginning of our scientific history and the friendship we still carry along with us. Bea, for briefly being my lab companion and then becoming my movie mate. Carles for all the help you gave me while I was away, I am deeply grateful, and for making it so that for the last 6 months of my research I was not an only child anymore.

I want to thank Hector, who since we met has always been there to help me with my worst enemy (bioinformatics) and to share a laugh even after 5 hours of lab meeting. I have relied on you for

the past three years and you have never said no to any of my desperate requests. You have been a great friend to me and I hope one day we will be able to work together again.

I want to thank Alba for talking to me on our first day of class in 2014 and then deciding we didn't need any other friends. And then for resuming our friendship as colleagues, as if we never stopped sharing that class bench and notes (mostly hers). I am grateful for our friendship, mostly because you understand all the parts of my personality and because your soul is Italian.

I want to mention Alysha, we started hanging out together only because I could communicate with you in English and look at us now! Thank you for the love and support you show me all the time and being part of my (almost) everyday life. Thank you for caring and be part of my adopted family.

Angela and Mariona. You were my flatmates for the past three years, but we shared much more than just a living room and a kitchen. You were like my family, and for that I am grateful.
Impossible not to mention my team. All the girls from the rugby that have come and gone over the past six years, and in particular my hermanas from this very season. Girls, I am so grateful to have shared my greatest passion with you and thank you for making it my safe space.

Although they were not present in person, I would like to mention some people that were very important to me while I was living abroad and in many different ways contributed to my personal growth. So, I want to mention my Italian family of Casa Garibaldi (Bologna), the people who lived there and the people who would just show up for a family lunch on Sundays. Michi, Chiara, Candy, Mati, and Sofu. You guys have been a milestone in my life and although it took a global crisis to make us use technology to see each other, I love you.
Finally, I would like to mention Kate who seven years ago helped me find the courage to start my journey and has borne with me and supported me throughout the closure of this chapter.

ABSTRACT

Conservation actions aim to preserve and recover animal and plant species using *in-situ* or *ex-situ* strategies. The first, aims to protect and sustain populations in their natural habitat, the second are implemented when local populations are extinct or are about to be. Conservation genetics can provide important insights into the dynamics of endangered populations facilitating their management.

This thesis uses traditional markers and new generation sequencing to improve conservation management of the loggerhead (*Caretta caretta*) and the green turtle (*Chelonia mydas*). In the first 2 chapters we used microsatellites and mtDNA to assess the outcome of a reintroduction program of green turtles in the Cayman Islands and the status of the reintroduced wild population. We found that 90% of adult wild females and 80% of wild F1 hatchlings were related to the captive population, proving the program successful. This relatedness did affect negatively the fitness of the wild population. Moreover, we found that after only one generation, genetic differentiation between the populations was significant. Our results suggest that assisted colonisation is a viable solution to the global decline of biodiversity. The third chapter explores the potential of 2b-RAD methodology in the field of non-model species population genomics and provides guidelines to optimise protocol and decision making using 2b-RAD. We discovered that, given the big genome size of the loggerhead turtle, a selective-base ligation should be used to obtain an overall depth of coverage of 20x and make the study cost-effective. The fourth chapter studies the population structure and local adaptation of 9 Eastern Mediterranean loggerhead turtle rookeries using 2b-RAD genomic sequencing. We found a high level of population structure and no overlapping among rookeries. Bayesian clustering indicated our individuals to be grouped in nine genetic clusters, which correspond to the distribution found in the PCoA. We found that atmospheric temperature and geographic location of the rookery have a significant impact on population structure, as outlier loci were found associated with these predictors. These results aim to use fine scale genetic information of the Eastern Mediterranean Sea to inform and improve conservation management of loggerhead turtle rookeries.

Sea turtle conservation:
genetics and genomics for a better management

INDEX

Introduction 1
Threats to sea turtle conservation 1
Sea turtle conservation strategy 5
In-situ conservation 7
Ex-situ conservation 8
Science and conservation 9
A brief history of conservation genetics 9
Genetic markers 10
The role of genetics and genomics in conservation 11
Thesis structure 12

Objectives 15

Advisors' Report 17

Chapter 1: How many came home? Evaluating ex-situ conservation of green turtles in the Cayman Islands 19
Supplementary Information 35

Chapter 2: Architecture of assisted colonisation in sea turtles 51
Methods 65
Supplementary Information 71

Chapter 3: Helping decision making for a reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species 77
Supplementary Information 89

Chapter 4: Population structure and local adaptation of the Eastern Mediterranean loggerhead turtle (*Caretta caretta*) rookeries 105
Supplementary Information 135

Discussion 141
Genetic analysis for *ex-situ* conservation planning 141
Genomic analysis for *in-situ* conservation planning 156
Final remarks 152

Conclusions 153

References 155

INTRODUCTION

In the past few decades, conservation of biodiversity and ecosystems has become a priority on a global scale. The causes of the ongoing process of biodiversity collapse lie mostly in the collateral effects of anthropogenic activities (Dietz and Adger, 2003): plastic pollution (Pawar *et al*., 2016), global warming (Midgley *et al*., 2002), and invasive species (Doherty *et al*., 2016), are only three of the many sources of threats for biodiversity and ecosystems. Conservation actions aim to preserve and recover animal and plant species that are being or have been eradicated from their natural habitat. Due to the enduring pressure caused by anthropogenic activities, conservation strategies are often difficult to plan and implement. In fact, conservation and socio-economic progress are hardly balanced in a battle where too often conservation comes out as loser (Dietz and Adger, 2003; Barlow *et al.*, 2016). Depending on the level of stress acting on a habitat, conservation strategies can take place *in-situ* (Blanco *et al*., 2019; Mooney *et al.,* 2020) or *ex-situ* (Michaels *et al.,* 2014; Kumar *et al.,* 2017). The first, aims to protect and sustain populations in their natural habitat, by preventing or eliminating threats and by individual care. *In-situ* conservation does not only act directly on the target species, but also on restoring and protecting its environment aiming towards self-sustainability. These techniques are generally recognised as more secure and financially efficient. *Ex-situ* strategies, on the other hand, are implemented when local populations are already extinct or are about to be. These strategies, such as captive breeding programs, aim to recover populations outside of their natural habitat to avoid the pressure of threats and can be followed by a reintroduction of captive individuals in their natural habitat or by a relocation of both wild and captive individuals to a safer area. Although *ex-situ* projects can be extremely costly, they have become an important conservation tool to face biodiversity loss (Fischer and Lindenmayer, 2000; Storfer, 1999). These strategies, though, always need a proper monitoring driven by key questions to improve efficiency on active conservation (Nichols and Williams, 2006).

Threats to sea turtle conservation

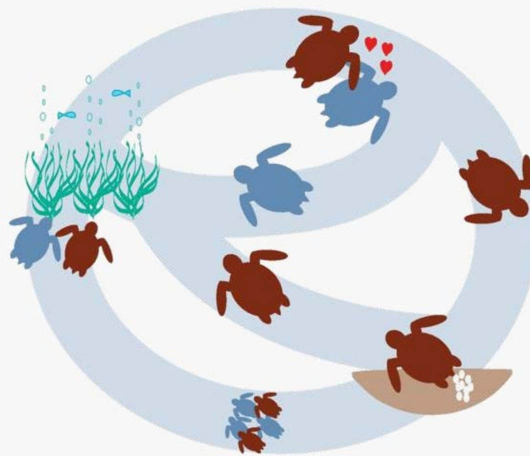
Sea turtles are circumtropical species which live in temperate waters across the five continents. Due to their range of distribution and to their characteristic and complex life cycle (Box 1) almost all sea turtles are species of conservation concern on a global scale. The high level of anthropogenic impact weighting on these animals derives from the high number of threats that individuals of these species may encounter during their life (Denkinger *et al.,* 2013; de Carvalho *et al.,* 2015; Guebert *et al.,* 2013).

BOX 1

Sea turtle life cycle

All species of sea turtles are characterised by long-life cycle and long generation times. Sea turtles can, in fact, live up to 80-100 years of age.

Nesting season starts in the early Summer and continues throughout the warmest months until early autumn. Adult females approach the shore at night to build nests and lay about 100 eggs per clutch (Bjorndal and Carr, 1989; Broderick *et al.,* 2003). The incubation lasts between 50 and 60 days, and then offspring hatch at night and reach the sea following the light of the moon.



The first stages of sea turtles' lives are characterised by very high rates of mortality, due to the fragility of hatchlings and the high level of predation. In fact, approximately 1 in 1000 hatchlings survives to reach sexual maturity (Frazer, 1986). The stages of a hatchling and juvenile turtle are called "lost years" since the tracking of survivals to sexual maturity is almost impossible and therefore there is not much information about the individuals. During this period, the individuals can travel thousands of kilometres during their developmental migration until maturity, when they return to their natal beaches (Bolten *et al.,* 1998, Bowen, et al., 1995). The age of sexual maturity varies depending on the species, ranging approximately between 15 and 30 years of age. Adult females mate and nest several times during their life. The interval between two nesting seasons is usually estimated between 2 and 6 years (Broderick *et al.,* 2003; Troëng and Chaloupka, 2007), but it can vary depending on the effort made by the female during the last nesting season. Between nesting seasons, males and females migrate from the nesting beach to foraging grounds, where individuals from multiple rookeries gather during these periods. During one nesting season a female can lay numerous nests with an overall average of 13 days between one nest and the next (Broderick *et al.,* 2002). Female turtles have the capability of storing sperm to fertilise more than one clutch during a season and mate with more than one male during reproduction time. Therefore, it is possible to find multiple paternity inside one same clutch.

The sex of sea turtles cannot be assessed externally until they reach adulthood, since they do not display clear sexual dimorphism, and even when adult it is very hard to distinguish the sexes in open water. Females approach the shore only during nesting seasons, while males remain in the shallow and almost never leave the water.

An indirect effect of anthropogenic activities on sea turtles is the impact of global warming (Zhai *et al.,* 2018). In 2011 Fuentes et al. predicted the consequences of this climatic process on turtle nesting grounds due to turtle Temperature-dependent Sex Determination (TSD) (Box 2). In fact, given the seriousness of the global warming threat, it is critical to understand the rate at which sand temperatures are likely to change and the extent to which associated hatching success and sex ratio will vary spatially as climate change progresses (Fuentes *et al.,* 2009). The rise of sand temperatures not only affects the viability of the nest, exciding the incubation temperature range (i.e. 25 to 33°C) (Miller, 1985), but can also causes the production of a sex ratio strongly skewed towards females, resulting in a future feminization of the adult population, as already happened in Australia (Jensen *et al.,* 2018). The shifting of geographical nesting range has been suggested as a possible natural solution to nesting beaches warming, although it is impossible to know how long this process will take and if it would be too late for the recovery of the population. Nevertheless, in the past few years the Western Mediterranean Sea has been experiencing a colonisation of loggerhead turtle females nesting sporadically along the coast of Spain, France and Italy (Figure 1). This is probably a response to the rising temperatures in the Southern-East Mediterranean where all the rookeries are based (Carreras *et al.,* 2018).

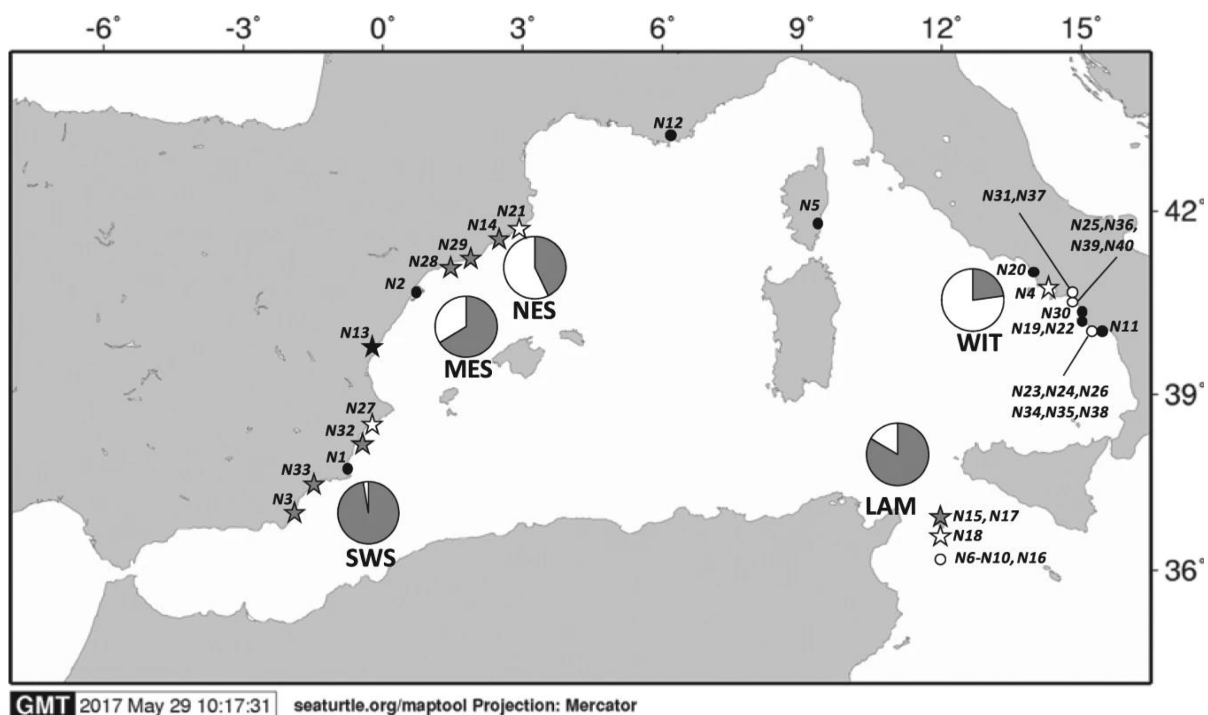


Figure 1. Sporadic nesting in the Mediterranean Basin. Stars and dots indicate nesting events as specified in Carreras et al. (2018). Pie graphs show the percentage of Atlantic (grey) and Mediterranean (white) turtles visiting foraging grounds located near sporadic nesting events. SWS: south Western Spain; MES: mid-Eastern Spain; NES; north Eastern Spain; WIT: Western Italy; LAM: Lampedusa. Image from Carreras et al. (2018).

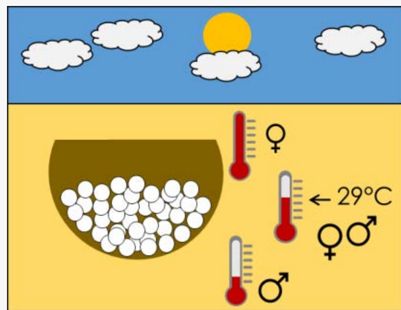
BOX 2

Sea turtles' complex behaviour

Sea turtles are considered a charismatic species, not only for their cute appearances but also for a number of complex behaviours such as natal homing (or philopatric behaviour), temperature-dependent sex determination and their capability to migrate for thousands of kilometres.

Natal homing

Also known as philopatric behaviour, natal homing means that adult sea turtles go back to their natal beach to nest once they reach sexual maturity (Greenwood, 1980). Both females and males are capable to find their way back thanks to chemical and mostly magnetic cues (Lohmann *et al.,* 2013). For females the accuracy of natal homing is high, while male have been suggested to have a lower degree of philopatry. This behaviour causes the development of geographically and genetically separate populations within a very short distance (relative to sea turtle mobility) (Lee *et al.,* 2007).



Temperature-dependent Sex Determination

See turtles: highly migratory species

TSD

As for other reptiles, the sex of sea turtle embryos is not defined by specific chromosomes, but by the temperature of the clutch incubation (Janzen and Paukstis, 1991). For instance, the temperature range for the success of a viable clutch in the loggerhead sea turtle is between 25-33°C, and 29°C is the pivotal temperature at which in a stable environment the clutch would result in half of the individuals being male and half females (Mrosovsky and Pieau, 1991). Nest conditions are therefore a key element for the development of offspring and sex determination, including not only the mere temperature of the sand but also moisture and salinity of the environment (Lolavar and Wyneken, 2020).

Migrations

Sea turtles are capable of swimming across oceans for several thousands of kilometres to reach foraging or breeding grounds. This behaviour is found in both juveniles (developmental migration) and adult (foraging migration) sea turtles. Loggerhead turtles of the Caribbean Sea, for instance, cross the Atlantic Ocean to the Mediterranean Sea during juvenile stages of life to find foraging grounds (Bolten et al., 1998). They then go back to the Caribbean once they reach adulthood to mate and nest. Other similar migrations have been studied across The Pacific Ocean among other species of sea turtles (Shillinger *et al.,* 2008). Turtles from different origins may share the same developmental or foraging grounds and form 'mixed stocks' (Clusa *et al*., 2014).

One of the main human activities directly affecting sea turtles is the fishing industry. In several countries the intentional capture of sea turtles for meat consumption is still legal at date, representing a threat to conservation (Milner-Gulland et al. 2003). In some regions the harvest and trade of turtle meat is, in fact, the major cause of local extinction of sea turtle populations (Nada and Casale, 2010; Mancini & Koch, 2009). For this reason, these species commercialization is now regulated by CITES (Convention on International Trade of Endangered Species) (Aiken *et al.,* 2001; Seminoff, 2004). On the other hand, although in most countries turtle fishing is illegal, the accidental bycatch represents one of the main causes of mortality (James *et al.,* 2005; Lewison and Crowder, 2007; Casale *et al.,* 2010). Casale (2011) estimated that only in the Mediterranean Sea 132000 turtles are captured by fishing gear every year, with possibly at least 44000 deaths.

Sea turtle conservation strategy

The use of genetics in conservation studies led to the use of genetic information to define of management and conservation units in either Evolutionary Significant Units (ESUs, evolutionary independent units) or Management Units (MUs: genetically different units) (Moritz, 1994). In order to deal with such mobile and complex species, the IUCN Marine Turtle Specialist Group (MTSG) decided to define a specific conservation unit, the Regional Management Units (RMUs) (i.e., spatially explicit population segments defined by biogeographical data of marine turtle species) as a framework for defining population segments for assessments (Wallace *et al.,* 2010). These units were generated based on geographic boundaries to distributions derived from studies on genetics, tag returns, satellite telemetry, and other data. Wallace et al (2010), identified 10 RMUs for *Caretta caretta* and 17 RMUs for *Chelonia mydas* (Figure 2). These units cover vast geographical areas, including several nesting and foraging grounds, making it challenging to manage the whole unit as a singular block. In the case of RMUs such the Mediterranean and Atlantic Northwest region, they also comprise a very high number of countries which have different regulations and cultural approaches towards the environment. This implies an even harder task for conservation management of such large geographical ranges. In 2014 Casale and Mariani suggested the identification of sub-Regional Management Units (sub-RMus) to facilitate turtle conservation through a relative small-scale international approach, based on within region dispersal patterns. The complexity of sea turtle behaviour and the uniqueness of each nesting population status and conditions, in fact, cannot be grouped together in large scale blocks and managed all in the same way with no distinction. For this reason, in 2018 Laurent et al. suggested to consider each nesting area of the Mediterranean Sea as an independent demographic identity and to manage each one separately at all life history stages. Gathering population level

information such as the effective population size and connectivity of these MUs is crucial to plan future management strategies effectively.

During the past century conservation management of sea turtles has been carried out on a global scale (Wallace et al., 2010; Wallace et al., 2011), but the complexity of this species behaviour represents multiple obstacles for their study and conservation. Nevertheless, philopatry is an advantage in the study and conservation of nesting females, nests and hatchlings, since it allows to group individuals into distinct reproductive populations (Chesser, 1991).

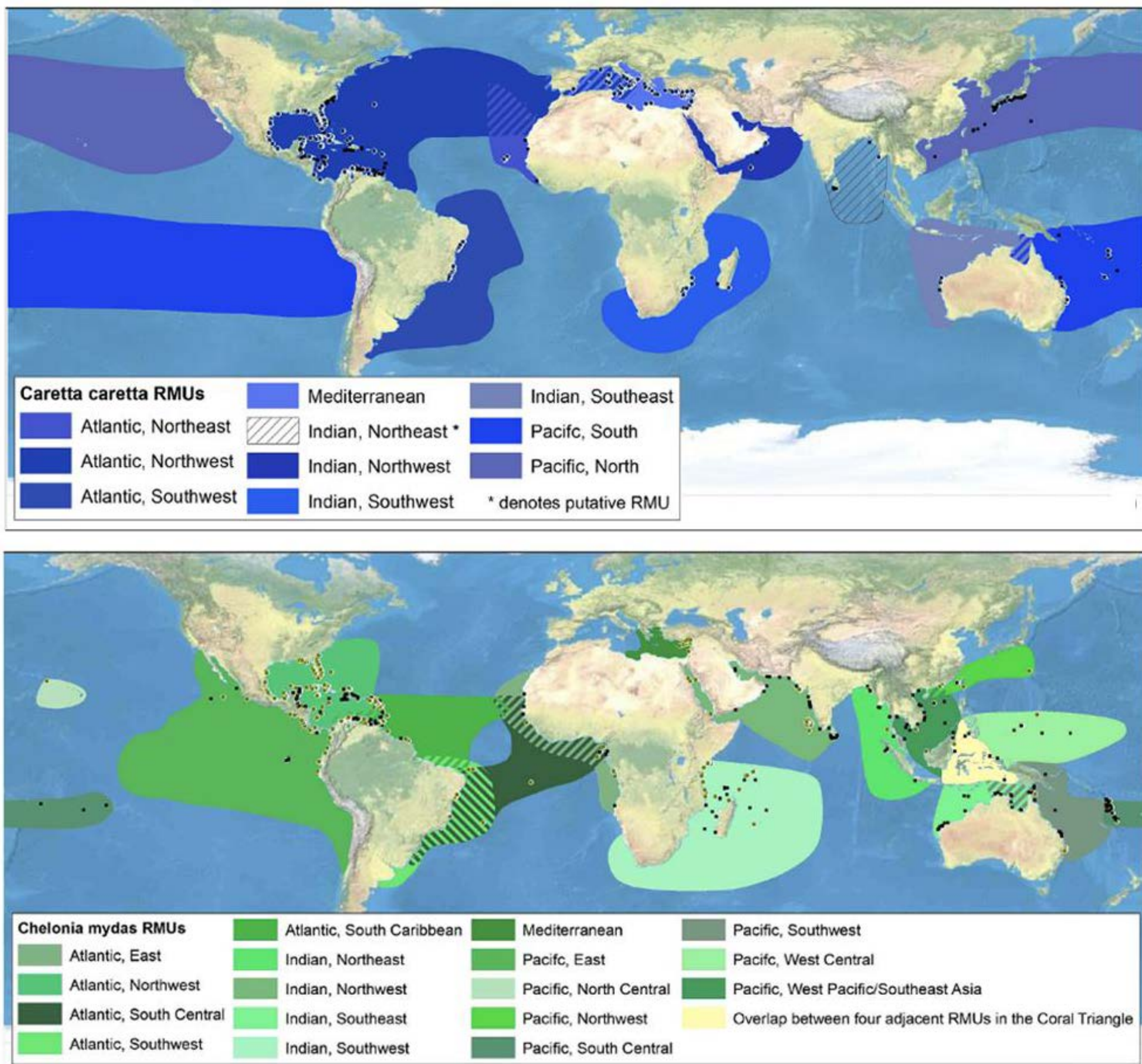


Figure 2. Global Regional Management Units (RMUs) for the loggerhead and green turtle. Loggerhead turtle RMUs are shown in shades of blue in the top graph; green turtle RMUs are shown in shades of green in the bottom graph. RMUs were identified by georeferencing data on marine turtle biogeography, including nesting sites, population abundances and trends, population genetics, and satellite telemetry. Figures from Wallace et al. (2010).

For this clear distinction of nesting populations and the accessibility to nesting females, the large majority of conservation effort for sea turtles is carried out on nesting individuals (Dutton *et al.,* 2005; Engeman *et al.,* 2003; Engeman *et al.,* 2005; Hawkes *et al.,* 2005), and many studies regarding turtle conservation rely on nesting activity and nesting-related data (Broderick *et al.,* 2002; McClenachan *et al.,* 2006). Males, juveniles and non-nesting adult females, on the other hand, are more complicated to track, monitor and sample since they do not approach the shore throughout the year (Roos *et al.,* 2005). For this reason, data on yearlings, juveniles and adult males is scarce and, therefore, conservation actions towards these groups are almost impossible to be planned. Conservation management of these species is connected to four fundamental factors which any conservation plan has to take into account: the industry (management and monitoring), universities (research), the government (legislation) and the general public (public education) (Figure 3). These are the four pillars on which conservation of biodiversity stands and all of them should be in constant communication with each other. Working together these branches can increase the possibilities of success of the conservation project in plan, and be beneficial for both in-*situ* and *ex-situ* strategies. These two are both used in the conservation of sea turtles in order to maximise the probability of recovery of endangered populations.

In-situ conservation

Most conservation actions to preserve sea turtles take place on nesting beaches. As mentioned above, the pelagic stages of these species, both as juveniles and as adults, are very difficult to track. For this reason, most efforts concentrate on preserving nests and on monitoring populations based on nesting females' data. Adult females and hatchlings can be sampled, tagged and monitored over time to estimate population indices. Population size, for instance, is calculated inferring the total number of adults from a census of nests or nesting females (Broderick *et al.,* 2002; Dutton *et al.,* 2005), but given the possibility of missing individuals, even in nesting beaches with an intense monitoring effort, this value can only be an estimate (Chassin-Noria *et al.,* 2017). For this reason, nests are commonly used as a proxy to estimate the size of a population (Bjorndal *et al.,* 1999), by performing simple population models as in Casale and Heppell (2016). Nevertheless, the parameters used to transform nest count on adult individuals have a geographical variation (e.g. Casale *et al.,* 2018) and recent research has found that calculations of breeding adults have been biased and thus need additional corrections (Casale and Ceriani, 2020). *In-situ* conservation actions also involve the management of fisheries, the control of turtle harvest and the improvement of awareness among local communities and tourists. To ensure a good management of all these aspects of conservation, the identification of fine scale MUs is essential to put in place correct conservation actions.

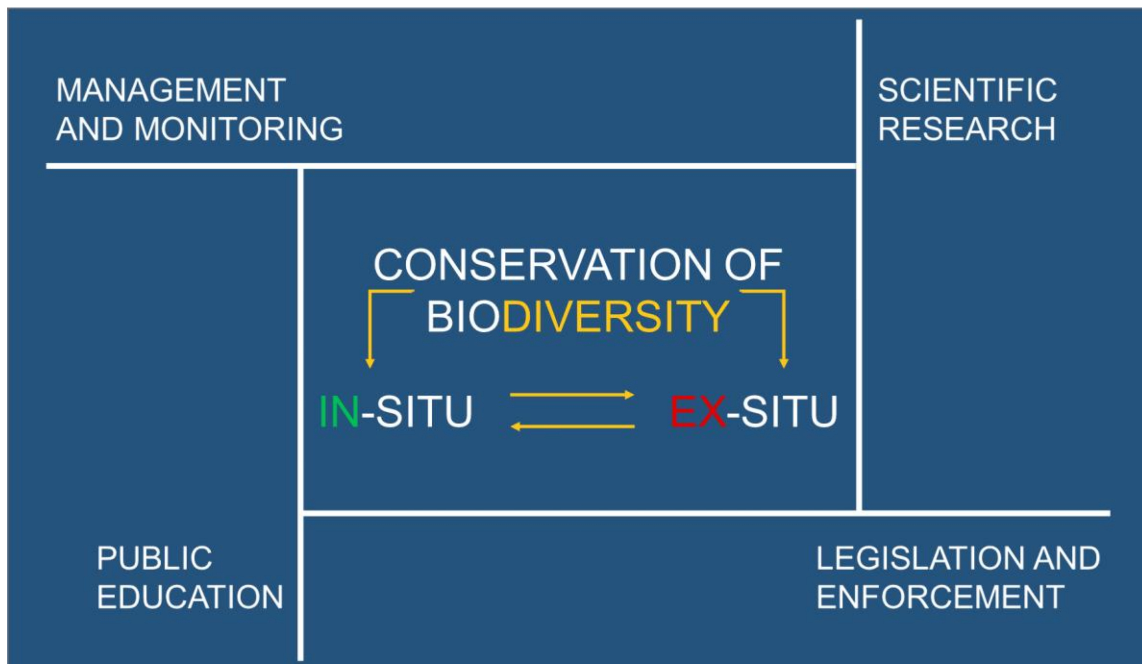


Figure 3. Conservation of biodiversity network. This figure shows the main fields influencing and interacting with conservation of biodiversity. All these four areas play a crucial role in the positive outcome of conservation action. The figure also shows the relationship between *in-situ* and *ex-situ* conservation, as the interaction between the two is a vital part of conservation management and it is often overlooked.

Ex-situ conservation

Ex-situ strategies are fairly new to the conservation of sea turtles. Captive breeding and reintroduction are the two main *ex-situ* approaches applied to these individuals. These techniques have become crucial tools for the recovery of locally extinct species of other taxa, such as the Przewalski horse (*Equuferus przewalskii*) in Mongolia or the Yellow-shouldered Amazon Parrot (*Amazona barbadensis*) in Margarita Island (Venezuela) (Sanz and Grajal, 1998; Van Dierendonck *et al.,* 1996). For sea turtles, only two programs managed a consistent long-term captive breeding and the reintroduction of individuals: the case of the Kemp's ridley sea turtle (*Lepidochelys kempii*) in Texas (Fontaine, 2005) and the case of the green turtle (*Chelonia mydas*) in the Cayman Islands (Bell *et al.,* 2005). In sea turtles, the reintroduction of captive individuals, is based on turtles' philopatry: this behaviour will lead the reintroduced individuals back to their release area to nest, once sexual maturity is reached (Cury, 1994; Greenwood, 1980; Mayr, 1963). Both reintroduction programs turned out to be a success after many years of captive breeding and headstarting (Heppell *et al.,* 1996; Mitrus, 2005), demonstrating that these techniques are feasible for sea turtles, although the impact of the reintroduction on the recovery process was never assessed. Nevertheless, captive breeding has rose concerns related with

human health, animal welfare and the efficiency of conservation activities (Warwick *et al.,* 2013), therefore short- and long-term monitoring of these programs is necessary to understand their actual contribution to the recovery of the wild populations and its impact on other geographically close populations.

Science and conservation

In the past decades, science has taken a step into the study of past and upcoming extinctions, focusing on understanding the causes and finding solutions (McLaughlin *et al.,* 2002; Gurevitch and Padilla, 2004; Bellard *et al.,* 2016). Several fields of research are now interacting to update conservation with scientific information (Figure 3), implementing complex techniques such as behavioural studies (Putman *et al.,* 2012), stable isotopes analysis (Lemons *et al.,* 2011) and genetics (Jensen *et al.,* 2013). The use of science in conservation has allowed to have a better understanding of the functioning of ecosystems (Flynn *et al.,* 2011), of the ecology of individuals (Carrión-Cortez *et al.,* 2010) and of dynamics and interactions between populations (Hays and Scott, 2013), among others. Thanks to these studies, conservation decision making and actions have greatly improved in the past few years, reverting the decline of the endangered species such as the loggerhead turtle (*Caretta caretta)* in the Mediterranean Sea (Casale, 2015).

The study of genetic composition of biological populations is called population genetics. Population genetics, thus, focuses on the dynamics that influence the evolutionary path of a population, the status of the individuals and their contribution to the population, and the interaction between populations in matters of migration, connectivity and structure (Hartl *et al.,* 1997). Two populations of the same species are considered genetically different if their gene flow is significantly low. This can be measured comparing the genetic variability of individuals of a population with the other, based on allele frequencies. Conservation genetics is the application of population genetics to address, answer and inform conservation issues (Frankham *et al.,* 2002).

A brief history of conservation genetics

Already in the 1980's genetics was addressed as a field of extreme importance to manage conservation activities, since genetic aspects of populations were considered fundamental in conservation programs to maximise probability of long-term survival and continued adaptability (Meffe, 1986; Hedrick and Miller, 1992). Genetics has been therefore applied in conservation to improve management and inform decision-making. Some of these studies focused, for instance, on fishery management through the assessment of acceptable gene flow (Ryman, 1991), population structuring of isolated populations with zoogeographic models (Maffe & Vrijenhoek, 1988), management of dispersal in fragmented populations to design captive breeding programs

(Vrijenhoek, 1998) and the effect of habitat fragmentation on gene flow and genetic variation (González *et al.,* 1998). In the past three decades, genetics has been spanning in any possible direction gradually increasing the facets of this field that could be applied to conservation of endangered species and ecosystems (DeSalle and Amato, 2004). In a few years, the rapid ascending of technology introduced genetics to high-throughput sequencing, driving the field of conservation genetics to a proper revolution (Allendorf *et al.,* 2010).

Genetic markers

Since the early 1990s, mitochondrial DNA and nuclear microsatellite loci have been the tools of choice in molecular studies in ecology and evolution for answering population-level questions (Morin *et al.,* 2004). MtDNA has been largely used in this field to assess and detect population structure based on haplotypes, as in the case of the shortnose sturgeon (*Acipenser brevirostrum*) (Grunwald *et al.,* 2002) and in loggerhead turtles mtDNA was used to date the Mediterranean colonisation of the Pleistocene (Clusa *et al.,* 2013). Since mitochondria are only transmitted from mother to offspring, the information provided by this marker is partial and female based. For this reason, in recent studies mtDNA was coupled with microsatellites to complement the population study. Microsatellites are nuclear markers and explain variability of both sexes in a given population. These markers have been used to understand how geographical and environmental features structure genetic variation (Manel et al., 2003), to assess interpopulation differences in genetic variation in black bears (*Ursus americanus*) (Paetkau & Strobeck, 1994) and to detect fine scale genetic structure in brown trout (*Slamo trutta*) (Carlsson *et al.,* 1999). In the loggerhead turtle, microsatellites were essential to revealed that both females and males show philopatric behaviour (Clusa *et al.,* 2018), and have been used worldwide to run sea turtles' population genetics (Bowen and Karl, 2007). In addition, the identification of MUs was recently found highly related to the number of markers used for population studies (Bradshaw *et al.,* 2018; Carreras *et al.,* 2007; Clusa *et al.,* 2018), therefore technologies that allow to increase that number would be more suited for such studies. The application of genetics in sea turtles has been crucial to improve our knowledge on the biology of the species and their conservation (Komoroske *et al.,* 2017).

Recent improvements in the speed, cost and accuracy of next generation sequencing (NGS) are revolutionizing the opportunities for generating genetic resources in non-model organisms. This is driving a shift from mtDNA and microsatellites markers to the analyses of genome-wide markers (Helyar *et al.,* 2011). Until a few years ago, the idea of sequencing a whole genome or working with thousands of markers was just idealistic in non-model organisms, either because the technology did not exist, or, later on, because of the enormous costs of the new technology. In

the last decade, the cost of NGS has dropped and laboratory and bioinformatic analysis have progressed exponentially. As a consequence, several studies have explored the potential of genomic analysis in non-model species in fields such as population structuring (Carreras *et al.,* 2020), inbreeding depression (Hoffman *et al.,* 2014), local adaptation (Savolainen *et al.,* 2013) or hybridization (Hohenlohe *et al.,* 2011). These new technologies providing an exponentially higher number of markers, make population studies more reliable and allow to address questions that remained inaccessible due to the lack of power of previous genetic markers. For these reasons, in the past few years the field of sea turtle conservation studies has begun to shift towards NGS technologies to have a deeper understanding of these species' behaviour and status in order to improve their conservation (Chow *et al.,* 2019; Hurtado *et al.,* 2016;; Komoroske *et al.,* 2019).

The role of genetics and genomics in conservation

Genetics and genomics are not two separate entities but are highly integrated and interdependent (McMahon *et al.,* 2014). Genetic, and hence genomic diversity is recognised as one of the most fundamental levels of biodiversity (Genome 10K, 2009) and the question on what should be conserved and what matters the most for species survival, can only be solved with more genetic data and genomic techniques (Allendorf *et al.,* 2010). Genetic and genomic studies have more and more taken an important role in the management of endangered species conservation, and although conservation still does not rely on genetic studies as much as it should, the assessment of population genetics has become a key factor for good management and animal welfare (Shafer *et al.,* 2015). Conservation genetics can provide important insights into the dynamics of endangered populations facilitating the understanding of processes such as inbreeding and genetic drift (Hoglund, 2009). Additionally, genetics can contribute to conservation by providing crucial population parameters such as metapopulation structure, geneflow, effective population size and evolutionary history (Hoglund, 2009; Puechmaille *et al.,* 2011). On the other hand, genomics can highly contribute to conservation science bringing important insights on local adaptation and more accurate estimates of effective population size (McMahon *et al.,* 2014). Applying genomics to conservation would also mean gathering a higher number of markers per individual, decreasing the necessity for enormous number of individuals to be sampled, and therefore reducing sampling effort and costs (Funk *et al.,* 2012). Genomics should be employed but must be as cost-effective and designed in the best way possible to address a specific set of scientific questions.

Here we want to show how genetics can be used to construct and improve both *in-situ* and *ex-situ* conservation actions in sea turtles. A deeper knowledge of endangered species genetic

structure can only benefit and advance the field of conservation management, preventing uninformed decisions that can lead to poor actions.

Thesis structure

This thesis uses both traditional markers and new generation sequencing to answer key ecological questions in order to improve conservation management of two species of sea turtles, the loggerhead (*Caretta caretta*) and the green turtle (*Chelonia mydas*). We divided the study into two blocks: the first block focuses on the application of genetic analysis in ex-situ conservation of green turtles; the second block centres on the power of genomics in in-situ conservation of loggerhead turtles (Figure 4). Each block is made of two chapters.

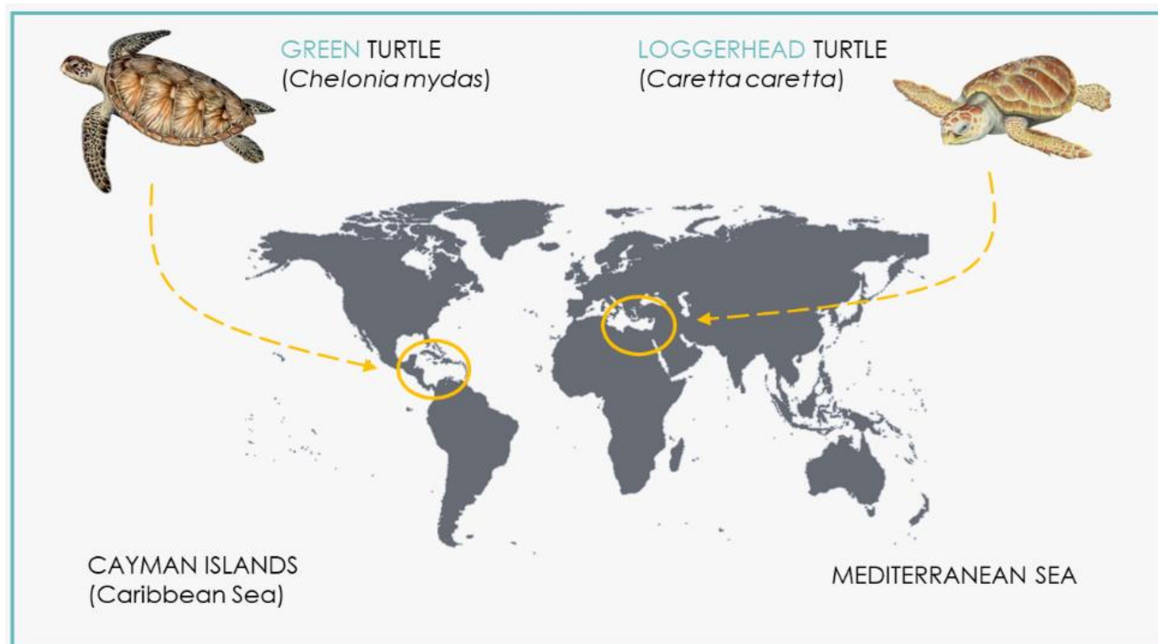


Figure 4. Thesis Structure. This thesis is divided in two main blocks, focusing on green turtles of the Cayman Islands and loggerhead turtles of the Mediterranean Sea.

Population genetics in ex-situ *conservation management*

The first two chapters focus on an *ex-situ* conservation strategy, showing how genetic studies are essential in every step of the structuring and implementation of such programs. In these chapters we use microsatellites and mitochondrial DNA to unveil the outcome of a reintroduction program of green turtles in the Cayman Islands and to assess the status of the new reintroduced wild population. We used traditional markers for two main reasons: first, to be able to compare genetic variability with other wild populations, which had been measured using mtDNA; and second, because microsatellites have been used and found to be reliable markers for paternity studies in

this species (Wright *et al.,* 2012). These chapters show the importance of the use of genetic tools in combination with other sources of data in *ex-situ* conservation, assessing the present population structure of a captive breeding program to detect signs of inbreeding, evaluating the outcome of the long-term reintroduction of green turtles, and identifying possible negative impacts of the reintroduction on future wild generations of green turtles.

Population genomics in in-situ *conservation management*

The third and fourth chapters focus on the development and use of genomic analysis to understand the population structure of loggerhead turtles in the Eastern Mediterranean Sea and unravel how environmental factors may impact their distribution. The third chapter explores the possibilities given by 2b-RAD methodology in the field of non-model species population genomics. This study also highlights how this methodology can be suitable for sea turtle studies, considering the status of high degradation which turtle samples are frequently found. The fourth and last chapter studies the population structure of Eastern Mediterranean loggerhead turtle rookeries using 2b-RAD genomic sequencing. In this chapter we also investigate possible migrations and connectivity among these populations and how few key environmental variables can affect their population structure. This study aims to gather fine scale genetic information about this area of the Mediterranean Sea to inform and improve conservation management of loggerhead turtle rookeries.

OBJECTIVES

Sea turtle conservation: genetics and genomics for a better management.

The main objective of this thesis is to apply genetic and genomic resources for *ex-situ* and *in-situ* conservation to inform and improve sea turtles' management actions. The thesis evaluates sea turtle populations' structure and relatedness patterns to answer specific ecological questions using population genetics and genomics.

Ex-situ - *Genetics*

- Evaluate the effect of the Cayman Turtle Centre reintroduction project in the natural population to estimate the potential of such ex-situ strategy on conservation

- Identify the population structure of both captive and wild green turtle populations to evaluate their potential negative impacts on other wild populations of the Caribbean Sea.

- Study up close the foundation process and differentiation of new populations in a long living and philopatric species

- Assess the effect of the Cayman Turtle Centre reintroduction on fitness of wild new-borns, and the impact on neighbour rookeries (i.e. Little Cayman Island).

In-situ - *Genomics*

- Test and optimise 2b-RAD sequencing methodology protocols and post-genotyping analysis for non-model organisms.

- Create a set of guidelines to follow in order to reduce cost of sequencing and facilitate decision making when using 2b-RAD in non-model organisms.

- Refine population structure of Eastern Mediterranean rookeries improving the power and reliability of population analysis using 2b-RAD genomic sequencing.

- Assess levels gene flow between populations and estimate effective population size.

- Understand the role of key environmental factors in genetic structuring and the possible future impact of global warming.

ADVISORS' REPORT

As directors of the doctoral thesis of Anna Barbanti entitled "Sea turtle conservation: genetics and genomics for a better management" we certify that the doctoral candidate has actively participated in designing and conducting the experimental work included in this thesis. She has analysed and discussed the results, and prepared the manuscripts from the first draft to the final edits in collaboration with co-authors.

Chapter 1: How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands.

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.

Molecular Ecology 2019; 28: 1637– 1651. https://doi.org/10.1111/mec.15017

Impact factor: (2019 JCR Science Edition) 5.163 Q1 ECOLOGY (19/168)

Participation of the candidate: Laboratory and data analyses. Manuscript drafting, tables and figures production and final edits.

Chapter 2: Architecture of assisted colonisation in sea turtles

Barbanti, A.; Blumenthal, J. M.; Broderick, A. C.; Godley, B. J.; Turmo, M.; Pascual, M.; Carreras, C.

Publication status: Manuscript in final revision by the co-authors and pending submission

Participation of the candidate: Study design, laboratory and data analyses. Manuscript drafting, tables and figures production and final edits.

Chapter 3: Helping decision making for a reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species.

Barbanti, A.*; Torrado, H.*; Macpherson, E.; Bargelloni, L., Franch, R.; Carreras, C.; Pascual, M.

Molecular Ecology Resources 2020; 20: 795– 806. https://doi.org/10.1111/1755-0998.13144

Impact factor: (2019 JCR Science Edition) 6.286 D1 ECOLOGY (13/168)

*equally contributed to this work

Participation of the candidate: Study design, laboratory and data analyses and bioinformatic simulations of *Caretta caretta*. Manuscript drafting, tables and figures production and final edits. This paper is also part of the thesis of Hector Torrado that carried out all the analyses for *Diplodus puntazzo*.

Chapter 4: Population structure and local adaptation of the Eastern Mediterranean loggerhead turtle (*Caretta caretta*) rookeries

<u>Barbanti, A.;</u> Margaritoulis, D.; Turkozan, O.; Khalil, M.; Snape, R.; Demetropoulos, A.; Hamza, A. A.; Ulger, C.; Pascual, M.; Carreras, C.

Publication status: Manuscript in preparation

Participation of the candidate: Study design, laboratory and data analyses. Manuscript drafting, tables and figures production and final edits.

Barcelona, 20th July 2020

Thesis supervisors:

Dr. Marta Pascual Berniola

Dr. Carlos Carreras Huergo

CHAPTER 1

Received: 27 April 2018 | Revised: 31 December 2018 | Accepted: 2 January 2019

DOI: 10.1111/mec.15017

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands

Anna Barbanti[1] | Clara Martin[1] | Janice M. Blumenthal[2] | Jack Boyle[2] | Annette C. Broderick[3] | Lucy Collyer[2] | Gina Ebanks-Petrie[2] | Brendan J. Godley[3] | Walter Mustin[4] | Víctor Ordóñez[1] | Marta Pascual[1]* | Carlos Carreras[1]*

[1]Department of Genetics, Microbiology and Statistics and IRBio, Universitat de Barcelona, Barcelona, Spain

[2]Department of Environment, Grand Cayman, Cayman Islands

[3]Centre for Ecology and Conservation, University of Exeter, Penryn, UK

[4]Cayman Turtle Centre, West Bay, Cayman Islands

Correspondence
Carlos Carreras, Department of Genetics, Microbiology and Statistics and IRBio, Universitat de Barcelona, Barcelona, Spain.
Email: carreras@ub.edu

Funding information
Agència de Gestió d'Ajuts Universitaris i de Recerca, Grant/Award Number: 2017 FI_B 00997; Fundació Bosch i Gimpera, Grant/Award Number: 9013 and 10084; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: CTM2013-48163 and CTM2017-88080; Darwin Plus Project, Grant/Award Number: DPLUS019; Generalitat de Catalunya, Grant/Award Number: SGR2017-1120

Abstract

Ex situ management is an important conservation tool that allows the preservation of biological diversity outside natural habitats while supporting survival in the wild. Captive breeding followed by re-introduction is a possible approach for endangered species conservation and preservation of genetic variability. The Cayman Turtle Centre Ltd was established in 1968 to market green turtle (*Chelonia mydas*) meat and other products and replenish wild populations, thought to be locally extirpated, through captive breeding. We evaluated the effects of this re-introduction programmme using molecular markers (13 microsatellites, 800-bp D-loop and simple tandem repeat mitochondrial DNA sequences) from captive breeders (N = 257) and wild nesting females (N = 57) (sampling period: 2013–2015). We divided the captive breeders into three groups: founders (from the original stock), and then two subdivisions of F_1 individuals corresponding to two different management strategies, cohort 1995 ("C1995") and multicohort F_1 ("MCF1"). Loss of genetic variability and increased relatedness was observed in the captive stock over time. We found no significant differences in diversity among captive and wild groups, and similar or higher levels of haplotype variability when compared to other natural populations. Using parentage and sibship assignment, we determined that 90% of the wild individuals were related to the captive stock. Our results suggest a strong impact of the re-introduction programmme on the present recovery of the wild green turtle population nesting in the Cayman Islands. Moreover, genetic relatedness analyses of captive populations are necessary to improve future management actions to maintain genetic diversity in the long term and avoid inbreeding depression.

KEYWORDS

Chelonia mydas, ex situ conservation, microsatellites, mtDNA, parentage analysis

*These authors contributed equally as senior researchers.

 © 2019 John Wiley & Sons Ltd

1 | INTRODUCTION

Over the past two decades, biodiversity loss has become a pressing global issue (Barnosky et al., 2011; Dirzo & Raven, 2003; Hooper et al., 2012; Mora & Sale, 2011). Deforestation (Barlow et al., 2016; Gibson et al.., 2013; Turner, 1996), overexploitation (Coleman & Williams, 2002), agricultural expansion (Allan et al., 2015) and invasive species (Ceballos & Ehrlich, 2002; Doherty, Glen, Nimmo, Ritchie, & Dickman, 2016) are some of the factors driving species and populations to experience severe declines and are negatively influencing the functionality of food webs (Dunne, Williams, & Martinez, 2002) and ecosystem sustainability (Hooper et al., 2012; Worm et al., 2006).

Ex situ strategies (i.e., conservation measures applied away from the natural habitat of the target species) such as captive breeding and re-introduction have become an important conservation tool used to combat biodiversity loss by recovering locally extinct populations (Fischer & Lindenmayer, 2000; Storfer, 1999). The release of captive-bred individuals into the wild has been identified as an instrument for the conservation of threatened populations (re-introduction) and for the establishment of new ones (introduction) (IUCN, 1987). Captive breeding programmes followed by re-introductions, although controversial (Jule, Leaver, & Lea, 2008), are one of the most commonly used ex situ conservation strategies (Fischer & Lindenmayer, 2000). Captive breeders may include local individuals and/or individuals belonging to other wild populations, depending on the status of the population to be recovered. Some species of conservation concern, such as the Przewalski horse (*Equuferus przewalskii*) in Mongolia or the yellow-shouldered Amazon parrot (*Amazona barbadensis*) on Margarita Island (Venezuela), have successfully recovered to self-sustaining populations after captive breeding and re-introduction programmes (Sanz & Grajal, 1998; Van Dierendonck, Bandi, Batdorj, Dügerlham, & Munkhtsog, 1996). Re-introductions from captive breeding programmes may, however, produce individuals incapable of long-term survival in the wild due to feeding incompetence (i.e., an inability to hunt or find food resources in the natural habitat), unsuccessful predator/competitor avoidance and disease (Jule et al., 2008). During the 1990s, several studies highlighted the need for monitoring after the release of individuals (Armstrong, Soderquist, & Southgate, 1994; Sarrazin & Barbault, 1996; Sutherland et al., 2010), and that this monitoring should be driven by key questions to improve efficiency on active conservation (Nichols & Williams, 2006). Nonetheless, outcomes often remain unknown and the causes of failures are rarely understood (Rees et al., 2016; Weeks et al., 2011) as a result of the paucity of monitoring and/or the time lag necessary to detect actual failure/success (Fischer & Lindenmayer, 2000).

The origin and number of breeders in ex situ conservation programmes should be considered to reduce potential negative impacts during re-introductions, such as the generation of weak hybrid offspring as a consequence of outbreeding depression (Edmands, 2007; Weeks et al., 2011; Witzenberger & Hochkirch, 2011) or the loss of genetic variability and inbreeding depression due to a low number of founders (Hedrick & Fredrickson, 2008; Hedrick, Miller, Geffen, & Wayne, 1997; Ralls & Ballou, 1986; Witzenberger & Hochkirch, 2011). Re-introduction programmes may show differential success, ranging from total failure to complete replacement by re-introduced individuals and extirpation of the wild local population (Sutherland et al., 2010; Sweeting, Beamish, Noakes, & Neville, 2003).

Monitoring re-introduction programmes can be challenging, in particular for species with high dispersal rates and long generation times (Canessa et al., 2016); therefore, several methodologies, from tracking using electronic devices to the use of biological markers, have been adopted in different species. Telemetry was used to monitor dispersal patterns of an endangered freshwater fish (the trout cod *Maccullochella macquariensis*) in Australia (Ebner & Thiem, 2009), whilst growth rates and survival indices were used in the management of the re-introduced peninsular bighorn sheep (*Ovis canadensis*) in California (Ostermann, Deforge, & Edge, 2001). Nuclear genetic markers, such as microsatellites, have been valuable for assessing the effectiveness of re-introduction programmes and measuring their impact on natural populations (DeMay, Becker, Rachlow, & Waits, 2017; Koelewijn et al., 2010; Stenglein, Waits, Ausband, Zager, & Mack, 2010). Similarly, mitochondrial DNA (mtDNA) has been successfully used to monitor re-introduction (Godoy, Negro, Hiraldo, & Donazar, 2004) and captive breeding programmes (Kitanishi et al., 2013). Moreover, combining different types of genetic markers can help to obtain diverse and complementary information about the same sample set (Kim et al., 2011; Puckett et al., 2014).

To date, the long-term Cayman Turtle Farm (CTF) green turtle (*Chelonia mydas*) re-introduction programmme has not been evaluated genetically. Green turtles play an important ecological role in the maintenance of seagrass beds, as grazing stimulates new growth (Aragones, Lawler, Foley, & Marsh, 2006). The large nesting population of green turtles historically present in the Cayman Islands served as a key fishery resource (Aiken et al., 2001; Bass, Epperly, & Braun-McNeill, 2006), and was exposed to massive anthropogenic perturbations by the commercial harvesting of nesting females for meat consumption. The decline of green turtle nesting populations worldwide led this species to be listed as Endangered in 1975 by the IUCN (International Union for Conservation of Nature) and its commercialization was regulated by CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) (Aiken et al., 2001; Seminoff, 2004). In the 1980s several studies concluded that the green turtle nesting population of the Cayman Islands was extinct (King, 1982; Stoddart, 1980), although the presence of some green turtles was reported in the waters surrounding the Islands (Brunt & Davies, 2012). In 1968 a private company, the CTF (up to 1983 known as Mariculture Ltd and now called Cayman Turtle Centre Ltd), started a green turtle captive breeding programmme to restore a population nesting in the Cayman Islands whilst providing an alternative source of turtle meat to alleviate harvest from the wild population (Cayman Turtle Farm, 2002). The project consisted of importing adult turtles and eggs from other populations to breed

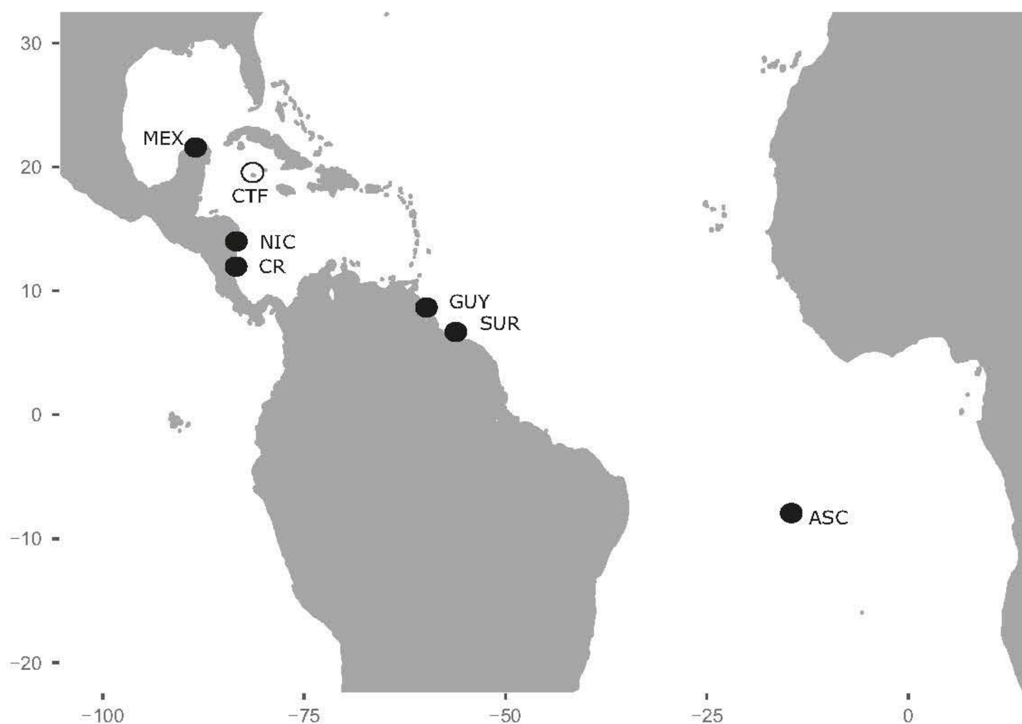


FIGURE 1 Map of founders of the Cayman Turtle Farm (CTF). Wild adult turtles and eggs were taken from populations in Mexico (MEX), Costa Rica (CR), Guyana (GUY), Suriname (SUR) and Ascension Island (ASC) and from the foraging area of Nicaragua (NIC) (for details on adults and eggs see Supporting Information Tables S1 and S2). Locations of the origin of founders are marked with black circles and the Cayman Turtle Farm is marked with an empty circle

in captivity, and raised future generations for the re-introduction and as a harvest resource. Individuals representing the F_1 generation were typically grown in the farm up to 4–6 years of age and then arbitrarily chosen to be part of the breeding stock, to be released or to be slaughtered for meat consumption. By contrast, individuals of the F_2 generation were only released or used as a source of meat, but not as part of the captive breeding stock.

Adult turtles and eggs were collected from the nesting populations of Costa Rica, Suriname, Guyana, Ascension Island and Mexico and from the foraging area of Nicaragua between 1968 and 1978 to form the founder stock of the CTF (Figure 1; see Supporting Information Tables S1 and S2). Given that individuals imported to the farm were gathered from widely separated areas, they probably belonged to genetically different populations, as shown by Naro-Maciel et al. (2014). Between 1980 and 2001, the CTF released ~30,000 captive-raised hatchlings and yearlings as part of the re-introduction programmme (Bell et al., 2005). Between 1982 and 1983, the farm reduced the size of the founder breeding stock by 46% (Supporting Information Table S3) to decrease management costs (Cayman Turtle Farm, 2002). In 2001, Hurricane Michelle caused major damage to the CTF, further decreasing the number of founder breeders with the loss of 81% of individuals, which either died or escaped (Supporting Information Table S3).

The re-introduction of marine turtles into the wild is based on the premise that they exhibit natal philopatry. This behaviour is described as the return of individuals to their natal site to reproduce (Cury, 1994; Greenwood, 1980; Mayr, 1963). Newborn hatchlings are thought to memorize different chemical and magnetic cues from the nesting beaches where they hatch and use this information in adulthood to find their natal nesting beaches to reproduce (Lohmann, Lohmann, Brothers, & Putman, 2013; Lohmann, Witherington, Lohmann, & Salmon, 1997; Meylan, Bowen, & Avise, 1990). Natal philopatry leads to genetic distinctiveness of populations, and thus geographically distant groups might have limited genetic exchange (Chesser, 1991; Lee, Luschi, & Hays, 2007). Therefore, the natural recovery of isolated populations on the verge of extinction may be difficult, as little migration would be expected from other populations to increase the number of mating adults. This philopatric behaviour, in both females and males (Clusa et al., 2018), is the basis for a rapid colonization of new potential nesting areas after the first arrival of marine turtles (Carreras et al., 2018).

Philopatry is also the basis of the success of the re-introduction programmme of Kemp's ridley sea turtle (*Lepidochelys kempii*) in Texas, through a headstarting programmme started in 1978 (Fontaine, 2005). Headstarting consists of the rearing of the offspring in captivity up to a certain size before their release, to prevent high rates of mortality typical of the early stages of life (Heppell, Crowder, & Crouse, 1996; Mitrus, 2005). However, headstarting, as a conservation measure, has been questioned over the last 20 years because of

the expected poor survival of the released turtles. They have been found to have nutritional deficiencies and behavioural modifications resulting from factors associated with captivity, including insufficient exercise, lack of stimuli or lack of feeding skills (Heppell, 1998; Heppell et al., 1996; Moll & Moll, 2000). Furthermore, the probability of surviving to adulthood increases exponentially with age, and therefore the population dynamics of organisms such as turtles are driven more strongly by changes in annual juvenile survival than by survival in their first year of life (Heppell et al., 1996).

The headstarting re-introduction programmme of the CTF has also raised some concerns about its utility and possible negative impacts. Some of these concerns are related to human health, animal welfare and conservation activities (Warwick, Arena, & Steedman, 2013). Moreover, re-introduction programmes also have possible genetic consequences such as the alteration of genetic variability of natural populations caused by the introduction of hatchery-bred individuals (Horreo, de la Hoz, Pola, Machado-Schiaffino, & Garcia-Vazquez, 2012). Farm releases in the Caribbean region of individuals hatched from a founder stock that includes South Atlantic genetic material are thus a potential source of outbreeding depression (Narum, Arnsberg, Talbot, & Powell, 2007). Despite these concerns, the wild population of green turtles nesting in the Cayman Islands has increased and the number of nesting females is increasing despite the long generation time of the species (Aiken et al., 2001; Cayman Islands DoE, unpublished data). To date, the exact role of the CTF breeding programmme in this recovery is unknown, but the application of living tags (created by the transplantation of a 4-mm-diameter disc of plastron to the carapace) has shown that at least some of the released hatchlings survived to adulthood and reproduced on nesting beaches in the Cayman Islands (Bell et al., 2005).

Genetic evaluation and monitoring of the success of the CTF re-introduction programmme is necessary to understand its contribution to the recovery of the wild populations and its impact on the local gene pool. Using a set of 13 microsatellites, a fragment of the D-loop mtDNA (800 bp) and four mitochondrial simple tandem repeat (STR) markers, we analysed the genetic diversity and genetic structure of 257 captive and 57 wild green turtle females nesting on the islands of Grand Cayman. This study aims to reconstruct the farm population structure and evaluate the re-introduction programme, specifically (a) estimate the genetic diversity of the farm breeding stock; (b) assess parentage and sibship relationships between the farm and wild population; and (c) identify the genetic structuring of the farm breeding stock and of the wild population in relation to other wild green turtle populations. We aim to provide novel insights and guidelines for future re-introduction actions using our results as a case study.

2 | MATERIALS AND METHODS

2.1 | Sampling and DNA extraction

The study was conducted using samples from wild green turtle females nesting on Grand Cayman (Cayman Islands) and from breeding females of the CTF. Tissue biopsies were taken from all females of the farm breeding stock from 2012 to 2014 (N = 257) and from all wild nesting females encountered during 2013 and 2014 (N = 57). Tissue samples were taken from the neck or from the rear flippers with a scalpel blade and stored in 100% ethanol. All individuals were PIT-tagged (passive integrated transponder; Bjorndal, Reich, & Bolten, 2010) to avoid pseudoreplication. We also obtained information about the origin of the farm breeders or year of birth from the farm databases when available (Supporting Information Dataset S1), which indicated that the breeding stock consisted of original founder and captive F_1 individuals. Based on this background data, we identified three sample groups within the farm breeding stock, as they represent different stages of the re-introduction: founders, C1995 and multicohort F_1 breeders (MCF_1). The group "Founders" includes individuals known to belong to the original stock and to come from distinct populations (N = 25). The group "C1995" consists of F_1 individuals born in the farm in 1995 and kept to increase the number of breeders after hurricane Michelle (N = 189). The group "MCF1" (Multicohort F_1 breeders)are F_1 females born from 1986 up to 2002 and used for routine replacement of the original founder stock in order to maintain management census sizes (N = 43). These two F_1 groups were considered separately because they are the result of two different management strategies (a single large replacement, the first, vs. continuous small replacements, the latter).

The DNA of all samples was extracted using a QIAamp Blood and Tissue Kit (Qiagen) or using an E.Z.N.A. Tissue DNA kit (OMEGA Bio-tek), following the manufacturers' protocols. DNA was suspended in 100 µl of deionized water.

2.2 | Laboratory analysis

All samples were genotyped at 13 microsatellite loci, originally designed for different species of sea turtles that amplify and are polymorphic in green turtles (Wright et al., 2012). Additionally, we sequenced an 800-bp fragment of the mtDNA D-Loop region (Abreu-Grobois et al., 2006) and four $(AT)_n$ mtDNA STRs (Tikochinski et al., 2012) in all wild individuals and a selection of the farm animals. The selection of farm samples was based on the known origin of the animals coupled with our microsatellite results, in order to characterize the founder stock and to confirm parentage assignments (see Results). Amplification PCR conditions for each marker are given in Supporting Information Table S4.

One of the primers for each microsatellite (Supporting Information Table S5) was labelled with a fluorescent dye (6-FAM, HEX or NED). Microsatellite loci were amplified with two multiplex PCR sets as described by Wright et al. (2012) and carried out with a GenAmp PCR System 2700 (Applied Biosystems). Each multiplex was amplified in a final volume of 5 µl, with 2.5 µl of Multiplex PCR Master Mix (Qiagen), 1.5 µl of primer mix (as detailed in Bradshaw et al., 2018) and 1 µl of DNA. After amplification, 15 µl of ultrapure H_2O Ecolab was added in each reaction tube and amplification success was assessed in an agarose gel. Microsatellite allele sizes were estimated in 2 µl of diluted amplified DNA, 0.5 µl of GeneScan 500 Liz Size standard (Applied Biosystems) and 12.5 µl of deionized

formamide on an ABI 3730 DNA Analyzer (Applied Biosystems) at the Serveis Científico-Tècnics of the Universitat de Barcelona, and alleles were assigned using GENEMAPPER software (version 3.7, Applied Biosystems). To check for genotyping errors, 27 samples were randomly selected and genotyped twice, resulting in a genotyping error <0.2%.

Mitochondrial D-Loop sequences (800 bp long) were amplified in 142 individuals (Supporting Information Dataset S1). The final reaction volume was 15 µl containing 5.08 µl of deionized water, 3 µl of PCR buffer 5× (GoTaq Promega), 1.8 µl of dNTPs (1 mm), 0.6 µl of $MgCl_2$ (25 mm), 1.8 µl of bvine serum albumin, 0.3 µl of forward primer (10 µm), 0.3 µl of reverse primer (10 µm), 0.12 µl of GoTaq G2 Flexi DNA Polymerase (Promega, 5 U/µl), and 2 µl of DNA. Mitochondrial STR amplifications were conducted for the same individuals (Supporting Information Dataset S1). The final reaction volume was 15 µl, with 5.48 µl of deionized water, 3 µl of PCR buffer 5× (GoTaq Promega), 1.8 µl of dNTPs (1 mm), 0.6 µl of $MgCl_2$ (50 mm), 1 µl of forward primer (10 µm), 1 µl of reverse primer (10 µm), 0.12 µl of GoTaq G2 Flexi DNA Polymerase (Promega 5 U/µl), and 2 µl of DNA. The amplified DNAs (3 µl) of both mtDNA markers were purified with Exo-SAP (2 µl containing 0.4 U of EXO and 0.4 U of TSAP) using a single cycle of 37°C for 15 min and 80°C for 15 min. Then, 1 µl (5 µm) of the corresponding forward primer was added to the purified product (LCM15382 for D-loop and CM-D-1 for STRs) and dried at 80°C for 30 min in order to be sequenced on an ABI 3730 automated DNA analyser (Applied Biosystems) at the Serveis Científico-Tècnics from Universitat de Barcelona.

2.3 | Data analysis: microsatellites

We checked for null alleles using the program MICRO-CHECKER (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004). Pairwise linkage disequilibrium and deviation from Hardy–Weinberg equilibrium were assessed using GENEPOP version 4.3 software (Raymond & Rousset, 2004). To correct for multiple comparisons, we used the Benjamini–Yekutieli (B-Y) FDR (false discovery rate) correction (Narum, 2006). To test for inbreeding through observed (H_O) and expected (H_E) heterozygosity we used GENETIX version 4.05.2 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 2004) software on all groups of individuals. Allelic richness was computed with rarefaction using the R package DiveRsity (Keenan, Mcginnity, Cross, Crozier, & Prodöhl, 2013; RStudio Team, 2016). Pairwise genetic distances (F_{ST}) among our groups were calculated using GENALEX 6.503 (Peakall & Smouse, 2012). We used the same program to calculate the relatedness estimator of Lynch and Ritland (1999) among individuals within each group. Because we are analysing samples with potentially very distinct origin, we estimated relatedness values for each subset separately based on the allele frequencies obtained within the subset. We also tested whether relatedness values for each subset deviate significantly from those randomly obtained by 9,999 permutations, as implemented in GENALEX 6.503 (Peakall & Smouse, 2012). For this last analysis we considered all the samples together to calculate baseline allele frequencies.

We identified the most probable number of genetic groups among the founder individuals of the captive stock, as they could come from geographically distant areas, while the remaining samples from the breeding stock belong to the F_1 of the captive breeding and therefore are the result of a mix of genetic material. We used the software STRUCTURE version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) and performed 10 repetitions of each independent *K* value from 1 to 10; burn-in length was set to 50,000 with 500,000 Markov chain Monte Carlo (MCMC) runs. To select the best *K* we calculated the log probability of the data with STRUCTURE HARVESTER (Earl, 2012).

We conducted sibship and maternity analysis using three different programs based on maximum-likelihood: COLONY version 2.3 (Jones & Wang, 2010), CERVUS version 3.0.7 (Marshall, Slate, Kruuk, & Pemberton, 1998) and ML-RELATE (Kalinowski, Wagner, & Taper, 2006). COLONY performs parentage assignment, sibship analysis and reconstructs genotypes of unsampled parents. COLONY also generates the best cluster in which the program infers unsampled mothers and fathers and allows for the formation of family groups. We set the parameters to long run, high precision and error rate = 0.0001. All individuals were included as offspring and motherhood input data. To refine the analysis and to minimize the error we excluded as mothers all wild individuals, as they could not have sired any of the farm individuals, and all captive individuals born in 1995 (cohort C1995) or later, according to the information provided by the CTF (Supporting Information Dataset S1), as they would be too young to be mothers of the other breeding females. CERVUS performs parentage analysis using strict a confidence level set at 95%. ML-RELATE estimates the relationship among individuals from codominant genetic data. We computed the log-likelihood of relatedness for all pairs of individuals and produced a confidence interval of 95% after 999 simulations per test. The outputs of the three programs were then combined to identify for maternity and sibship relationships in our sample set. Maternity outputs of CERVUS and COLONY were also compared using PEDAGREE (version 1.06), software which can be used to assess accuracy and congruence for genetically reconstructed pedigree relationships from these two programs (Coombs, Letcher, & Nislow, 2010).

2.4 | Data analysis: mtDNA

D-Loop sequences were aligned, cut and compared with published haplotype sequences found in the database maintained by the Archie Carr Center for Sea Turtle Research (http://accstr.ufl.edu/) using BIOEDIT software (Hall, 1999). STRs were scored by counting the number of $(AT)_n$ repeats in each of the four loci of the sequence described in the literature and haplotypes were named using the four-number barcoding system (Tikochinski et al., 2012). MEGA7 software (Kumar, Stecher, & Tamura, 2016) was used to create a Neighbor-Joining tree (Saitou & Nei, 1987) to identify the phylogeny of D-Loop haplotypes by maximum likelihood (Tamura, Nei, & Kumar, 2004) with 999 bootstrap replicates. The tree was rooted in the middle of the longest branch. We also created a haplotype network using Median Joining calculation (Bandelt, Forster, & Röhl, 1999) as implemented in

NETWORK 5.0 software (www.fluxus-engineering.com). Each D-Loop haplotype was assigned to a lineage by comparison to the lineages identified by Naro-Maciel et al. (2014). We calculated haplotype and nucleotide diversity for wild and captive individuals separately using ARLEQUIN 3.5 (Excoffier & Lischer, 2010) and DNASP version5 (Librado & Rozas, 2009). Both D-loop and STR sequences were also used to confirm maternal assignments resulting from microsatellite analysis. Using D-loop sequences we performed analysis of wild and captive populations compared to other wild populations of the Caribbean (Shamblin, Bagley et al., 2015), African (Patrício et al., 2017; Shamblin, Dutton et al., 2015) and Mediterranean (Bradshaw et al., 2018) regions. Incorporating D-loop sequences from these populations, we tested for genetic structuring (F_{ST}) and genetic diversity.

We performed Mixed-Stock analysis (MSA) of the Founders subset against a baseline of rookeries with short (400-bp) D-loop sequences, consisting of North Caribbean, South Caribbean and South Atlantic populations (23 populations in total). We used short sequences to be able to include data from populations of known origin of the founders according to CTF background records. We used the program BAYES (Pella & Masuda, 2001), with 40,000 MCMC runs for each potentially contributing nesting site with prior expectations of 0.978 for a particular nesting site and 0.001 for the 22 other nesting sites. Lack of convergence was assessed with the shrink factor of Gelman and Rubin (1992). The contribution of each rookery to the founder group was estimated from the mean of chains after 20,000 burn-in steps.

The results of both mtDNA and microsatellite analysis of wild and captive animals were used together to define the farm population structure. For this, we combined D-loop and STRs in a haplotypic system as in Shamblin, Dutton et al. (2015) to perform F_{ST} tests among all groups.

3 | RESULTS

3.1 | Genetic diversity

All 314 individuals were genotyped with the 13 microsatellite markers showing a number of alleles ranging from six (B123) to 23 (Cc2) (Supporting Information Table S5). Four markers were found to be in Hardy–Weinberg equilibrium, one (D2) was not at equilibrium in both groups, and the remaining eight yielded different results depending on the sample group considered. However, we did not discard any of the markers that were not in Hardy–Weinberg equilibrium for two reasons. First, previous studies concluded that none of these markers deviates from Hardy–Weinberg equilibrium in other wild populations (Bradshaw et al., 2018; Wright et al., 2012), suggesting that the detected deviations are not due to the properties of the marker. Second, these deviations are expected to be found both in the captive individuals, considering the process of founding from different natural populations, and in the wild population, due to the re-introduction process. Furthermore, analyses were run without these markers and the results did not change substantially. Expected heterozygosity (H_E) decreased from founders to C1995 and from C1995 to wild individuals (Table 1), although differences were not significant as assessed with a Wilcoxon matched pairs test. Mean observed

	Farm			
	Founders	C1995	MCF1	Wild
Microsatellites				
N	25	189	43	57
Ar	8.538 (0.592)	7.644 (0.181)	7.885 (0.369)	7.821 (0.399)
H_E	0.717 (0.037)	0.702 (0.034)	0.719 (0.029)	0.693 (0.042)
H_O	0.681 (0.04)	0.72 (0.038)	0.751 (0.038)	0.664 (0.046)
Rel	−0.021 (0.0023)	−0.003# (0.0004)	−0.012 (0.0018)	−0.009 (0.0016)
F_{IS}	0.05*	−0.025*	−0.042	0.045*
mtDNA				
N	25	41	19	57
D-Loop				
Haplo	8	8	7	12
H	0.703	0.578	0.602	0.573
π	0.0069	0.0038	0.0043	0.0039
STRs				
Haplo	10	13	9	16
H	0.877	0.806	0.848	0.814
D-Loop + STRs				
Haplo	13	16	11	19
H	0.703	0.578	0.602	0.573

TABLE 1 Genetic diversity values of each sample subset. The farm breeding stock was subdivided into Founders (from the original founder stock), MCF1 (individuals of the breeding stock born at the Cayman Turtle Farm) and C1995 (individuals born in 1995 at the Cayman Turtle Farm). The table shows number of samples used with each marker (N), allelic richness (Ar), expected (H_E) and observed (H_O) heterozygosities, degree of relatedness (Rel) (values significantly higher than those obtained by random permutations are marked with #), inbreeding coefficient (F_{IS}) (values significant for Hardy–Weinberg disequilibrium are marked with *), number of haplotypes (Haplo), haplotype diversity (H) and nucleotide diversity (π). For microsatellite values, standard errors are given in parentheses

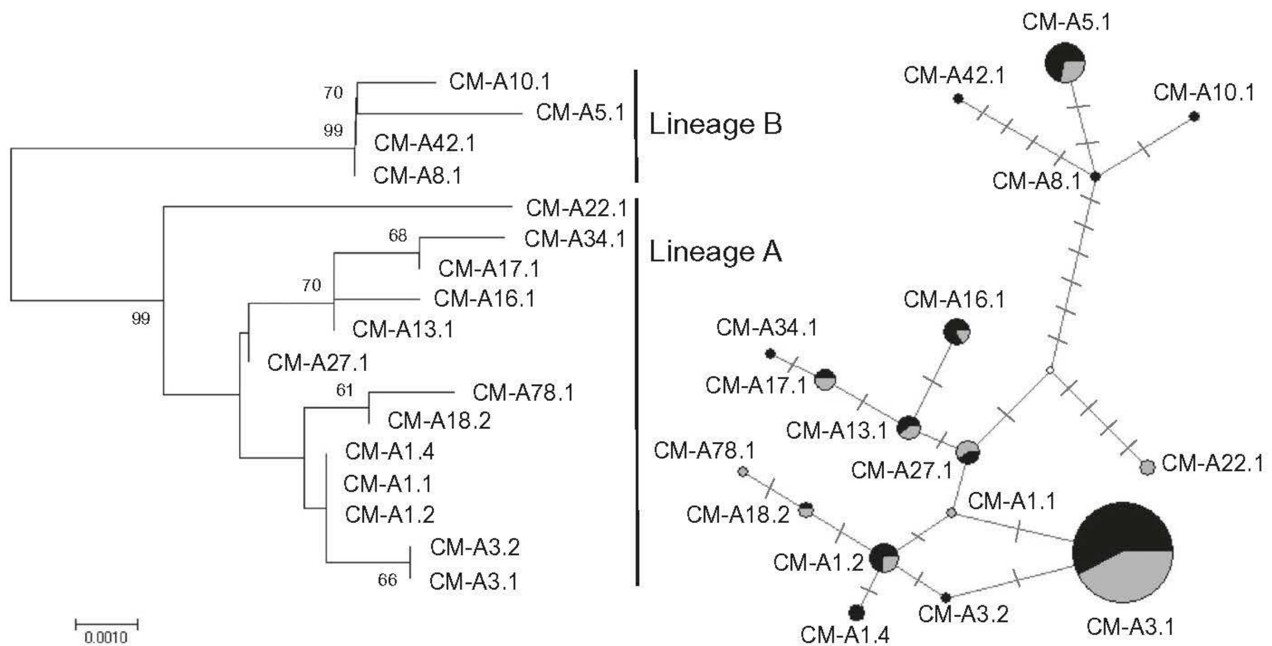


FIGURE 2 Genetic relationships between the haplotypes found in the farm and wild samples of the Cayman Islands. Left: Neighbor-Joining haplotype tree, middle-rooted at the longest branch, indicating maximum likelihood bootstrap values higher than 60%. Top branch represents lineage B, and bottom branch represents lineage A. The new haplotype found (CM-A78.1) belongs to lineage A. Right haplotype network of the individuals of the Cayman Islands. Connecting lines show single mutational changes between haplotypes. The red dot represents an unsampled intermediate haplotype connecting sampled haplotypes. The size of the pies represents haplotype frequencies of farm (blue) and wild (grey) individuals

heterozygosity (H_O) was highest in individuals in cohort C1995, although differences were not significant. We obtained a total of 17 D-Loop haplotypes (Figure 2; Supporting Information Table S6). All but one had been previously described in populations in the Caribbean Sea, South America, South Atlantic (Ascension Island; Formia, Godley, Dontaine, & Bruford, 2006) and Africa (Shamblin, Bagley et al., 2015; Shamblin, Dutton et al., 2015). The haplotypes found in our samples belonged to different lineages as defined in the literature (Naro-Maciel et al., 2014): most haplotypes belonged to lineage A (84%), which is typically found in the Caribbean, while the rest belonged to lineage B (Figure 2) typically found in South America, South Atlantic and Africa (Patrício et al., 2017; Shamblin, Bagley et al., 2015; Shamblin, Dutton et al., 2015). CM-A5.1 is the only haplotype of lineage B found shared by both the captive and the wild populations; however, this haplotype is not exclusive of the South Atlantic region and can be found in other wild populations of the Caribbean region (Naro-Maciel et al., 2014). The new haplotype (CM-A78.1) (GenBank accession no.: MH177873) belongs to lineage A (Figure 2). Haplotype diversity (*H*) and nucleotide diversity (π) decreased from the founder generation to cohort 1995 and to wild females (Table 1). We found 23 different STR haplotypes, with the highest haplotype diversity in the founder generation. When considering the two mitochondrial markers together, 32 haplotypes were obtained (Supporting Information Table S6), also with the highest diversity in the founder generation (Table 1).

3.2 | Relatedness reconstruction among wild and captive individuals

Individuals of the original founder stock (N = 25) showed the lowest degree of relatedness (r = −0.021), while the cohort 1995 (N = 189) showed the highest value (r = −0.003, Table 1). Only C1995 presented relatedness values significantly higher than those expected considering the permutation analysis (p = 0.0001).

COLONY identified a total of 82 mothers and 54 fathers (both assigned and inferred), differentially contributing to the F_1 generation (Supporting Information Figure S1), while almost all founder individuals were sired by different males and females. Mother/father sex ratio proportions for parents assigned to each subset increased from Founders (sex ratio = 0.96), to MCF1 (sex ratio = 1.25) to C1995 (sex ratio = 1.52). The proportion offspring/mother (Founders = 1.05, MCF1 = 1.22 and C1995 = 2.82) and offspring/father increased in the same way, in accordance with the increase of relatedness levels in each group. A total of 43 mothers and 34 fathers were identified for wild females with a sex ratio similar to MCF1 individuals (sex ratio = 1.26).

COLONY identified 40 parent–offspring pairs consisting of seven mothers of the captive breeding stock and 36 captive offspring plus four wild offspring. ML-RELATE found 45 parent–offspring pairs consisting of 27 mothers of the captive breeding stock and 33 captive offspring plus 12 wild offspring. Finally, CERVUS assigned a possible mother to each one of the offspring in the sample set, so we only

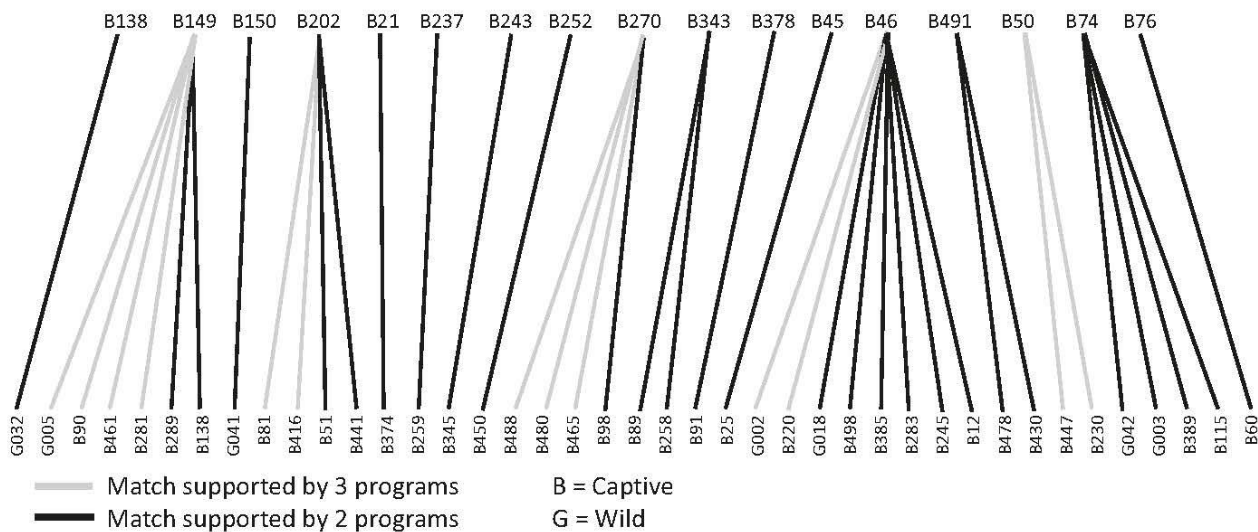


FIGURE 3 Pedigree of individuals of the Cayman Islands Turtle farm breeding stock. Parent–Offspring pairs were inferred by COLONY (Jones & Wang, 2010), CERVUS (Marshall et al., 1998) and ML-RELATE (Kalinowski et al., 2006). The top row consists of captive individuals inferred as mothers; the bottom row consists of wild and captive individuals, assigned to the mother from the farm. Black lines represent matches supported by all three programs, while grey lines represent matches supported by two of the programs

considered the pairs involving a mother already found in at least one of the other two programs as an additional support of the results. Sibship relationships were assigned prioritizing the following order: Parent–Offspring, Full-Siblings, Half-Siblings and Unrelated. Comparison of the three programs found a total of 17 mothers and 41 offspring, of which seven individuals were wild and 34 were from the farm (Figure 3). Of the identified mothers, 12 were parents of captive offspring only, two were parents of wild offspring only and three were parents of both farm and wild offspring. Only 13.2% of farm individuals were assigned to a captive mother. Six wild individuals were full-siblings of one or more captive individuals as estimated using two programs. All Parent–Offspring and Full-Siblings relationships between a wild and a captive individual were consistent with D-loop and STR haplotypes (Supporting Information Table S7). A total of 90% of the wild individuals were found to be related to the farm by at least two of the three programs used, either as offspring or as sibling (Figure 4).

3.3 | Population differentiation

Pairwise F_{ST} values from microsatellite data identified significant genetic differentiation between wild and two farm subsets (Founders and C1995) and also between these two subsets (Table 2). However, pairwise F_{ST} analysis based on the combination of both mtDNA markers, D-loop and STR sequences did not show any significant differentiation.

Moreover, based on published D-loop sequence data from other wild populations of the Caribbean, South Atlantic and Mediterranean Sea (Supporting Information Table S8), we found significant genetic differentiation of both wild and farm Cayman populations to all the other populations of the Atlantic and Mediterranean with two exceptions

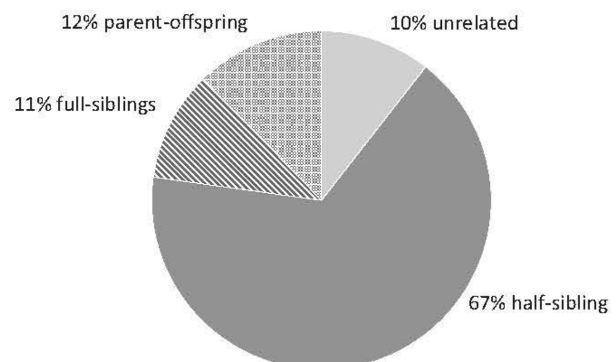


FIGURE 4 Parentage assignment of wild individuals to the farm breeding stock. All assignments are supported by at least two of the three programs used – COLONY (Jones & Wang, 2010), CERVUS (Marshall et al., 1998) and ML-RELATE (Kalinowski et al., 2006)

TABLE 2 Pairwise F_{ST} values among the wild population and the different groups of the farm. Microsatellite results are shown below the diagonal and the results of the combination of D-loop and STR markers are shown above the diagonal. Values in bold are significantly different after FDR correction ($FDR_{0.05}$ = 0.020)

	Wild	Founders	C1995	MCF1
Wild	0	−0.0075	−0.0001	−0.0114
Founders	**0.012**	0	−0.0049	−0.0156
C1995	**0.005**	**0.015**	0	−0.0035
MCF1	0.007	0.033	−0.017	0

within the Caribbean (Supporting Information Table S9). The Dry Tortugas (DRT) population is similar to the farm and wild Cayman Island turtles and the Tequesta (TEQ) population is similar to wild only,

but in both cases the sample size of previous studies was low and several haplotypes present in both farm and wild Cayman Island turtles are absent in DRT and TEQ. Moreover, we found that all Cayman Island sampling groups have high haplotype diversity when compared to the other wild populations of the Caribbean, the South Atlantic and the Mediterranean Sea (Supporting Information Figure S2). MSA of the Founders subset identified that the highest contributions were from Cuba (22.74%), Singer Island, Florida (USA) (14.2%), Mexico (11.76%) and Aves Island, Venezuela (11.76%) (Figure 5). This result is consistent with the known contributions to the founder stock, which are limited to five nesting locations and the foraging area of Nicaragua, potentially hosting individuals from all the Caribbean (Figure 1). On the other hand, only one genetic group was identified by bayesian clustering using STRUCTURE (Supporting Information Figure S3).

4 | DISCUSSION

Biodiversity loss has become a major problem on a global scale and ex situ conservation programmes are a useful tool to preserve biodiversity in a wide range of taxa (Barnosky et al., 2011). It has been estimated that in the next 200 years between 4,000 and 6,000 species of terrestrial vertebrates will require captive breeding and re-introduction to avoid extinction (Frankham, Ballou, & Briscoe, 2010). Ex situ conservation actions require a scientifically informed management strategy throughout the different stages of the process, to establish self-sustained wild populations following the re-introduction. In this study, we have combined the potential of genetic analysis with background information of captive individuals across different generations, to demonstrate how the Cayman Island's re-introduction programmme has contributed to restore the wild population of green turtles. We have also shown how the different farm management strategies have conditioned the genetic composition of the breeding stock with added genetic value for the continuous small replacements of breeders.

4.1 | Farm structure

When the CTF was founded, eggs and adults from different populations in the Caribbean Sea and the South Atlantic Ocean were taken

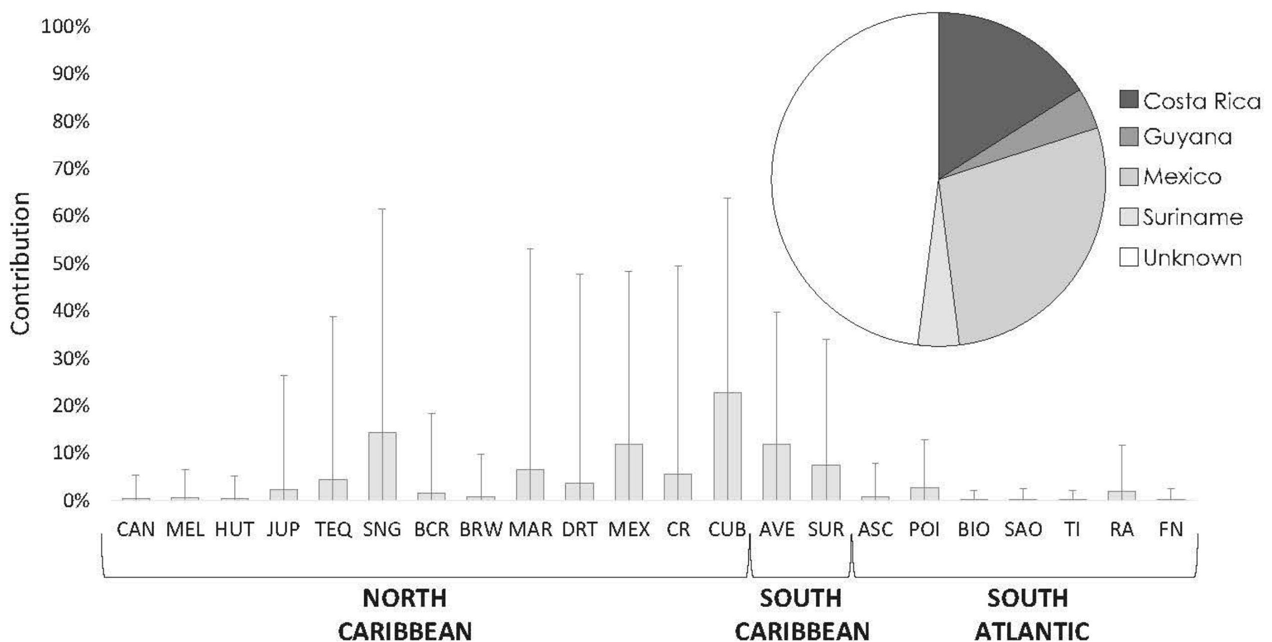


FIGURE 5 Mixed-Stock Analysis of the Founder subset against wild populations based on short (400-bp) D-loop sequences to include samples of known origin of the founders for which long sequences were not available (i.e., Mexico). The highest contribution appears to be from Cuba (CUB), Singer Island (SNG) (Florida), Mexico (MEX) and Aves Island (AVE), corresponding to the putative geographical areas of origin of these individuals. The pie graph represents the origin of founder individuals as reported in the Cayman Turtle Farm background data; the "unknown" category comprises individuals that are known to have wild origin but lack the information on the specific collection site. North Caribbean populations are CAN = Canaveral National Seashores, MEL = Melbourne Beach, Archie Carr National Wildlife Refuge, HUT = Southern Hutchinson Island, JUP = Northern Jupiter Island, TEQ = Tequesta (Southern Jupiter Island), SNG = Singer Island, BCR = Boca Raton, BRW = Hillsboro, MAR = Key West Archie Carr National Wildlife Refuge and DRT = Dry Tortugas National Park (all in Florida, USA) (Shamblin, Bagley et al., 2015), MEX = Quintana Roo (Mexico) (Encalada et al., 1996), CR = Tortuguero (Costa Rica) (Bjorndal, Bolten, & Troeng, 2005) and CUB = Cuba (Ruiz-Urquiola et al., 2010). South Caribbean populations are AVE = Aves Island (Venezuela) and SUR = Matapica (Suriname) (Bolker, Okuyama, Bjorndal, & Bolten, 2007). South Atlantic populations are ASC = Ascension Island, BIO = Bioko Island (Equatorial Guinea), SAO = Sao Tome (Formia et al., 2006), TI = Trinidad Island (Trinidad y Tobago), RA = Rocas Atoll and FN = Fernando de Noronha (Brazil), and POI = Poilao (Guinea Bissau) (Patrício et al., 2017; Shamblin, Dutton et al., 2015)

to the farm to divide the impact of the removal of individuals among different populations (Cayman Turtle Farm, 2002). However, this strategy had an additional unexpected effect, because later studies demonstrated the profound genetic structuring among Atlantic nesting beaches (Naro-Maciel et al., 2014, and references therein). Farm haplotypes in fact belong to both lineage A and lineage B, described by Naro-Maciel et al. (2014) from the Caribbean and from the South Atlantic/Africa region, respectively, consistent with the reported origin of the founder stock.

The founder stock was thus characterized by high initial diversity coupled with an expected low relatedness among individuals. However, the breeding stock suffered a reduction in numbers in October 2001 due to Hurricane Michelle when it was reduced from 355 to 87 adult individuals, only 34 belonging to the initial founder stock. Individuals born on the farm, mainly from the 1995 cohort, were kept for breeding purposes in order to increase the size of the breeding stock after the hurricane, now representing 72.9% of the present captive breeding stock. The high percentage of F_1 breeders not assigned to any of the current founder females (80.2%) shows the contribution of the adult turtles lost in the hurricane or in previous management actions. Thus, F_1 breeders remain a potentially valuable source of diversity to the wild population.

The reduction of the breeding stock caused by the hurricane and the subsequent use of a large number of individuals (189) of one cohort (C1995) in the breeding stock reduced the farm genetic variability at nuclear and mitochondrial markers. Moreover, this management strategy has increased the degree of genetic relatedness within farm individuals (Table 1) due to the higher proportion of breeders sired by the same parents when using this many individuals from a single generation. In contrast, the levels of variability of the MCF1 group are higher, with no signals of inbreeding and lower relatedness values. This suggests that continuous small replacements of the breeding stock using individuals from different cohorts is a better strategy to maintain diversity, when possible. Regardless, the loss of variability and increased relatedness are expected consequences of any captive breeding programmme due to genetic drift, especially in those programmes lacking genetic management (Ralls & Ballou, 1986; Witzenberger & Hochkirch, 2011). Furthermore, the higher observed (H_O) than expected (H_E) heterozygosity in C1995 and the MCF1 fits the expected outcome when individuals from different populations reproduce (Witzenberger & Hochkirch, 2011), as their parents belong to the original founder stock. The observed variations in diversity provide valuable knowledge for future management actions in the farm, for instance, while deciding which individuals to keep for the breeding stock or as a basis for a directed reproduction programme. The correct management of captive stock meant for re-introduction is a critical point for any ex situ programme, as the selection of captive breeders will reflect in the future wild re-introduced population. Therefore, the genetic balance of the captive stock has to be taken under consideration not only at the beginning, but also throughout the whole project to ensure a genetic combination that is as optimal as possible.

4.2 | Relationship with the wild population

During the past 40 years, the CTF has been releasing hatchling and yearling turtles following the headstarting method in order to avoid the high rates of mortality during their early life stages (Bell et al., 2005). Although in the 1980s several studies declared the former wild population extinct (King, 1982; Stoddart, 1980), the Cayman Islands currently hosts a nesting population. Fifty-seven of these nesting females were captured and sampled, but ongoing tagging studies suggest that there are around 100–150 nesting females (ongoing data collection). Our sibship reconstruction showed that the farm provided a significant contribution to the wild population as 90% of the wild nesting females in Grand Cayman are offspring, full- or half-siblings of female captive breeders. Consequently, most mothers and fathers (assigned or inferred by the program) of wild breeding females were either permanently captive in the farm or had escaped from the captive breeding stock (Supporting Information Table S10). The contribution of the farm to the wild population should be considered a minimum, as potential captive parents for the wild breeders and current wild nesters might be part of the farm breeding stock lost in 2001. Furthermore, the contribution of the younger breeders has not yet had an impact on the population, due to the long life cycle of marine turtle, with the released individuals needing between 15 and 19 years to reach maturity, depending on stage of release.

Considering the large number of related individuals detected among captive and wild populations it is not surprising to find no significant differences in haplotype frequencies even for the two mitochondrial markers combined. The two groups share the highly frequent CM-A3.1_6-8-4-4 (30%) haplotype, but also some rare haplotypes, such as CM-A13.1_5-7-7-4 (<5%) and CM-A27.1_5-9-4-4 (<5%), which further reinforces the relatedness between captive and wild populations. By contrast, microsatellites show significant differences between wild and two farm subsets, C1995 and founders, which could be due to the contribution of males to nuclear markers. Therefore, although the success of the re-introduction has already been determined by the outcome of the present analysis, the genotyping of male individuals or the reconstruction of male genotypes (Phillips, Mortimer, Jolliffe, Jorgensen, & Richardson, 2014; Wright et al., 2012) could refine the actual contribution of the farm to the wild population.

Due to the lack of historical samples of the original wild Cayman nesting population for genetic analysis, it is not possible to know the extent of the impact of the farm re-introduction programmme on the recovery of the population but our results indicate two possible scenarios. On the one hand, the original wild population could have been completely replaced by captive individuals and thus the 10% of unrelated wild individuals could be some of the captive individuals lost in 2001 during hurricane Michelle or their descendants or siblings. As evidence for this, the four South Atlantic exclusive haplotypes in the wild population are found in individuals related to the farm as full- or half-siblings. Therefore, these haplotypes may have been inherited from captive

individuals not present in our breeding sample. In fact, the possible escape of captive individuals caused by the hurricane could be considered an accidental re-introduction. On the other hand, the few wild females with no relationship to captive turtles could be the remains of the original wild population. In fact, these non-related individuals presented haplotypes typically found in other Caribbean populations. In the context of a captive breeding or re-introduction programme, these scenarios highlight the importance of collecting samples from wild individuals of a population on the edge of extinction, whenever possible. In fact, the gathering of original wild samples would facilitate the identification of original and re-introduced individuals of the future recovered population, resulting in more accurate management decisions.

4.3 | Comparison with other natural populations

Any re-introduction programmme is usually associated with a decrease in genetic diversity due to the reduced size of the captive stock and to the maintenance of the captive population that may lead to major problems caused by inbreeding depression (Edmands, 2007; Witzenberger & Hochkirch, 2011). Although in 1980 the number of farm founder breeders (208) doubled the optimum suggested by Witzenberger and Hochkirch (2011) to avoid inbreeding and loss of genetic diversity, subsequent deaths in captivity and escapes as a result of the hurricane caused a drop in the number of founders, potentially increasing the risk of inbreeding depression. On the other hand, the different origins of these individuals might trigger the loss of individual fitness due to outbreeding, as a result of negative interpopulation hybridization (Edmands, 2007). Using MSA we showed that the present founder stock still includes individuals from the North Caribbean region (Mexico, Costa Rica and Nicaragua) and the South Caribbean region (Guyana and Suriname), although the contribution of the South Atlantic region (Ascension) remains unknown (Figure 5). However, we found the African haplotypes CM-A8.1 and CM-A42.1 in the C1995 subset. Haplotype CM-A8.1 is the most abundant in Ascension Island (Naro-Maciel et al., 2014), one of the source populations of the founder stock. Haplotype CM-A42.1 is exclusive to Poilao (Patrício et al., 2017) where it coexists at low frequency with CM-A8.1. As the populations of Poilao and Ascension are genetically similar (Patrício et al., 2017), finding the CM-A42.1 haplotype in the farm would imply that this haplotype is also found in Ascension Island but has yet to be discovered there. Considering that after the hurricane catastrophe only 28 founder females out of 148 survived, this reduction probably resulted in extensive loss of haplotypes in the founder stock but that were transmitted to the F_1 and potentially also to the re-introduced individuals (Supporting Information Table S3). Therefore, the former founder stock could have presented African and south Atlantic haplotypes, now not detected in the founders, that could be found in the future in wild breeders if admixture does not compromise their fitness.

The levels of variability of D-loop mtDNA found in captive and wild females are similar to or higher than in other populations of green turtles from the Atlantic Ocean or the Mediterranean Sea analysed in other studies (Figure 5). The explanation of the high diversity found in the CTF may relate to the large number and high diversity of origins of the farm breeders' founder stock. This diverse origin can be easily detected by an increase in observed heterozygosity among the individuals that resulted from the admixture of the founders (MCF1 and C1995); any offspring from parents of different origin are much more likely to have high levels of heterozygosity, due to the parents not sharing common alleles. However, F_{IS} values of wild Cayman females are positive and significant despite their high relatedness to the farm. The admixture of individuals from genetically differentiated units can affect the fitness and reproductive capacity of the offspring because of outbreeding depression (Weeks et al., 2011), by disrupting fine-scale local adaptation or epistatic interactions (Weber et al., 2012). Tentative evidence has been proposed for such an inbreeding–outbreeding tension in an Indian Ocean population of hawksbill turtles (Phillips, Jorgensen, Jolliffe, & Richardson, 2017). In the Cayman Islands, both admixed breeding farm females and sampled wild females seem to be fully capable of reproduction, suggesting that outbreeding depression is not relevant. Nonetheless, the monitoring of diversity along with the study of the reproductive success of the wild population, as well as the farm, is extremely important, in order to evaluate any long-term impact on natural populations. Monitoring studies rarely evaluate re-introduction effects of F_2 or F_3 generations, despite the fact that some of the negative effects of outbreeding may appear in late generations (Edmands, 2007). For instance, a study on artificially translocated pink salmon detected outbreeding depression in F_2 hybrids resulting from spatially separated populations (Gilk et al., 2004). Therefore, when forming a founder stock for captive breeding, although the gathering of individuals from distinct genetic populations is reasonable, the genetic composition of the populations should be previously tested to minimize the risk of outbreeding depression. Therefore, continuous genetic monitoring of wild Cayman nesting events (including fertility and variability records) would be crucial to investigate fitness consequences after different genetic groups have mixed (Edmands, 2007).

4.4 | Concluding remarks

In this study, we have shown that the re-introduction programmme of green turtles in the Cayman Islands has greatly impacted the recovery of the wild population as 90% of the wild population are related to the farm turtles. This re-introduction has been fuelled by a high genetic diversity due to the diverse origin of the founders used to start the captive population. Considering these results, we suggest scientifically controlling the future mating of the captive breeding stock to avoid outbreeding or inbreeding in the captive population while recording fitness values of fecundity and survival. The success of the re-introduction programmme opens new challenges for the future management of the wild population. Further monitoring should assess whether the recovered population is self-sustainable and monitoring is essential to detect and prevent eventual negative impacts on natural populations in the Caribbean. This monitoring is necessary because in species with long life cycles, such as green turtles, potential

shifts in fitness can only be detected in the long term. In this study, we evaluated a re-introduction programmme 40 years after its implementation. However, the ideal scenario for any re-introduction programmme would be to incorporate genetic studies from the beginning. Future captive breeding programmes with re-introduction purposes can benefit from following a few recommendations that arise from this study. First, founder stock individuals should be collected from the genetic region of re-introduction, to avoid the mixing of unrelated genetic groups and the risk of outbreeding. Second, genetic pedigrees could be used to program appropriate breeding strategies to maintain genetic diversity, minimize inbreeding in the captive stock and select individuals for the re-introduction. Finally, temporal monitoring of the wild population should be performed including information regarding its status before re-introduction. Scientifically informed ex situ conservation actions might have higher chances of success in the recovery of endangered species.

ACKNOWLEDGEMENTS

This work was supported by the projects CTM2013-48163 and CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER), by the Darwin Plus Project DPLUS019 "Socio-economic aspects of turtle conservation in the Cayman Islands" funded through the Department for Environment, Food and Rural Affairs (Defra) in the UK and by the projects 9013 and 10084 of the Fundació Bosch I Gimpera-UB. C.C. and M.P. are part of the research group SGR2017-1120 of the Generalitat de Catalunya. A.B. was supported by grant 2017 FI_B 00997 of the Generalitat de Catalunya-AGAUR.

AUTHOR CONTRIBUTIONS

B.G., A.C.B., C.C., M.P. and J.M.B. conceived and designed the study. W.M. sampled all of the Cayman Turtle Farm (CTF) breeders and provided data from the CTF databases. G.E.P., J.M.B., L.C. and J.B. coordinated the sampling of the wild nesting females. A.B., C.M., V.O. and C.C. did the laboratory analysis. A.B., C.M., C.C. and M.P. conducted the data analysis with inputs from A.C.B., B.G. and J.M.B. A.B., C.C., V.O. and M.P. wrote the manuscript with input all of the authors.

DATA ACCESSIBILITY

A list of all the sampled individuals that includes the sample code, population (farm or wild), mtDNA haplotype, STR repeats and microsatellite genotypes is available as Supporting Information Dataset S1.

ORCID

Carlos Carreras https://orcid.org/0000-0002-2478-6445

REFERENCES

Abreu-Grobois, F. A., Horrocks, J., Formia, A., Dutton, P., LeRoux, R., Vélez-Zuazo, X., … Meylan, P. (2006). *New mtDNA Dloop primers which work for a variety of marine turtle species may increase the resolution of mixed stock analyses*. In Proceedings of the 26th annual symposium on sea turtle biology. Island of Crete, Greece: ISTS.

Aiken, J. J., Godley, B. J., Broderick, A. C., Austin, T., Ebanks-Petrie, G., & Hays, G. C. (2001). Two hundred years after a commercial marine turtle fishery: The current status of marine turtles nesting in the Cayman Islands. *Oryx*, *35*(2), 145–151. https://doi.org/10.1017/S0030605300031707

Allan, E., Manning, P., Alt, F., Binkenstein, J., Blaser, S., Blüthgen, N., … Fischer, M. (2015). Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecology Letters*, *18*(8), 834–843. https://doi.org/10.1111/ele.12469

Aragones, L. V., Lawler, I. R., Foley, W. J., & Marsh, H. (2006). Dugong grazing and turtle cropping: Grazing optimization in tropical seagrass systems? *Oecologia*, *149*(4), 635–647.

Armstrong, D., Soderquist, T., & Southgate, R. (1994). Designing experimental reintroductions as experiments. In M. Serena (Ed.), *Reintroduction Biology of Australian and New Zealand Fauna* (pp. 27–29). Chipping Norton, New South Wales: Surrey Beatty & Sons.

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barlow, J., Lennox, G. D., Ferreira, J., Berenguer, E., Lees, A. C., Mac Nally, R., … Oliveira, V. H. F. (2016). Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature*, *535*(7610), 144–147.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., … Maguire, K. C. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*(7336), 51–57.

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

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). *GENETIX 4.05, POPULATION genetics software for Windows TM*. Montpellier, France: Université de Montpellier II.

Bell, C. D., Parsons, J., Austin, T. J., Broderick, A. C., Ebanks-Petrie, G., & Godley, B. J. (2005). Some of them came home: The Cayman Turtle Farm headstarting project for the green turtle *Chelonia mydas*. *Oryx*, *39*(2), 137–148. https://doi.org/10.1017/S0030605305000372

Bjorndal, K. A., Bolten, A. B., & Troeng, 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., Reich, K. J., & Bolten, A. B. (2010). Effect of repeated tissue sampling on growth rates of juvenile loggerhead turtles *Caretta caretta*. *Diseases of Aquatic Organisms*, *88*(3), 271–273. https://doi.org/10.3354/dao02175

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

Bradshaw, P. J., Broderick, A. C., Carreras, C., Fuller, W., Snape, R., Wright, L. I., & Godley, B. J. (2018). Defining conservation units with enhanced molecular tools to reveal fine scale structuring among

Mediterranean green turtle rookeries. *Biological Conservation*, *222*(2018), 253–260. https://doi.org/10.1016/j.biocon.2017.12.014

Brunt, M., & Davies, J. (2012). *The Cayman Islands: Natural history and biogeography*. New York, NY: Springer Science & Business Media.

Canessa, S., Genta, P., Jesu, R., Lamagni, L., Oneto, F., Salvidio, S., & Ottonello, D. (2016). Challenges of monitoring reintroduction outcomes: Insights from the conservation breeding program of an endangered turtle in Italy. *Biological Conservation*, *204*, 128–133. https://doi.org/10.1016/j.biocon.2016.05.003

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., … Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). *Scientific Reports*, *8*(1), 1435. https://doi.org/10.1038/s41598-018-19887-w

Cayman Turtle Farm (2002). *Application to register a captive breeding operation involving Chelonia mydas on Grand Cayman, Cayman Islands*. Submitted by the CITES Management Authority of the United Kingdom of Great Britain and Northern Ireland pursuant to Conference Resolution, 11, 14.

Ceballos, G., & Ehrlich, P. R. (2002). Mammal population losses and the extinction crisis. *Science*, *296*(5569), 904–907.

Chesser, R. K. (1991). Gene diversity and female philopatry. *Genetics*, *127*(2), 437–447.

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. f., … Pascual, M. (2018). Philopatry in loggerhead turtles *Caretta caretta*: Beyond the gender paradigm. *Marine Ecology Progress Series*, *588*, 201–213. https://doi.org/10.3354/meps12448

Coleman, F. C., & Williams, S. L. (2002). Overexploiting marine ecosystem engineers: Potential consequences for biodiversity. *Trends in Ecology & Evolution*, *17*(1), 40–44. https://doi.org/10.1016/S0169-5347(01)02330-8

Coombs, J., Letcher, B., & Nislow, K. (2010). PedAgree: Software to quantify error and assess accuracy and congruence for genetically reconstructed pedigree relationships. *Conservation Genetics Resources*, *2*(1), 147–150. https://doi.org/10.1007/s12686-010-9202-9

Cury, P. (1994). Obstinate nature: An ecology of individuals. Thoughts on reproductive behavior and biodiversity. *Canadian Journal of Fisheries and Aquatic Sciences*, *51*(7), 1664–1673. https://doi.org/10.1139/f94-167

DeMay, S. M., Becker, P. A., Rachlow, J. L., & Waits, L. P. (2017). Genetic monitoring of an endangered species recovery: Demographic and genetic trends for reintroduced pygmy rabbits (*Brachylagus idahoensis*). *Journal of Mammalogy*, *98*(2), 350–364. https://doi.org/10.1093/jmammal/gyw197

Dirzo, R., & Raven, P. H. (2003). Global state of biodiversity and loss. *Annual Review of Environment and Resources*, *28*(1), 137–167. https://doi.org/10.1146/annurev.energy.28.050302.105532

Doherty, T. S., Glen, A. S., Nimmo, D. G., Ritchie, E. G., & Dickman, C. R. (2016). Invasive predators and global biodiversity loss. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(40), 11261–11265. https://doi.org/10.1073/pnas.1602480113

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, *5*(4), 558–567. https://doi.org/10.1046/j.1461-0248.2002.00354.x

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ebner, B., & Thiem, J. (2009). Monitoring by telemetry reveals differences in movement and survival following hatchery or wild rearing of an endangered fish. *Marine and Freshwater Research*, *60*(1), 45–57.

Edmands, S. (2007). Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, *16*(3), 463–475.

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/10.1046/j.1365-294x.1996.00108.x

Excoffier, L., & Lischer, H. E. (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. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fischer, J., & Lindenmayer, D. B. (2000). An assessment of the published results of animal relocations. *Biological Conservation*, *96*(1), 1–11. https://doi.org/10.1016/s0006-3207(00)00048-3

Fontaine, C. (2005). Head-starting the Kemp's ridley sea turtle, *Lepidochelys kempii*, at the NMFS Galveston Laboratory, 1978–1992: A review. *Chelonian Conservation Biology*, *4*, 838–845.

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.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). *Introduction to conservation genetics* (2nd ed.). Cambridge, UK: Cambridge University Press.

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

Gibson, L., Lynam, A. J., Bradshaw, C. J., He, F., Bickford, D. P., Woodruff, D. S., … Laurance, W. F. (2013). Near-complete extinction of native small mammal fauna 25 years after forest fragmentation. *Science*, *341*(6153), 1508–1510.

Gilk, S. E., Wang, I. A., Hoover, C. L., Smoker, W. W., Taylor, S. G., Gray, A. K., & Gharrett, A. J. (2004). Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhynchus gorbuscha*, populations: Marine survival, homing ability, and variability in family size. *Environmental Biology of Fishes*, *69*(1–4), 287–297. https://doi.org/10.1007/978-94-007-0983-6_23

Godoy, J. A., Negro, J. J., Hiraldo, F., & Donazar, J. A. (2004). Phylogeography, genetic structure and diversity in the endangered bearded vulture (*Gypaetus barbatus*, L.) as revealed by mitochondrial DNA. *Molecular Ecology*, *13*(2), 371–390. https://doi.org/10.1046/j.1365-294x.2003.02075.x

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, *28*(4), 1140–1162. https://doi.org/10.1016/S0003-3472(80)80103-5

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95.

Hedrick, P. W., & Fredrickson, R. (2008). Captive breeding and the reintroduction of Mexican and red wolves. *Molecular Ecology*, *17*(1), 344–350. https://doi.org/10.1111/j.1365-294X.2007.03400.x

Hedrick, P. W., Miller, P. S., Geffen, E., & Wayne, R. (1997). Genetic evaluation of the three captive Mexican wolf lineages. *Zoo Biology*, *16*(1), 47–69. https://doi.org/10.1002/(SICI)1098-2361(1997)16:1<47:AID-ZOO7>3.0.CO;2-B

Heppell, S. S. (1998). Application of life-history theory and population model analysis to turtle conservation. *Copeia*, *1998*, 367–375. https://doi.org/10.2307/1447430

Heppell, S. S., Crowder, L. B., & Crouse, D. T. (1996). Models to evaluate headstarting as a management tool for long-lived turtles. *Ecological Applications*, *6*(2), 556–565. https://doi.org/10.2307/2269391

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E., Hungate, B. A., Matulich, K. L., … O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105–108.

Horreo, J. L., de la Hoz, J., Pola, I. G., Machado-Schiaffino, G., & Garcia-Vazquez, E. (2012). Ecological and economic costs of supportive

breeding: Atlantic salmon (*Salmo salar*) as a case study. *Aquaculture, 356*, 1–6. https://doi.org/10.1016/j.aquaculture.2012.06.006

IUCN (1987). *IUCN position statement on traslocation of living organisms: Introductions, re-introductions and re-stocking*. 22nd meeting of the IUCN Council, Gland, Switzerland.

Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources, 10*(3), 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Jule, K. R., Leaver, L. A., & Lea, S. E. (2008). The effects of captive experience on reintroduction survival in carnivores: A review and analysis. *Biological Conservation, 141*(2), 355–363. https://doi.org/10.1016/j.biocon.2007.11.007

Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: A computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Resources, 6*(2), 576–579.

Keenan, K., Mcginnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution, 4*(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kim, Y.-K., Hong, Y.-J., Min, M.-S., Kim, K. S., Kim, Y.-J., Voloshina, I., ... Lee, H. (2011). Genetic status of Asiatic black bear (*Ursus thibetanus*) reintroduced into South Korea based on mitochondrial DNA and microsatellite loci analysis. *Journal of Heredity, 102*(2), 165–174. https://doi.org/10.1093/jhered/esq121

King, F. W. (1982). *Historical review of the decline of the green turtle and the hawksbill*. Washington, DC: Smithsonian Institution Press.

Kitanishi, S., Nishio, M., Uehara, K., Ogawa, R., Yokoyama, T., & Edo, K. (2013). Patterns of genetic diversity of mitochondrial DNA within captive populations of the endangered itasenpara bitterling: Implications for a reintroduction program. *Environmental Biology of Fishes, 96*(4), 567–572. https://doi.org/10.1007/s10641-012-0044-z

Koelewijn, H., Pérez-Haro, M., Jansman, H., Boerwinkel, M., Bovenschen, J., Lammertsma, D., ... Kuiters, A. (2010). The reintroduction of the Eurasian otter (*Lutra lutra*) into the Netherlands: Hidden life revealed by noninvasive genetic monitoring. *Conservation Genetics, 11*(2), 601–614.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*, 1870–1874.

Lee, P. L., Luschi, P., & Hays, G. C. (2007). Detecting female precise natal philopatry in green turtles using assignment methods. *Molecular Ecology, 16*(1), 61–74.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*(11), 1451–1452.

Lohmann, K. J., Lohmann, C. M., Brothers, J. R., & Putman, N. F. (2013). Natal homing and imprinting in sea turtles. *The Biology of Sea Turtles, 3*, 59–77.

Lohmann, K., Witherington, B. E., Lohmann, C. M., & Salmon, M. (1997). Orientation, navigation, and natal beach homing in sea turtles. *The Biology of Sea Turtles, 1*, 107–136.

Lynch, M., & Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Genetics, 152*(4), 1753–1766.

Marshall, T., Slate, J., Kruuk, L., & Pemberton, J. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology, 7*(5), 639–655.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Belknap Press of Harvard University Press.

Meylan, A. B., Bowen, B. W., & Avise, J. C. (1990). A genetic test of the natal homing versus social facilitation models for green turtle migration. *Science, 248*, 724–727.

Mitrus, S. (2005). Headstarting in European pond turtles (*Emys orbicularis*): Does it work? *Amphibia-Reptilia, 26*(3), 333–341.

Moll, E., & Moll, D. (2000). Conservation of river turtles. In M. W. Klemens (Ed.), *Turtle conservation* (pp. 126–155). Washington, DC: Smithsonian Institution Press.

Mora, C., & Sale, P. F. (2011). Ongoing global biodiversity loss and the need to move beyond protected areas: A review of the technical and practical shortcomings of protected areas on land and sea. *Marine Ecology Progress Series, 434*, 251–266. https://doi.org/10.3354/meps09214

Naro-Maciel, E., Reid, B. N., Alter, S. E., Amato, G., Bjorndal, K. A., Bolten, A. 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

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics, 7*(5), 783–787. https://doi.org/10.1007/s10592-005-9056-y

Narum, S. R., Arnsberg, W. D., Talbot, A. J., & Powell, M. S. (2007). Reproductive isolation following reintroduction of Chinook salmon with alternative life histories. *Conservation Genetics, 8*(5), 1123. https://doi.org/10.1007/s10592-006-9268-9

Nichols, J. D., & Williams, B. K. (2006). Monitoring for conservation. *Trends in Ecology & Evolution, 21*(12), 668–673. https://doi.org/10.1016/j.tree.2006.08.007

Ostermann, S. D., Deforge, J. R., & Edge, W. D. (2001). Captive breeding and reintroduction evaluation criteria: A case study of peninsular bighorn sheep. *Conservation Biology, 15*(3), 749–760. https://doi.org/10.1046/j.1523-1739.2001.015003749.x

Patrício, A. R., Formia, A., Barbosa, C., Broderick, A. C., Bruford, M., Carreras, C., ... 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

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics, 28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

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

Phillips, K. P., Jorgensen, T. H., Jolliffe, K. G., & Richardson, D. S. (2017). Evidence of opposing fitness effects of parental heterozygosity and relatedness in a critically endangered marine turtle? *Journal of Evolutionary Biology, 30*(11), 1953–1965.

Phillips, K. P., Mortimer, J. A., Jolliffe, K. G., Jorgensen, T. H., & Richardson, D. S. (2014). Molecular techniques reveal cryptic life history and demographic processes of a critically endangered marine turtle. *Journal of Experimental Marine Biology and Ecology, 455*, 29–37.

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

Puckett, E. E., Kristensen, T. V., Wilton, C. M., Lyda, S. B., Noyce, K. V., Holahan, P. M., ... White, D. (2014). Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA, 50 years after translocation. *Molecular Ecology, 23*(10), 2414–2427.

Ralls, K., & Ballou, J. (1986). Captive breeding programs for populations with a small number of founders. *Trends in Ecology & Evolution, 1*(1), 19–22.

Raymond, M., & Rousset, F. (2004). *GENEPOP (version 3.4): Population genetics software for exact tests and ecumenicism*. Montpellier, France: Laboratoire de Genetique et Environment.

Rees, A., Alfaro-Shigueto, J., Barata, P., Bjorndal, K., Bolten, A., Bourjea, J., ... Carreras, C. (2016). Are we working towards global research priorities for management and conservation of sea turtles? *Endangered Species Research, 31*, 337–382.

RStudio Team (2016). *RStudio: Integrated development for R*. Boston, MA: RStudio Inc.

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

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Sanz, V., & Grajal, A. (1998). Successful reintroduction of captive-raised yellow-shouldered Amazon parrots on Margarita Island, Venezuela. *Conservation Biology*, *12*(2), 430–441.

Sarrazin, F., & Barbault, R. (1996). Reintroduction: Challenges and lessons for basic ecology. *Trends in Ecology & Evolution*, *11*(11), 474–478.

Seminoff, J. (2004). *Chelonia mydas*. IUCN Red List of Threatened Species. Version 2011.

Shamblin, B. M., Bagley, D. A., Ehrhart, L. M., Desjardin, N. A., Martin, R. E., Hart, K. M., ... Sobel, D. (2015). Genetic structure of Florida green turtle rookeries as indicated by mitochondrial DNA control region sequences. *Conservation Genetics*, *16*(3), 673–685.

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

Stenglein, J. L., Waits, L. P., Ausband, D. E., Zager, P., & Mack, C. M. (2010). Efficient, noninvasive genetic sampling for monitoring reintroduced wolves. *Journal of Wildlife Management*, *74*(5), 1050–1058.

Stoddart, D. R. (1980). Little Cayman: Ecology and significance. *Atoll Research Bulletin*, *241*(16), 171.

Storfer, A. (1999). Gene flow and endangered species translocations: A topic revisited. *Biological Conservation*, *87*(2), 173–180.

Sutherland, W. J., Armstrong, D., Butchart, S. H., Earnhardt, J. M., Ewen, J., Jamieson, I., ... Nichols, J. D. (2010). Standards for documenting and monitoring bird reintroduction projects. *Conservation Letters*, *3*(4), 229–235.

Sweeting, R., Beamish, R., Noakes, D., & Neville, C. (2003). Replacement of wild coho salmon by hatchery-reared coho salmon in the Strait of Georgia over the past three decades. *North American Journal of Fisheries Management*, *23*(2), 492–502. https://doi.org/10.1577/1548-8675(2003)023<0492:ROWCSB>2.0.CO;2

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 11030–11035. https://doi.org/10.1073/pnas.0404206101

Tikochinski, Y., Bendelac, R., Barash, A., Daya, A., Levy, Y., & Friedmann, A. (2012). Mitochondrial DNA STR analysis as a tool for studying the green sea turtle (Chelonia mydas) populations: The Mediterranean Sea case study. *Marine Genomics*, *6*, 17–24. https://doi.org/10.1016/j.margen.2012.01.002

Turner, I. (1996). Species loss in fragments of tropical rain forest: A review of the evidence. *Journal of Applied Ecology*, 200–209. https://doi.org/10.2307/2404743

Van Dierendonck, M., Bandi, N., Batdorj, D., Dügerlham, S., & Munkhtsog, B. (1996). Behavioural observations of reintroduced Takhi or Przewalski horses (*Equus ferus przewalskii*) in Mongolia. *Applied Animal Behaviour Science*, *50*(2), 95–114. https://doi.org/10.1016/0168-1591(96)01089-1

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources*, *4*(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Warwick, C., Arena, P. C., & Steedman, C. (2013). Health implications associated with exposure to farmed and wild sea turtles. *JRSM Short Reports*, *4*(1), 1–7. https://doi.org/10.1177/2042533313475574

Weber, S. B., Broderick, A. C., Groothuis, T. G., Ellick, J., Godley, B. J., & Blount, J. D. (2012). Fine-scale thermal adaptation in a green turtle nesting population. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1731), 1077–1084. https://doi.org/10.1098/rspb.2011.1238

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., ... Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evolutionary Applications*, *4*(6), 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Witzenberger, K. A., & Hochkirch, A. (2011). Ex situ conservation genetics: A review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. *Biodiversity and Conservation*, *20*(9), 1843–1861. https://doi.org/10.1007/s10531-011-0074-4

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., ... Palumbi, S. R. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science*, *314*(5800), 787–790.

Wright, L. I., Fuller, W. J., Godley, B. J., McGowan, A., Tregenza, T., & Broderick, A. C. (2012). Reconstruction of paternal genotypes over multiple breeding seasons reveals male green turtles do not breed annually. *Molecular Ecology*, *21*(14), 3625–3635. https://doi.org/10.1111/j.1365-294X.2012.05616.x

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Barbanti A, Martin C, Blumenthal JM, et al. How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Mol Ecol*. 2019;28:1637–1651. https://doi.org/10.1111/mec.15017

MOLECULAR ECOLOGY

Supplementary Information for:

How many came home? An evaluation of ex-situ conservation of green turtles in the Cayman Islands.

Anna Barbanti, Clara Martin, Janice M. Blumenthal, Jack Boyle, Annette C. Broderick, Lucy Collyer, Gina Ebanks-Petrie, Brendan J. Godley, Walter Mustin, Víctor Ordóñez, Marta Pascual, Carlos Carreras.

Table of Contents:

Table S1	Page 2
Table S2	Page 3
Table S3	Page 4
Table S4	Page 5
Table S5	Page 6
Table S6	Page 7
Table S7	Page 8
Table S8	Page 9
Table S9	Page 10
Figure S1	Page 11
Figure S2	Page 12
Figure S3	Page 13
References	Page 14

Table S1. Eggs collected between 1968 and 1978 and part of the Cayman Turtle Farm founder stock. On average, 64% of eggs successfully hatched. One percent of hatched eggs were released at the collection sites as one-year old animals (Cayman Turtle farm, 2002).

Country	1968	1969	1970	1971	1972	1973	1974	1976	1977	1978
Costa Rica	15000	15000	15000	-	14958	14803	-	-	-	-
Ascension Island	-	15000	15000	-	16746	19105	19814	-	-	-
Guyana	-	-	5000	-	-	-	-	-	-	-
Suriname	-	-	24000	30000	29582	63404	60650	42830	33609	28173

Table S2. Adult, sub-adult and juvenile turtles collected between 1968 and 1977 to form the Cayman Turtle Farm founder stock. Individuals were taken from Costa Rica, Suriname, Guyana, Ascension Island, Mexico and Nicaragua. These individuals cannot be precisely divided by life stage or origin because for most of them one of these parameters or both are not specified in the literature (Cayman Turtle Farm, 2002).

Year	Turtles
1968	350
1968-1973	60
1971	31
1976-1977	117

Table S3. Number of breeding founder individuals in the Cayman Turtle Farm from 1973 to 2001 before (1) and after (2) hurricane Michelle (Cayman Turtle Farm, 2002).

Year	Males	Females	Total
1973	10	65	75
1974	13	75	88
1975	22	89	111
1976	17	125	142
1977	42	172	214
1978	84	401	485
1979	90	450	540
1980	90	442	532
1981	62	437	499
1982	62	306	368
1983	41	228	269
1984	48	219	267
1985	45	212	257
1986	44	212	256
1987	41	210	251
1988	38	208	246
1989	32	204	236
1990	32	204	236
1991	26	195	221
1992	32	191	223
1993	34	205	239
1994	41	174	215
1995	38	171	209
1996	38	169	207
1997	37	165	202
1998	35	156	191
1999	35	155	190
2000	34	150	184
2001(1)	34	148	182
2001(2)	6	28	34

Table S4. PCR amplification conditions used for each marker type. The primers used for amplification are described in the references.

Marker	Reference	PCR		
		Step 1	**Step2**	**Step 3**
Microsatellites	See Supplementary Table 5	95°C 15min	35 cycles of 94°C 30s, 58°C 90s, 72°C 60s	60°C 30min
mtDNA D-Loop 800bp	Abreus-Grobois et al., 2006 Clusa et al., 2013	94°C 5min	40 cycles of 94°C 60s, 52°C 60s, 72°C 90s	72°C 10min
mtDNA STRs	Tikochinski et al., 2012	95°C 15min	35 cycles of 94°C 60s, 52°C 60s, 72°C 90s	72°C 10min

Table S5. Characterization of microsatellites loci in green turtles (*Chelonia mydas*) of the Cayman Islands Farm and Wild individuals sampled in the Cayman Islands. He = expected heterozygosity; Ho = observed heterozygosity. The primers used for amplification are described in the references.

					Farm (n = 257)		Wild (n = 57)	
Locus	**Multiplex**	**Fluorescent dye**	**Reference**	**No of alleles**	**He**	**Ho**	**He**	**Ho**
Cm3	1	NED	FitzSimmons et al., 1995	10	0.741	0.757	0.745	1.000
Cc28	1	HEX	Monzon-Arguello et al., 2008	11	0.840	0.843	0.836	0.785
Cc7E11	1	NED	Shamblin et al., 2007	8	0.591	0.603	0.636	0.666
CcP7D04	1	6-FAM	Shamblin et al., 2009	16	0.884	0.906	0.901	0.781
Klk314	2	6-FAM	Shamblin et al., 2009	6	0.734	0.852	0.791	0.821
Cm58	2	NED	FitzSimmons et al., 1995	10	0.765	0.766	0.687	0.719
B103	2	HEX	Dutton & Frey 2009	10	0.757	0.743	0.723	0.578
Cc2	2	NED	Monzon-Arguello et al., 2008	23	0.832	0.879	0.818	0.767
C102	2	6-FAM	Dutton & Frey 2009	9	0.479	0.505	0.43	0.339
B123	2	HEX	Dutton & Frey 2009	6	0.540	0.529	0.362	0.333
A6	1	HEX	Dutton & Frey 2009	10	0.727	0.742	0.716	0.772
D2	1	HEX	Dutton & Frey 2009	16	0.666	0.612	0.67	0.625
Or7	1	6-FAM	Aggarwal et al., 2004	9	0.689	0.634	0.711	0.627

6

Table S6. Mitochondrial haplotype distribution considering the D-Loop haplotype and the number of *AT* STR repeats among farm and wild individuals of the Cayman Islands. "A" and "B" lineages correspond to the lineages described by Naro-Maciel et al., 2014. (†) This sample presents a (T) insertion in STR 2 that divides this STR in two fragments.

ID	D-loop	STRs	Lineage	Farm	Wild
H1_CM-A1.1_7-7-4-4	CM-A1.1	7-7-4-4	A	0	1
H2_CM-A1.2_6-8-4-4	CM-A1.2	6-8-4-4	A	2	1
H3_CM-A1.2_7-7-4-4	CM-A1.2	7-7-4-4	A	1	0
H4_CM-A1.2_8-10-4-4	CM-A1.2	8-10-4-4	A	3	1
H5_CM-A1.4_6-8-4-4	CM-A1.4	6-8-4-4	A	1	0
H6_CM-A1.4_7-8-4-4	CM-A1.4	7-8-4-4	A	1	0
H7_CM-A3.1_5-8-4-4	CM-A3.1	5-8-4-4	A	2	0
H8_CM-A3.1_5-9-4-4	CM-A3.1	5-9-4-4	A	6	2
H9_CM-A3.1_6-8-4-4	CM-A3.1	6-8-4-4	A	26	19
H10_CM-A3.1_6-8-4-5	CM-A3.1	6-8-4-5	A	1	0
H11_CM-A3.1_7-7-4-4	CM-A3.1	7-7-4-4	A	0	1
H12_CM-A3.1_7-7-4-5	CM-A3.1	7-7-4-5	A	2	0
H13_CM-A3.1_7-8-4-4	CM-A3.1	7-8-4-4	A	13	13
H14_CM-A3.1_7-8-4-6	CM-A3.1	7-8-4-6	A	0	2
H15_CM-A3.2_6-8-4-4	CM-A3.2	6-8-4-4	A	1	0
H16_CM-A5.1_6-12-4-4	CM-A5.1	6-12-4-4	B	1	1
H17_CM-A5.1_6-13-4-4	CM-A5.1	6-13-4-4	B	4	1
H18_CM-A5.1_6-14-4-4	CM-A5.1	6-14-4-4	B	1	0
H19_CM-A5.1_7-11-4-4	CM-A5.1	7-11-4-4	B	1	0
H20_CM-A5.1_7-12-4-4	CM-A5.1	7-12-4-4	B	3	2
H21_CM-A8.1_7-12-4-4	CM-A8.1	7-12-4-4	B	1	0
H22_CM-A10.1_7-16-4-4	CM-A10.1	7-16-4-4	B	1	0
H23_CM-A13.1_5-7-7-4	CM-A13.1	5-7-7-4	A	3	2
H24_CM-A16.1_5-11-6-5	CM-A16.1	5-11-6-5	A	0	1
H25_CM-A16.1_5-8-6-4	CM-A16.1	5-8-6-4	A	5	0
H26_CM-A17.1_7-10-7-4	CM-A17.1	7-10-7-4	A	2	2
H27_CM-A18.2_7-6-5-4	CM-A18.2	7-6-5-4	A	1	1
H28_CM-A22.1_5-8-4-5	CM-A22.1	5-8-4-5	A	0	2
H29_CM-A27.1_5-9-4-4	CM-A27.1	5-9-4-4	A	2	3
H30_CM-A34.1_7-7+6-6-4	CM-A34.1	7-(7+6)-6-4†	A	0	1†
H31_CM-A42.1_7-12-4-4	CM-A42.1	7-12-4-4	B	1	0
H32_CM-A78.1_8-7-4-4	NEW	8-7-4-4	A	0	1

Table S7. Parent-offspring (PO) and full-siblings (FS) relationships between wild and farm individuals based on three programs: COLONY, CERVUS and ML-RELATE. The table only shows pairs found by a minimum of two out of the three programs. D-Loop and STR columns show the corresponding haplotypes of the pair. NA indicates not available, as CERVUS only reports parentage relationships.

Wild ID	Farm ID	Relationship	COLONY	CERVUS	ML-RELATE	D-loop	STR
G005	B149	PO	1	1	1	CM-A3.1	5-9-4-4
G002	B46	PO	1	1	1	CM-A3.1	6-8-4-4
G042	B74	PO	0	1	1	CM-A3.1	6-8-4-4
G041	B150	PO	0	1	1	CM-A3.1	6-8-4-4
G032	B138	PO	0	1	1	CM-A3.1	7-8-4-4
G018	B46	PO	1	1	0	CM-A3.1	6-8-4-4
G003	B74	PO	0	1	1	CM-A3.1	6-8-4-4
G056	B390	FS	1	NA	1	CM-A17.1	7-10-7-4
G053	B301	FS	1	NA	1	CM-A1.2	8-10-4-4
G050	B415	FS	1	NA	1	CM-A3.1	7-8-4-4
G034	B415	FS	1	NA	1	CM-A3.1	7-8-4-4
G027	B415	FS	1	NA	1	CM-A3.1	7-8-4-4
G001	B415	FS	1	NA	1	CM-A3.1	6-8-4-4

Table S8. D-Loop 800bp haplotype frequencies for the green turtle populations. GAN: Genbank Accession Number. The populations include the Cayman Island populations from the present study (FAR=Farm and WIL=Wild), and published wild populations of the Caribbean Sea: CAN = Canaveral National Seashores, MEL = Melbourne Beach, Archie Carr National Wildlife Refuge, HUT = Southern Hutchinson Island, JUP = Northern Jupiter Island, TEQ = Tequesta (Southern Jupiter Island), SNG = Singer Island, BCR = Boca Raton, BRW = Hillsboro, MAR = Key West Archie Carr National Wildlife Refuge and DRT = Dry Tortugas National Park (all in Florida, USA) (Shamblin et al., 2015a). South Atlantic region:TI = Trinidad Island (Trinidad y Tobago), RA = Rocas Atoll and FN = Fernando de Noronha (Brazil), and POI = Poilao (Guinea Bissau) (Shamblin et al., 2015b; Patrício et al., 2017). Mediterranean region: SKA = South Karpaz, NKA = North Karpaz, ALA = Alagadi and AKD = Akdeniz (Cyprus) (Bradshaw et al., 2018).

HAPLOTYPE	CAN	MEL	HUT	JUP	TEQ	SNG	BCR	BRW	MAR	DRT	TI	RA	FN	POI	SKA	NKA	ALA	AKD	FAR	WIL	TOTAL	GAN
A1.1	22	150	14	10	3	1														1	**201**	JF308465
A1.2		10	1	3			4	3		3									6	2	**32**	JF308466
A1.4																			2		**2**	KT581617
A2.1		4	2	1	1		2	3		3											**16**	JX306006
A3.1	8	76	8	29	9	25	21	14	15	14									50	37	**306**	JN632497
A3.2																			1		**1**	JN632497
A5.1		2								4									10	4	**20**	JN632498
A8.1			1								55	27	14	169					1		**267**	JF308472
A8.3											12	1		1							**14**	JF308474
A9.1											19	2									**21**	JF308475
A10.1												2							1		**3**	JF308476
A11.1											1										**1**	KT232136
A12.1												3									**3**	JF308482
A13.1		5	2	1	1										33	52	226	83	3	2	**408**	JX306007
A14.1															11	2	8	1			**22**	KR011755
A16.1		2				1													5	1	**9**	JN632500
A17.1				1	1														2	2	**6**	JQ420802
A18.2		1						1											1	1	**4**	JX306008
A22.1																				2	**2**	KT581619
A23.1											6										**6**	JF308478
A24.1											1										**1**	JF308479
A25.1												1	2								**3**	JF308483
A27.1																			2	3	**5**	**MH025956**
A28.1		2					1	1		1											**5**	JX306009
A32.1											5	1									**6**	JF308480
A34.1																				1	**1**	
A42.1														1					1		**2**	JF308481
A53.1						1	1	1													**3**	JX306010
A60.1															1						**1**	KR011754
A62.1															1						**1**	
A78.1																				1	**1**	MH177873
TOTAL	30	252	28	45	15	28	29	23	15	25	99	37	16	171	46	54	234	84	85	57	1373	

Table S9. Population pairwise F_{ST} values using D-Loop 800bp sequences for green turtle. Population codes are the same as in Supplementary table S8. FDR correction: $FDR_{0.05}$ = 0.0085. In bold are indicated non-significant values after FDR correction.

	FAR	WIL	CAN	MEL	HUT	JUP	TEQ	SNG	BCR	BRW	MAR	DRT	TI	RA	FN	POI	SKA	NKA	ALA	AKD	Reference
FAR	-																				Present study
WIL	**0.002**	-																			Present study
CAN	0.105	0.104	-																		Shambiln et al., 2015a
MEL	0.167	0.146	**-0.01**	-																	Shambiln et al., 2015a
HUT	0.056	0.05	**0.001**	**-0.003**	-																Shambiln et al., 2015a
JUP	0.072	0.035	0.164	0.108	0.095	-															Shambiln et al., 2015a
TEQ	0.029	**-0.007**	0.15	0.086	0.032	**-0.017**	-														Shambiln et al., 2015a
SNG	0.094	0.065	0.509	0.316	0.299	0.09	0.125	-													Shambiln et al., 2015a
BCR	0.073	0.042	0.31	0.215	0.176	0.01	0.038	0.033	-												Shambiln et al., 2015a
BRW	0.056	0.029	0.213	0.156	0.102	**0.000**	**0.007**	0.086	**-0.023**	-											Shambiln et al., 2015a
MAR	0.091	0.068	0.658	0.361	0.338	0.141	0.202	**-0.015**	0.066	0.12	-										Shambiln et al., 2015a
DRT	**-0.014**	**0.004**	0.149	0.218	0.067	0.101	0.041	0.138	0.089	0.055	0.122	-									Shambiln et al., 2015a
TI	0.708	0.799	0.91	0.898	0.875	0.902	0.893	0.916	0.904	0.895	0.919	0.797	-								Shambiln et al., 2015b
RA	0.651	0.753	0.914	0.899	0.857	0.898	0.88	0.919	0.899	0.882	0.924	0.733	0.05	-							Shambiln et al., 2015b
FN	0.615	0.71	0.885	0.896	0.808	0.872	0.818	0.889	0.863	0.834	0.883	0.65	0.155	0.033	-						Shambiln et al., 2015b
POI	0.814	0.898	0.987	0.937	0.967	0.974	0.982	0.988	0.98	0.977	0.994	0.92	0.168	0.067	0.371	-					Patricio et al., 2017
SKA	0.396	0.474	0.706	0.703	0.588	0.694	0.613	0.752	0.712	0.673	0.764	0.527	0.886	0.873	0.842	0.963	-				Bradshaw et al., 2018
NKA	0.419	0.525	0.879	0.726	0.731	0.815	0.809	0.904	0.855	0.825	0.946	0.625	0.922	0.936	0.922	0.988	0.173	-			Bradshaw et al., 2018
ALA	0.606	0.713	0.903	0.785	0.841	0.881	0.878	0.925	0.906	0.894	0.941	0.808	0.95	0.962	0.961	0.983	0.301	**-0.011**	-		Bradshaw et al., 2018
AKD	0.474	0.594	0.929	0.749	0.808	0.867	0.884	0.945	0.905	0.885	0.98	0.712	0.938	0.957	0.951	0.992	0.277	**0.000**	**0.001**	-	Bradshaw et al., 2018

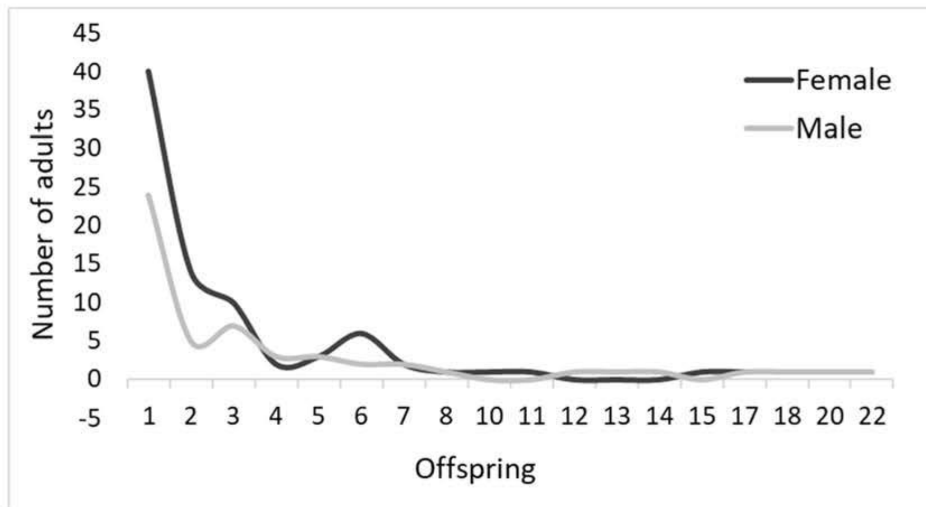


Figure S1. Individual contribution of males and females to the farm F_1 generation. This graph reports the number of offspring for every male and female turtle detected or inferred by COLONY v2 (Jones & Wang, 2010) as parent of farm individuals.

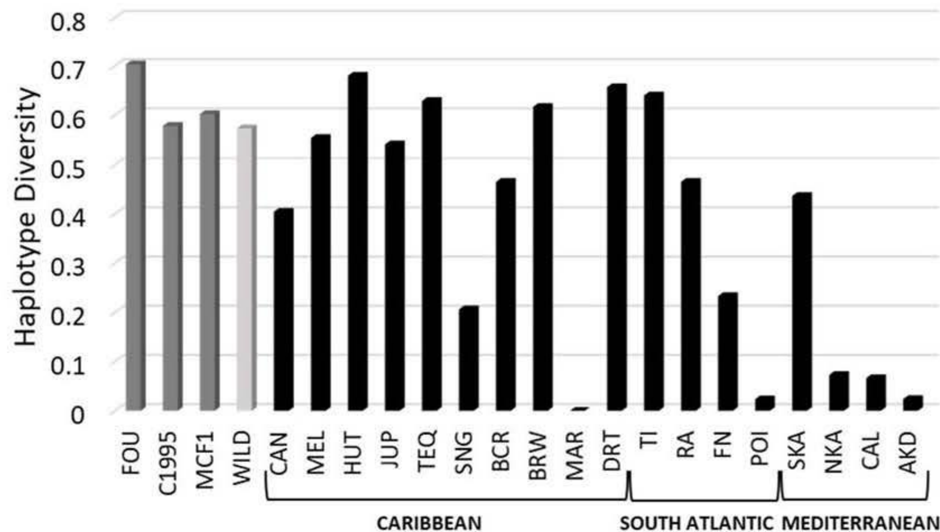


Figure S2. Haplotype diversity in D-loop sequences (800bp) from the Cayman Island compared to other wild populations. Haplotype diversity of farm breeding stock subgroups in dark grey, Cayman wild population in light grey and other wild populations in black. Farm subgroups are FOU = founders, MCF1= breeders born in captivity and C1995 = breeders born in captivity in 1995 (Present study). Caribbean populations are CAN = Canaveral National Seashores, MEL = Melbourne Beach, Archie Carr National Wildlife Refuge, HUT = Southern Hutchinson Island, JUP = Northern Jupiter Island, TEQ = Tequesta (Southern Jupiter Island), SNG = Singer Island, BCR = Boca Raton, BRW = Hillsboro, MAR = Key West Archie Carr National Wildlife Refuge and DRT = Dry Tortugas National Park (all in Florida, USA)(Shamblin et al., 2015a). South Atlantic populations are TI = Trinidad Island (Trinidad y Tobago), RA = Rocas Atoll and FN = Fernando de Noronha (Brazil), and POI = Poilao (Guinea Bissau) (Shamblin et al., 2015b; Patrício et al., 2017). Mediterranean populations are SKA = South Karpaz, NKA = North Karpaz, ALA = Alagadi and AKD = Akdeniz (Cyprus) (Bradshaw et al., 2018).

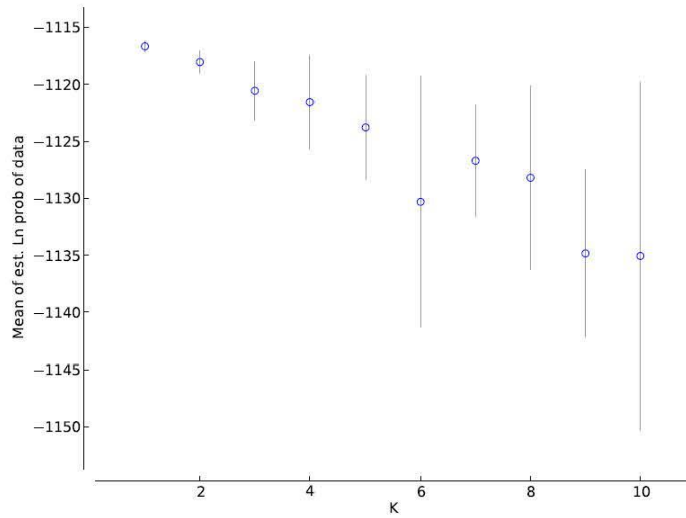


Figure S3. Mean Ln probability across 20 runs at different number of genetic clusters (K) for the Founder subset. The highest value is K=1, which means that the program STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000) does not divide these individuals into distinct populations.

References

Abreu-Grobois, F. A., Horrocks, J., Formia, A., Dutton, P., LeRoux, R., Vélez-Zuazo, X., ... Meylan, P. (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. Island of Crete, Greece: ISTS.

Aggarwal, R. K., Velavan, T. P., Udaykumar, D., Hendre, P. S., Shanker, K., Choudhury, B. C., & Singh, L. (2004). Development and characterization of novel microsatellite markers from the olive ridley sea turtle (*Lepidochelys olivacea*). *Molecular Ecology Resources*, *4*(1), 77-79.

Bjorndal, K. A., Bolten, A. B. & Troeng, 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.

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.

Bradshaw, P.J., Broderick, A.C., Carreras, C. Fuller, W., Snape, R., Wright, L. I. & Godley, B. J. (2018). Defining conservation units with enhanced molecular tools to reveal fine scale structuring among Mediterranean green turtle rookeries, *Biological Conservation*, 222(2018), 253-260. https://doi.org/10.1016/j.biocon.2017.12.014

Cayman Turtle Farm. (2002). Application to register a captive breeding operation involving Chelonia mydas on Grand Cayman, Cayman Islands. Submitted by the CITES Management Authority of the United Kingdom of Great Britain and Northern Ireland pursuant to Conference Resolution, 11, 14.

Clusa, M., Carreras, C., Pascual, M., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... Türkozan, O. (2013). Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). *Journal of Experimental Marine Biology and Ecology*, *439*, 15-24.

Dutton, P. H., & Frey, A. (2009). Characterization of polymorphic microsatellite markers for the green turtle (*Chelonia mydas*). *Molecular Ecology Resources*, 9(1), 354-356.

FitzSimmons, N. N., Moritz, C., & Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution*, *12*(3), 432-440.

Jones, O. R. & Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551-555.

Monzón-Argüello, C., Muñoz, J., Marco, A., López-Jurado, L. F., & Rico, C. (2008). Twelve new polymorphic microsatellite markers from the loggerhead sea turtle (*Caretta caretta*) and

cross-species amplification on other marine turtle species. *Conservation Genetics*, *9*(4), 1045-1049.

Naro-Maciel, E., Reid, B. N., Alter, S. E., Amato, G., Bjorndal, K. A., Bolten, A. B., ... Pineda-Catalan, O. (2014). From refugia to rookeries: phylogeography of Atlantic green turtles. *Journal of Experimental Marine Biology and Ecology*, 461, 306-316.

Patricio, 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, doi: 10.3354/meps12078.

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

Shamblin, B. M., Faircloth, B. C., Dodd, M., Wood-Jones, A., Castleberry, S. B., Carroll, J. P., & Nairn, C. J. (2007). Tetranucleotide microsatellites from the loggerhead sea turtle (*Caretta caretta*). *Molecular Ecology Resources*, *7*(5), 784-787.

Shamblin, B. M., Faircloth, B. C., Dodd, M. G., Bagley, D. A., Ehrhart, L. M., Dutton, P. H., ... & Nairn, C. J. (2009). Tetranucleotide markers from the loggerhead sea turtle (*Caretta caretta*) and their cross-amplification in other marine turtle species. *Conservation Genetics*, *10*(3), 577-580.

Shamblin, B. M., Bagley, D. A., Ehrhart, L. M., Desjardin, N. A., Martin, R. E., Hart, K. M., ... & Johnson, C. (2015a). Genetic structure of Florida green turtle rookeries as indicated by mitochondrial DNA control region sequences. *Conservation Genetics*, 16(3), 673-685.

Shamblin, B. M., Dutton, P. H., Bjorndal, K. A., Bolten, A. B., Naro-Maciel, E., Santos, A. J. B., ... & Nairn, C. J. (2015b). Deeper mitochondrial sequencing reveals cryptic diversity and structure in Brazilian green turtle rookeries. *Chelonian Conservation and Biology*, 14(2), 167-172.

Tikochinski, Y., Bendelac, R., Barash, A., Daya, A., Levy, Y. & Friedmann, A. (2012). Mitochondrial DNA STR analysis as a tool for studying the green sea turtle (*Chelonia mydas*) populations: the Mediterranean Sea case study. *Marine genomics*, 6, 17-24.

CHAPTER 2

Architecture of assisted colonisation in sea turtles

Anna Barbanti[1], Janice M. Blumenthal[2], Annete C. Broderick[3], Brendan J. Godley[3], Maria Turmo[1], Marta Pascual[1], Carlos Carreras[1]

Affiliations

1.Department of Genetics, Microbiology and Statistics and IRBio, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain.

2. Department of Environment, PO Box 10202, Grand Cayman KY1-1002 Cayman Islands

3. Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, TR10 9FE, UK

Abstract

Many species will have to adapt to changing environmental conditions or occupy new suitable areas to avoid potential extinction in the biodiversity crisis our planet is facing. Long-lived animals, with limited colonising potential, are especially vulnerable and *ex-situ* conservation actions can provide solutions through assisted colonisations. However, there is little empirical evidence on the process of founding new populations for such species organisms, nor on the feasibility of assisted colonisations as a viable conservation measure. Here we combined genetics with reproductive data to study the rise of two populations of green turtle (*Chelonia mydas*) in the Cayman Islands as a result of a reintroduction program that started 50 years ago. The two reintroduced populations rapidly diverged from the captive population of origin due to genetic drift, although direct relatedness between individuals could still be detected. Individuals from the reintroduced populations showed high levels of nest fidelity, indicating that philopatry may help reinforce the success of new populations. Additionally, we show that reintroduction from captive populations has not undermined the reproductive fitness of the individuals, and finally, that reintroduction programs of sea turtles can be very successful in establishing new populations.

The rate of biodiversity loss has accelerated during the last decade[1]. Anthropogenic impacts such as global warming, habitat alteration and human-mediated dispersal are some of the main causes of the biodiversity crisis facing animal and plant species at a global scale[2]. Ecosystems are being dramatically altered to the point that they are no longer suitable for some of the organisms they contain[3]. Consequently, species have to either adapt or move to new suitable habitats to avoid extinction. For this reason, several species are changing their distributions by founding new populations worldwide[4]. Under these circumstances, species with limited potential to adapt their distributions to the new climatic conditions are potentially more vulnerable and thus more likely to become extinct[5]. While adaptation is difficult to predict, range expansions can be detected and even facilitated through reintroductions from *ex-situ* conservation programs[6–8]. Unfortunately, reintroductions from captive breeding programs are rarely evaluated to assess their longer term success[9–11]. This evaluation is crucial, as newly founded populations can suffer reductions of genetic variability due to the founder effect, or detrimental effects in the reproductive behaviour of the reintroduced individuals[12]. The foundation process of new populations has been theoretically described in the past century[13], however few studies have provided empirical data, and most focus on short-lived organisms[14,15].

The study of founding processes in long-lived vertebrates is very challenging, but essential in the current era of global biodiversity decline, as these species are potentially highly vulnerable to habitat alterations and can have slow responses to environmental change. As reptiles, sea turtles are highly affected by temperature[16,17], and have Temperature-dependent Sex Determination (TSD) with rising temperatures causing the feminization of nesting populations[18]. Furthermore, modelling studies have predicted a mid-term collapse of existing nesting populations worldwide due to environmental changes in current nesting areas while new potential areas would become optimal for nesting[19,20]. While sea turtles are highly migratory species[21], their potential to colonise new nesting areas is limited due to their philopatric behaviour[22,23]. To date only few cases of change in distribution of sea turtle nesting areas have been detected[24] and for this reason, assisted colonisation has been proposed as a promising conservation tool to conserve populations threatened by anthropogenic activities or to reinforce natural expansion processes[25]. The Cayman Island green turtle reintroduction program offers a unique opportunity to study the process and consequences of an assisted colonisation of sea turtles.

The Cayman Islands (Figure 1a) green turtle (*Chelonia mydas*) nesting populations was considered nearly extinct[26], however over the past 20 years it has increased exponentially in Grand Cayman (Figure S1) potentially in part as a result of the reintroduction program initiated in 1983 from the now named Cayman Turtle Center (CTC)[27,28]. This reintroduction was based on

headstarting (i.e. the rearing of offspring in captivity for the first few years) and yearlings would be released to sea from Grand Cayman Island, where the CTC is based. This strategy increases survival and, because of philopatry, ensures that the released yearlings would one day come back as adults to nest where they had been released. Since philopatry drives genetically apart geographically distant populations, this strategy would probably also cause the genetic isolation of the reintroduced population from others surrounding it. The process and rate of differentiation into genetically separated nesting grounds has never been observed in a newly founded sea turtle population. For this reason, we have studied the foundation and differentiation process of the Grand Cayman wild population (where the CTC is based) and also the possible impact of the CTC on the nearby population of Little Cayman (distant 108.4 km), the role of philopatry during foundation, and effect of the reintroduction on fitness that may result in long-term consequences.

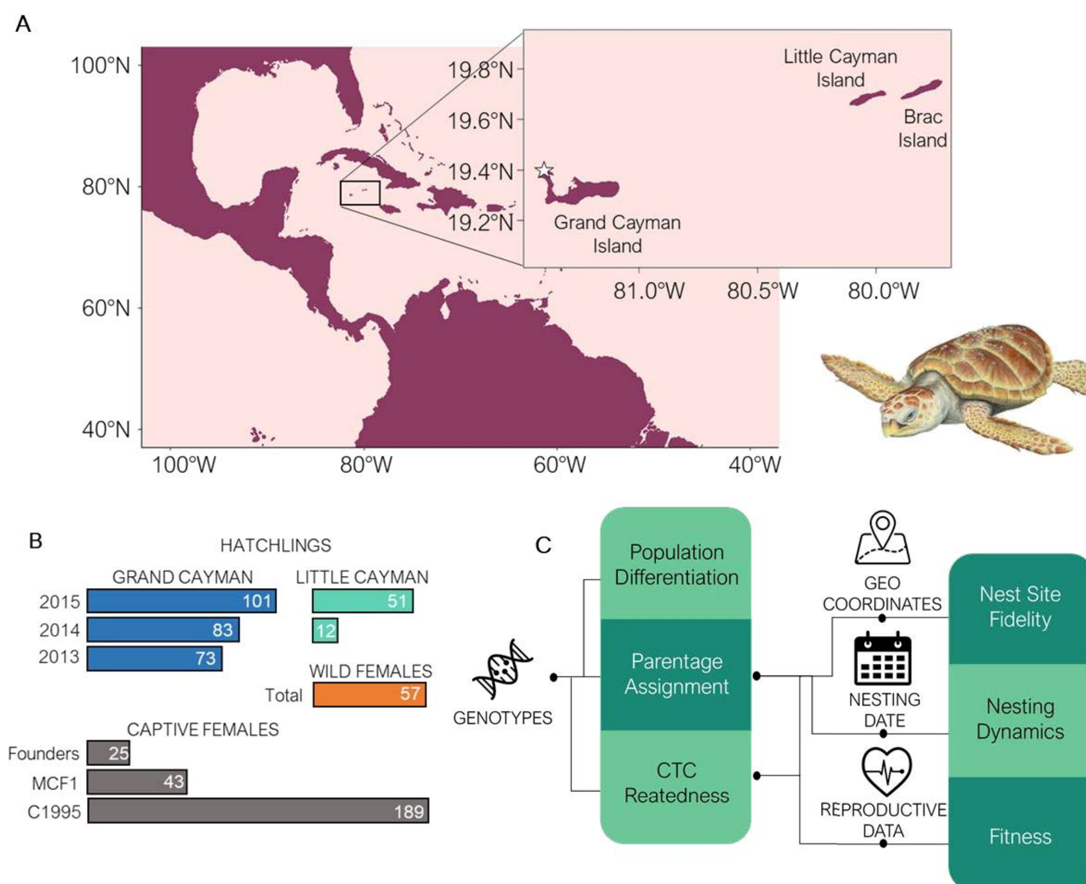


Figure 1. Study design. **A**, Location of the Cayman Islands in the Caribbean Sea, the white star shows the location of the Cayman Turtle Centre. **B**, Samples used for the study. In dark blue, number of hatchlings analysed during 2013, 2014, and 2015 from Grand Cayman Island, in light blue number of hatchlings analysed during 2014, and 2015 in Little Cayman Island. In orange wild adult females sampled during boat surveys in the waters of the Cayman Islands, and in grey Cayman Turtle Centre female breeders, including original founders of the captive population, a multi cohort of F1 breeders (MCF1) and the cohort of 1995 of F1 breeders (C1995). Wild and

captive females' genotypes are from a previous study[29]. C, Flowchart of computational analysis performed in our study. Each genotyped hatchling was collected from a different nest and combined with female genotypes to perform a parentage analysis. Genotypes from hatchlings and females were also combined with CTC breeders' genotypes to assess relatedness to the CTC. Geographic coordinates, nesting dates and reproductive data were collected for each nest and combined with parentage and relatedness analysis to assess nest-site fidelity, nesting dynamics and fitness.

Population diversification from the captive population

Here we present a study that uses genetic data from 634 green turtles, including two generations of CTC breeders[27] and two generations of wild individuals (Figure 1b, Methods). This genetic information is used in combination with individual nest information from Grand Cayman and Little Cayman beaches gathered across three consecutive nesting seasons (Figure 1c). Similar levels of observed heterozygosity were found between wild females and wild hatchlings of the two islands (Table S1). We reconstructed the female breeding population by running maternity analysis using the genotypes of hatchlings and wild females (Figure 1c). This analysis not only linked our sampled females with clutches laid (n=140) but also inferred potential unsampled mothers of the remaining clutches (n=171) (Figure 2a). By identifying mother-offspring pairs, progeny from more than 43% of nests was found to be related to the turtles in the CTC, and by adding the results of relatedness analysis between wild hatchlings and CTC breeders, the number of related progeny (r-value > 0.3070) increased to 88.1%, for an overall total of 282 CTC related hatchlings. We could therefore conclude that 79.4% of Little Cayman hatchlings and 90.3% of Grand Cayman hatchlings were related to the adults in the CTC (Figure 2b), with no significant difference between the proportion for the two islands (Chi-squared = 0.259, p-value = 0.610). These results confirm that the nesting populations of these two Islands are mainly the result of an assisted colonisation through individuals reintroduced from the captive breeding program.

Despite the high degree of relatedness of both nesting populations with the CTC, significant genetic differences were found among the three groups, especially with biparentally inherited markers (Figure 2c). This was also observed with the limited overlap of the three groups in the Discriminant Analysis of Principal Components (Figure 2d). In particular Little Cayman hatchlings showed less genetic overlap to the CTC female breeders than Grand Cayman hatchlings with both nuclear and mitochondrial markers (Figure 2c, Figure 2d). This result is consistent with the higher geographic distance from where captive individuals were released but also with the lower level of relatedness found (Figure 2b).

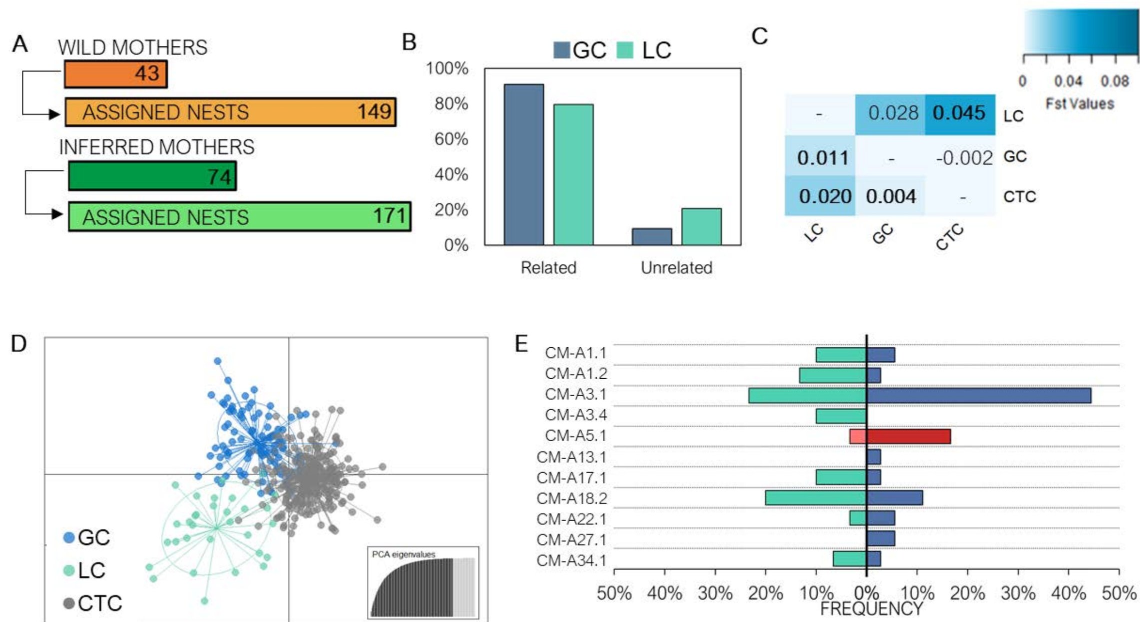


Figure 2. Population differentiation from the captive population. **A**, Number of nests assigned to wild and inferred females according to parentage analysis using COLONY. **B**, Percentage of nests in the Grand Cayman (GC) and Little Cayman (LC) found to be related or unrelated to the Cayman Turtle Farm (CTC) breeding stock as shown by Coancestry and ML-Relate. **C**, Genetic differentiation among LC, GC and CTC showed as a heatmap of pairwise FST values obtained with mitochondrial markers (above diagonal) or microsatellite markers (below diagonal). Significant p-values after FDR correction values are shown in bold. Only one nest per female per nesting season and island was considered in FST computations. **D**, Discriminant Analysis of Principal Components of GC, LC and CTC individuals. The small inset shows the cumulative eigenvalues (dark grey) of the 128 retained PCAs. **E**, Frequency of D-Loop haplotypes found in GC (dark blue) and LC (light blue) nests. Blue shaded haplotypes belong to the Northern lineage and the red shaded haplotypes to the Southern lineage as defined in a previous study[30].

This can be the result of the contribution of remnant individuals from the original Little Cayman population but could also be caused by the breeding of founders from other locations, or of CTC breeders that could not be assessed genetically. Genetic differentiation between hatchlings from Grand and Little Cayman was statistically significant at the nuclear level and can be also appreciated as shifts in haplotype frequencies (Figure 2e), meaning that they should also be considered as two different and separated rookeries. The same haplotypes, from Caribbean and South Atlantic lineages[30], were present in the CTC captive population but were found at different frequencies. The degree of differentiation found between these three populations suggests that genetic drift that results from a founder effect is a strong force able to drive genetic differentiation on a short time and geographic scale. Previous studies suggested that philopatry is one of the main drivers of the deep genetic structuring found in sea turtles, as this behaviour prevents gene

flow. Therefore, mutation and genetic drift across many generations would generate differentiation among sea turtle populations at large evolutionary scales[31]. In this study we demonstrate that genetic drift during founding processes can have also an important role in generating significant genetic structuring in sea turtles in only one generation after the foundation of new populations. Philopatry is then expected to increase this initial differentiation in future generations.

The role of nest-site fidelity in founding new populations

Philopatry limits the colonising potential of specific organisms but ensures re-utilisation of a suitable habitat, reinforcing population growth[32]. Furthermore, if a new population is established, philopatry will accelerate its growth during the following generations[33]. The CTC reintroduction program was based on the premise that the released animals would be philopatric to the new areas, as shown for other species[34], and that the individuals of the new population would maintain this successful evolutionary behaviour. We analysed the breeding dispersal (i.e. displacements between different breeding episodes[35]) on Grand Cayman individuals to assess the degree of nest-site fidelity by combining parentage analyses and nesting information. Using nest geographic coordinates, we calculated the distance between temporally consecutive nests in wild females laying more than three clutches in the same season (Figure 3a). The majority of females exhibited high nest-site fidelity within a nesting season, with 85.1% of mean distances between nests being less than 5 km (Figure 3b). We also measured the distance between the two most distant nests for the same female within a season, finding that 77.7% of observations covered less than 5 km. These results show that females of Grand Cayman have a high degree of nest-site fidelity (Figure 3c) despite coming from a reintroduction program. Nevertheless, one wild female was found nesting on both Little and Grand Cayman, covering a minimum distance of 141.7 km when moving between islands, but showing strong nest-site fidelity when nesting in Little Cayman (mean distance between nests = 484 m). Using parentage assignment analyses, a further 8 females and 13 males were inferred by the program contributing to both Little Cayman and Grand Cayman rookeries. Females nesting on the two islands could reflect an actual failure in finding the natal beach or be a consequence of external disturbance during nesting, which could affect philopatry even in different nesting seasons. Nonetheless, long distance nesting (either on the same or different islands) could also be an evolutionary strategy sea turtles developed to maintain the income of gene flow into a population and avoid collapse due to extreme philopatry. Recent research on within-season nest-site fidelity reports that long distance nesting appears to be more common than previously described[36], unravelling the possibility of an evolutionary strategy behind this behaviour. Similarly, sporadic nesting far away from natal beaches could be an opportunistic

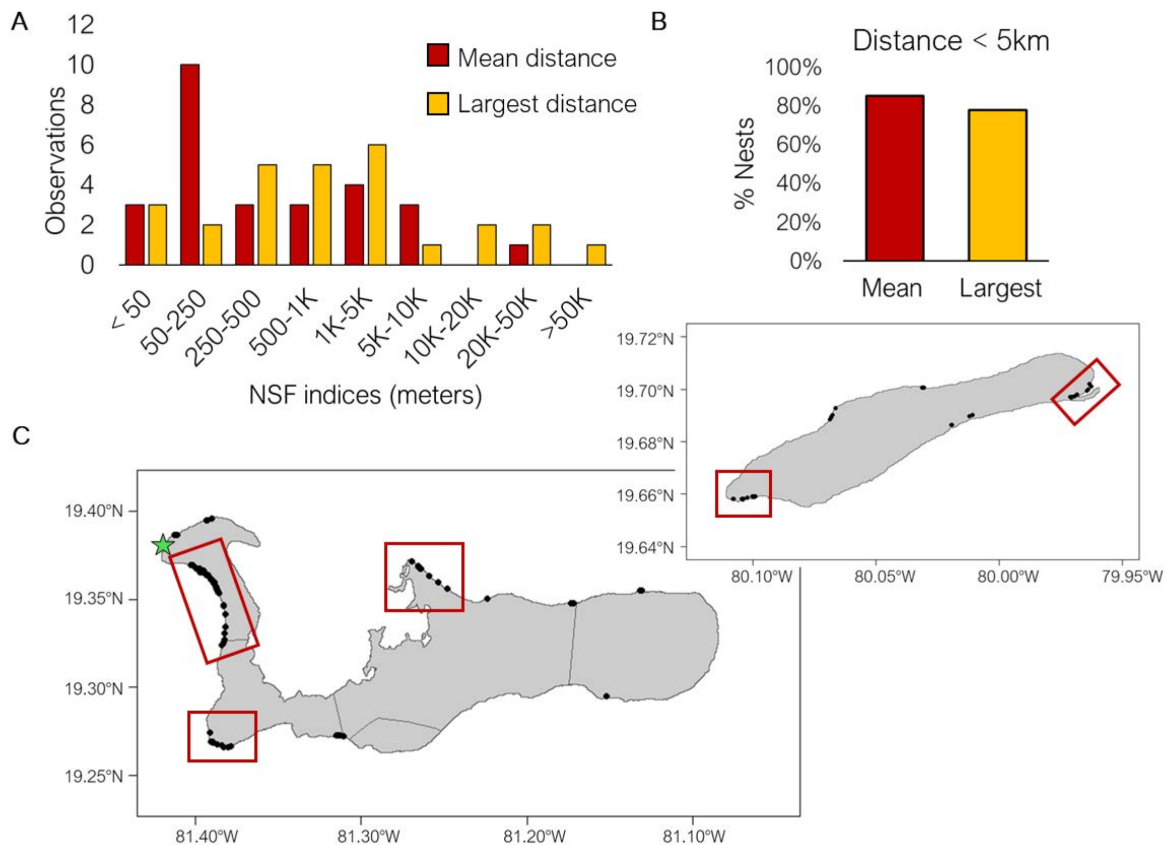


Figure 3. Nest-site fidelity. A, Nest-site fidelity (NSF) of wild females laying more than 3 clutches per season (N=27). Red shows the mean distance between consecutive nests, while yellow shows the distance between the two most distant analysed clutches laid by the same female within a nesting season. B, Percentage of females nesting within 5km, showing the high nesting fidelity of the populations recovered after the reintroduction. C, Map detailing the exact geographical locations of the 320 sampled nests, indicating in red rectangles the major nesting sites on Grand Cayman and Little Cayman Islands. The CTC is marked with a green star on Grand Cayman Island.

behaviour to escape philopatry and promote long-term species survival[33]. We did not find any significant impact of female heterozygosity (adjusted R^2 = 0.010, p-value = 0.267) or of their CTC relatedness (Kruskal-Wallis chi-squared = 0.410, p-value = 0.814) on MND, or on LND (adjusted R^2 = 0.019, p-value = 0.230, Kruskal-Wallis chi-squared = 0.739, p-value = 0.691) meaning that longer distance nesting events are not genetically determined and probably the result of stochastic processes.

Long-term effects of the reintroduction

In the last few years, the male/female proportion of sea turtle populations has become a cause of concern due to ocean warming[16,18,37,38]. Sea turtles, as with many other reptile species, exhibit Temperature-dependent Sex Determination (TSD) with a greater proportion of female offspring

produced at temperatures above the pivotal temperature (~29°C)[39,40] and a greater proportion of males produced at temperatures below this. In new Cayman wild breeding adults, we did not detect female skewed sex ratios that could indicate a potential feminisation due to temperature changes. The genetically estimated breeding population size of the Cayman Islands was of 119 females and 115 males, with a sex ratio (male/female) of 0.96 not significantly different from the 50-50 proportion expected (Chi-square = 0.008, p-value = 0.926). Similar non-significant results were obtained when each island was analysed separately (Grand Cayman sex ratio= 1.020 and Little Cayman sex ratio = 0.935). Recently, studies on several Caribbean green turtle populations indicated a higher proportion of females than males at primary sex ration as inferred using incubation temperatures[38], but our analyses show that in the Cayman Islands the sex ratio of the adult breeding population is still balanced. Nevertheless, since adult sex ratios represent sand incubation temperatures of at least 15 years ago, future wild generations should be monitored to detect any effects of global warming on sex ratio shifting over time.

We used Linear Mixed Effects Models to detect the effect of hatchling heterozygosity, adult female heterozygosity, and relatedness to the CTC on fitness variables. As described in previous studies[27,41] we found that larger females lay a significant higher number of eggs per clutch and have higher fecundity (i.e. number of eggs that developed an embryo) (Table S5). Moreover, nests with higher fecundity showed higher viability (i.e. number of eggs hatched) regardless of the CTC relatedness (Figure 4c, Table S5)). Females' and hatchlings' heterozygosity and CTC relatedness did not have any significant effect on fecundity or viability. This suggests that the individuals coming from the CTC program are not affecting the fitness of the new population. These results only refer to a first generation of wild hatchlings (hatchlings from reintroduced individuals), however, and thus population fitness analyses should be repeated in the future to monitor potential drops due to outbreeding of the different genetic lineages that conformed the initial captive population[42]. In fact, although the reproductive fitness is not affected, hatchling relatedness with the CTC, had an influence on their heterozygosity (Table S5), since hatchlings not related to the CTC had significantly higher observed heterozygosity than hatchlings not related (Figure 4d).

Wilcoxon-Mann-Whitney test comparing heterozygosity values of hatchlings related and unrelated to the CTC, with the adult wild population and the subgroups of the CTC (i.e. Founders, MCF1 and C1995)[27], showed that related hatchlings had similar low heterozygosity levels to the wild population and the founder group. On the other hand, unrelated hatchlings had similar values to the outbred individuals within the CTC breeders (MCF1 and C1995). Thus CTC breeders have

high levels of heterozygosity since both groups resulted from the mating of the founders, which included individuals from genetically differentiated populations[27]. The hatchlings of the wild populations not related to the CT could be the result of the mating of individuals coming from three different groups: i) individuals of external contribution (i.e. migrations from other populations); ii) individuals of the original wild population still living in the waters of the Cayman Islands; or iii) adult captive turtles escaped after Hurricane Michelle damaged the facilities in 2001. Unfortunately, these different hypotheses must remain untested, since genetic data from the original population and the initial CTC founder stock are not available.

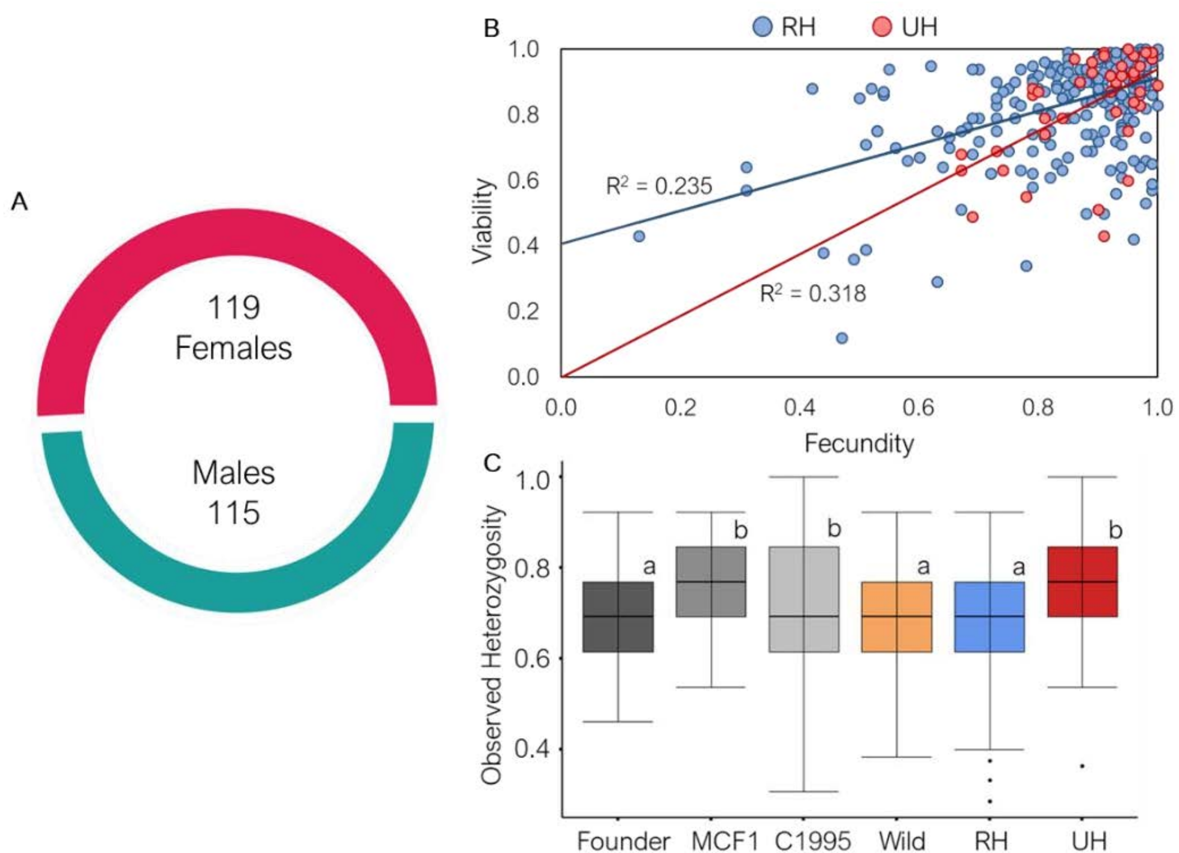


Figure 4. Analyses of biological parameters. A, Genetically estimated breeding population size of the Cayman Islands. Females count both wild individuals and females inferred by parentage analysis, while males were all inferred by parentage analysis. **B**, Considering all clutches, fecundity (i.e. number of eggs that developed an embryo) has a significant influence on viability (i.e. number of eggs hatched). This happens in clutches related (blue) and unrelated (red) to the CTC. **C**, Boxplots of observed heterozygosity values for the CTC breeding stock subgroups, the wild females and the nests related (RH) or unrelated (UH) to the CTC. Similar letters show statistically similar level of heterozygosity among groups after post hoc Wilcoxon sum rank tests. Nests related to the CTC are grouped wild females and to the founder CTC breeding stock while the nests non-related to the CTC show significant higher values, like F1 captive breeding females.

Assisted colonisation as a possible conservation measure

This study provides evidence of success of an assisted colonisation as a conservation strategy for an endangered species. Our results show that, at least in sea turtles, assisted colonisations could possibly help to mitigate the predicted loss of nesting habitats due to climate change. Consequently, assisted colonisations might soon become the most viable conservation strategy to prevent species extinction where habitat degradation undermines species short-term survival[25]. In this context, the study of the foundation of new populations using a multidisciplinary approach, is crucial to improve assisted colonisations and to tailor conservation action plans to the target species. Here, careful analyses have shown how this attempt started almost 50 years ago has been successful even though it was subject to controversy. Furthermore, we show how the study of the foundation of new populations in complex vertebrate species can provide information relevant at evolutionary scales such as the time required to establish a genetically differentiated new population or potential alterations of the fitness. Assisted colonisation has, therefore, potential in conservation of sea turtles and in future application on other complex and highly migratory species that require new habitats due to changing environmental conditions.

References

1. Díaz, S. *et al. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. (2019).
2. Driscoll, D. A. *et al.* A biodiversity-crisis hierarchy to evaluate and refine conservation indicators. *Nature Ecology and Evolution* vol. 2 775–781 (2018).
3. Henson, S. A. *et al.* Rapid emergence of climate change in environmental drivers of marine ecosystems. *Nat. Commun.* **8**, 1–9 (2017).
4. Pecl, G. T. *et al.* Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* vol. 355 (2017).
5. Kelly, M. W., Sanford, E. & Grosberg, R. K. Limited potential for adaptation to climate change in a broadly distributed marine crustacean. *Proc. R. Soc. B Biol. Sci.* **279**, 349–356 (2012).
6. Fischer, J. & Lindenmayer, D. B. An assessment of the published results of animal relocations. *Biological Conservation* vol. 96 1–11 (2000).
7. Storfer, A. Gene flow and endangered species translocations: A topic revisited. *Biol. Conserv.* **87**, 173–180 (1999).
8. IUCN. IUCN position statement on traslocation of living organisms: introductions, re-introductions and re-stocking. *22nd Meet. IUCN Counc. Gland. Switz.* (1987).
9. Sarrazin, F. & Barbault, R. Reintroduction: Challenges and lessons for basic ecology. *Trends Ecol. Evol.* **11**, 474–478 (1996).
10. Sutherland, W. J. *et al.* Standards for documenting and monitoring bird reintroduction projects. *Conserv. Lett.* **3**, 229–235 (2010).
11. Armstrong, D., Soderquist, T. & Southgate, R. Designing experimental reintroductions as experiments. *Reintroduction Biol. Aust. New Zeal. Fauna* 27–29 (1994).
12. Mayr, E. *Systematics and the Origin of Species*. (Columbia Univ. Press, New York, 1942).
13. Gavrilets, S. & Hastings, A. Founder effect speciation: A theoretical reassessment. *Am. Nat.* **147**, 466–491 (1996).
14. Kolbe, J. J., Leal, M., Schoener, T. W., Spiller, D. A. & Losos, J. B. Founder effects persist despite adaptive differentiation: A field experiment with lizards. *Science (80-.).* **335**, 1086–1089 (2012).
15. Alsos, I. G. *et al.* Long-distance plant dispersal to North Atlantic islands: colonization routes and founder effect. *AoB Plants* **7**, (2015).
16. Hawkes, L. A., Broderick, A. C., Godfrey, M. H. & Godley, B. J. Investigating the potential impacts of climate change on a marine turtle population. *Glob. Chang. Biol.* **13**, 923–932 (2007).
17. Jensen, M. P. *et al.* Environmental warming and feminization of one of the largest sea turtle populations in the world. *Curr. Biol.* **28**, 154-159.e4 (2018).

18. Jensen, M. P. *et al.* Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the World. *Curr. Biol.* 28, 154-159.e4 (2018).

19. Pike, D. A. Forecasting the viability of sea turtle eggs in a warming world. *Glob. Chang. Biol.* 20, 7–15 (2014).

20. Witt, M. J., Hawkes, L. A., Godfrey, M. H., Godley, B. J. & Broderick, A. C. Predicting the impacts of climate change on a globally distributed species: the case of the loggerhead turtle. *J. Exp. Biol.* 213, 901–911 (2010).

21. Luschi, P., Papi, F., Liew, H. C., Chan, E. H. & Bonadonna, F. Long-distance migration and homing after displacement in the green turtle (Chelonia mydas): A satellite tracking study. *J. Comp. Physiol. A Sensory, Neural, Behav. Physiol.* 178, 447–452 (1996).

22. Limpus, C. J., Fleay, A. & Baker, V. The flatback turtle, chelortia depressa, in queensland: Reproductive periodicity, philopatry and recruitment. *Wildl. Res.* 11, 579–587 (1984).

23. Nishizawa, H., Abe, O., Okuyama, J., Kobayashi, M. & Arai, N. Population genetic structure and implications for natal philopatry of nesting green turtles Chelonia mydas in the Yaeyama Islands, Japan. *Endanger. Species Res.* 14, 141–148 (2011).

24. Carreras, C. *et al.* Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (Caretta caretta). *Sci. Rep.* 8, 1435 (2018).

25. Hoegh-Guldberg, O. *et al.* ECOLOGY: Assisted Colonization and Rapid Climate Change. *Science (80-.).* 321, 345–346 (2008).

26. King, F. W. *Historical review of the decline of the green turtle and the hawksbill.* (Smithsonian Institution Press, Washington, DC, 1982).

27. Barbanti, A. *et al.* How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Mol. Ecol.* 28, 1637–1651 (2019).

28. Cayman Turtle Farm. Application to register a captive breeding operation involving Chelonia mydas on Grand Cayman, Cayman Islands. Submitted by the CITES Management Authority of the United Kingdom of Great Britain and Northern Ireland pursuant to Conferen. (2002).

29. Barbanti, A. *et al.* How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Mol. Ecol.* 28, 1637–1651 (2019).

30. Naro-maciel, E. *et al.* Journal of Experimental Marine Biology and Ecology From refugia to rookeries : Phylogeography of Atlantic green turtles. 461, 306–316 (2014).

31. Chakraborty, R. & Nei, M. Genetic differentiation of quantitative characters between populations or species: I. Mutation and random genetic drift. *Genet. Res.* 39, 303–314 (1982).

32. LePage, C. & Cury, P. How spatial heterogeneity influences population dynamics: Simulations in SeaLab. *Adapt. Behav.* 4, 255–281 (1996).

33. Carreras, C. *et al.* Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (Caretta caretta). *Sci. Rep.* 8, 1435 (2018).

34. Fontaine, C. & Shaver, D. Head-starting the Kemp's ridley sea turtle, Lepidochelys kempii, at the NMFS Galveston Laboratory, 1978-1992: A review. *Chelonian Conserv. Biol.* 4, 838–845 (2005).

35. Greenwood, P. J. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* **28**, 1140–1162 (1980).

36. Shamblin, B. M. *et al.* Improved female abundance and reproductive parameter estimates through subpopulation-scale genetic capture-recapture of loggerhead turtles. *Mar. Biol.* **164**, 1–14 (2017).

37. Candan, O. & Kolankaya, D. Sex Ratio of Green Turtle (Chelonia mydas) Hatchlings at Sugözü, Turkey: Higher Accuracy with Pivotal Incubation Duration . *Chelonian Conserv. Biol.* **15**, 102–108 (2016).

38. Laloë, J. O., Esteban, N., Berkel, J. & Hays, G. C. Sand temperatures for nesting sea turtles in the Caribbean: Implications for hatchling sex ratios in the face of climate change. *J. Exp. Mar. Bio. Ecol.* **474**, 92–99 (2016).

39. Janzen, F. J. & Paukstis, G. L. Environmental sex determination in reptiles: Ecology, evolution, and experimental design. *Q. Rev. Biol.* **66**, 149–179 (1991).

40. Mrosovsky, N. & Pieau, C. Transitional range of temperature, pivotal temperatures and thermosensitive stages for sex determination in reptiles. *Amphib. Reptil.* **12**, 169–179 (1991).

41. Broderick, A. C., Glen, F., Godley, B. J. & Hays, G. C. Variation in reproductive output of marine turtles. *J. Exp. Mar. Bio. Ecol.* **288**, 95–109 (2003).

42. Frankham, R. *et al.* Predicting the Probability of Outbreeding Depression. *Conserv. Biol.* **25**, 465–475 (2011).

METHODS

Sampling and data collection

Sampling and data collection were performed during the program of beach monitoring carried out by the Department of Environment of the Cayman Islands Government to assess the populations of marine turtles on the islands of Little Cayman and Grand Cayman, in the Caribbean. Data and samples were collected from 320 nests laid in Grand Cayman and Little Cayman (Figure 1a) during 2013, 2014 and 2015 nesting seasons, from June to September (Figure 1b). This sampling effort corresponded to 58% of the 552 nests reported during these nesting seasons and locations. For each nest, we recorded the nesting date and its exact location with GPS, as well as the following reproductive parameters (Supplementary datafile): number of eggs, egg fecundity and egg viability, following standard procedures[1]. If the female was present at the nest discovery, we recorded its identification by using Passive Integrated Transponders (PIT tags) and we measured its Curved Carapace Length (CCL). All known female-offspring pairs were used as control for genetic parentage identification. Samples were taken from the rear flipper of dead hatchlings found in the nest after excavation and only one hatchling per nest was sampled. Samples were obtained with a scalpel blade and stored in 100% ethanol. For the statistical analysis, in addition to our hatchling samples, we used genetic data from a previous study[2] that included genotypes from 57 wild green turtle females nesting on Grand Cayman in 2013 and 2014, as well as 257 females belonging to the Cayman Turtle Centre (CTC) breeding stock. This CTC breeding stock included original founders of the captive population (n=25), a first generation (F1) cohort of breeders born in 1995 corresponding to a single cohort breeders replacement strategy (C1995, n= 189) and a multicohort group of F1 breeders corresponding to a continuous breeders replacement strategy (MCF1, n= 43) (Figure 1b).

Laboratory analysis and genotyping

The DNA of all samples was extracted using the QIAamp Blood and Tissue Kit (Qiagen®) or using E.Z.N.A.® Tissue DNA kit (OMEGA Bio-tek), following the manufacturer protocols. All samples were genotyped at 13 microsatellite loci (Supplementary datafile), originally designed for different species of sea turtles that amplify and are polymorphic in green turtles[3] using protocols previously described[2]. A selection of individuals from 67 independent nests, as determined by the parentage analyses described below, was sequenced for 800 bp of the D-Loop mitochondrial DNA using published protocols[2]. Haplotypes were assessed (Supplementary datafile) using Bioedit[4] by comparison to the haplotype database maintained by the Archie Carr Center for Sea Turtle research (https://accstr.ufl.edu/).

We used GENALEX[5] to compute within group observed and expected heterozygosity (Table S1).

Parentage analysis

We performed maternity analysis using COLONY v2 software[6], which performs parentage assignment and reconstructs genotypes for unsampled parents, allowing the identification of family groups with sampled and unsampled females and males. We set the parameters to long run, high precision and error rate = 0.0001. All hatchlings were included in the analysis as offspring and the genotypes of 57 wild adult females from a previous study[7] were included as mothers. We checked the accuracy of COLONY by comparison with 30 female-offspring known pairs recorded during field observations (Table S2). We detected 25 concordant matches between the field observation and the genetic assignment. In one case, the real and assigned mother yielded equal probabilities of assignation, indicating that probably both females were close relatives. The remaining four nests were assumed to be the result of a tagging mistake during data collection as the genotypes of the hatchling and the tagged mother were not compatible (i.e. alleles of the mother not found in the hatchling for several loci). In this case we considered the inferred mother for further analyses In addition to providing parentage and sibship relationships, the output of COLONY was used to perform a genetic census based on the number of males and females identified or inferred by the program as parents of the analysed hatchlings.

In order to understand the impact of the CTC reintroduction program on the two populations we computed Queller and Goodnight relatedness estimator[8] using the program Coancestry[9] between the 320 hatchlings collected on the two islands and the 257 CTC individuals genotyped in a previous study[2]. A pair of individuals was considered unrelated if its lower bound of 95% confidence interval was lower than 0.0001 and its r value was less than or equal to 0.3069[10]. ML-Relate[11] was also used to estimate the relationship between individuals using a log-likelihood approach. We only accepted pairs of individuals found as related by both programs. Hatchlings of known relationship with the CTC (because having a genotyped wild mother related to the CTC[7]) were included in this analysis as control to assess the reliability of the programs. A total of 131 hatchlings had assigned a wild mother previously found to be related to the CTC. Of these, 120 hatchlings were confirmed as related to the CTC by the two programs, and the remaining 11 hatchlings were scored as half-siblings by ML-Relate. We calculated the proportion of hatchlings related to the CTC by both programs for Little Cayman and Grand Cayman separately and we tested for significant differences among islands with a Chi-squared test with Yates' continuity correction with R [12].

Genetic differentiation

To identify early signs of genetic structuring, we tested the level of genetic differentiation between the hatchlings sampled in Grand Cayman and Little Cayman and the CTC adult females using both nuclear and mitochondrial markers. Using microsatellites we calculated pairwise F_{ST} and statistical significance through 999 permutations using GENALEX[5]. Then, we performed a Discriminant Analysis of Principal Components using the R package adegenet[13], retaining 128 PCAs. As several nests were found to be laid by the same female (see Results), we used only one random nest per female laid during the same nesting season to avoid pseudoreplication. The mitochondrial haplotype frequencies extracted from 67 independent nests (nests laid by different females as indicated by the parentage analysis) were also used to calculate pairwise F_{ST} values between groups of samples, and significance was assessed through an exact test using Arlequin[14]. A Benjamini–Yekutieli (B-Y) FDR correction[15] was applied to p-values in all multiple comparisons, with an initial threshold before correction of $p<0.05$.

Nesting fidelity and reproductive fitness

With the results of the parentage analysis, we were able to link all the data collected in the field for each nest with the identified mother. We added to the dataset information on 16 female-offspring pairs recorded during night patrols but not genotyped (Table S3). With this data we evaluated female nesting fidelity by considering the geographic locations of all the nests laid per female. We analysed intra-seasonal nest-site fidelity (NSF) of wild females using geographic coordinates of their nests within the same season. We only considered for the analysis nesting females that laid three or more nests in the same season (N=27). Distances between nests were obtained by measuring the coastline between the geographic coordinates of consecutive nests using Daft Logic (https://www.daftlogic.com/projects-google-maps-distance-calculator.htm#), an online tool to calculate distances with Google satellite maps. We used two different measures of NSF: the mean distance between consecutive nests and the distance between the two most distant nests laid by the same female within a season.

In order to assess any potential impact of the reintroduction program or inbreeding on NSF we performed a linear regression in R[12] between wild female heterozygosity and the mean and largest distance between their nests. We also performed a Kruskal-Wallis test to detect any impact of the female relatedness to the CTC on mean and largest distance. We plotted bar plots of MND and LND distributions and the map of the main nesting sites of Little Cayman Island and Grand Cayman Island using R package ggplot2 (Wickham, 2016).

The parentage analysis performed by COLONY also provided the number of males and females that produced the offspring sampled on both Cayman Islands along the different nesting seasons. This data allowed us to estimate a minimum breeding population across seasons and the degree of mobility of nesting events per female and male between islands. We added to the analysis two females recorded during night patrols whose nests were not genotyped (Table S3). We calculated the sex ratio of the whole population and of each island separately, and performed a Chi squared test with Yates' continuity correction[15] with R[12] to evaluate significant difference in the number of males and females that would indicate skewed sex ratios.

We carried out Linear Mixed-Effects Models to detect possible impacts on nest fitness caused by the reintroduction program as measured by female heterozygosity and relatedness with the CTC. We performed six different models using as response variables clutch size, fecundity, viability and nest heterozygosity (Table S5). Mother ID and the year of nesting season were set as random factors in all the models. We only considered data belonging to wild sampled females due to the lack of some parameters of the inferred females (Curved Carapace Length, Mother Heterozygosity and Mother relatedness to the farm). We considered the nesting date as the quartile of the nesting season in which the nest was laid. As nesting seasons can shift slightly in different year, we first calculated the duration of the nesting season as the period between the first and the last recorded nest of the season, and then we the divided this period into quartiles to know in which quartile of the nesting season was laid a particular nest was laid. The models were performed using the R package 'lme4' [16] and significance of categorical values were assessed with the package 'car'[17]. Samples from 2014 nesting season of Little Cayman were excluded due to non-available field data. Finally, we evaluated the levels of observed heterozygosity of hatchlings related and unrelated to the CTC, and we compared them with observed heterozygosity values of wild sampled females and CTC subgroups[2] using a Wilcoxon sum rank test as implemented in R[12].

Methods references

1. Broderick, A. C., Glen, F., Godley, B. J. & Hays, G. C. Variation in reproductive output of marine turtles. *J. Exp. Mar. Bio. Ecol.* **288,** 95–109 (2003).
2. Barbanti, A. *et al.* How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Mol. Ecol.* **28,** 1637–1651 (2019).
3. Wright, L. I. *et al.* Reconstruction of paternal genotypes over multiple breeding seasons reveals male green turtles do not breed annually. *Mol. Ecol.* **21,** 3625–3635 (2012).
4. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98 (1999).
5. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28,** 2537–2539 (2012).
6. Jones, O. R. & Wang, J. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* **10,** 551–555 (2010).
7. Barbanti, A. *et al.* How many came home? Evaluating ex situ conservation of green turtles in the Cayman Islands. *Mol. Ecol.* **28,** 1637–1651 (2019).
8. Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution (N. Y).* **43,** 258–275 (1989).
9. Wang, J. Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* **11,** 141–145 (2011).
10. Alderson, G., Gibbs, H. & Sealy, S. Parentage and kinship studies in an obligate brood parasitic bird, the brown-headed cowbird (Molothrus ater), using microsatellite DNA markers. *J. Hered.* **90,** 182–190 (1999).
11. Kalinowski, S. T., Wagner, A. P. & Taper, M. L. ML-RELATE: A computer program for maximum likelihood estimation of relatedness and relationship. *Mol. Ecol. Notes* **6,** 576–579 (2006).
12. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *R Core Team (2019). — European Environment Agency* (2019). at <http://www.r-project.org/index.html>
13. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24,** 1403–1405 (2008).
14. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–567 (2010).
15. White, T., van der Ende, J. & Nichols, T. E. Beyond Bonferroni revisited: concerns over inflated false positive research findings in the fields of conservation genetics, biology, and medicine. *Conserv. Genet.* **20,** 927–937 (2019).
16. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).
17. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, 2019). at <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

Acknowledgements

This work was supported by the project CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER), by the Darwin Plus Project DPLUS019 "Socio-economic aspects of turtle conservation in the Cayman Islands" funded through the Department for Environment, Food and Rural Affairs (Defra) in the UK and by the projects 9013 and 10084 of the Fundació Bosch I Gimpera-UB. C.C. and M.P. are part of the research group SGR2017-1120 of the Generalitat de Catalunya. A.B. was supported by grant 2017 FI_B 00997 of the Generalitat de Catalunya-AGAUR.

SUPPLEMENTARY INFORMATION

Table S1. Genetic diversity values for the studied group of samples (Figure 1b). Total number of samples (N), expected heterozygosity (He) and observed heterozygosity (Ho). The letter "a" marks groups genetically similar to Little Cayman and "b" to Grand Cayman Island as assessed with Wilcoxon-Mann-Whitney test on Ho.

Samples	N	He	Ho
Wild Females	57	0.693	0.664ab
CTC	257	0.712	0.721
*Founders	25	0.717	0.681ab
*C1995	189	0.702	0.72a
*MCF1	43	0.719	0.751
Little C.	63	0.691	0.679a
Grand C.	257	0.717	0.699b

Table S2. Mother-offspring pairs recorded during night patrols while nesting and used as control to test the efficiency of COLONY parentage assignations. Columns show the year of nesting event, ID of the tagged mother during night patrols, ID of the nest where the hatchlings where sampled, COLONY results and the diagnosis of the program failure based on our interpretation. Equal probability: the real and assigned mother yielded equal probabilities of assignation. Tag mistake: assumed to be the result of a tagging mistake during data collection.

Year	Tagged	Nest	COLONY Match	Diagnostic
2014	G025	307	Yes	
2014	G025	324	Yes	
2014	G025	382	Yes	
2014	G026	224	#32	Tag mistake
2014	G027	368	Yes	
2014	G031	889	Yes	
2014	G033	385	Yes	
2014	G034	391	Yes	
2014	G038	424	Yes	
2014	G043	487	Yes	
2014	G044	494	Yes	
2015	G001	321	Yes	
2015	G008	179	#21	Tag mistake
2015	G048	281	Yes	
2015	G048	324	Yes	
2015	G048	428	Yes	
2015	G049	421	Yes	
2015	G050	277	Yes	
2015	G051	292	Yes	
2015	G052	438	#55	Tag mistake
2015	G052	293	Yes	
2015	G053	432	#36	Tag mistake
2015	G053	380	Yes	
2015	G053	369	Yes	
2015	G054	375	Yes	
2015	G055	403	Yes	
2015	G055	427	Yes	
2015	G056	429	Yes	
2015	G056	409	Yes	
2015	G058	431	G049	Equal probability

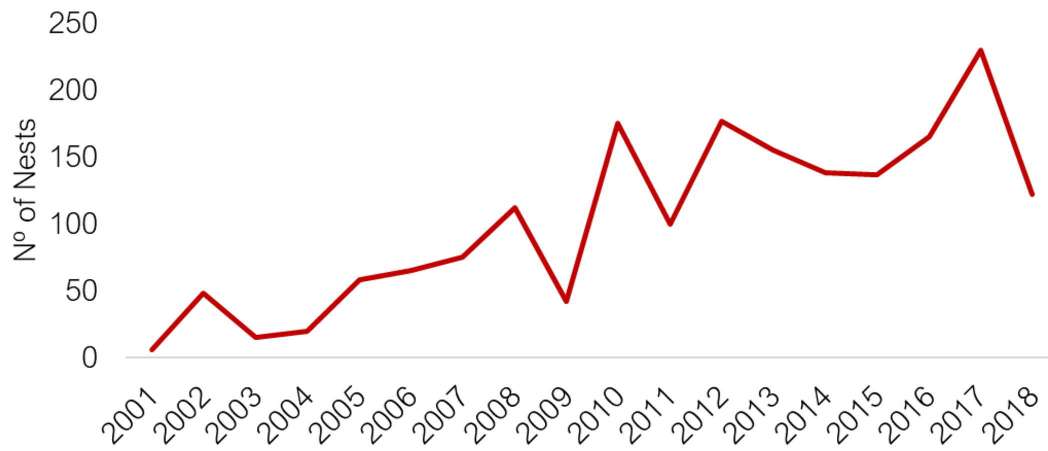
Table S3. Female-nest pairs recorded during night patrols while nesting. Some females laid more than one nest therefore the ID is repeated. No hatchling was genotyped from these nests. Na indicates nests from which the exact geographical position was not recorded.

Year	Female ID	Nest ID	Latitude	Longitude
2013	G001	80	19.366360	-81.396160
2013	G001	112	19.366850	-81.396910
2013	G002	81	19.366700	-81.396600
2013	G002	116	na	na
2013	G003	85	19.364080	-81.393240
2013	G003	117	19.367140	-81.397350
2013	G005	125	19.366440	-81.396270
2013	G005	327	19.363430	-81.392370
2013	G006	115	19.369480	-81.402030
2013	G008	165	19.356590	-81.387120
2013	G008	242	19.355920	-81.386870
2013	G014	241	19.386630	-81.411200
2013	G016	267	19.354780	-81.386420
2013	G020	329	19.363590	-81.392600
2013	G022	396	19.386270	-81.412280
2014	G027	296	na	na

Table S4. Details of the six Linear Mixed-Effects Models performed on the nests that were assigned to wild sampled females according to parentage analysis (n = 149). For each model we detail the Response variable and the Random and Fixed effects considered. P-values in bold are statistically significant after FDR correction.

Model	Response	Random Effects	Fixed Effects	p-value
1	*Viability*	Mother ID	Mother Heterozygosity	0.870
		Year	Nest Heterozygosity	0.493
			Fecundity	**0.000**
			CCL	0.146
2	*Viability*	Mother ID	Mother – Farm relatedness	0.920
		Year	Nest – Farm relatedness	0.133
			Nesting date (Quartile)	0.103
3	*Nest Heterozygosity*	Mother ID	Mother – Farm relatedness	0.899
		Year	Nest – Farm relatedness	**0.014**
			Mother – Farm rel. * Mother Heteroz.	0.781
4	*Fecundity*	Mother ID	Mother Heterozygosity	0.393
		Year	Nest Heterozygosity	0.765
			CCL	**0.045**
			Clutch Size	0.741
5	*Fecundity*	Mother ID	Mother – Farm relatedness	0.739
		Year	Nest – Farm relatedness	0.281
			Nesting date (Quartile)	0.153
6	*Clutch size*	Mother ID	Mother Heterozygosity	0.857
		Year	Mother – Farm relatedness	0.834
			Nesting date (Quartile)	0.579
			CCL	**0.001**

Figure S1. Number of nests laid per nesting season in the Cayman Islands in the past 20 years. Data provided by the Department of Environment of the Cayman Islands Government.



CHAPTER 3

Received: 30 October 2019 | Revised: 4 February 2020 | Accepted: 10 February 2020
DOI: 10.1111/1755-0998.13144

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species

Anna Barbanti[1] | Hector Torrado[1,2] | Enrique Macpherson[2] | Luca Bargelloni[3] | Rafaella Franch[3] | Carlos Carreras[1] | Marta Pascual[1]

[1]Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Barcelona, Spain

[2]Center for Advanced Studies of Blanes (CEAB-CSIC), Blanes, Girona, Spain

[3]Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Italy

Correspondence
Carlos Carreras, Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Barcelona, Spain.
Email: carreras@ub.edu

Funding information
Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER), Grant/Award Number: CTM2017-88080 ; Generalitat de Catalunya, Grant/Award Number: SGR2017-1120 and SGR2017-378 ; Agència de Gestió d'Ajuts Universitaris i de Recerca, Grant/Award Number: FI_B 00997; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: FPU15/02390

Abstract

High-throughput sequencing has revolutionized population and conservation genetics. RAD sequencing methods, such as 2b-RAD, can be used on species lacking a reference genome. However, transferring protocols across taxa can potentially lead to poor results. We tested two different IIB enzymes (AlfI and CspCI) on two species with different genome sizes (the loggerhead turtle *Caretta caretta* and the sharpsnout seabream *Diplodus puntazzo*) to build a set of guidelines to improve 2b-RAD protocols on non-model organisms while optimising costs. Good results were obtained even with degraded samples, showing the value of 2b-RAD in studies with poor DNA quality. However, library quality was found to be a critical parameter on the number of reads and loci obtained for genotyping. Resampling analyses with different number of reads per individual showed a trade-off between number of loci and number of reads per sample. The resulting accumulation curves can be used as a tool to calculate the number of sequences per individual needed to reach a mean depth ≥20 reads to acquire good genotyping results. Finally, we demonstrated that selective-base ligation does not affect genomic differentiation between individuals, indicating that this technique can be used in species with large genome sizes to adjust the number of loci to the study scope, to reduce sequencing costs and to maintain suitable sequencing depth for a reliable genotyping without compromising the results. Here, we provide a set of guidelines to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping decision-making for a reliable and cost-effective genotyping.

KEYWORDS
Caretta caretta, conservation genomics, *Diplodus puntazzo*, genotyping-by-sequencing, high-throughput sequencing, sequencing simulations

Anna Barbanti and Hector Torrado contributed equally to the manuscript.

Carlos Carreras and Marta Pascual authors are considered senior authors.

 © 2020 John Wiley & Sons Ltd

1 | INTRODUCTION

High-throughput sequencing technologies have revolutionized the fields of population and conservation genetics during the last 10 years by providing easy access to genomic data from virtually any taxonomic group (Andrews & Luikart, 2014; Bellin et al., 2009; Davey & Blaxter, 2011). Many studies have explored the potential of genomic analysis to address a variety of questions, such as population structuring (Girault, Blouin, Vergnaud, & Derzelle, 2014), inbreeding depression (Hoffman et al., 2014), local adaptation (Savolainen, Lascoux, & Merilä, 2013) or hybridization (Hohenlohe, Amish, Catchen, Allendorf, & Luikart, 2011). Restriction site associated techniques (RAD) are based on massive sequencing after enzymatically reducing the fraction of the genome being analysed and can identify and score thousands of genetic markers, randomly distributed across the genome in many individuals simultaneously (Davey & Blaxter, 2011; Pecoraro et al., 2016). The advantage of these methodologies is that they can be carried out with no or limited previous sequence knowledge, since RAD tags can be analysed using pipelines for de novo loci identification if a reference genome is not available (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Davey & Blaxter, 2011; Hapke & Thiele, 2016; Lu, Glaubitz, Harriman, Casstevens, & Elshire, 2012). These methods allow parallel and multiplexed sample sequencing of tag libraries, with a rapid and very cost-effective procedures resulting in high genome coverage (Baird et al., 2008; Pecoraro et al., 2016). The RAD marker approach has the flexibility to assay different number of markers depending on the restriction enzyme of choice (Baird et al., 2008).

Many studies focusing on population structure in non-model organisms have implemented different RAD technologies, such as RADseq (Lim et al., 2017; Xu et al., 2014), ddRAD (Lavretsky, DaCosta, Sorenson, McCracken, & Peters, 2019; Portnoy et al., 2015), GBS (Carreras et al., 2017; Hess et al., 2015), and 2b-RAD (Boscari et al., 2019; Galaska, Sands, Santos, Mahon, & Halanych, 2017). By shifting the realms of genomics from laboratory-based studies of model species towards studies of natural populations of ecologically well-characterized organisms, researchers can now start to address important ecological and evolutionary questions on a scale and precision that, only a few years ago, was unrealistic (Ekblom & Galindo, 2011). As for all genotyping-by-sequencing methodologies, the mean number of reads per locus (mean depth of coverage) is crucial to consider reliable the quality of markers and their genotypes (Sims, Sudbery, Ilott, Heger, & Ponting, 2014). Some recent population studies prioritised the number of sequenced individuals over depth of coverage or used improved bioinformatics pipelines to extract information from low coverage data (Buerkle & Gompert, 2013; Maruki & Lynch, 2014). However, when depth is generally low, statistical uncertainty of individual sequence data is high and calling of genotypes is difficult (Maruki & Lynch, 2017). Although probabilistic genotyping methods are thought to overcome shortcomings of low-depth sequencing data, they may behave unpredictably when compared to high-depth data (Hendricks et al., 2018). Thus, any analysis involving individual genotypes is going to be limited by coverage (Chow, Anderson, & Shedlock, 2019). For this reason, RAD sequencing techniques and laboratory protocols should be adjusted to target enough sequencing depth to obtain reliable genotypes while optimising sequencing costs.

2b-RAD is a RAD methodology that uses IIB restriction endonucleases, which cleave genomic DNA upstream and downstream of the target sites producing 32–34 bp fragments (Wang, Meyer, McKay, & Matz, 2012). This method is simple and provides a cost-effective alternative to existing reduced representation genotyping methods, allowing its use in routine experimental laboratory (Baird et al., 2008; Luo et al., 2017; Wang et al., 2012). One of the most interesting features of 2b-RAD is that the number of loci/marker density can be adjusted by using selective adaptors (Wang et al., 2012) to reduce the number of expected markers and increase the coverage per locus for a given sequencing effort. This RAD sequencing technique has been used to identify candidate genes associated with specific traits (Luo et al., 2017), to construct ultra-high density genetic maps (Fu, Liu, Yu, & Tong, 2016), to identify genomic regions under selection in population genetic studies (Pecoraro et al., 2016), and to perform genomic prediction for relevant traits in agricultural species (Palaiokostas, Ferraresso, Franch, Houston, & Bargelloni, 2016). It has also been extended to microbial ecology (Pauletto et al., 2016).

In this paper, we provide a protocol for laboratory and bioinformatic analyses to optimise studies using 2b-RAD sequencing on different non-model organisms. We focused our study on the sharpsnout seabream *Diplodus puntazzo* Walbaum, 1792 and the loggerhead turtle *Caretta caretta* Linnaeus, 1758 characterized by very different genome sizes. This study aims to unveil key elements to adapt library building of non-model organisms to best profit from this genomic method. Specifically, we focused our analyses on five main objectives. (a) Assess the effect of initial DNA quality and concentration on sequencing results. (b) Evaluate the performance of different IIB enzymes (i.e., AlfI and CspCI) on genomic library construction in the two species. (c) Calculate the optimum number of raw sequences needed per each combination of species and enzyme in order to achieve the maximum number of loci with an optimum depth per locus for a correct genotyping. (d) Assess if selective base ligation protocols have an impact on genetic differentiation among individuals. (e) Set guidelines for new population genomic studies on non-model organisms. Our study provides useful information for future studies on non-model species with different genome sizes, helping decision-making to obtain a reliable and cost-effective genotyping.

2 | MATERIALS AND METHODS

2.1 | Samples

We analysed two species with approximately three-fold different genome sizes. We consider the sharpsnout seabream (*D. puntazzo*) genome size to be similar to that of *Diplodus anularis* (0.9 Gb), its

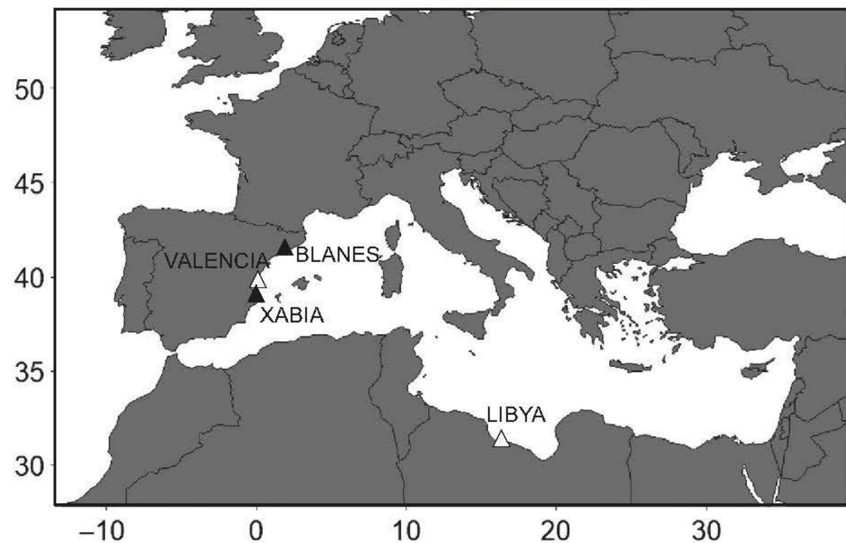


FIGURE 1 Sampling sites. White triangles show sampling sites for *Caretta caretta*, Libya is a nesting ground while Valencia is a foraging ground. Black triangles show sampling sites for *Diplodus puntazzo*

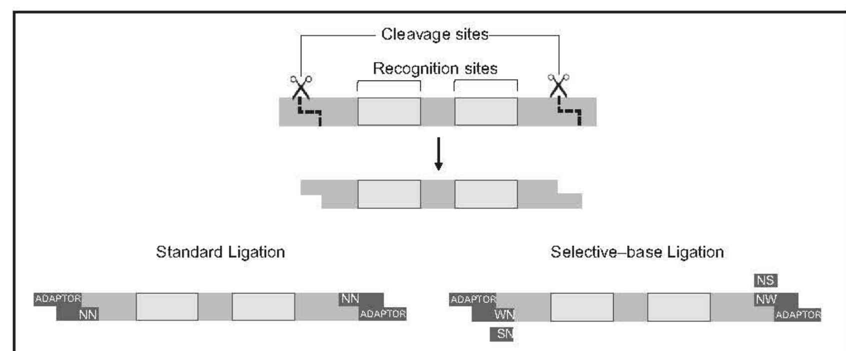


FIGURE 2 Selective-base ligation. In 2b-RAD protocol, after IIB enzyme digestion, specific fragments can be selected to reduce the density of markers to be amplified by designing customised adaptors with one fully degenerated base (N) and one partially degenerated base (S = G and C bases, W = A and T bases)

closest relative's sequenced genome (www.genomesize.com). The loggerhead turtle (*C. caretta*) genome size was considered to be similar to the genome of *Chelonia mydas* (Wang et al., 2013), which measures 2.24 Gb.

Juveniles of *D. puntazzo* were collected in Blanes (N = 12) and Xabia (N = 12) (Spain) during recruitment using hand nets (Figure 1). Samples of *C. caretta* were taken from bycaught juveniles at the foraging ground off Valencia (Spain) (N = 9; Figure 1) and from dead hatchlings at the nesting ground west of Sirte (Lybia) (N = 14; Clusa et al., 2018). We also added a sample collected from a live female turtle nesting in Pulpí (Spain) as positive control (Carreras et al., 2018). Consequently, our study included 24 samples per species. All samples were stored in 96% ethanol.

2.2 | DNA extraction and library construction

Genomic DNA was extracted using Qiagen Qiamp blood and tissue kit following the manufacturer's protocol. DNA concentration was measured with Nanodrop or Qubit, and DNA degradation assessed in 1% agarose gels. This information was recorded to be used in further statistical analysis. We coded the level of degradation as high if the DNA was mostly located at the bottom of the run in the agarose gel or the smear intensity increased in direction top-to-bottom, and as low if the DNA was mostly located at the top of the gel or the smear intensity faded in direction top-to-bottom. When possible we included samples that presented degraded DNA or low DNA concentration to test 2b-RAD efficiency for population genomics, as DNA degradation is a common issue when sampling non-model organisms (e.g., marine turtle studies targeting stranded individuals or dead embryos found after excavation of nests). A total of 24 individual libraries were constructed with each enzyme per species. Individual libraries were prepared adjusting the protocol from Wang et al. (2012). In brief, the construction of 2b-RAD libraries consisted of four main steps (for detailed protocol, see Annex S1). (a) Genomic DNA was digested by a IIB restriction enzyme providing short (32–34 bp) sequences. Each individual sample was digested with *AlfI* and *CspcI* enzymes separately. (b) During ligation, adaptors were attached to the sticky ends of the digested sequences. This step is crucial in the library preparation process because at this point, adaptors can be customised to attach to any sticky end or to attach only to sticky ends with specific sequences, based on the last two bases of the adaptor. For this study we used degenerated bases (5′-NN-3′) for our adaptors (Figure 2). (c) In the amplification step, barcodes and Illumina primers were attached to the adaptors and sequences were amplified by PCR. At the end of this step the resulting fragment is expected to measure

~165 bp. Library products were run through a 1.8% agarose gel to check amplification success. The library DNA quality of each sample was coded as 'good' when the band of the agarose gel was bright or 'bad' when it was faint (Appendix S1: Figure S1). (d) Purification was performed using magnetic beads to remove primers and sequences longer and shorter than 165 bp. At the end of this step, 2b-RAD libraries were ready to be sequenced. The DNA concentration of purified libraries was quantified using a Real Time PCR. The 48 libraries of each species were pooled for SR50 single read sequencing (one species per lane) with a HiSeq 4000 Illumina at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center.

2.3 | Genotyping

Sequences were processed using customized scripts (Annex S2). First, raw sequences were trimmed to eliminate ligation adaptors and then cut down to the same length (i.e., 32 bp for CspCI and 34 bp for AlfI). Processed sequences were used for genotyping using the STACKS versus 1.47 pipeline (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen et al., 2013). To construct a loci catalogue we used Stacks function *denovo_map.pl* setting the following parameters: a minimum depth of three reads to consider a stack within an individual (m = 3), up to three mismatches allowed between stacks (putative alleles) to merge them into a putative locus within an individual (*M* = 3), and two mismatches allowed between stacks (putative loci) during construction of the catalogue (*n* = 2). Individual genotypes were outputted as haplotype loci VCF files. We used five main filters to process loci found in our samples. We removed individual genotypes based on less than five reads, loci present in <70% of individuals, loci with outlier values of mean depth across all individuals (those above the upper whisker of the R boxplot, corresponding to 1.5 times the interquartile range from the data), loci with a major allele frequency higher than 99% and loci out of Hardy-Weinberg equilibrium (HWE) in at least one of the populations. In the case of *C. caretta* HWE was considered only for Libya, since Valencia is a feeding aggregation of individuals from different populations, and thus deviations from HWE are expected (Clusa et al., 2014). Filtering was performed with VCFTOOLS vs 1.12 (Danecek et al., 2011), with the exception of loci with a major allele frequency higher than 99%, which were identified by the function isPoly from the package ADEGENET (Jombart, 2008) and the assessment of HWE, computed with the function hw.test from the package pegas (Paradis, 2010) in R (R Core Team, 2018). We performed linear regression and Wilcoxon-Mann-Whitney test in R to assess whether initial and library DNA concentrations, initial DNA degradation and library quality influenced the number of total sequences and the final number of loci of each sample.

2.4 | Resampling analysis

We used bioinformatic simulations for each species and enzyme to obtain several sample sets, each one presenting a different number of reads per individual. We used a customised script to create new sample sets with different number of reads per sample by performing a random selection with replacement of the real data up to different target numbers of raw reads per sample (Annex S2). We performed 10 iterations for each target number. Target numbers varied for each species to accommodate the data points to the expected accumulation curve results for the different genome sizes. For *D. puntazzo* we simulated 0.5, 1, 2, 4, 8 and 10 million raw reads per sample for CspCI and AlfI enzymes. For *C. caretta* we simulated 4, 8, 12, 16 and 20 million raw reads per sample for each enzyme. Each resampled set underwent the same process of loci identification and filtering as explained above with the exception of the filter removing loci out of Hardy-Weinberg equilibrium. This filter was not applied because loci genotyping could be biased in the low depth data sets, artificially creating loci out of H-W equilibrium, since resampling was done with replacement. For this reason, this technique should not be used to artificially increase locus depth for a proper genotyping, as these genotyping errors are going to persist in the extended data sets. We calculated the formula that best fitted the accumulation curve for each species and enzyme and plotted the curve with R package 'GGPLOT2' (Wickham, 2016). We calculated the number of reads per individual needed to obtain a mean depth of coverage of 20×, since this threshold of quality is used in 2b.RAD studies (Resh, Galaska, & Mahon, 2018; Whelan et al., 2019). We also estimated values for a coverage of 25× (Warmuth & Ellegren, 2019) to evaluate if with higher coverage we can detect an improvement in the number of total loci.

2.5 | Selective-base ligation simulation

We assessed the potential impact of reducing the number of loci by selective-base-ligation in population genomic analyses. We bioinformatically selected trimmed reads of the corresponding combination of nucleotides to simulate the use of customised adaptors for selective-base ligation on each combination of species and enzyme (Annex S2). This type of ligation is usually performed in the laboratory by designing adaptors that will attach only to reads having the target base at both sticky ends (Figure 2). The simulation of a selective-base-ligation aims to test whether the processing of a proportionally lower number of loci per individual results in the same genetic differentiation as for the whole sample set. We removed from this analysis all samples that had a final mean depth per locus <10 to eliminate errors given by low depth of coverage. For *D. puntazzo* no samples were removed, while for *C. caretta* 5 samples were removed from the AlfI sample set and 7 form CspCI sample set. We used a customized script simulating the effects of building libraries with adaptors ending in 5′-WN-3′ (W = A and T) or 5′-SN-3′ (S = G and C) instead of 5′-NN-3′. These simulations aimed to select trimmed sequences by their first and last base and allocate them in separate folders. These selected sequences were then analysed with Stacks and loci were filtered with the same process as explained above for the whole data set. We calculated the genetic differentiation between pairs of

individuals using Prevosti distance with the R function prevosti.dist from the package poppr 2.8.0 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014) for the data set containing all combinations (NN) and for the two simulated selective-base-ligation data sets. The pairwise genetic distance matrixes among individuals for each selective-base-ligation subset were compared to the original NN matrix with a Mantel test using GENALEX v6.503 (Peakall & Smouse, 2012), then for each matrix we ran a Principal Coordinate Analysis (PCoA) to evaluate whether individuals maintained the same clustering pattern among subsets, using the same program. To detect the eventual decrease of heterozygosity in the subsets compared to their original set of loci we calculated individual observed heterozygosities for the three data sets with VCFTOOLS and used R to perform a Kruskal-Wallis test for each species and enzyme.

3 | RESULTS

3.1 | Library construction and loci identification in *Caretta caretta*

In *C. caretta* extracted DNA ranged from 17.3 to 133.5 ng/μl, and showed high level of degradation in 38% of the samples probably due to the bad condition of the tissue used (Appendix S1: Table S1). After adaptor ligation and amplification by PCR we observed generally good results with AlfI but much lower amplification success with CspCI with 46% of faint bands, as assessed with gel electrophoresis (Appendix S1: Table S1). After purification, library DNA concentration was similar for the two enzymes ranging between 6.7 and 52.3 ng/μl. The mean number of reads per sample was higher for AlfI digested samples, 7.6×10^6 reads per sample (max 10.1×10^6, min 4.0×10^6), than for CspCI digested samples, 6.6×10^6 reads per sample (max 10.7×10^6, min 2.6×10^6) mostly because some samples had low number of reads (Appendix S1: Table S1). The trimming process discarded all the sequences that were shorter than 34 bp for AlfI and 32 bp for CspCI or missed the chosen restriction site, with an average (±*SE*) lower loss per sample in AlfI (19.2 ± 2.1%) than in CspCI (41.9 ± 4.7%) (Table 1). After the loci calling, *C. caretta* showed higher total number of loci with AlfI (66,907 loci) than CspCI (25,416 loci). The mean number of loci retained after all filtering steps were slightly higher for AlfI (72.9 ± 0.4%) than for CspCI (69.4 ± 0.9%), although their final mean depth was smaller (Table 1).

3.2 | Library construction and loci identification in *Diplodus puntazzo*

In *D. puntazzo* starting concentrations ranged from 22.3 to 43.1 ng/μl and none of the samples was degraded. Adaptor ligation and amplification yielded successful amplifications with both enzymes although 17% of the samples digested with CspCI had faint bands (Appendix S1: Table S2). After purification, library DNA concentration was slightly higher for AlfI ranging between 13.6 and 109.63 ng/μl. As for *C. caretta* the sequencing of AlfI in *D. puntazzo* resulted in slightly higher mean number of reads per sample than for CspCI (Table 1). After the loci calling and filtering higher number of loci were also found for *D. puntazzo* for AlfI (84,382 loci) than for CspCI (31,111 loci). The mean number of loci retained after all filtering steps was similar for AlfI (90.6 ± 0.1%) than for CspCI (90.8 ± 0.1%), although their final mean depth was almost double in the latter (Table 1).

3.3 | Quality predictors of sequencing success

In the two species analysed and for both restriction enzymes the number of raw reads was significantly correlated to the final number of loci (Table 2). For *D. puntazzo*, initial DNA concentration, DNA degradation and library DNA quality had no significant effect in the number of raw reads or number of loci. However, for CspCI in *C. caretta*, the initial DNA concentration showed a significant impact on number of reads and loci, and on library concentration (Table 2). The library DNA concentration explained sequencing success in both species since the regression between library DNA concentration and the number of reads and loci was significant in most cases, with the exception of AlfI in *C. caretta* and the number of loci with CspCI in *D. puntazzo* (Table 2). The impact of DNA degradation on sequencing success was only assessed in *C. caretta* since in *D. puntazzo* DNA had initial good quality (Appendix S1: Tables S1 and S2). Interestingly, initial DNA degradation

TABLE 1 Summary of sequencing outcome

	Caretta caretta		*Diplodus puntazzo*	
Species enzyme	AlfI	CspcI	AlfI	CspcI
TR	$7.6 \pm 0.3 \times 10^6$	$6.6 \pm 0.4 \times 10^6$	$7.1 \pm 0.3 \times 10^6$	$6.5 \pm 0.3 \times 10^6$
TMR	$6.2 \pm 0.4 \times 10^6$	$4.2 \pm 0.5 \times 10^6$	$5.3 \pm 0.2 \times 10^6$	$4.3 \pm 0.2 \times 10^6$
IL	48,740 ± 1,489	17,811 ± 1,010	75,971 ± 130	27,989 ± 40
FL	35,576 ± 1,124	12,455 ± 732	68,978 ± 115	25,421 ± 27
RL	72.9 ± 0.4%	69.4 ± 0.9%	90.6 ± 0.1%	90.8 ± 0.1%
FMD	11.5 ± 0.7	19.3 ± 2.4	29.2 ± 1.4	52.2 ± 2.3

Note: Mean (±*SE*) values per individual are given for FMD, final mean depth of coverage per locus; FL, final number of loci after filtering; IL, initial number of loci; RL, percentage of loci retained after filtering; TR, total number of reads; TMR, number of trimmed reads.

TABLE 2 Statistical analyses of potential quality predictors

Explanatory variable	Response variable	Test	*Caretta caretta*				*Diplodus puntazzo*			
			CspCI		*AlfI*		*CspCI*		*AlfI*	
			F or W	p-Value	F or W	p-Value	F or W	p-Value	F or W	p-Value
Raw reads	Final loci	Linear regression	17.7	**.000**	30.4	**.000**	4.7	**.041**	34.4	**.000**
Initial DNA concentration	Raw reads	Linear regression	15.4	**.001**	2.7	.115	0.5	.469	0.4	.522
	Final loci	Linear regression	5.2	**.032**	2.1	.159	0.0	.959	1.5	.236
	Library DNA concentration	Linear regression	15.8	**.001**	0.0	.986	3.2	.087	0.3	.611
Initial DNA degradation	Raw reads	Wilcoxon-Mann-Whitney	60.0	.682	61.0	.726	na	na	na	na
	Final loci	Wilcoxon-Mann-Whitney	44.0	.170	45.0	.194	na	na	na	na
	Library DNA concentration	Wilcoxon-Mann-Whitney	34.0	**.048**	44.0	.174	na	na	na	na
Library DNA concentration	Raw reads	Linear regression	14.2	**.001**	2.0	.174	22.6	**.000**	6.3	**.020**
	Final loci	Linear regression	20.3	**.000**	3.2	.086	0.4	.559	6.2	**.021**
Library DNA quality	Raw reads	Wilcoxon-Mann-Whitney	19.0	**.002**	13.0	**.037**	na	na	25.0	.261
	Final loci	Wilcoxon-Mann-Whitney	12.5	**.001**	6.0	**.005**	na	na	9.0	**.018**

Note: Significant *p*-values after FDR correction are shown in bold. na, tests not available due to insufficient samples with bad initial DNA quality or low library DNA quality.

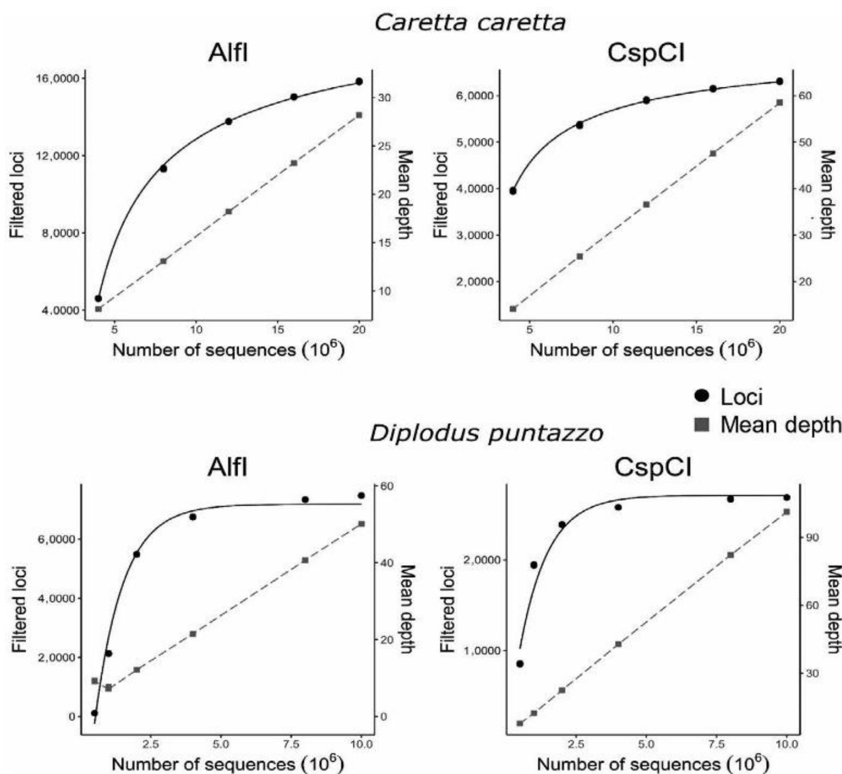


FIGURE 3 Accumulation curves resulting from the resampling analysis. The graphs show the number of final loci (circles) and the mean depth per locus (squares) obtained after filtering, for *Caretta caretta* (top) and *Diplodus puntazzo* (bottom)

was not a good predictor of neither the number of reads nor loci (Table 2). However, library DNA quality and thus amplification success assessed in an agarose gel significantly increased the number of raw reads and final number of loci (Table 2).

3.4 | Resampling analysis

We simulated the sequencing of different target number of reads per sample set and we obtained the total number of loci and mean

TABLE 3 Estimated number of loci and reads needed to obtain different mean depth per locus as derived from the accumulation curve

		Caretta caretta		*Diplodus puntazzo*	
		AlfI	CspCI	AlfI	CspCI
20×	Reads (10^6)	13.5	6.1	3.5	1.7
	Loci	142,910	49,588	68,079	22,225
25×	Reads (10^6)	17.4	7.9	4.6	2.2
	Loci	152,998	53,842	70,571	24,173

Note: The table shows the number of reads per individual and the total number of loci per set corresponding to a mean depth of coverage of 20× and 25× for each species and enzyme.

depth for each simulation (Figure 3; Appendix S1: Table S3). In all simulations, the mean depth of coverage was highly correlated to the number of reads per individual with an R^2 > .99. Based on the accumulation curve (Figure 3) we estimated the mean number of reads per individual and the corresponding number of loci for two mean depth of coverage, 20× and 25× (Table 3). For both species, AlfI needed a much higher number of reads per individual than CspCI to reach the desired coverage of 20×, due to the higher number of loci obtained with this enzyme. We found that, using a coverage of 25×, the total number of final loci improved in AlfI by 4% and by 7% for *D. puntazzo* and *C. caretta*, respectively, and by 9% in CspCI for both species.

3.5 | Selective-base ligation simulation

The selective-base ligation subsets obtained from *C. caretta* retained between 22.2% and 31.5% of the total loci from their original sample sets (Appendix S1: Table S4). In *D. puntazzo* the amount of loci retained was more variable between the two tested subsets (Appendix S1: Table S4), ranging from 19.8% to 43.4%. In this species we also found that for CspCI enzyme the subsets presented lower coverage than the original set, which could be a consequence of the base composition of the regions where this enzyme is cutting and related with the characteristics of the genomes that make the results species specific (Seetharam & Stuart, 2013). Mantel tests in both species showed high correlation between the pairwise genetic distances among individuals assessed with all loci and assessed with a selective base ligation, for both CspCI and AlfI enzymes (Figure 4). This was also reflected in the PCoA, as *C. caretta* samples do not have the exact same pattern among subsets whereas *D. puntazzo* patterns match perfectly despite the lower number of loci retained in the different data sets (Appendix S1: Figure S2). The Kruskal-Wallis test showed no significant differences in observed heterozygosity among any of the subsets and the original set of loci for both species and enzymes (Appendix S1: Table S5).

3.6 | Protocol optimization

We used the results obtained from these simulations to refine the laboratory protocol for 2b-RAD libraries preparation and

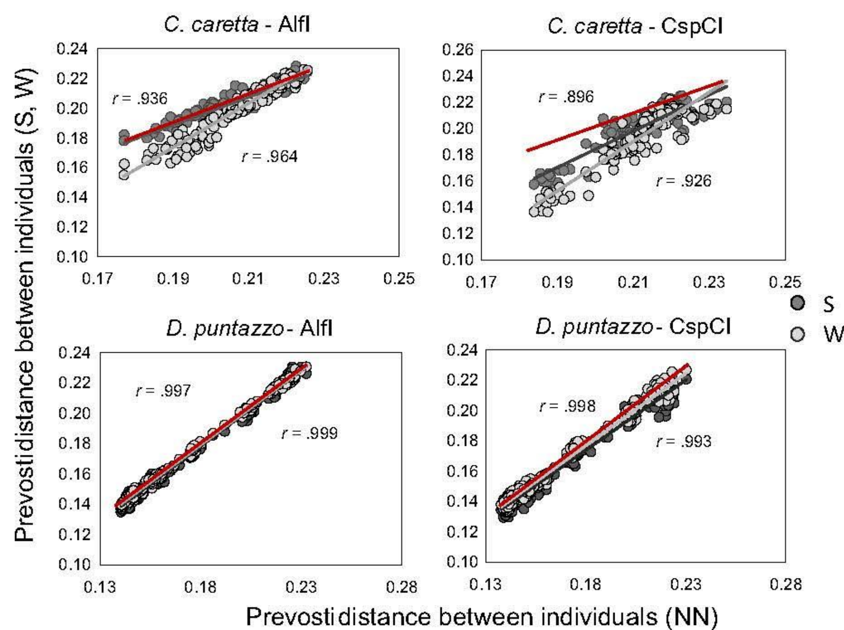


FIGURE 4 Mantel test of genetic differentiation between selective-base subsets and original sets. *X*-axes show Prevosti distance between pairs of individuals for each one of the four original sample sets (with fully degenerated bases –NN-). *Y*-axis show Prevosti distance between the same pairs of individuals for subsets obtained from bioinformatic simulations of selective base ligation (either –SN- or –WN-) for each species and enzyme. Dark grey shows genetic differentiation for S (G and C bases) subsets and light grey for W (A and T) subsets. Correlation coefficient (*r*) is given for each test above the lines for S and below for W. The red line represents the expected correlation function when no deviation in genetic distances is found in the selective-base subsets compared to NN [Colour figure can be viewed at wileyonlinelibrary.com]

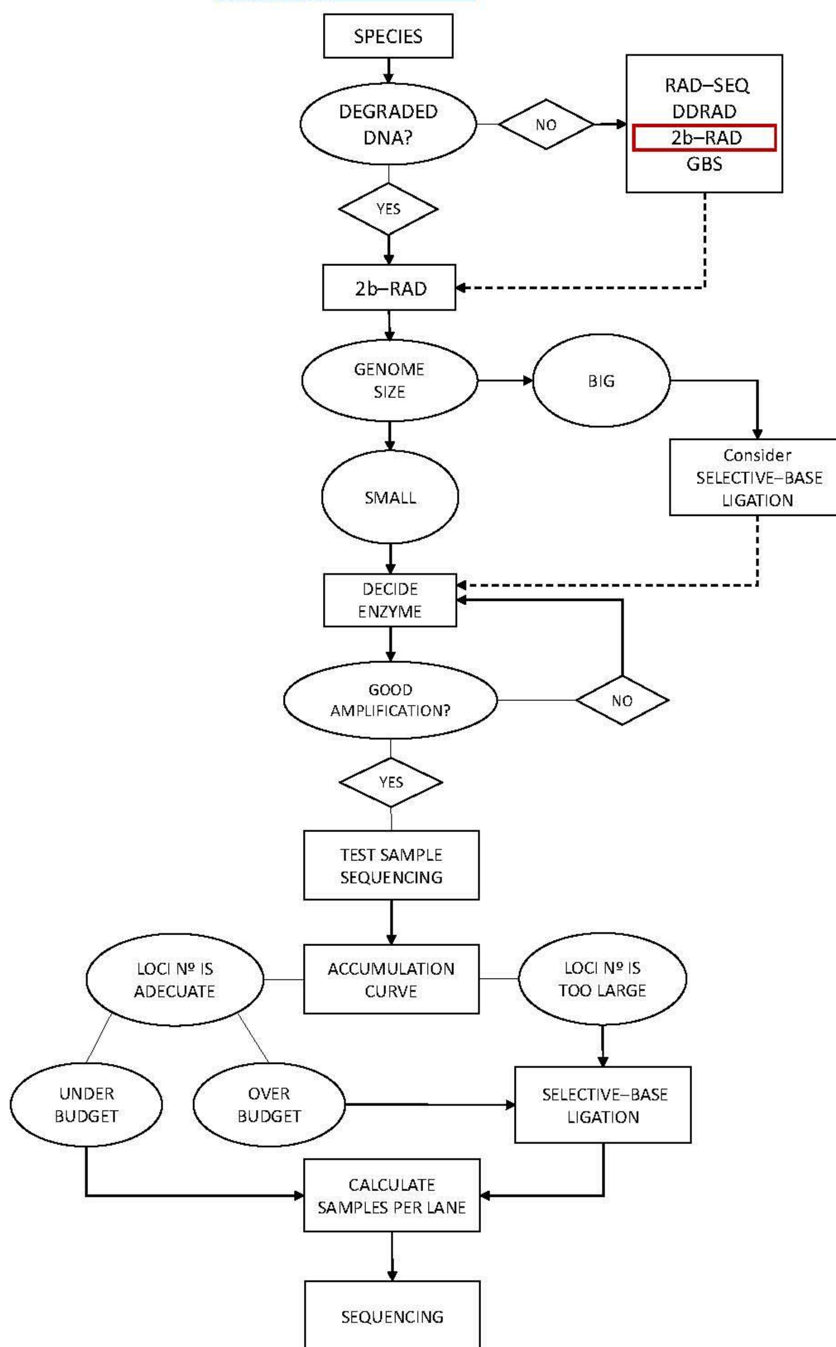


FIGURE 5 Flowchart for 2b-RAD laboratory protocol. This flowchart is meant to aid decision making for 2b-RAD laboratory protocols when studying non-model species. Together with the guidelines listed above this chart aims to make 2b-RAD studies not only easier but also more cost-effective [Colour figure can be viewed at wileyonlinelibrary.com]

sequencing. In fact, given the mean value of depth of coverage, the optimum number of loci and the size of the studied species genome, we can calculate the number of samples to be sequenced in one lane to optimize costs without compromising the results. To facilitate the decision-making process, based on our results, we constructed a flowchart (Figure 5) and a set of guidelines (Box 1) to help future studies design the most efficient and cost effective protocol to reach their goals.

4 | DISCUSSION

In this study, we have shown that 2b-RAD protocol provides efficient results even with degraded samples and we demonstrated how this protocol can be optimised for population genomics of non-model species with different genome sizes. To prove this point, we analysed the sharpsnout seabream *D. puntazzo* and the loggerhead turtle *C. caretta* with two different enzymes, AlfI and CspCI, and performed

BOX 1. Guidelines for the optimisation of a 2b-RAD protocol with non-model species.

- Use 2b-RAD instead of other RAD sequencing techniques if you have degraded samples.
- If the target species has a big genome size, consider performing a selective-base ligation to retain 20%–40% of total loci.
- If the species genome is small, proceed without selective base-ligation.
- Test different IIB enzymes with the target species.
- Use library quality and concentration as predictors of sequencing success.
- Sequence the test samples with conservative conditions to obtain good coverage.
- Calculate an accumulation curve in a preliminary analysis with the test samples to identify the number of reads needed per individual and the total number of loci corresponding to a coverage ≥20×.
- If the total number of loci is adequate for the selected type of study, proceed to sequence the rest of your samples to obtain the mean number of reads needed according to the curve.
- If the total number of loci is too high for the selected study, use a selective base ligation for library building to reduce the amount of loci.
- The number of samples to be sequenced in the same lane is a trade-off between the number of reads per individual, the number of reads provided per lane and available budget.
- If the total number of loci is adequate but the cost of sequencing is over budget, use a selective base ligation for further 2b-RAD library building to reduce the amount of reads needed per sample and therefore fit more samples in one lane.

bioinformatic simulations. Our simulations allow estimating the mean number of reads needed per individual to obtain a reliable genotyping and the corresponding expected number of loci. Moreover, our results indicate that selective-base ligation can be used without compromising pairwise genetic distances among individuals.

In the case of the loggerhead turtle, where several samples had highly degraded DNA, we found that the quality of the initial DNA did not affect the number of raw reads nor the final number of loci, for both enzymes. In fact, the DNA short length for proper IIB enzyme functioning (i.e.,32–34 bp digested fragment) reduces the probability of missing loci even in highly degraded samples. This is a highly valuable characteristic of 2b-RAD methodology, since not all studies can easily access high quality samples. For instance, marine turtle genetic studies usually rely on sampling of stranded individuals (Clusa et al., 2016) or dead embryos found at nests after excavation (Clusa et al., 2018), due to the complexity of their behaviours and the paucity of individuals. In such cases, a genomic protocol capable of providing optimal results with degraded samples is invaluable.

The library quality after adaptor ligation and amplification was a good predictor of sequencing success. The electrophoresis gel after the library amplification of the loggerhead turtle clearly showed that AlfI resulted in a better amplification than CspCI, which failed to yield a clear band in 46% of individuals. Moreover, the sequencing success was poor for samples with faint amplification bands, which resulted in lower number of reads per individual and thus lower number of loci. We thus suggest discarding samples with poor library DNA quality to help optimising sequencing costs. In the case of the sharpsnout seabream, both enzymes showed good results after the amplification, although a few individuals yielded poorer amplification that resulted in significantly lower number of loci, as observed also in the loggerhead turtle were the difference in library quality with the two enzymes was even greater. Moreover, AlfI provided higher number of loci than CspCI in both species as expected, since AlfI recognition sequence has six fixed nucleotides, while CspCI has seven fixed nucleotides. Therefore, AlfI is expected to have a greater density of restriction sites across any genome than CspCI, and potentially yield more loci as observed in the kissing bug *Rhodnius ecuadoriensis* (Hernandez-Castro et al., 2017).

Obtaining more loci, though, reduces depth of coverage per locus for the same mean number of reads per individual. As expected, when using CspCI enzyme our sample sets showed higher values of mean depth then when using AlfI in both species, despite poorer amplification success for CspCI in the loggerhead turtle. A low mean depth per locus leads to less accurate genotype calling and thus higher percentage of missing data across loci (Casso, Turon, & Pascual, 2019; Chow et al., 2019; Maruki & Lynch, 2017), and for this reason a good depth coverage is important to consider data reliable. Since library construction and sequencing produces a variable number of reads per locus, a mean depth of 20× would guarantee that the minimum of five reads per genotype is consistently achieved across most loci for each sample. This would result in fewer genotypes lost and thus more loci retained over all samples. Our simulations on resampling analyses, allowed the construction of the accumulation curve relating the number of reads per sample and the resulting number of loci as well as the linear correlation between the mean depth per locus and the number of reads per individual. Based on the combination of these two functions the number of individuals to be sequenced in one lane can be calculated easily, simplifying decision-making and analysis design for optimizing population genomic studies at the lowest cost. The amount of reads per individual required by the sharpsnout seabream would allow including a fair number of individuals per lane for each enzymes, since both yielded good library DNA quality across samples. However, in the case of the loggerhead turtle, only AlfI enzyme should be used according to library DNA quality. In this case, the amount of reads needed to achieve an adequate

coverage would be very large and the number of loci obtained very high, due to the size of the genome. Under these circumstances, the number of individuals of loggerhead turtle to be included in one sequencing lane would be too small and not affordable by most research groups.

The difference between the two species is mostly related to the crucial role played by the genome size. Species with large genomes will probably produce more loci (due to a greater number of regions yielding the enzyme recognition site) and would need a greater sequencing effort to reach the suitable number of reads per sample for an adequate genotyping. Using a selective-base ligation the number of individuals can be adjusted to the needs of the study considering the number of loci projected by the accumulation curve. Our simulations of customized adaptors with selective base ligation, which extremities would end in –WN or –SN, proved that this type of reduction in the number of loci does not affect genetic differentiation between pairs of individuals. Therefore, the use of a selection of sequences for each sample instead of the whole set, would allow reducing costs by fitting more samples in one lane without compromising overall genetic differentiation. In both species we found that the subsets from the simulated selective-base ligation had a proportionally similar lower number of raw sequences and final loci than the original sets (~25%). However, some differences were observed according to the base and enzyme used in each species suggesting that the species' genome base composition may affect the outcome. Nonetheless, the high levels of correlation that we found between the subsets and the original sets, regardless of the number of loci retained, indicate that they are reliable sources of information. In fact, the slightly lower correlation in genetic distances of *C. caretta* and its differences in PCoAs patterns among subsets were probably a consequence of the bigger genome size of the species, resulting in a lower coverage. This type of selective ligation would be particularly interesting in the case of species with large genomes such as *C. caretta*. Considering the size of this species genome (2.24 Gb) and referring to our resampling simulation, we would need 13.5–17.4 million reads per sample to achieve 20×–25× of coverage, therefore only 20–25 samples could be sequenced in the same lane of a platform providing 340 million reads per run as in the present study. A selective-base ligation would allow reducing the costs of sequencing while ensuring good loci coverage, without influencing the outcome. In fact, since the selective-base ligated set would need only ~ 25% of the original set, between 3.4 and 4.4 million reads per sample are expected to reach the adequate coverage (Warmuth & Ellegren, 2019). Therefore, as much as 78–100 samples could fit in the same Illumina lane, greatly reducing costs without compromising genetic differentiation between individuals. Nevertheless, the number of loci required for a study depends on the scope, the type of analysis performed, and the target species. For instance, selective-base ligation would be less powerful for studies aiming to identify adaptation, since the probability of finding candidate genes can decrease when analysing only a small fraction of the genome (Ahrens et al., 2018).

Finally, we show that 2b-RAD methodologies can be reliable even for degraded DNA samples. Following our set of guidelines, researchers can optimize effort, time, and sequencing cost of 2b-RAD library building for non-model species while maintaining good sequencing depth for a proper genotyping (Box 1, Figure 5).

In conclusion, genomic population studies are increasing in species without reference genomes that rely on restriction-site associated DNA sequencing techniques, although some protocols require good quality DNA. Moreover, transferring protocols across taxa can potentially lead to poor results, such as low number of recovered markers or inadequate genotyping due to differential genomic features. Researchers working with species with large genome sizes or needing lower number of markers can adjust the number of loci by performing selective-base ligation, allowing the sequencing of a larger number of samples, without altering genomic differentiation between individuals as observed by our simulations. The optimal number of samples per lane can, therefore, be adjusted as a trade off with the desired target number of loci and the species genome size for an adequate mean depth of coverage for a correct genotyping. Our results and guidelines aim to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping initial decision-making for a reliable, faster and cost-effective genotyping for population genomic studies.

ACKNOWLEDGEMENTS

M. Babucci (University of Padova) for helping with the initial trimming and preliminary analysis of sequenced data and J. Abril (University of Barcelona) for helping with script building for the simulation analysis with replacement. L. Cardona (University of Barcelona), J. Tomás (University of Valencia) and A.A. Hamza (Alfateh University and University Malaysia Terengganu) for providing tissue samples of *Caretta caretta* from the Valencian feeding ground and the Libyan rookery. This work was supported by the project CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER). CC and MP are part of the research group SGR2017-1120, and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390). AB was supported by grant 2017 FI_B 00997 of the Catalan Government-AGAUR.

AUTHOR CONTRIBUTIONS

A.B., H.T., E.M., C.C., and M.P. conceived and designed the study. A.B., and H.T. did the laboratory analysis with inputs from L.B., and R.F. A.B., and H.T. conducted the data analysis. A.B., and H.T. wrote the manuscript with input from all authors.

DATA AVAILABILITY STATEMENT

Raw reads from all individuals, including information of location of all samples, can be found in the SRA Bioproject PRJNA604507.

All customised scripts (.sh files) can be found in the Supporting Information customised_scripts.zip.

ORCID

Carlos Carreras https://orcid.org/0000-0002-2478-6445
Marta Pascual https://orcid.org/0000-0002-6189-0612

REFERENCES

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D., & Dudaniec, R. Y. (2018). The search for loci under selection: Trends, biases and progress. *Molecular Ecology*, *27*(6), 1342–1356.

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular Ecology*, *23*(7), 1661–1667.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376.

Bellin, D., Ferrarini, A., Chimento, A., Kaiser, O., Levenkova, N., Bouffard, P., & Delledonne, M. (2009). Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics*, *10*(1), 555.

Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., … Fraschetti, S. (2019). A population genomics insight by 2b-RAD reveals populations' uniqueness along the Italian coastline in *Leptopsammia pruvoti* (Scleractinia, Dendrophylliidae). *Diversity and Distributions*, *25*, 1101–1117.

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22*(11), 3028–3035.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation. *Scientific Reports*, *7*, 43417.

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., … Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). *Scientific Reports*, *8*(1), 1435.

Casso, M., Turon, X., & Pascual, M. (2019). Single zooids, multiple loci: Independent colonisations revealed by population genomics of a global invader. *Biological Invasions*, *21*(12), 3575–3592. https://doi.org/10.1007/s10530-019-02069-8

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182.

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.

Chow, J. C., Anderson, P. E., & Shedlock, A. M. (2019). Sea turtle population genomic discovery: Global and locus-specific signatures of polymorphism, selection, and adaptive potential. *Genome Biology and Evolution*, *11*(10), 2797–2806.

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., … Pascual, M. (2018). Philopatry in loggerhead turtles *Caretta caretta*: Beyond the gender paradigm. *Marine Ecology Progress Series*, *588*, 201–213.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Avolio, D., … Cardona, L. (2016). Potential bycatch impact on distinct sea turtle populations is dependent on fishing ground rather than gear type in the Mediterranean Sea. *Marine Biology*, *163*(5), 122.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Giacoma, C., … Cardona, L. (2014). Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. *Marine Biology*, *161*(3), 509–519.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFTOOLS. *Bioinformatics*, *27*(15), 2156–2158.

Davey, J., & Blaxter, M. L. (2011). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics*, *9*, 108.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1.

Fu, B., Liu, H., Yu, X., & Tong, J. (2016). A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). *Scientific Reports*, *6*, 28679.

Galaska, M. P., Sands, C. J., Santos, S. R., Mahon, A. R., & Halanych, K. M. (2017). Geographic structure in the Southern Ocean circumpolar brittle star *Ophionotus victoriae* (Ophiuridae) revealed from mt DNA and single-nucleotide polymorphism data. *Ecology and Evolution*, *7*(2), 475–485.

Girault, G., Blouin, Y., Vergnaud, G., & Derzelle, S. (2014). High-throughput sequencing of *Bacillus anthracis* in France: Investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics*, *15*(1), 288.

Hapke, A., & Thiele, D. (2016). GIBPSS: A toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular Ecology Resources*, *16*(4), 979–990.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., … Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197–1211.

Hernandez-Castro, L. E., Paterno, M., Villacís, A. G., Andersson, B., Costales, J. A., De Noia, M., … Llewellyn, M. S. (2017). 2b-RAD genotyping for population genomic studies of Chagas disease vectors: *Rhodnius ecuadoriensis* in Ecuador. *PLOS Neglected Tropical Diseases*, *11*(7), e0005710.

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., … Narum, S. R. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, *15*(1), 187–202.

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., … Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, *111*, 3775–3780.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *11*, 117–122.

Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*, 208.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). POPPR: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*, 2594–2609.

Lim, H. C., Gawin, D. F., Shakya, S. B., Harvey, M. G., Rahman, M. A., & Sheldon, F. H. (2017). Sundaland's east-west rain forest population structure: Variable manifestations in four polytypic bird species examined using RAD-Seq and plumage analyses. *Journal of Biogeography*, *44*(10), 2259–2271.

Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., & Elshire, R. (2012). TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. *White Paper*, *2012*, 1–12.

Luo, X., Shi, X., Yuan, C., Ai, M., Ge, C., Hu, M., ... Yang, X. (2017). Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. *Parasites & Vectors*, *10*(1), 31.

Maruki, T., & Lynch, M. (2014). Genome-wide estimation of linkage disequilibrium from population-level high-throughput sequencing data. *Genetics*, *197*(4), 1303–1313.

Maruki, T., & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes, Genomes, Genetics*, *7*(5), 1393–1404.

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D., & Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 (Bethesda)*, *6*(11), 3693–3700. https://doi.org/10.1534/g3.116.035220

Paradis, E. (2010). PEGAS: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420.

Pauletto, M., Carraro, L., Babbucci, M., Lucchini, R., Bargelloni, L., & Cardazzo, B. (2016). Extending RAD tag analysis to microbial ecology: A comparison between MultiLocus Sequence Typing and 2b-RAD to investigate Listeria monocytogenes genetic structure. *Molecular Ecology Resources*, *16*(3), 823–835. https://doi.org/10.1111/1755-0998.12495

Peakall, R., & Smouse, P. E. (2012). GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*, 2537–2539.

Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., ... Cariani, A. (2016). Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Marine Genomics*, *25*, 43–48.

Portnoy, D. S., Puritz, J. B., Hollenbeck, C. M., Gelsleichter, J., Chapman, D., & Gold, J. R. (2015). Selection and sex-biased dispersal in a coastal shark: The influence of philopatry on adaptive variation. *Molecular Ecology*, *24*(23), 5877–5885.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Resh, C. A., Galaska, M. P., & Mahon, A. R. (2018). Genomic analyses of Northern snakehead (*Channa argus*) populations in North America. *PeerJ*, *6*, e4581.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807.

Seetharam, A. S., & Stuart, G. W. (2013). Whole genome phylogeny for 21 Drosophila species using predicted 2b-RAD fragments. *PeerJ*, *1*, e226. https://doi.org/10.7717/peerj.226

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8), 808.

Warmuth, V. M., & Ellegren, H. (2019). Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Molecular Ecology Resources*, *19*, 586–596.

Whelan, N. V., Galaska, M. P., Sipley, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms, B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened Round Rocksnail, *Leptoxis Ampla*. *Molecular Ecology*, *28*(7), 1593–1610.

Wickham, H. (2016). GGPLOT2: *Elegant graphics for data analysis*. Berlin, Germany: Springer.

Xu, P., Xu, S., Wu, X., Tao, Y. E., Wang, B., Wang, S., ... Li, G. (2014). Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. *The Plant Journal*, *77*(3), 430–442.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Barbanti A, Torrado H, Macpherson E, et al. Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Mol Ecol Resour*. 2020;20:795–806. https://doi.org/10.1111/1755-0998.13144

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species

Anna Barbanti1*, Hector Torrado1,2*, Enrique Macpherson2, Luca Bargelloni3, Rafaella Franch3, Carlos Carreras1§, Marta Pascual1§.

1.Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Diagonal 643, 08028 Barcelona, Spain.
2.Center for Advanced Studies of Blanes (CEAB-CSIC). C/ d'accés a la Cala St. Francesc, 14, 17300 Blanes, Girona, Spain.
3.Department of Comparative Biomedicine and Food Science. University of Padova, Viale dell'Universita' 16, I-35020 Legnaro, Italy.

* These authors contributed equally to the manuscript
§ These author are considered senior authors

Table of Contents:

Table S1	Page 2
Table S2	Page 4
Table S3	Page 6
Table S4	Page 8
Table S5	Page 9
Figure S1	Page 10
Figure S2	Page 11
Annex I	Page 12
Annex II	Page 14

Table S1. ***Caretta caretta*** **individual information of DNA quality, number of reads and loci for the two 2b-RAD assayed enzymes.** ID: individual identifier; conc = initial DNA concentration; degradation = initial DNA degradation in an agarose gel; Q-AlfI = AlfI library DNA quality amplification assessed with an agarose gel; conc-AlfI=DNA concentration values for AlfI after individual library building; Q-CpcCl = CspCl library DNA quality amplification assessed with an agarose gel; conc-CspCl = DNA concentration values for CspCl after individual library building; TR = total number of reads; TMR = Number of Trimmed Reads; IL= initial number of loci; FL = final number of loci; FMD = mean depth of retained loci

ID	Sampled area	DNA Quality						AlfI					CspCl				
		conc	degradation	Q-AlfI	conc-AlfI	Q-CspCl	conc-CspCl	TR	TMR	IL	FL	FMD	TR	TMR	IL	FL	FMD
Fa	Libya	45.4	Low	Bad	34.8	Bad	7.8	4981427	3383737	40938	27513	6.1	3897562	1153261	13788	8605	5.1
L12	Libya	59.0	High	Good	36.3	Good	22.2	8503920	7407106	53736	38924	13.8	10762107	8485899	21461	14686	38.6
L13	Libya	133.5	Low	Good	26.4	Good	34	8691255	7764809	54181	39044	14.5	9176352	8146930	21426	14644	36.6
L14	Libya	107	Low	Good	33.4	Good	34.9	9791491	8766155	54774	39240	16.2	9536007	7792121	21376	14659	35.3
L15	Libya	29.4	Low	Good	26.8	Bad	19.1	6838207	5865178	51074	37725	10.8	6795486	4451626	20735	14625	20.7
L17	Libya	25.5	High	Bad	11.2	Bad	10.1	5238093	2439548	24673	16470	3.8	5211919	515087	4883	2600	1.8
Lib1	Libya	36.4	Low	Good	31.4	Bad	19.0	7188673	6142948	51654	38102	11.6	6711673	4544145	20785	14638	21.4
Lib10	Libya	32.7	High	Good	13.7	Bad	8.7	7656346	6039881	50169	37387	11.4	5968403	2405693	17837	13522	11.9
Lib12	Libya	78.4	Low	Bad	11.8	Good	39.0	4027319	3646508	40379	30240	6.6	6495805	5216573	21110	14625	23.4
Lib13	Libya	45.6	Low	Good	25.1	Good	36.5	8282366	7203959	52432	38478	12.0	5578612	4246097	20755	14573	18.5
Lib2	Libya	70.4	High	Good	44.8	Good	36.3	7334074	6472629	52019	38392	12.1	7526553	5845812	21099	14627	26.9
lib3	Libya	27.8	Low	Good	43.5	Good	22.6	8845532	7668269	54410	39016	14.1	7587741	5417232	21161	14641	24.9
Lib4	Libya	70.2	Low	Good	36.7	Good	41.4	10082899	8932214	55136	39288	16.6	10524401	8567348	21406	14634	38.6
Lib5	Libya	42.3	High	Good	23.9	Bad	20.5	8684819	7013970	51675	38045	13	8321646	3920769	20043	14489	18.7
Lib6	Libya	62	High	Good	21.2	Good	20.2	10199714	8651217	55150	39193	15.9	8638922	5523854	21171	14640	25.4
Lib8	Libya	78.8	Low	Good	25.8	Good	38.0	8213153	7366455	53925	38963	13.7	8500414	7124259	21321	14645	32.3
T12	Valencia	40.9	Low	Good	30.2	Good	29.6	7359966	6153472	51325	38108	11.7	6212987	4028229	20469	14587	19.1
T13	Valencia	26.7	Low	Good	29.0	Good	30.7	7670958	6545305	52432	38587	12.2	5222785	3335228	20116	14534	15.8
T2a	Valencia	41	Low	Bad	27.4	Good	25.6	8409824	6980169	49708	37080	12.6	6258105	3988243	19550	14385	18.8

T36	Valencia	28.0	High	Good	18.8	Bad	6.7	5626480	3631987	37369	27686	6.7	4746026	897692	9269	6256	4.0
T38a	Valencia	24.4	Low	Good	52.3	Bad	8.5	7315948	5588810	46206	34455	10.1	4257887	1911553	14943	11216	8.8
T4	Valencia	17.9	Low	Good	26.5	Bad	16.4	7510600	5735800	50195	37436	10.8	4208941	1012260	10452	7216	4.5
T44	Valencia	37.0	High	Good	37.7	Bad	9.2	6277604	4454643	40844	30191	8.5	2672161	1104035	10626	7449	5.1
T50	Valencia	40.5	High	Good	17.9	Bad	12.6	7485230	5387302	45367	34258	10.5	4088163	1197188	11675	8418	5.7

Table S2. ***Diplodus puntazzo*** **individual information of DNA quality, number of reads and loci for the two 2b-RAD assayed enzymes.** ID: individual identifier; conc = initial DNA concentration; degradation = initial DNA degradation in an agarose gel; Q-AlfI = AlfI library DNA quality amplification assessed with an agarose gel; conc-AlfI=DNA concentration values for AlfI after individual library building; Q-CpcCI = CspCI library DNA quality amplification assessed with an agarose gel; conc-CspCI = DNA concentration values for CspCI after individual library building; TR = total number of reads; TMR = Number of Trimmed Reads; IL= initial number of loci; FL = final number of loci; FMD = mean depth of retained loci

ID	Sampled area	DNA Concetration ng/ul						AlfI					CspCI				
		conc	degradation	Q-AlfI	conc-AlfI	Q-CspCI	conc-CspCI	TR	TMR	IL	FL	FMD	TR	TMR	IL	FL	FMD
Bla01	Blanes	37.6	Low	Good	85.1	Good	27.3	5324128	3992915	75542	68333	21.6	7427275	4433067	28093	25511	52.9
Bla02	Blanes	37.1	Low	Good	58.2	Good	36.0	8700118	6505955	76307	69175	35.3	5498932	3361477	27813	25311	39.9
Bla03	Blanes	36.6	Low	Good	54.0	Good	28.9	7771627	6009317	76262	69235	32.6	6871194	4442372	27996	25490	53.0
Bla04	Blanes	43.1	Low	Good	69.1	Good	33.3	7920370	6120075	76094	69114	33.2	7355511	4856652	27959	25500	58.9
Bla07	Blanes	37.6	Low	Good	52.0	Good	19.3	7160027	5762100	75679	68878	31.2	7018614	4744338	27824	25368	56.9
Bla18	Blanes	37.4	Low	Good	67.2	Good	34.4	7578818	6029578	77225	69426	32.8	7671834	5595596	28585	25641	69.3
Bla19	Blanes	37.7	Low	Good	58.2	Good	70.3	10119330	8148373	76907	69461	44.2	4890047	3587440	28114	25464	44.1
Bla20	Blanes	38.6	Low	Good	88.2	Good	73.8	6182499	4786426	75759	68883	26.0	5497955	3971877	28023	25538	49.2
Bla21	Blanes	38.6	Low	Good	77.8	Good	65.5	6753023	5222713	75826	68991	28.3	4848127	3564442	27915	25452	43.9
Bla22	Blanes	38.5	Low	Good	78.1	Good	90.3	5814910	4465047	75607	68679	24.2	4113480	2946595	27881	25383	36.3
Bla23	Blanes	35.0	Low	Good	89.8	Good	80.7	5942181	4679485	75552	68689	25.4	4176015	3015593	27804	25312	37.3
Bla25	Blanes	40.9	Low	Good	72.1	Good	54.7	5877649	4303486	75603	68707	23.3	6266242	4115257	27925	25420	49.8
Xab01	Xabia	35.5	Low	Good	63.9	Good	23.3	5805863	4220838	75676	68643	22.8	7875510	5134996	28103	25538	61.6
Xab02	Xabia	22.3	Low	Good	33.7	Good	14.0	10024412	7212432	76325	69208	38.8	7272840	4947218	27985	25493	59.8
Xab03	Xabia	37.8	Low	Good	45.6	Bad	14.6	7181639	5107720	75998	68878	27.7	5814862	3268040	27666	25160	38.1
Xab04	Xabia	34.0	Low	Good	63.2	Bad	13.6	6440979	4755949	75932	68307	25.3	6351337	3717435	27802	25086	42.7
Xab09	Xabia	36.2	Low	Good	65.1	Bad	18.6	6478589	4955504	75788	68873	26.8	5903444	3707734	27872	25389	43.9
Xab18	Xabia	40.1	Low	Good	92.4	Good	20.2	8354162	6670262	76506	69285	36.2	8390215	5379501	28184	25580	65.6
Xab19	Xabia	27.5	Low	Good	68.5	Good	18.3	9311164	6403999	76534	69283	34.7	8569552	5713713	28253	25575	70.0

Xab20	Xabia	32.3	Low	Good	100.1	Good	38.9	7759509	6095935	76210	68984	32.8	6886576	5005121	28090	25488	61.3
Xab21	Xabia	39.4	Low	Good	82.2	Good	22.4	8895766	6858713	76940	68965	36.5	9081702	6254933	28218	25385	75.7
Xab23	Xabia	32.2	Low	Good	109.6	Good	21.0	5212297	3780980	75319	67785	20.2	6525446	4240660	27966	25291	51.2
Xab24	Xabia	29.1	Low	Good	95.1	Good	17.9	4511470	3184710	74107	66913	17.2	6710301	4054563	27877	25394	48.3
Xab25	Xabia	35.6	Low	Good	63.9	Bad	19.2	5530558	4236331	75597	68457	22.9	6123340	3599048	27802	25333	42.8

Table S3. Number of loci and mean depth per locus after filtering of the ten simulated runs for each initial number of reads per individual. The simulated number of reads varies between the two species according to their different genome sizes. Simulations for *C. caretta* (2.2 Gb) range between 4x10^{6}-20x10^{6} reads and for *D. puntazzo* (0.9 Gb) range between 0.5x10^{6}-10x10^{6}.

Caretta caretta						***Diplodus puntazzo***					
		AlfI		*CspCI*				*AlfI*		*CspCI*	
Reads (x10^{6})	RUN	Loci	MD	Loci	MD	Reads (x10^{6})	RUN	Loci	MD	Loci	MD
4	1	46083	8.1	39308	14.2	0.5	1	1115	9.1	8459	7.8
4	2	46380	8.1	39720	14.1	0.5	2	1151	9.2	8548	7.8
4	3	46339	8.1	39734	14.1	0.5	3	1111	9.1	8508	7.8
4	4	46157	8.2	39373	14.1	0.5	4	1150	9.1	8603	7.8
4	5	46340	8.1	39582	14.1	0.5	5	1157	9.1	8510	7.8
4	6	45948	8.2	39713	14.1	0.5	6	1148	9.2	8521	7.8
4	7	46387	8.2	39657	14.1	0.5	7	1119	9.3	8557	7.8
4	8	46206	8.1	39627	14.2	0.5	8	1138	9.4	8551	7.8
4	9	45818	8.1	39499	14.1	0.5	9	1091	9.2	8527	7.8
4	10	46075	8.1	39706	14.1	0.5	10	1145	9.1	8609	7.8
8	1	113685	13.1	53510	25.5	1	1	21375	7.1	19521	12.4
8	2	112937	13.1	53802	25.4	1	2	21200	7.7	19409	12.4
8	3	113427	13.1	53664	25.5	1	3	21288	7.1	19461	12.4
8	4	113452	13.1	53870	25.4	1	4	21271	7.1	19426	12.4
8	5	113279	13.1	53596	25.5	1	5	21409	7.1	19463	12.4
8	6	113311	13.1	53413	25.5	1	6	21145	7.1	19433	12.4
8	7	112776	13.1	53570	25.5	1	7	21204	7.1	19445	12.4
8	8	113500	13.1	53675	25.4	1	8	21287	7.1	19410	12.4
8	9	113636	13.1	53521	25.4	1	9	21183	7.1	19428	12.4
8	10	113153	13.1	53682	25.5	1	10	21169	7.1	19358	12.4
12	1	137681	18.2	58829	36.5	2	1	54705	12.2	23888	22.5
12	2	137503	18.2	58990	36.6	2	2	54989	12.2	23953	22.5
12	3	137875	18.2	58922	36.6	2	3	54817	12.2	23890	22.5
12	4	137789	18.2	59093	36.6	2	4	54925	12.2	23883	22.5
12	5	137956	18.2	59035	36.6	2	5	54928	12.2	23950	22.5
12	6	137720	18.2	59063	36.6	2	6	54840	12.2	23866	22.5
12	7	137445	18.2	58910	36.6	2	7	54838	12.2	23828	22.5
12	8	137647	18.2	58843	36.6	2	8	54861	12.2	23904	22.5
12	9	137839	18.2	58904	36.7	2	9	55024	12.2	23914	22.5
12	10	137635	18.2	59209	36.6	2	10	54807	12.2	23892	22.5
16	1	150107	23.3	61443	47.6	4	1	67398	21.5	25827	42.7
16	2	150559	23.2	61562	47.6	4	2	67344	21.5	25819	42.7
16	3	150159	23.2	61622	47.6	4	3	67440	21.5	25779	42.8
16	4	150510	23.2	61668	47.6	4	4	67680	21.5	25818	42.8
16	5	150550	23.2	61569	47.6	4	5	67482	21.5	25809	42.7
16	6	150708	23.2	61535	47.5	4	6	67497	21.5	25804	42.7
16	7	150237	23.2	61639	47.5	4	7	67459	21.5	25785	42.7
16	8	150282	23.2	61469	47.6	4	8	67459	21.5	25775	42.8
16	9	150334	23.3	61619	47.6	4	9	67535	21.5	25812	42.7
16	10	150081	23.3	61349	47.6	4	10	67453	21.5	25784	42.7
20	1	158271	28.2	63011	58.4	8	1	73463	40.7	26665	82.2

20	2	158016	28.2	62979	58.5
20	3	158256	28.2	62965	58.5
20	4	158118	28.2	63187	58.5
20	5	158723	28.2	62948	58.5
20	6	158557	28.2	63001	58.6
20	7	158416	28.2	62972	58.5
20	8	158281	28.2	63208	58.5
20	9	158386	28.2	63019	58.5
20	10	158596	28.2	63054	58.4

8	2	73443	40.7	26735	82.2
8	3	73391	40.7	26723	82.2
8	4	73439	40.7	26739	82.2
8	5	73453	40.7	26693	82.1
8	6	73356	40.7	26693	82.1
8	7	73292	40.7	26765	82.0
8	8	73443	40.7	26711	82.2
8	9	73490	40.7	26748	82.2
8	10	73355	40.7	26742	82.1
10	1	74664	50.1	26849	101.1
10	2	74686	50.1	26868	101.3
10	3	74679	50.2	26939	101.2
10	4	74645	50.1	26909	101.2
10	5	74653	50.1	26878	101.2
10	6	74837	50.1	26850	101.3
10	7	74818	50.1	26907	101.2
10	8	74810	50.1	26853	101.2
10	9	74660	50.2	26886	101.1
10	10	74678	50.1	26875	101.3

Table S4. Summary of selective-base ligation simulation for *Caretta caretta* and *Diplodus puntazzo* for the 2b-RAD enzymes (AlfI and CspcI). The symbols of the selective bases simulated for ligation of adapters correspond to the IUPAC notation: S=C+G and W=A+T, N=all bases. TR: total number of reads matching the base criteria. TMR: number of trimmed reads. IL: initial number of loci. FL: final number of loci. FMD: mean depth of coverage per locus after all the filtering process.

Species	Enzyme	Base	TR	TMR	IL	FL	FMD
Caretta caretta	Alf	N	182215898	149242071	66907	40319	11.5
		S	60699692	32301347	17065	9903	11.0
		W	69557740	40145869	15756	8943	10.4
	CspcI	N	158900658	100831134	25416	16332	19.3
		S	28296438	23381824	10442	4965	15.4
		W	29059846	24998483	10910	5149	15.3
Diplodus puntazzo	Alf	N	170651088	129508843	84382	66786	29.5
		S	26519654	25397694	17340	13209	27.2
		W	37242552	36219043	25703	19761	27.7
	CspcI	N	157140351	103657668	31111	24401	52.9
		S	22561494	19068008	11772	7372	29.2
		W	31887386	28060229	16177	10600	32.6

Table S5. Kruskal-Wallis test of heterozygosity values. We found no significant difference between heterozygosity values of selective-base ligation subsets and their original set of loci.

	Caretta caretta		*Diplodus puntazzo*	
	AlfI	CspCI	AlfI	CspCI
Kruskal-Wallis test	4.512	3.232	1.647	3.229
p-value	0.105	0.198	0.438	0.198

Figure S1. Example of library DNA quality identification. This agarose gel of *C. caretta* shows how library DNA was labelled as 'Good' (G) if the band was bright or 'Bad' (B) if the band was faint.

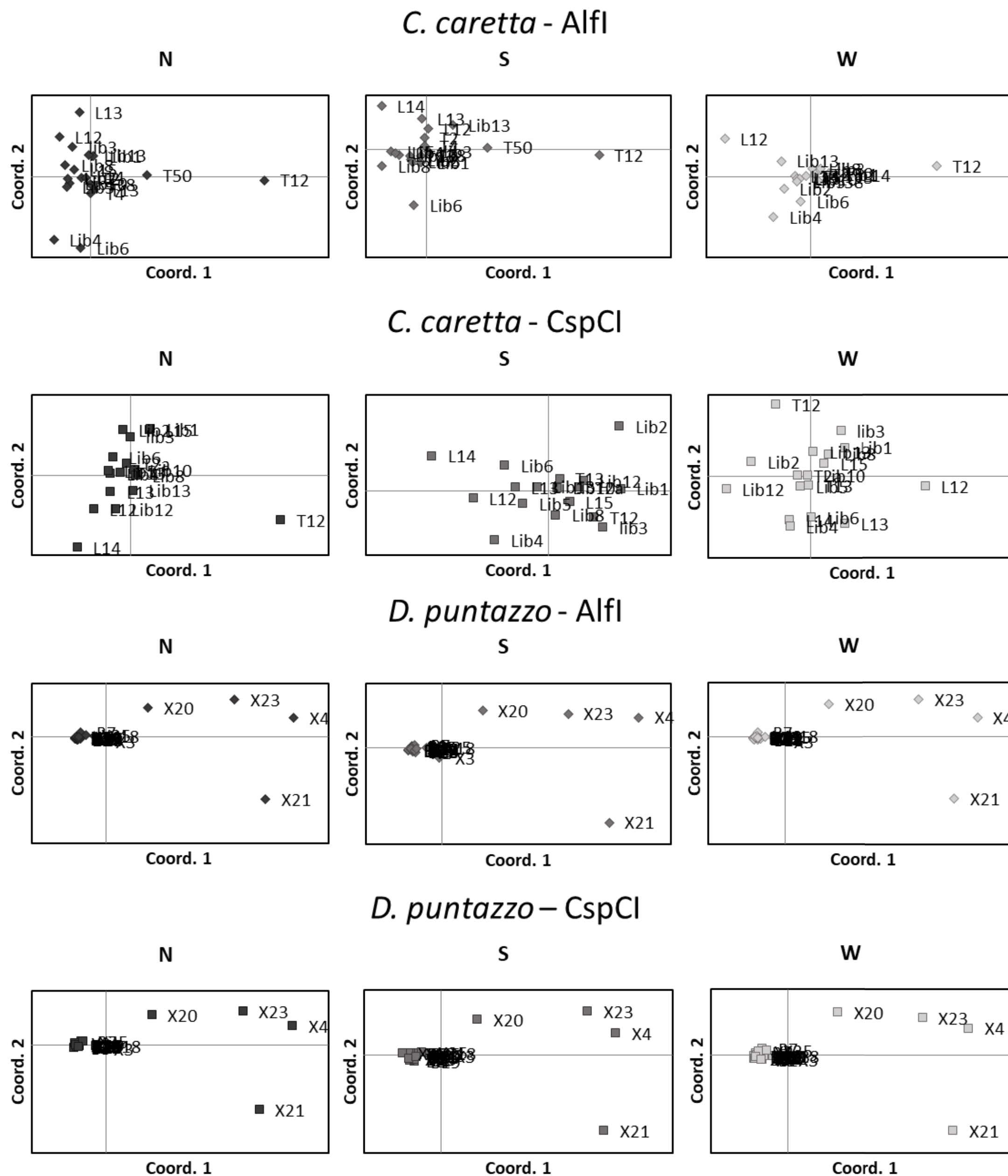


Figure S2. Principal Coordinates Analysis of different sets of loci. The PCoAs show the distribution of genetic distances of subsets and original sets of loci. In the case of *D. puntazzo* the distribution follows the same pattern among all enzymes and data sets. In *C. caretta* patterns do not match probably due to the low depth of coverage of the original sequences. Diamonds show data for AlfI and squares for CspCI enzyme. Individuals' points for the original data set (N) are in black, For CG selection (S) in dark Grey and for AT selection (W) in light grey.

Annex I

Library building protocol

Genomic DNA of each sample is diluted with DNA-free water to obtain 7µl containing 180ng of DNA.

i)Digestion

Digestion of each sample is prepared using one of the following protocols depending on the enzyme used. a) 1µl Enzyme Buffer R (10x), 1µl SAM (S-adenosyl-methionine, 100µM), 1µl AlfI enzyme (2u/µl), 3µl of sterile water and 4µl of DNA, or b) 1µl Enzyme CutSmart (10x), 1µl SAM (100µM), 1µl CspCI enzyme (5u/µl), 3µl of sterile water and 4µl of DNA. All steps are performed with a GenAmp PCR System 2700 (Applied Biosystems®). Digesting conditions are 37.0 °C for 60 min, 65.0 °C for 20 min.

Enzyme	Fragment length and recognition sequence (5'-3')
AlfI	(N_{10-12}) **GCA** (N_6) **TGC** (N_{12-10})
CspCI	(N_{10-11}) **CAA** (N_5) **GTGG** (N_{12-13})

ii)Ligation

The two adaptors are prepared by hybridization of different oligos: adaptor 2 is obtained by hybridising the oligos 1 and 2 and adaptor 3 is obtained by hybridizing the oligos 1 and 3. Mix 22µl of oligo 1 (100µM) with 22µl of oligo 2 (or 3) (100µM) and 206µl sterile water. Annealing conditions are: 65.0 °C for 30 min, 83 cycles of 65.0 °C for 20 sec and 64.7 °C for 20 sec, with temperature decreasing by 0.3°C for each cycle, and a final hold of 15.0 °C for 10 min.

Oligo	Sequence (5'-3')
1	AGA TCG GAA GAGC
2	CTA CAC GAC GCT CTT CCG ATC TNN
3	CAG ACG TGT GCT CTT CCG ATC TNN

Ligation is prepared using the following protocol: 3µl sterile water, 1.5µl T4 Ligase Buffer (10x), 2.5µl adaptor 2 (4µM), 2.5µl adaptor 3 (4µM), 0.5µl ATP (10mM), 5µl T4 Ligase (200u/µl) and 10µl of digested DNA. Ligation conditions are: 16.0 °C for 180 min and 65.0 °C for 10 min.

iii)Amplification

Amplification is prepared using the following protocol: 25.15µl sterile water, 12µl Taq HF Buffer (5x), 0.75µl dNTPs (25mM), 1.2µl Amplification primer F (10µM), 1.2µl Amplification primer R (10µM), 3µl primer FOR (10µM), 1.2µl Taq phusion, 3µl Barcode primer (10µM) and 12.5µl of ligated DNA. Each 60 µl sample amplification is split in 3 microplate wells containing 20µl of amplification mix to optimise amplification outcome. Amplification PCR conditions are: 98.0 °C for 5 min, 14 cycles of 98.0 °C for 5 sec, 60.0 °C for 20 sec and 72.0 °C for 5 sec, and a final extension step of 72.0 °C for 5 min.

Primer	Sequence (5'-3')
Amplification primer F	AAT GAT ACG GCG ACC ACC GA
Amplification primer R	CAA GCA GAA GAC GGC ATA CGA
Primer FOR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATCT
Barcode primer (N_7)	CAA GCA GAA GAC GGC ATA CGA GAT**(N_7)**GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC

iv)Purification

Purification is carried out mixing 60µl of amplified product with 90µl of magnetic beads (Beckman Coulter®) and placing the plate on a magnetic plate. After removing the supernatant, the product was washed with 180µl of 85% ethanol. The ethanol was removed and the beads with the attached DNA were left to dry. Finally the DNA was resuspended in 25µl of sterile water and 20µl of supernatant without beads were stored for sequencing.

v) Pool

Prepare the pool by pipetting 180ng of DNA for each purified library, based on each sample's library concentration.

Annex II

AlfI trimming

Cut the adapter form the raw data using **cut_adapter_AlfI.sh.**

The input files are raw reads in fastq format. This script cuts the adapter from the sequences using the program cutadapt-1.5 (Martin, 2011) and finally converts the sequences in fasta format. The output files must be used as input files for the next step.

Trim raw sequences using **bRad_AlfI.sh**.

The input files are raw reads in fasta format. This script identifies the enzyme recognition site, cut all sequences down to a same length, flip sequences to have them all orientated in the same direction and check for duplicates. This script uses the following programs: TruncateFastq.pl, 2b_Extract.pl (http://people.oregonstate.edu/~meyere/tools.html), fnafile (NEWBLERtools), SHRiMP v2.2.3 (David, Dzamba, Lister, Ilie, & Brudno, 2011), revcompl.pl (https://github.com/KorfLab/Perl_utils) and CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012). The output file includes the sequences ready to be used in Stacks.

CspCI trimming

Cut the adapter form the raw data using **cut_adapter_CspCI.sh.**

The input files are raw reads in fastq format. This script cuts the adapter from the sequences and finally converts the sequences in fasta format. The output files must be used as input files for the next step.

Trim raw sequences using **recsite_CspCI_subset.sh**.

The input files are raw reads in fasta format. This script identifies the enzyme recognition site, cut all sequences down to a same length, flip sequences to have them all orientated in the same direction and check for duplicates. The output file includes the sequences ready to be used in Stacks.

Resampling simulation for accumulation curve

Perform a resampling simulation with replacement using **resampling_fasta.sh**.

The input files are raw reads in fasta format. This script resamples a given number of raw sequences per individual (N_{SEQ}), with replacement.

Selective-base Ligation

Remove the adaptor sequence from each read using either **cut_adaptor_AlfI.sh** or **cut_adaptor_CspCI.sh.** The input files are raw reads in fasta format. The output file is used as input for the simulation of selective-base ligation.

Perform a simulation of selective-base ligation using **select_bases_fasta_2.0.sh.**

This script extracts sequences selected by first and last base to simulate a selective-base ligation. The output has to be trimmed with one of the scripts mentioned above depending on the enzyme.

Each script is annotated with information on how to set the script parameters.

All scripts will be available upon request.

References

David, M., Dzamba, M., Lister, D., Ilie, L., & Brudno, M. (2011). SHRiMP2: sensitive yet practical short read mapping. *Bioinformatics*, 27(7), 1011-1012.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), 10-12.

CHAPTER 4

Population structure and local adaptation of the Eastern Mediterranean loggerhead turtle (*Caretta caretta*) rookeries

Barbanti A[1]., Margaritoulis D.[2], Turkozan O.[3], Khalil M.[4], Snape R.[5,6], Demetropoulos A.[7], Hamza A. A.[8,9], Ulger C. [3], Pascual M[1]. and Carreras C[1].

Affiliations

1.Department of Genetics, Microbiology and Statistics and IRBio, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain.

2. ARCHELON, the Sea Turtle Protection Society of Greece, Solomou 57, 10432 Athens, Greece

3. Adnan Menderes University, Faculty of Science and Arts, Department of Biology, Aydin, 09010, Turkey

4. MEDASSET, PO Box 19, Tyre, Lebanon

5. Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Penryn TR10 9FE, UK

6. Society for Protection of Turtles, Kyrenia, Cyprus

7. Cyprus Wildlife Society, PO Box 24281, 1703 Nicosia, Cyprus

8. Libyan Seaturtle Program, Environment General Authority, Alfateh University, P.O. Box 13793, Tripoli, Libya,

9. Faculty of Marine Science and Environment, University Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Malaysia.

Abstract

The loggerhead turtle (*Caretta caretta*) is a highly migratory species which interestingly fits in with its philopatric behaviour. Feminisation of populations for global warming, mortality by anthropogenic activities and declining hatchling survival rates are threatening current nesting and foraging areas. Conservation measures in the eastern Mediterranean Sea are crucial to support and monitor nesting populations, and should be based on population structure, not yet resolved by previous research. Here we obtained genomic data from 2b-RAD libraries of 199 individuals of 9 loggerhead turtle populations from the eastern Mediterranean. We calculated the effective population size (*Ne*) from thousands of markers compared with the number of adult breeders to understand the status of the studied populations. We found a strong correlation between these estimates, although Turkish populations had higher *Ne* values, probably due to a past large population size. We analysed the population structure of this area using principal coordinate analysis (PCoA), pairwise F_{ST} and Bayesian clustering analysis and found strong genetic differentiation among all rookeries. Finally, we used outlier analysis and environmental association analysis to identify candidate loci for local adaptation and test the potential role of temperature in the population genetic structure of the region. We found several loci associated to temperature, independently of the geographic location of the nesting beaches. Our results show the ability of 2bRAD to refine the population structure of endangered species such as marine turtles and identify signals of local adaptation. These findings provide the baseline for future studies on sea turtle genomics for conservation.

Introduction

The marine ecosystem is one of the environments most affected by the pressure of anthropogenic activities and climate change (Dietz and Adger, 2003; Hillebard *et al*., 2018). The study and monitoring of marine and coastal environments has therefore become crucial for the preservation of wildlife (Harris *et al*., 2019; Watanabe *et al*., 2019). Accurate knowledge of populations' structure is fundamental to design management units (MUs) and to detect genetically isolated populations, which require independent conservation actions (Casale and Mariani, 2014). The design and conservation management of MUs not only depends on the mere structure of the target population of study, but considers several other factors that shape and impact population structure, connectivity and self-sustainability. For instance, the adaptability of a species to the constant changing environment is critical for conservation planning since it can shape population structure. Similarly, small effective population size is a crucial factor for populations' self-sustainability as it increases vulnerability for populations of conservation concern. Genetic assessments can provide this information to advise and assist conservation planning of MUs. Conservation of marine species should therefore ideally include a complete genetic assessment of population structuring, including the role of adaptation, and estimates of effective population sizes.

Sea turtles are species of conservation concern that would benefit from a complete genetic assessment. These highly migratory species have been affected by the current biodiversity crisis on a global scale and their populations are listed as vulnerable, endangered or critically endangered by the IUCN red list worldwide (IUCN 2019). Although in the past 60 years conservation actions have been looking after these species in the more critical regions, the study and monitoring of their populations are still crucial to ensure the recovery and preservation of sea turtles in the long term. Sea turtles can migrate thousands of kilometres, moving between foraging grounds and nesting beaches (Mansfield *et al*., 2009; Nichols *et al*., 2000). Although they are highly mobile individuals, they display philopatric behaviour, called natal-homing, by which sexually mature females (i.e. approximately age from 15 to 30 years old (Casale *et al*., 2011)) return to their natal beach to nest (Limpus *et al*., 1984; Lohmann *et al*., 2013). This behaviour leads geographically separated rookeries to be genetically isolated within a relatively short distance (e.g. 30 kilometres (Nishizawa *et al*., 2018)). Nonetheless, females can nest sporadically in very distant areas (Carreras *et al*., 2018). Recently adult males have also been discovered to present a certain degree of natal homing (Clusa *et al*., 2018), although due to the difficulties in studying male individuals of sea turtle species, the extent of this behaviour is still uncertain.

The loggerhead turtle (*Caretta caretta*) is a charismatic species distributed worldwide throughout the tropics and characterised by temperature-dependent sex determination (TSD), where the pivotal temperature is around 29°C at which half of the clutch will result in female individuals and half in males (Janzen and Paukstis, 1991; Mrosovsky *et al*., 2002). The Mediterranean Sea is considered a regional management unit (RMU) for loggerhead turtles (Wallace *et al*., 2010), with rookeries mostly located in the eastern region (Casale *et al*., 2018), where the water is warmer and temperatures have been historically more suitable for nesting than the western region (Pike, 2013). The western area is characterised by the presence of developmental foraging grounds, were mostly juvenile turtles from different regional management units gather (Carreras *et al*., 2011, Clusa *et al*., 2014). The Mediterranean population has increased but is still dependent on conservation actions (Casale, 2015). The two main conservation concerns linked to Mediterranean populations are global warming and fishery activity (Casale *et al*., 2018). On one hand, global warming is considered a global threat to marine turtles due to the temperature sex determination (Hawkes *et al*., 2009), and recent reports indicate that the Mediterranean Sea is going to be one of the places most affected by increasing temperatures on the planet (Zhai *et al*., 2018). In this context, current Mediterranean nesting areas have been predicted to be heavily impacted (Hawkes *et al*., 2007: Witt *et al*., 2010) resulting in a feminisation of populations and in a decrease of hatchling viability (Pike, 2014). For this reason, information on the genomic response of these populations to temperature can be highly informative for a better understanding of the potential adaptation of the species to global warming. On the other hand, accidental bycatch is still one of the main threats to juveniles and adults in foraging areas (Casale *et al*., 2018). Assessing the area of origin of the animals captured at sea (e.g. Mediterranean or Atlantic) is crucial to assign threats to the populations affected (Clusa *et al*., 2016) but high resolution of the genetic markers is necessary to assign each individual to its own population within the region. Therefore, understanding the genetic structure and dynamics of the nesting populations in this area is critical to inform conservation management in both nesting and foraging areas in order to improve conservation actions.

Eastern Mediterranean loggerhead nesting populations have been previously analysed using genetic markers, although important questions remain to be answered. The use of 15 microsatellite markers revealed the existence of 5 different genetic units (Clusa *et al*., 2018), while the maternally inherited mitochondrial DNA suggested at least 7 differentiated units (Shamblin *et al*., 2014). Furthermore, the current nesting populations of loggerhead turtles in this region are likely the result of at least two independent colonisation events from the Atlantic: first in Libya (65,000 years ago) and more recently in Calabria (15,000 years ago) (Clusa *et al.*, 2013). Early

studies with microsatellites suggested that only females were philopatric, and that male mediated gene flow helped homogenize some of the nesting populations in terms of nuclear DNA (Bowen *et al*., 2005: Carreras *et al*., 2007). The use of larger sets of microsatellites provided general support for male philopatry while other hypothesis were provided to explain the genetic similarities among geographically distant rookeries (i.e. opportunistic mating either in foraging grounds or on migratory routes) (Clusa *et al*., 2018). However, these genetic similarities could be due to the lack of power of the markers used, considering the observed ability to identify genetic differentiation among some nesting areas when increasing the number of markers (Carreras *et al*., 2007; Clusa *et al*., 2018). As a result, the extent of male philopatry and genetic structuring has yet to be resolved with a higher number of genetic markers. Moreover, it has been hypothesized that the 3-dimensional variations of the mitochondrial genes ND1 and ND3 within the Mediterranean may be related to thermal adaptation (Novelletto *et al*., 2016), although a genome wide assessment of the role of temperature in shaping population differentiation has not yet been performed.

All these unresolved questions that rely on using a large number of markers can be assessed using genomic tools. Nowadays, high-throughput sequencing technologies can be applied to non-model species at the individual level, to allow scoring many markers across the whole genome, and identifying candidate loci for local adaptation (Carreras *et al*., 2020). A growing number of population genomic studies in the Mediterranean Sea have focused on non-model marine species, providing both an improved resolution for population structure (Boscari *et al*., 2019; Casso *et al*., 2019) and genome wide signals of adaptation (Carreras *et al*., 2017, Carreras *et al*., 2020, Torrado *et al*., 2020). Most population genomic studies rely on library construction technologies that are only feasible with good quality and quantity of DNA, which is often hard to obtain for sea turtle samples (e.g. dead hatchlings or stranded individuals). However, 2b-RAD library construction and sequencing has been proved to work successfully with poor quality loggerhead turtle samples, ensuring a good trade-off between cost and outcome, opening the field of sea turtle population genomic studies (Barbanti *et al*., 2020). Implementing genomic approaches in studying the population genetic structure of Mediterranean loggerhead turtles can provide robust results for identifying population differentiation, including adaptation signals, to help conservation policies. Moreover, genomics can also provide reliable estimates of effective population sizes, a fundamental concept in conservation to evaluate population viability and the genetic risk associated to small populations, affected by the number of mating individuals, sex ratio, reproductive success, age structure, migration and other demographic factors (Waples *et al*., 2016). While census sizes are normally used for conservation assessments worldwide (IUCN, 2012), genetic deleterious effects are related to the effective population size, defined as the

number of breeders of a theoretical population that will show the same impact of random genetic drift as the real population (Wright, 1938). Thus, potential genetic threats of natural populations, such as loss of alleles by genetic drift or inbreeding depression, are related to the effective population size rather than the adult census size (N_a). Even though the concept of effective population size is fundamental in conservation genetics, this estimate is rarely used and is assumed to be constantly related to N_a across populations.

Here, we aim to identify the diversity and population structure of loggerhead turtle Mediterranean rookeries from a genomic perspective, to elucidate connectivity and adaptation patterns, and to assess effective population sizes to improve conservation decision-making. We hypothesize that if both males and females are philopatric we will detect strong genetic differentiation among rookeries. However, if male mediated geneflow is acting among geographically distant sites as previously suggested, increasing the number of markers will not result in significant genetic differentiation among them. Furthermore, considering the potential vulnerability of this species to global warming, we expect a strong genomic signal associated to temperature in Mediterranean nesting populations. In this study we had the following specific objectives: i) assess the diversity and effective population size of Mediterranean rookeries, ii) identify population differentiation between rookeries, and iii) understand the role of environmental factors shaping population genomic structure and local adaptation. This study will serve as a baseline for future studies on sea turtle conservation.

Methods

Sampling

Tissue samples of 199 dead loggerhead hatchlings were collected from 9 rookeries across the Eastern Mediterranean basin (Figure 1). For Sirte, El Mansouri and Akamas, we reanalysed samples from Clusa *et al.*, (2018), for Belek and Dalyan samples from Yilmaz *et al*., (2011) and for Kyparissia samples from Carreras *et al*., (2014). The three remaining sites were sampled between 2016 and 2018 (Table 1). To avoid pseudoreplication (e.g. sampling hatchlings from two nests of the same female), sampled females were tagged with external flipper tags or subcutaneous PIT tags. When this was not possible, samples from different clutches were considered to be from different females if laid within a 14-day window, or in two consecutive nesting seasons, as in previous studies (Carreras *et al*., 2007). One hatchling per nest was sampled and the tissue was stored in 96% ethanol.

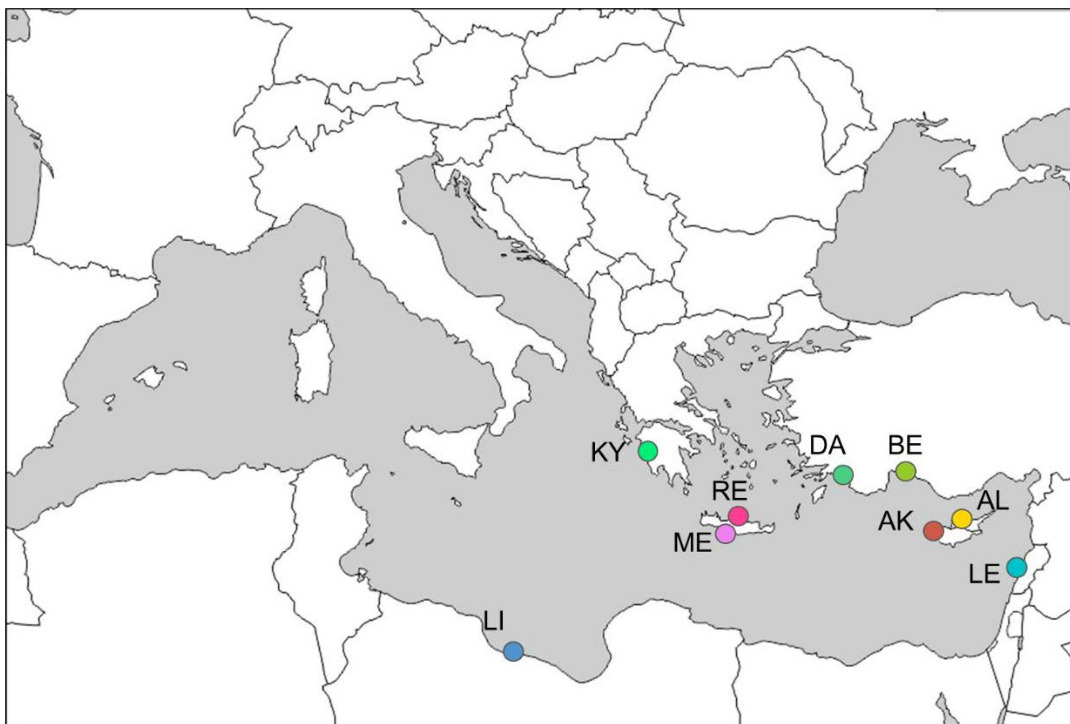


Figure 1. Rookeries sampled in the Eastern Mediterranean basin. The map shows the location of the study sites. Colours are the same as in the further analyses. LI = Sirte, LE = El Mansouri, AL = Alagadi, AK = Akamas, DA = Dalyan, BE = Belek, RE = Rethymno, ME = Messara and KY = Kyparissia.

Table 1. Basic statistics per rookery. Each population was analysed separately to maximise the number of markers used (Nomiss datasets). Columns show study site, code for this study, year of sampling, number of samples, number of loci, observed and expected heterozygosity, and coefficient of inbreeding.

Site	Code	Sampling year	n	Loci	Ho	He	F_{IS}
Sirte, Libya	LI	2005-2006	23	7455	0.245	0.237	-0.035
El Mansouri, Lebanon	LE	2004-2006	19	8393	0.261	0.250	-0.044
Alagadi, Northern Cyprus	AL	2016-2017	25	4706	0.218	0.228	0.043
Akamas, Cyprus	AK	2005	25	8100	0.235	0.230	-0.020
Belek, Turkey	BE	2007-2008	24	8145	0.241	0.232	-0.038
Dalyan, Turkey	DA	2004-2005	24	7575	0.244	0.236	-0.033
Rethymno, Greece	RE	2018	23	6493	0.243	0.230	-0.056
Messara, Greece	ME	2018	11	8637	0.300	0.289	-0.036
Kyparrissia, Greece	KY	2012	25	9821	0.246	0.239	-0.027

DNA extraction and library building

Genomic DNA was extracted using Qiagen® Gentra Puregene blood cell kit adjusting the manufacturer's protocol to our samples. DNA concentration was measured with Nanodrop® or PicoGreen®, and DNA degradation assessed in 1% agarose gels. 2b-RAD libraries were prepared adjusting the protocol from Wang *et al*., (2012) as in Barbanti *et al*., (2020). We performed DNA digestion using AlfI enzyme and customised adaptors were then attached to the digested sequences with sticky ends 5'-WN-3'. A previous pilot study simulating selective base ligation in 2b-RAD protocols showed that reducing the number of sequences simulating selective base ligation (W = A or T) results in the same genetic differentiation as using fully degenerated bases, allowing a more cost-effective sequencing (Barbanti *et al*., 2020). Therefore, these adaptors were used to select and sequence only a target portion of all the possible sequences across the genome, allowing balancing mean depth of coverage, number of final markers and number of samples in an Illumina plate. Barcodes and Illumina primers were attached to the adaptors, sequences were amplified by PCR and then purified using magnetic beads to remove primers and sequences longer and shorter than 165 bp. The DNA concentration of purified libraries was quantified using PicoGreen®. Libraries were pooled so that no more than 48 samples were sequenced in the same lane, to ensure good depth of coverage for our data. We performed single read 50bp sequencing per lane with a HiSeq 2500 Illumina at the Center for Genomic Regulations (CRG) of Barcelona.

Genotyping and filtering

Raw sequences were processed using customised scripts (Barbanti *et al*., 2020). First, sequences were trimmed to eliminate ligation adaptors and then cut down to 34bp. Trimmed sequences were used for genotyping using the STACKS v1.47 pipeline (Catchen *et al*., 2011; Catchen *et al*., 2013). To construct a locus catalogue we used the Stacks function denovo_map.pl, setting the following parameters: a minimum depth of 2 reads to consider a stack within an individual (m = 2), up to 3 mismatches allowed between stacks (putative alleles) to merge them into a putative locus within an individual (M = 3), and 4 mismatches allowed between stacks between individuals (n = 4). These parameters were optimised for our species following instructions as in Paris *et al*., (2017), using data of two populations with high number of reads per sample (i.e. Akamas and Kyparissia) (Supplementary Table S1). Individual genotypes were outputted as SNP (considering only the first SNP for each locus). Loci were filtered for minimum depth of 5 reads, minimum number of alleles = 2 and maximum minimum allele frequency = 2 (thus keeping only polymorphic loci). Loci with outlier values of mean depth across all individuals (above the upper whisker of the mean depth value boxplot, corresponding to the 95% CI) were removed as could be potential paralog DNA regions. Filtering was performed with VCFtools v1.12 (Danecek *et al*., 2011). We produced three datasets for statistical and bioinformatic analyses, which differ on the restrictiveness of missing data filters. First, we considered all populations together, filtering for 70% of missing data across all individuals (Base dataset). Since we detected high number of regional missing data related to the population of origin (see results), we produced a second type of dataset by treating each population separately and filtering for no missing data, as certain population analyses do not allow missing loci in the dataset (Nomiss datasets, one per population). Finally, for comparisons across rookeries we built a third dataset filtering all populations for 70% missing data within and across populations (HQ dataset).

Population diversity and effective population size

Within population analyses were carried out with the Nomiss datasets, that includes one dataset specific for each rookery. We computed basic genetic statistics for each population such as observed and expected heterozygosities (Ho and He) and inbreeding coefficient (F_{IS}) using VCFtools v1.12 (Danecek *et al*., 2011). We estimated the effective population size (*Ne*) of our populations using the function ldne from the R package 'strataG' (Archer *et al*., 2017). The R function ldne estimates *Ne* from linkage disequilibrium (LD) with a Pearson correlation approximation. The use of 'background' LD allows to estimate *Ne* using only one sample per population (instead of two or more temporally separated samples), but when using thousands of loci, physical LD can create bias to estimates such as Ne. for this reason this function follows

Waples *et al*., (2016) correction, which improves the *Ne* estimate precision by accounting for physical LD bias. We used the Nomiss datasets as the function used to calculate *Ne* requires no missing data. To obtain estimates number of adults (N_a) we gathered information on nests counts and population trends from the literature (Casale *et al*., 2018). Nest counts were then transformed to number of adults following the formula of Ad = nr/pd (Casale and Heppell 2016), where ***n*** is the number of nests, ***r*** is the remigration interval (years between consecutive nesting seasons for a given female) ***p*** is the proportion of females in the population and ***d*** is the number of nests per female and season. Values for these parameters were taken from the literature (r=2.3; p=0.4 and d=1.9, Casale and Heppell 2016).

Population genetic differentiation

All population structure analyses were performed using the HQ dataset for population comparison. We calculated pairwise genetic distances between individuals using Prevosti distance with the R function prevosti.dist from the package 'poppr' 2.8.0 (Kamvar *et al*., 2014; Kamvar *et al*, 2015). Using R package 'ape' (Paradis *et al*., 2004) we computed a Principal Coordinate Analysis (PCoA) to visualise the clustering of the studied populations based on prevosti distances and plotted the results using 'ggplot2' (Wickham, 2011). This analysis allows to visualise population structure at individual level, to identify the presence of clustering or random distribution, and to assess whether individuals cluster within the rookery they were collected from. Pairwise Fst values between populations were computed using Arlequin program (Excoffier and Lischer 2010). We analyzed the number of genetically distinct groups using the Bayesian assignment software STRUCTURE v2.3.4 (Pritchard *et al*., 2000). STRUCTURE assigns proportions of within samples genetic diversity to a priori selected number of genetic clusters (K). We performed 20 repetitions of each independent K value from 1 to 12; burn-in length was set to 50.000 MCMC steps and runs with 200.000 steps. We calculated the log probability of the data, LnP(K) and the rate of change in second-order derivates of the log probability between successive K values (IncK) (Evanno *et al*., 2005) with the aid of STRUCTURE HARVESTER (Earl, 2012). The 20 replicates per the most likely values of K were averaged using CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007). We carried out the analysis of Isolation by distance (IBD) to detect the correlation between geographic and genetic distance based on outlier markers. Euclidean geographic distance between sites was computed using the function distm from the R package 'geosphere' (Hijmans *et al*., 2017). We also ran the analysis using coastline distance, which was measured using Google maps (https://www.google.com/maps) (Supplementary Figure S1). We used mantel.randtest function from R package 'adegenet', using pairwise F_{ST} distances between individuals as genetic distances.

Detection of outlier loci

To identify candidate loci under selection in our dataset we used pcadapt function from R package 'pcadapt' (Luu *et al*., 2017). This package uses Bayesian factor models to determine both population structure and outlier loci. The pcadapt function assumes that markers excessively related to population structure are candidates for local adaptation. The number of principal components (PC)was selected by plotting the proportion of explained variance in relation to the number of PC. We ran this analysis using the HQ dataset. We calculated q-values using the qvalue function in 'qvalue' R package (Storey *et al*., 2020) which performs false discovery rate (FDR) estimation from a collection of p-values. We considered outliers all loci having a q-value lower than α = 0.1 (based on the distribution of p-values and q-values in test plots). We then divided our dataset in outlier and non-outlier markers and used R package 'ape' to perform a PCoA on both subsets separately to compare patterns of population structure. The analyses were plotted using 'ggplot2' package.

Influence of temperature on population structure

Atmospheric temperature was taken as a proxy of sand temperature, which directly influences the development of sea turtle embryos and determines their sex during incubation. We obtained data for monthly means of atmospheric temperature from 1948 to present (0.5° x 0.5° grid) from the NOAA Earth System Research Laboratory (ESRL) database (https://www.esrl.noaa.gov/). We considered data from such a long period of time so that the variable could be reliable as a possible impact factor on sea turtles' life cycle, considering the length of their life. We extracted minimum, maximum and mean temperatures for each one of our sampling sites, considering temperatures within a 10km distance ray for each site using R packages 'RANN' (Kemp and Jefferis, 2012) and 'ncdf4' (Pierce and Pierce, 2019). We collected approximate latitude and longitude of the sampling sites and tested the level of correlation among these two and the three temperature variables using linear regressions. We discarded variables that were significantly correlated (i.e. mean temperature) to avoid redundant data which could skew the analysis. To perform a Redundancy Analysis (RDA) we used the rda function of the package 'vegan' (Oksanen *et al*., 2007) and then plot the results using 'ggplot2' package in R. We considered as predictors only variables not highly correlated to each other (i.e. longitude, latitude, minimum and maximum temperature) and identified candidate outlier loci associated with the axes and significantly correlated to the predictors. All loci identified by RDA were also found by pcadapt (see results). We also performed a partial mantel test using mantel.partial function in R package 'vegan' to assess significant correlation between outlier markers found by pcadapt (n = 35) and mean temperature, removing geographic dependence. Here we used mean temperature because by

being significantly correlated to both minimum and maximum temperature, it was the best representation for our temperature data, and we performed analysis using both Euclidean and coastline geographic distance. Genetic distance was calculated using pairwise F_{ST} of outlier markers.

Results

Diversity

Illumina sequencing resulted in an average number of 6.1 million raw reads per individual and 5.3 million trimmed reads per individual. Genotyping provided more than 80 thousand genotype loci before filtering. After filtering the Base dataset counted 3685 loci and the mean depth of coverage was 22.21 sequences per locus. Given that a population-related missing data was found in this dataset (Supplementary Figure S2), within population analyses were carried out with the Nomiss datasets for each specific site. The number of polymorphic loci ranged between 4706-9821 with mean depth sequences per locus of 30.79 (Table 1).

The effective population size (*Ne*) of our populations ranged from 84 to 4754 individuals as shown in Table 2. The estimated effective population size was significantly correlated to the number of adults (Spearman rho, p=0.037, R=0.72) and was not correlated to the number of loci per population (Spearman rho, p=0.437, R=0.3). The populations of Belek and Dalyan yielded an effective population size greater than the adult census size (N_a) (Figure 2) while the remaining populations always yielded effective population sizes smaller than N_a (Table 2).

Table 2. Effective population size calculated with strataG R package using Nomiss datasets. Columns show the sampling site, effective population size and 95% confidence interval bounds, number of nests at two temporal periods (before 1999 and after 2000), adult census size (N_a) and the population trend calculated as the percentage of change in the number of nests. NA: data not available. Data on nest/year and population on the two temporal periods and the population trend is from Casale *et al*., (2018). N_a was calculated using the approach from Casale and Heppel 2016, as decribed in the methods.

Site	ldNe	Lower CI	Upper CI	Nests<1999	Nests >2000	N_a	Trend (%)
Sirte	432.4	411.8	455.1	NA	220	665.8	NA
El Mansouri	133.4	130.9	136	NA	55	166.4	NA
Alagadi	84.3	83.1	85.6	65.7	54.1	163.7	-17.7
Akamas	535.9	509.2	565.5	119.8	239.1	723.6	99.6
Belek	3585.7	2636.5	5598	129.7	638	1930.8	391.9
Dalyan	4754.6	3152.7	9649.8	165	269	814.1	63
Rethymno	110	108.2	111.9	387.3	275	832.2	-29
Messara	98	95.1	101	53.5	46.9	141.9	-12
Kyparissia	1670.1	1468.5	1935.5	580.7	987	2987	70

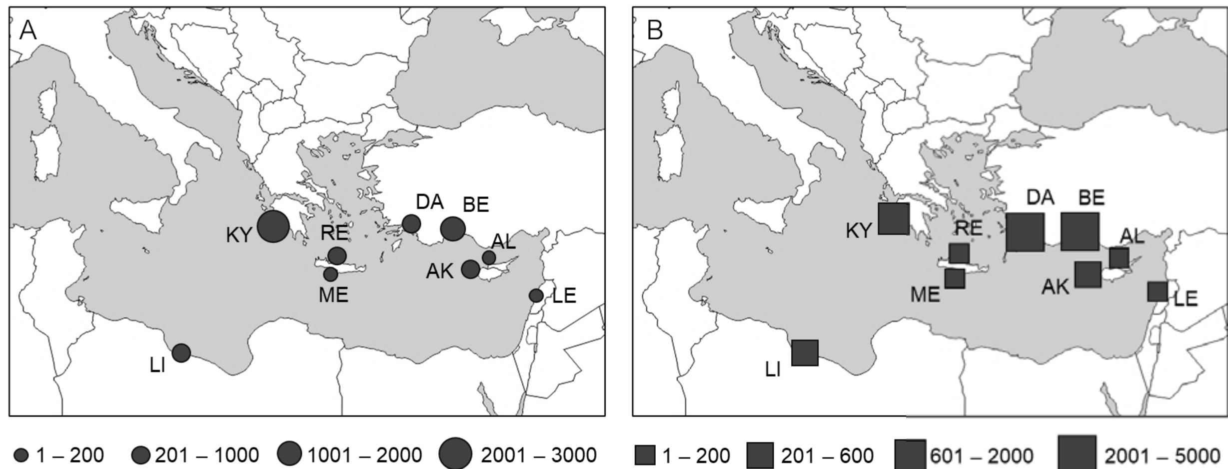


Figure 2. Population size estimates for Eastern Mediterranean sampled rookeries. A) adult census size (N_a) calculated based on nests counts from the literature (Casale *et al*., 2018). B) effective population sizes computed in this study based on genomic data.

Population structure

The HQ dataset resulted in 195 SNPs shared at least by 70% of the individuals within each locality and a mean depth of coverage of 22.10 and constitutes a high-quality dataset to carry out analyses sensible to missing data. Pairwise F_{ST} analyses showed that all populations are significantly different from each other (Supplementary Table S2). The Principal Coordinate Analysis based on Prevosti distances between pairs of individuals showed high level of genetic differentiation among the studied populations, with little overlap of hatchlings from different rookeries (Figure 3A). The first axis grouped Alagadi, Lebanon and Libya on one side and Kyparissia and Akamas on the opposite site. Belek and Dalyan were the most separated populations according to the second axis. Messara and Rethymno presented and intermediate location in the plot. We found that 5 individuals from Alagadi were near to individuals from Messara, as well as one individual of Libya and one of Rethymno. Bayesian clustering using STRUCTURE revealed that the highest L(K) value was for K=6. Similar results were found using the ΔK statistic (Evanno *et al*., 2005) with a clear peak on K=6 (Supplementary Figure S3). We plotted clump results for the 20 runs of the best K (Supplementary Figure S4) finding Alagadi grouped with Sirte and El Mansouri, except for five individuals which showed the same genetic clusters as Messara. Kyparissia shared its cluster with Akamas, which also showed the same clusters ass Alagadi and Messara. Isolation by distance analysis did not show significant correlation between genetic and geographic distance, neither expressed as Euclidean distance (simulated p-value = 0.761) nor as coastline distance (simulated p-value = 0.572).

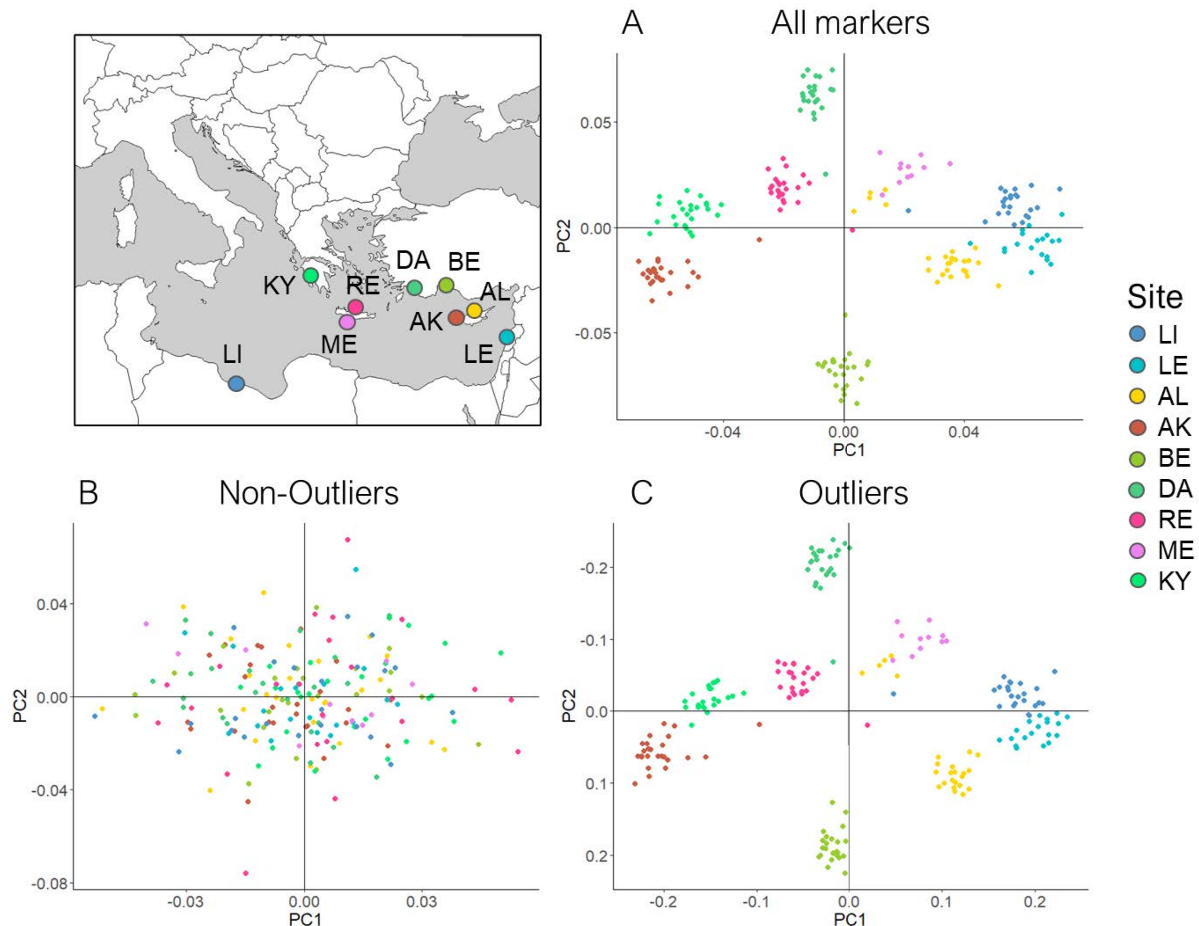


Figure 3. Reference map and Principal Coordinates Analysis. The map gives reference of the geographic location of the studied rookeries to be compared with PCoA distribution of individuals. Acronyms are as in Figure 1. A) shows PCoA for the whole HQ dataset (n=195), B) non-outlier markers only (n=160) and C) outlier markers only (n=35). A) and C) graphs show high level of population structure for our populations, while the distribution of individuals using non-outlier markers only is random and does not show any sign of clustering.

Local adaptation

We found 35 outlier loci using pcadapt and 16 of them were also candidate outlier loci significantly associated with the axes of the RDA. Each locus was significantly correlated with a different predictor (Figure 4). The Redundancy Analysis showed significant differentiation between populations under the effect of atmospheric temperature and geographic location (latitude and longitude). Our first RDA explained 40.27% of variance while the second RDA explained 26.06% of the remaining variance (Figure 4). Populations are distributed similarly as in the PCoA, with Sirte and El Mansouri on one side of the first axes and Akamas and Kyparissia on the other side. Belek is again isolated from the rest of populations, while Dalyan is now in between the two populations from Crete. The PCoA of non-outlier markers does not show any signal of population differentiation with random distribution and overlapping of individuals (Figure 3B). As expected,

the plot of outlier loci has a strong population structure showing the same pattern as for the whole dataset (Figure 3C).

The partial mantel test showed significant correlation between genetic distance and temperature removing the effect of geographic location with both Euclidean distance (r = 0.438, significance = 0.032) and coastline distance (r = 0.443, significance = 0.039).

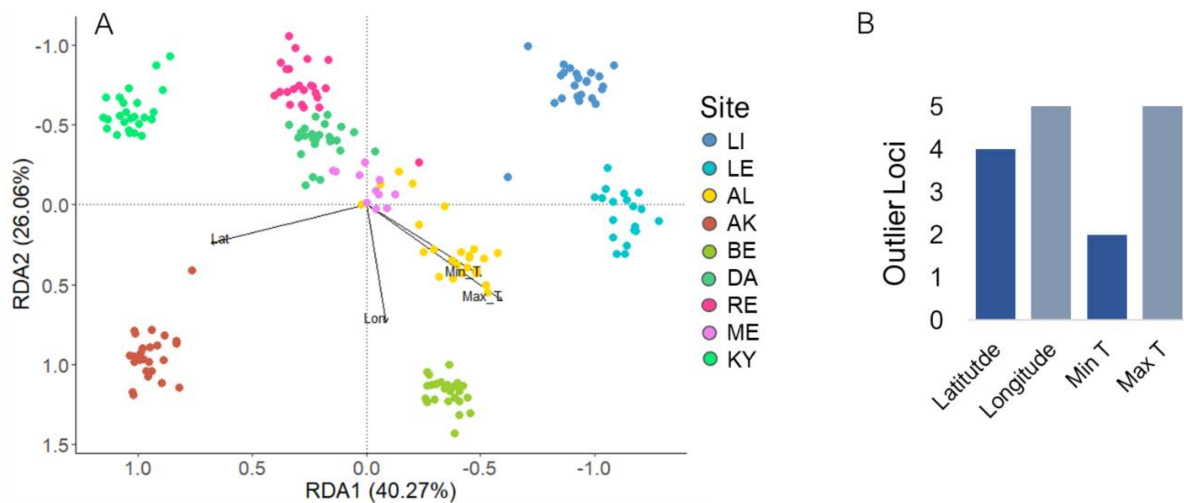


Figure 4. Redundancy analysis. A) shows the effect of environmental and geographic factors on genetic differentiation of studied populations. B) shows the number of outlier loci associated with RDA axes and correlated to the predictors of this analysis.

Discussion

Loggerhead turtle populations of the Mediterranean Sea have been recovering in the past years after the crucial intervention of conservation management on nesting beaches (Casale, 2015). Although individuals of this species are now increasing in numbers in this area, the level of anthropogenic impact, mostly caused by the fishing industry and by tourism, is still a major cause of mortality for adults and juveniles. Global warming is also predicted to have a great impact on Mediterranean nesting areas. For all these reasons, the species is still considered to be conservation dependent (Casale, 2015) as conservation actions and scientific studies are both essential to provide crucial information for conservation managers and governments to make optimal decisions for the health and longevity of these loggerhead populations. This study is the first to use population genomics on marine turtles to estimate the effective population size of eastern Mediterranean loggerhead rookeries, finding results that differ from the current adult census size used for IUCN red listing. We also refined the genetic structure of these populations, finding high genetic differentiation among all populations and low level of overlap of individuals from different rookeries. Finally, we tested for the influence of temperature on genetic structure, finding that this parameter has a significant impact on the differentiation of the studied populations.

Genotyping and missing data for population barcoding

The advent of NGS allowed to exponentially increase the power of genetic analyses (Andrews and Luikart, 2014). Using 2b-RAD methodology we were able to score thousands of markers in a species, the loggerhead turtle, that until now had only been analysed using a handful of nuclear markers. When filtering our dataset so that all loci would be present in 70% of individuals (Base dataset) we could still run analysis with more than three thousand high quality markers. When analysing our populations of interest separately we could still score between four and nine thousand loci present in 100% of individuals per dataset. These results show the power of this methodology in finding high quality markers for population genomic studies. Although the high number of markers found in the base dataset, we detected a pattern of missing loci which was not related to sequencing depth nor to plate related issues. The mean depth of coverage of the base dataset was in fact 22.21 which is a reliable value to support further population analyses as widely shown in previous studies (Resh *et al*., 2018; Whelan *et al*., 2019). The pattern of missing loci also does not reflect a plate related issue since populations sequenced in the same plate show different missing markers. Therefore, we suggest that the pattern of missing markers is actually population driven and could be informative to understand the heritage of the studied population. These regional missing loci can in fact be phylogenetically informative (Eaton *et al*.,

2017) as they can be the result of historical processes, mutations appeared at a certain locality that did not spread to others, strong regionally based genetic drift or selection (Carreras *et al*., 2020). In most cases in fact, the frequency of our markers does not follow a cline or a progressive gradient, but it stands at either sides of the spectrum, being fully present or fully absent in each population. This pattern can be used to interpret the population structure of an area and resolves in the concept of population barcoding, which relies on the visualisation of markers expressed as presence/absence for each studied population. Although for the sake of this study further analyses were performed using a dataset with reduced missing data (HQ dataset), regional missing loci could be exceptionally useful in future analyses involving population assignation of individuals. In the case of loggerhead turtle in fact, the identification of population of origin for individuals found in feeding grounds can be facilitated by this high diversification of marker frequencies among Mediterranean rookeries.

Effective population size and conservation

Understanding the effective population size of wild populations is critical to assess their status and to plan conservation measures. However, this parameter is rarely used for conservation purposes and adult census size (N_a) (of the whole population or of breeding individuals) is normally used for risk assessment (IUCN, 2012). Here we provided esteems of effective population sizes based on thousands of genome wide markers. Considering our results, the Turkish populations of Belek and Dalyan, followed by Kyparissia, had the highest effective population sizes, meaning that they are less vulnerable to detrimental genetic effects related to population sizes, such as the loss of alleles due to genetic drift. The Turkish populations are the only ones that exhibit an effective population size larger than adult census size (Figure 2). This result reflects that these populations have suffered a relatively recent reduction, and thus the effective population size is the reflection of genetic variability inherited from a past abundance. Individuals from Turkish populations use mainly the foraging areas of the Mediterranean Levantine (Casale and Mariani 2014; Clusa *et al*., 2014), an area which suffered an intense harvest of sea turtles for edible consumption in the XX century until 1970s (Hornell, 1935; Sella, 1982). Thus, these two populations were likely much larger than the present populations, suffered a decline at the end of the XX century, and are now recovering after intense conservation efforts (Table 2). On the contrary, Alagadi and the Island of Crete (Rethymno and Messara) populations showed the smallest effective sizes and are therefore more vulnerable to genetic drift. In the case of Alagadi and Messara, the low effective population size is clearly related to a low N_a, but not in the case of Rethymno, as this site hosts one of the largest populations in terms of nests per season.

This population exhibits the lowest relationship between effective population size and census of breeders of all our dataset, even though this population has showed signs of decrease in the past decades (Table 2). According to mtDNA haplotype composition, the populations of the two Crete and Cyprus have been proposed as colonised more recently in comparison of the populations of Greece, Libya and Turkey, and thus the lower effective population sizes can be the result of a more recent origin (Clusa *et al*., 2013).

Although we found a significant correlation between the effective population size and the estimated number of breeders, the population by population analysis showed that the relationship between these two parameters can be very variable. The parameters used to transform the nest counts into number of adults may vary across populations and thus the estimation of adult census sizes should be regarded with caution (Matsinos *et al*., 2008). However, these potential variations can hardly explain the wide range of variance in the ratio between effective population size and adult census size (from 0.13 to 5.84, Table 2). Consequently, although the adult census size is globally related to the effective size, it cannot be used for population risk assessment. For this reason, we recommend the use of effective population size based on a genome wide panel of markers, as a complementary measure of population size for conservation purposes. Previous studies suggest that an effective population of 500 is necessary to maintain equilibrium between loss of adaptative genetic variation due to genetic drift and its replacement by mutation (Franklin, 1980, Franklin and Frankham, 1998) and that the target *Ne* for conservation programs of endangered species should range from 500 to 1000 (Lynch and Lande, 1998). Although these estimates likely depend on the target species, previous studies on sea turtles considered populations with *Ne* of over a thousand breeders as healthy (Theissinger *et al*., 2009) and *Ne* ranging between 90-220 individuals as sign of vulnerability (Rivalan *et al*., 2006). Based on these different classes of extinction risk, our results support the management of Mediterranean rookeries as different units, given the range of *Ne* presented by the studied populations.

Genetic structure

We found that rookeries in the Eastern Mediterranean are more genetically isolated than previously found in studies based on microsatellites (Clusa *et al.,* 2018). This result is in agreement with the observation that increasing the number of loci improves the ability of identifying genetic differentiation (from Carreras *et al*., 2007; to Clusa *et al.,* 2018) and that the power of the genetic markers used is crucial to infer population genetic structuring (Bradshaw *et al.*, 2018). Previous studies suggested seven different units for management and conservation using mtDNA (Shamblin *et al*., 2014) and five units using microsatellites (Clusa *et al*., 2018). In our study we can conclude that at last 9 units should be considered, as every sampled population

is significatively different from the others. This highlights the importance of using markers with enhanced resolution for management and conservation and specifically the potential of genomics studies for delineating conservation units. However, this number may increase, as some of the populations considered isolated in previous studies (such as Calabria, Garofalo *et al*., 2013) have not been included in this study. As a consequence of this enhanced resolution, our results suggest that the degree of philopatry of both males and females in these populations is very strong. Previous studies (Clusa *et al*., 2018) suggested that the populations of Libya and Akamas belonged to the same genetic cluster despite being geographically distant. The authors suggested same male mediated gene flow in foraging areas or while migrating. These two populations were included in our study by using the same samples than in Clusa *et al*., 2018 and we found that they are genetically different. Thus, we can conclude that the lack of differentiation previously found between these two populations is due to a lack of resolution rather than male mediated gene flow. Another interesting result comes from the comparison between Kyparissia, Rethymno and Messara. Greek populations, including those on the island of Crete, have been postulated to be panmictic both in terms of mtDNA (Carreras *et al*., 2014) and nuclear DNA (Clusa *et al*., 2018). Despite the fact that we have no data for some of the populations analysed in previous studies (Lakonikos and Zakynthos), the genetic differentiation found between Kyparissia, Rethimno and Messara suggests the existence of at least three differentiated populations. In summary, our results indicate that on top of female's strong nest fidelity, males also present a strong degree of natal homing and probably most of the mating would happen in the areas surrounding nesting beaches. This extreme philopatry, coupled with reduced effective population sizes, would favour differentiation among populations due to genetic drift. Nevertheless, the misplacement of some individuals in the PCOAs and the results of STRUCTURE indicate that some very low-level of gene flow could be present within the Eastern Mediterranean to provide enough genetic variability to avoid inbreeding depression and the collapse of each rookery. We hypothesis that the individuals from Alagadi showing the same genetic clusters as Messara could in fact be descendants of individuals belonging to the Messara rookery, misplacing their nests. It is also possible though that these individuals were the result of the mating between Messara and Alagadi indviduals, since Messara is on the way from Alagadi to Libyan foraging grounds (Haywood *et al*., 2020).

The deep genetic structuring found in the Mediterranean populations has also implications on the study of foraging grounds. The assignation of individuals at sea to the populations of origin has been neglected in this species due to a lack of resolution, and thus could only be analysed at regional level (Carreras *et al*., 2011). The use of genomics is thus very promising for individual

assignations to assess the populations of origin of individuals incidentally captured by fisheries (Clusa *et al*., 2016) or individuals colonising new habitats (Carreras *et al*., 2018). Our results will provide the baseline for identifying the population of origin of individuals of this highly migratory species along its complex life cycle to adapt conservation strategies.

Temperature as driver for structuring

In addition to the extremely low degree of gene flow between the studied rookeries, temperature showed a significant impact on genetic differentiation. The effect of temperature on the genetic structuring is not an artifact of the geographical position of the nesting beaches, as shown by the partial mantel test and by the fact that the loci related to temperature were different from those related to either latitude or longitude in the redundancy analysis. Temperature is a crucial environmental factor during the development of the embryos, as fluctuations can bring to shifts in future sex ratio of adult breeders (Jensen *et al*., 2018) and can impact the viability and fertility of the clutch (Hawkes *et al*., 2007). As a result, it is not surprising to find potential adaptation genomic signals driven by temperature in marine turtles (Figure 3). The studied populations might have adapted to the different temperatures of their nesting areas, which results today in a strong genetic differentiation as found with the affected loci. Moreover, the fact that a higher number of markers was correlated with maximum than minimum temperature could be a sign that warmer climates have a stronger impact on sea turtles. In fact, the capacity to adapt to global warming could be crucial for the survival of these populations. A recent study has suggested the existence of sex-specific genotypes (Chow *et al*., 2019), stating that the coexistence of TSD and loci with sex specific genotypes may suggest that thermosensitivity has genetic basis and that certain genotypes may confer differential fitness benefits to the sexes.

Following this idea, the allele selected at population level may depend on the sex ratio, and therefore on incubation temperatures. Our results are a preliminary indication of the impact of temperature on population differentiation. Knowing which genes are affected, or how this genomic signal is related to the nesting environment is something that remains to be tested in future studies. We also found significant differentiation correlated to geographic locations of the rookeries. Several outlier markers candidate for adaptation were in fact correlated with latitude or longitude measurements of the sampling site. This can reflect how environmental factors linked with geographic gradients (either north-south or east-west) can have significant impact on local adaptation of each rookery to environmental conditions and therefore, enhance genetic differentiation. The understanding of nest adaptation to these changes could be of great help to the improvement of conservation measures, not only in the Mediterranean Sea but also globally and may aid the assessment of nest success in new colonised areas.

Conclusions

In summary, our findings allowed for a better understanding of the status and population dynamics of loggerhead turtles in the Eastern Mediterranean and revealed significant genetic differentiation among rookeries. Both males and females need to show a strong philopatric behaviour in order to create the observed population structure in this area. Furthermore, maximum temperature is a clear driver for genetic structuring in Mediterranean populations, highlighting the potential vulnerability of this species to global warming. Our results set the baseline for evaluating future conservation measures since the pressure loggerhead turtles face in the Mediterranean basin goes beyond the conservation of nesting beaches. For this reason, future population genomic studies should focus on other high-risk areas such as feeding grounds, to understand the population of origin of turtles congregating there and set appropriate conservation measures. This study shows that genomics represents a step forward in the field of population genetics, and the results can be highly beneficial to improve conservation management of endangered species. These methods not only apply to loggerhead turtles but can be used with all sea turtle species and more broadly transferred to any complex highly migratory species. We therefore suggest the application of genomic analysis for future studies focusing on marine organisms' population and conservation genetics, and, in specific cases, we suggest to consider reanalysing data published using traditional genetic markers, since the power of genomics could unravel past misleading results.

Acknowledgments

This work was supported by the project CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER). CC and MP are part of the research group SGR2017-1120. AB was supported by grant 2017 FI_B 00997 of the Catalan Government-AGAUR.

References

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. Molecular ecology, 23(7), 1661-1667.

Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources, 17(1), 5-11.

Barbanti, A., Torrado, H., Macpherson, E., Bargelloni, L., Franch, R., Carreras, C., & Pascual, M. (2020). Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. Molecular Ecology Resources.

Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., ... & Fraschetti, S. (2019). A population genomics insight by 2b-RAD reveals populations' uniqueness along the Italian coastline in *Leptopsammia pruvoti* (Scleractinia, Dendrophylliidae). Diversity and Distributions, 25(7), 1101-1117.

Bowen, B. W., Bass, A. L., Soares, L., & Toonen, R. J. (2005). Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). Molecular ecology, 14(8), 2389-2402.

Bradshaw, P. J., Broderick, A. C., Carreras, C., Fuller, W., Snape, R. T. E., Wright, L. I., & Godley, B. J. (2018). Defining conservation units with enhanced molecular tools to reveal fine scale structuring among Mediterranean green turtle rookeries. Biological Conservation, 222, 253-260.

Carreras, C., Pascual, M., Cardona, L., Aguilar, A., Margaritoulis, D., Rees, A., ... & Khalil, M. (2007). The genetic structure of the loggerhead sea turtle (*Caretta caretta*) in the Mediterranean as revealed by nuclear and mitochondrial DNA and its conservation implications. Conservation Genetics, 8(4), 761-775.

Carreras, C., Pascual, M., Cardona, L., Marco, A., Bellido, J. J., Castillo, J. J., ... & Aguilar, A. (2011). Living together but remaining apart: Atlantic and Mediterranean loggerhead sea turtles (*Caretta caretta*) in shared feeding grounds. Journal of Heredity, 102(6), 666-677.

Carreras, C., Rees, A. F., Broderick, A. C., Godley, B. J., & Margaritoulis, D. (2014). Mitochondrial DNA markers of loggerhead marine turtles (*Caretta caretta*)(Testudines: Cheloniidae) nesting at Kyparissia Bay, Greece, confirm the western Greece unit and regional structuring. Scientia Marina, 78(1), 115-124.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. Scientific Reports, 7(1), 1-12.

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., ... & Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). Scientific reports, 8(1), 1435.

Carreras, C., García-Cisneros, A., Wangensteen, O. S., Ordóñez, V., Palacín, C., Pascual, M., & Turon, X. (2020). East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). Diversity and Distributions, 26(3), 382-398.

Casale, P. (2015). Caretta caretta (Mediterranean subpopulation). The IUCN Red List of Threatened Species, 2015-4.

Casale, P., & Mariani, P. (2014). The first 'lost year'of Mediterranean sea turtles: dispersal patterns indicate subregional management units for conservation. Marine Ecology Progress Series, 498, 263-274.

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.

Casale, P., & Heppell, S. S. (2016). How much sea turtle bycatch is too much? A stationary age distribution model for simulating population abundance and potential biological removal in the Mediterranean. Endangered Species Research, 29(3), 239-254.

Casale, P., Broderick, A. C., Camiñas, J. A., Cardona, L., Carreras, C., Demetropoulos, A., ... & Lazar, B. (2018). Mediterranean sea turtles: current knowledge and priorities for conservation and research. Endangered species research, 36, 229-267.

Casso, M., Turon, X., & Pascual, M. (2019). Single zooids, multiple loci: independent colonisations revealed by population genomics of a global invader. Biological Invasions, 21(12), 3575-3592.

Clusa, M., Carreras, C., Pascual, M., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... & Türkozan, O. (2013). Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). Journal of experimental marine biology and ecology, 439, 15-24.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Giacoma, C., ... & Maffucci, F. (2014). Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine biology, 161(3), 509-519.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Avolio, D., … Cardona, L. (2016). Potential bycatch impact on distinct sea turtle populations is dependent on fishing ground rather than gear type in the Mediterranean Sea. Marine Biology, 163(5). https://doi.org/122 10.1007/s00227-016-2875-1

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... & Aguilar, A. (2018). Philopatry in loggerhead turtles *Caretta caretta*: beyond the gender paradigm. Marine Ecology Progress Series, 588, 201-213.

Dietz, S., & Adger, W. N. (2003). Economic growth, biodiversity loss and conservation effort. Journal of environmental management, 68(1), 23-35.

Eaton, D. A., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology, 66(3), 399-412.

Excoffier, L., & Lischer, H. E. (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.

Franklin. I. R., (1980). Evolutionary changes in small populations. Conservation Biology: An Evolutionary-Ecological Prospective. M. Soulé, B.A. Wilcox (Eds.), Sinauer Associates, Sunderland, MA, pp. 135-150

Franklin, I. R., & Frankham, R. (1998). How large must populations be to retain evolutionary potential?. Animal conservation, 1(1), 69-70.

Garofalo, L., Mastrogiacomo, A., Casale, P., Carlini, R., Eleni, C., Freggi, D., ... & Novarini, N. (2013). Genetic characterization of central Mediterranean stocks of the loggerhead turtle (*Caretta caretta*) using mitochondrial and nuclear markers, and conservation implications. Aquatic Conservation: Marine and Freshwater Ecosystems, 23(6), 868-884.

Harris, L. R., Bessinger, M., Dayaram, A., Holness, S., Kirkman, S., Livingstone, T. C., ... & Skowno, A. L. (2019). Advancing land-sea integration for ecologically meaningful coastal conservation and management. Biological Conservation, 237, 81-89.

Hawkes, L., Broderick, A., Godfrey, M. & Godley, B. (2007). Investigating the potential impacts of climate change on a marine turtle population. Global Change Biology, 13(5), 923-932.

Hawkes, L. A., Broderick, A. C., Godfrey, M. H., & Godley, B. J. (2009). Climate change and marine turtles. Endangered Species Research, 7(2), 137-154.

Haywood, J. C., Fuller, W. J., Godley, B. J., Margaritoulis, D., Shutler, J. D., Snape, R. T., ... & Broderick, A. C. (2020). Spatial ecology of loggerhead turtles: Insights from stable isotope markers and satellite telemetry. Diversity and Distributions, 26(3), 368-381.

Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2017). Package ‘geosphere’. Spherical trigonometry, 1, 7.

Hornell, J. (1935). Report on the fisheries of Palestine.

IUCN. (2012). IUCN Red List Categories and Criteria: Version 3.1. Second edition. Gland, Switzerland and Cambridge, UK: IUCN. iv + 32pp.

IUCN (2019). The IUCN Red List of Threatened Species. Version 2019-3. http://www.iucnredlist.org. Downloaded on 10 December 2019.

Janzen, F. J. & Paukstis, G. L. (1991). Environmental sex determination in reptiles: ecology, evolution, and experimental design. The Quarterly review of biology, 66(2), 149-179.4

Jensen, M. P., Allen, C. D., Eguchi, T., Bell, I. P., LaCasella, E. L., Hilton, W. A., ... 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. e154.

Kemp, S. E., & Jefferis, G. (2012). Package 'RANN'.

Limpus, C. J., Fleay, A. & Baker, V. (1984). The flatback turtle, *Chelonia depressa*, in Queensland: reproductive periodicity, philopatry and recruitment. Wildlife Research, 11(3), 579-587

Lohmann, K. J., Lohmann, C. M., Brothers, J. R., & Putman, N. F. (2013). Natal homing and imprinting in sea turtles. The biology of sea turtles, 3, 59-77.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular ecology resources, 17(1), 67-77.

Lynch, M., & Lande, R. (1998). The critical effective size for a genetically secure population. Animal Conservation, 1(1), 70-72.

Mansfield, K. L., Saba, V. S., Keinath, J. A., & Musick, J. A. (2009). Satellite tracking reveals a dichotomy in migration strategies among juvenile loggerhead turtles in the Northwest Atlantic. Marine Biology, 156(12), 2555-2570.

Matsinos, Y., Mazaris, A., & Pantis, J. (2008). Evaluating the effect of varying clutch frequency in nesting trend estimation of sea turtles. Amphibia-Reptilia, 29(3), 361-369.

Mrosovsky, N., Kamel, S., Rees, A. F., & Margaritoulis, D. (2002). Pivotal temperature for loggerhead turtles (*Caretta caretta*) from Kyparissia Bay, Greece. Canadian Journal of Zoology, 80(12), 2118-2124.

Nichols, W. J., Resendiz, A., Seminoff, J. A., & Resendiz, B. (2000). Transpacific migration of a loggerhead turtle monitored by satellite telemetry. Bulletin of Marine Science, 67(3), 937-947.

Nishizawa, H., Joseph, J., Chong, Y. K., Kadir, S. A. S., Isnain, I., Ganyai, T. A., ... Zhang, X. (2018). Comparison of the rookery connectivity and migratory connectivity: insight into movement and colonization of the green turtle (*Chelonia mydas*) in Pacific–Southeast Asia. Marine Biology, 165(4), 77.

Novelletto, A., Testa, L., Iacovelli, F., Blasi, P., Garofalo, L., Mingozzi, T., & Falconi, M. (2016). Polymorphism in mitochondrial coding regions of Mediterranean loggerhead turtles: evolutionary relevance and structural effects. Physiological and Biochemical Zoology, 89(6), 473-486.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. Community ecology package, 10, 631-637.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics, 20(2), 289-290.

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. Nature Precedings, 1-1.

Pierce, D., & Pierce, M. D. (2019). Package 'ncdf4'.

Pike, D. A. (2013). Climate influences the global distribution of sea turtle nesting. Global Ecology and Biogeography, 22(5), 555-566.

Pike, D. A. (2014). Forecasting the viability of sea turtle eggs in a warming world. Global change biology, 20(1), 7-15.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Resh, C. A., Galaska, M. P., & Mahon, A. R. (2018). Genomic analyses of Northern snakehead (Channa argus) populations in North America. PeerJ, 6, e4581.

Rivalan, P., Dutton, P. H., Baudry, E., Roden, S. E., & Girondot, M. (2006). Demographic scenario inferred from genetic data in leatherback turtles nesting in French Guiana and Suriname. Biological Conservation, 130(1), 1-9.

Sella, I. (1982). Sea turtles in the eastern Mediterranean and northern Red Sea. Biology and conservation of sea turtles, 4(1), 7-423.

Shamblin, B. M., Bolten, A. B., Abreu-Grobois, F. A., Bjorndal, K. A., Cardona, L., Carreras, C., ... & Nel, R. (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), e85956.

Storey JD, Bass AJ, Dabney A, Robinson D (2020). qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0, http://github.com/jdstorey/qvalue.

Theissinger, K., FitzSimmons, N. N., Limpus, C. J., Parmenter, C. J., & Phillott, A. D. (2009). Mating system, multiple paternity and effective population size in the endemic flatback turtle (*Natator depressus*) in Australia. Conservation Genetics, 10(2), 329-346.

Torrado, H.,, Carreras, C., Raventós, N., Macpherson, E., & Pascual, M. (In press) Individual-based population genomics reveal different drivers of adaptation in sympatric fish. Scientific Reports.

Wallace, B. P., DiMatteo, A. D., Hurley, B. J., Finkbeiner, E. M., Bolten, A. B., Chaloupka, M. Y., ... Bjorndal, K. A. (2010). Regional management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS One, 5(12), e15465.

Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, 117(4), 233-240.

Watanabe, J. I., Shao, Y., & Miura, N. (2019). Underwater and airborne monitoring of marine ecosystems and debris. Journal of Applied Remote Sensing, 13(4), 044509.

Whelan, N. V., Galaska, M. P., Sipley, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms, B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened Round Rocksnail, Leptoxis ampla. Molecular ecology, 28(7), 1593-1610.

Wickham, H. (2011). ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3(2), 180-185.

Witt, M. J., Hawkes, L. A., Godfrey, M. H., Godley, B. J., & Broderick, A. C. (2010). Predicting the impacts of climate change on a globally distributed species: the case of the loggerhead turtle. Journal of Experimental Biology, 213(6), 901-911.

Wright, S. (1938). Size of population and breeding structure in relation to evolution. Science, 87, 430-431.

Yilmaz, C., Turkozan, O., & Bardakci, F. (2011). Genetic structure of loggerhead turtle (*Caretta caretta*) populations in Turkey. Biochemical Systematics and Ecology, 39(4-6), 266-276.

Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., ... & Connors, S. (2018). Global Warming of 1.5 OC: An IPCC Special Report on the Impacts of Global Warming of 1.5° C Above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty (p. 32). V. Masson-Delmotte (Ed.). Geneva, Switzerland: World Meteorological Organization.

SUPPLEMENTARY INFORMATION

Table S1. Genoryping and filtering results from optimisation of STACKS parameters. The program was run analysing only Akamas and Kyparissia (n = 50) following instructions as in Paris et al (2017). The first table shows the optimisation for the minimum number of reads to consider a stack within an individual (m), the second table shows the optimisation for the number of mismatches allowed between stacks (putative alleles) within an individual to include them into a putative locus (M), and the third table shows the optimisation for number of mismatches allowed between stacks between individuals (n). In bold are values identified as optimum for our species. 'Full dataset' refers to the number of total loci in the dataset, and '80% of individuals' refers to the number of loci present in the 80% of individuals.

m	Full dataset	80% of individuals
1	17458	6472
2	**18202**	**6661**
3	16610	5957
4	15729	5604
5	15205	5467

m=2

M	Full dataset	80% of individuals
1	17194	6326
2	18202	6661
3	**18804**	**6892**

m=2
M=3

n	Full dataset	80% of individuals
2	19127	7035
3	19218	7173
4	**19528**	**7333**

Table S2. Pairwise FST computed using Arlequin. All pairwise comparisons are significantly different (p value < 0.000).

	LI	LE	AL	AK	BE	DA	RE	ME	KY
LI	0.000								
LE	0.118	0.000							
AL	0.166	0.125	0.000						
AK	0.273	0.286	0.215	0.000					
BE	0.211	0.181	0.214	0.226	0.000				
DA	0.180	0.199	0.204	0.207	0.238	0.000			
RE	0.216	0.245	0.183	0.203	0.249	0.194	0.000		
ME	0.172	0.180	0.163	0.243	0.261	0.147	0.243	0.000	
KY	0.231	0.260	0.241	0.198	0.225	0.194	0.212	0.240	0.000

Figure S1. Coastline considered for measurement between populations to calculate geographic distance. To calculate distance between Kyparissia and all other population except from Crete, we considered a mean value between northern and southern Crete path.

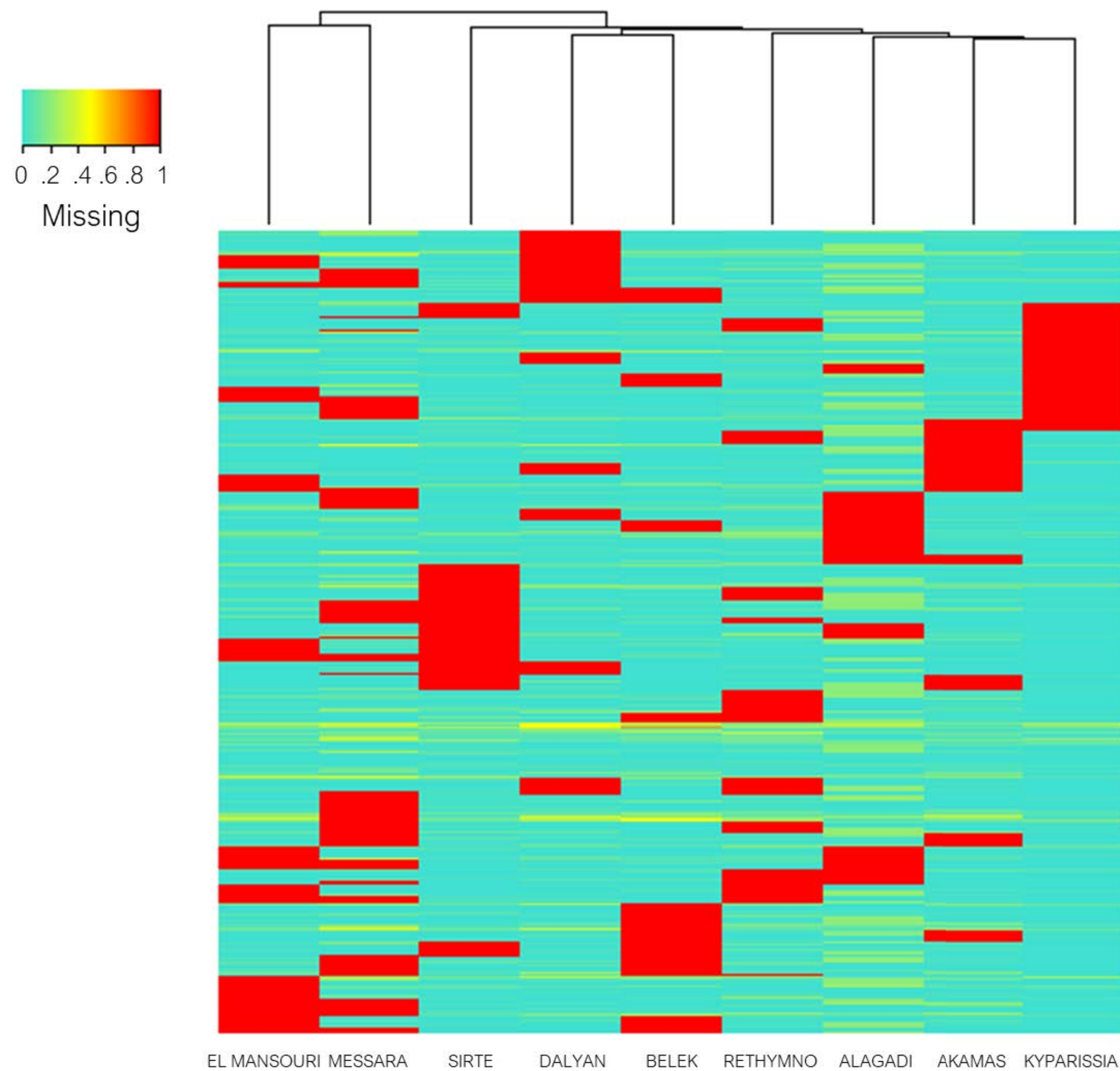


Figure S2. Heatmap showing the frequency of missing loci of the base dataset (n = 3685) per population. The frequency of markers does not follow a progressive gradient for most loci, but it stands at either sides of the spectrum, being fully present or fully absent in each population. This result can be interpreted to infer low connectivity among our populations and could be used in the future as population barcoding for assigning individuals.

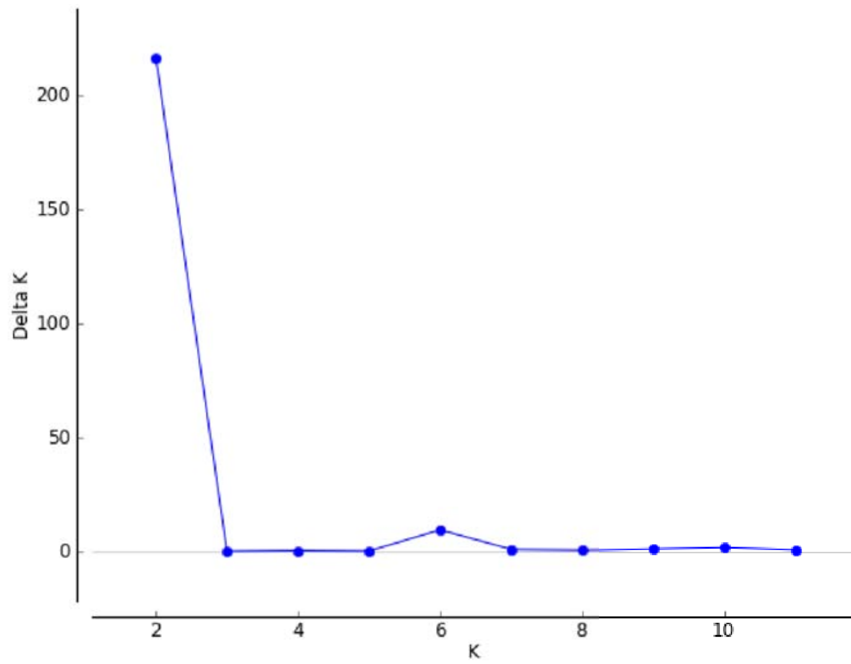


Figure S3. ΔK statistic (Evanno et al., 2005). Bayesian clustering using STRUCTURE revealed that best K value was 6.

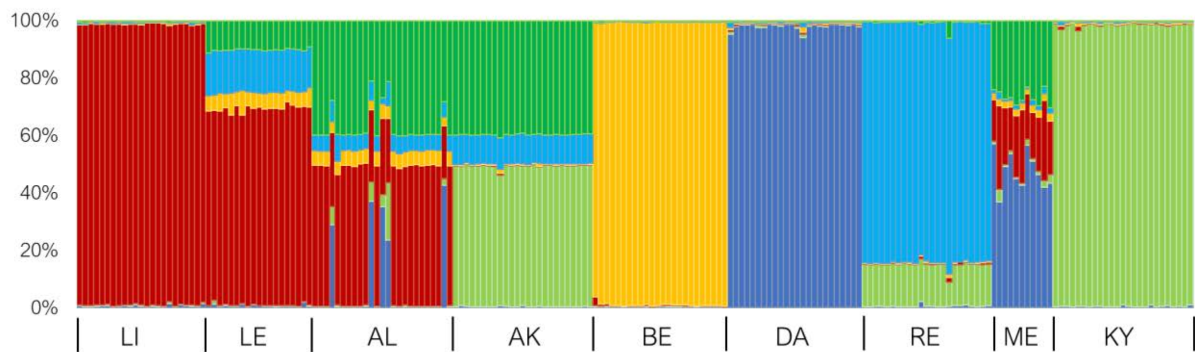


Figure S4. The graph shows Bayesian clustering analysis using STRUCTURE software. Following the estimation from Evanno the most likely number of clusters was K=6. Libya, Belek Kyparissia and Dalyan show four different clusters, while the remaining five populations show a mixture of the different clusters.

DISCUSSION

As climate change threatens thousands of species and populations throughout the world (Bálint *et al.,* 2011) it is critical that conservation management practices are undertaken in a manner that is both effective and cost efficient. Scientific research, and in particular genetic analysis, is a key factor for efficiency and effectiveness of conservation planning (McMahon *et al.,* 2014). This thesis explores traditional and new generation methodologies in two case studies to illustrate the critical role of genetic and genomic analysis for the management of both *ex-situ* and *in-situ* conservation programs. We thus provide new elements to draw conclusions about the most appropriate approaches on conservation genetics and to contribute to the understanding and conservation of the studied species.

After decades of conservation activities to restore their populations, sea turtles still need support and protection to avoid the definitive extinction of their species (Mazaris *et al.,* 2017). The ongoing threat of extinction places sea turtles in an unlucky group of species, most of them considered by the IUCN as endangered or critically endangered in their red list (IUCN 2020). The results of this thesis are applicable to all species in this group and their respective conservation management programs. With an increasing number of species facing extinction (Butchart *et al.,* 2010) and a declining pool of funds available for conservation projects on a per-species basis (Martin *et al.,* 2018), it is more important than ever to ensure that conservation management projects are designed in the most efficient and effective manner. The process of undertaking genetic analysis as described in this thesis allows for an efficient approach to making decisions for conservation management. The information gathered from genetic analyses can be used to improve conservation because the information achieved through this process can be translated into actual conservation management decisions. That is why this thesis can contribute to improve conservation of several non-model species under threat of extinction.

Genetic analysis for ex-situ conservation planning

Recently reintroduction programs have become a new powerful tool to deal with locally extinct populations (Russello and Jensen, 2018). Although this strategy is expensive and highly demanding management wise, in recent years has become more and more popular in order to deal with the global biodiversity crisis (Perzanowski *et al.,* 2020; Van Houtan *et al.,* 2020). Captive breeding, reintroduction and assisted colonisation therefore must be well planned and managed based on scientifically informed data at each step of the conservation plan. We carried out a monitoring study on the Cayman Island Turtle Centre (CTC), an *ex-situ* reintroduction program that has been breeding green turtles in captivity and releasing individuals to the wild for the last 50 years.

The translocation of individuals from nearby populations to form the founder captive stock is a crucial process that defines the quality of the whole program. In fact, the selection of the populations contributing should be done based on historical data regarding the relationship among them and between them and the now extinct populations, and most importantly, considering their genetic variability and genetic similarity (Witzenberger and Hochkirch, 2011). In the CTC reintroduction, the first founder group was gathered from several different populations in the Caribbean Sea and South Atlantic region in order to minimise the impact on a particular population by taking only a few individuals from multiple sources. As a consequence of this strategy, the founder stock had a high genetic diversity and a significant deficit of heterozygotes. However, this strategy also implied a potential risk associated with mixing genetically different groups, which could cause long term effects due to outbreeding on the new reintroduced population and potentially on nearby populations. This thesis though did not find any impact on fitness measurements for the first wild generations after the reintroduction, although further monitoring is crucial to detect long-term effects.

The correct management of the captive breeding stock is also a crucial step to set the program towards success. In fact, maintaining a high genetic diversity of the breeding stock over time is fundamental to avoid the risk of inbreeding in the short and long term. In this thesis we showed that different replacement strategies for the captive breeding population may have different consequences on the genetic diversity of the captive population. In the 1980s the original founder stock for captive breeding was drastically reduced supposedly in response to economic interests (i.e. the cost of the maintenance of the centre being too high due to the number of the individuals forming the breeding stock). The cut to the founder stock was performed arbitrarily without considering turtles' origin. This left the remaining captive stock with unknown genetic diversity. At this point, the population of breeders was maintained stable through small replacements of breeders born in the CTC once the original founders died. After the 2001, the captive population was reduced to less than 10% of its original numbers when the centre was partially destroyed by hurricane Michelle. The reduction of the breeding stock caused by the hurricane was compensated by the incorporation of a large number of individuals (189) of one single cohort (C1995) to the breeding stock. Our genetic analysis showed that this single cohort replacement decreased the CTC genetic variability at nuclear and mitochondrial markers because the use of many individuals from the same generation increased the degree of genetic relatedness within the CTC breeding stock. Thus, we show that continuous small replacements of the breeding stock using individuals from different cohorts is a better strategy to maintain diversity than a single replacement from the same cohort, since the levels of variability of the MCF1 group are higher,

with no signals of inbreeding and lower relatedness values. Preferably the replacement should be combined with genetic analyses to keep the diversity of the CTC high to ensure a long-term survival and avoidance of inbreeding.

In these conservation projects, the monitoring of the wild population after the release of individuals is necessary to assess the success of the reintroduction and to detect any possible short- and long-term effect on the population. In this way the captive breeding and reintroduction program can be adjusted and corrected if needed. Our results show that the CTC succeeded in reintroducing a new wild population of green turtles that is now safely nesting in both Little Cayman and Grand Cayman Islands. We found by microsatellite analyses that 90% of the wild population of green turtle nesting in Grand Cayman Island was related to the female population of the captive breeding stock. Therefore, most parents of wild breeding females were either permanently captive in the CTC or escaped in 2001 because of the hurricane. Additionally, we found that of the first generation of wild new-borns 79.4% of Little Cayman hatchlings and 90.3% of Grand Cayman hatchlings were related to the CTC. This result confirmed the success of the efforts carried out by the CTC in the past few decades and that the nesting events of green turtles in these two Islands are mainly the result of an assisted colonisation through individuals reintroduced from the captive breeding program.
We also scanned mitochondrial haplotypes in both captive and wild populations to assess their origin (i.e. belonging to the Caribbean or South Atlantic/African lineage). CTC haplotypes belong in fact to both lineage A and lineage B, described in Naro-Maciel et al. (2014) as from the Caribbean and from the South Atlantic/Africa region respectively. This was consistent with the reported origin of the founder stock. Although we did not find haplotypes from lineage B exclusive of the South Atlantic/Africa region in the wild population, captive individuals from lineage B might have contributed to the new wild population. These wild lineage B carriers might have not been detected at this stage because the contribution of the younger breeders has not yet shown its impact on the wild population, as the released individuals may need between 15 and 19 years to reach maturity.

Monitoring of reintroduction programs is not only necessary for the welfare of captive and reintroduced individuals, but also to assess the progress of the assisted colonisation in action and the fitness of the new wild population. In the case of the Cayman Islands new wild populations, we found that the reproductive fitness (i.e. size, fecundity and viability of the clutch) of these new populations was not affected by the relatedness with the CTC reintroduction program. In addition, hatchlings not related to the CTC showed higher heterozygosity than related hatchlings. These

results suggest that the new wild population is currently fit and does not tend towards an outbreeding scenario. Nevertheless, they only refer to a first generation of wild hatchlings and thus population fitness analyses should be repeated in the future to monitor potential drops due to outbreeding of the different genetic strains that were part of the initial captive population.

The study of the process and progress of an assisted colonisation is not only important for the feedback provided to management, but also because the foundation of new populations of long lived species is a process almost impossible to detect and study up close in nature while in action. Marine vertebrates in particular are difficult to track and challenging to study. Therefore, the CTC reintroduction offered a unique opportunity to understand how colonisation of new areas work for sea turtles. We discovered that during the assisted colonisation of Grand Cayman Island, some reintroduced individuals also reached Little Cayman Island which is located 108 km away. We also showed that after colonisation the populations were genetically different after just one generation. In particular Little Cayman hatchlings showed less genetic overlap to the CTC female breeders than Grand Cayman hatchlings with both nuclear and mitochondrial markers. This result is consistent with the higher geographic distance from where captive individuals were released but also with the lower level of relatedness found, which can be the result of the contribution of individuals from the original Little Cayman population. This means that, although turtles usually show philopatric behaviour, not all reintroduced individuals went back to their beach of release (i.e. Grand Cayman Island) to nest. Therefore, it is plausible that an unknown number of reintroduced turtles reached other nesting populations of the Caribbean Sea and mingled with their individuals. We cannot estimate what number of individuals might be involved in this multidirectional dispersion, but genetic analysis could allow in the future to test several populations of the area to assess the impact of the CTC. The possible mixing of turtles genetically belonging to the South Atlantic lineage with populations of the Caribbean Sea could, on one hand, enhance the genetic variability of those populations. On the other hand, this could also have negative impacts such as resulting in an outbreeding depression.

The multidirectional dispersion caused by the reintroduction program will be incredibly challenging to reconstruct. The assessment of the degree of nest site fidelity of the reintroduced females though could be used as a proxy to project an approximation of how many reintroduced adult females breed on other rookeries other than the Cayman Islands. Using nest geographic coordinates, we found that wild females nesting in Grand Cayman Island have a high degree of within season nest site fidelity. In fact, 85.1% of females showed a mean distance between nests of less than 5 km, and 77.7% showed a distance between the two most distant nests of less than

5 km. This was estimated to be the typical distance between successive nesting sites for adult females (Shroeder, 2003). On the other hand, we found 9 females and 13 males contributing to both Little Cayman and Grand Cayman rookeries. We hypothesise that long-distance nesting could be an evolutionary strategy sea turtles developed to maintain the income of gene flow into a population and avoid collapse due to extreme philopatry. Since we did not find any significant impact of female heterozygosity or of their relatedness to the CTC on mean distance between nests or on the two most distant nests, these long-distance nesting events are not genetically determined and probably are the result of random processes. This behaviour was already reported in the Mediterranean Sea for loggerhead turtles nesting sporadically in new areas, possibly as a form of response to the increasing temperatures of nesting beaches (Carreras *et al.,* 2018). The high degree of NSF found in the reintroduced females of Grand Cayman Island could suggest that only a low number of individuals moves to breed in other Caribbean rookeries. Of course, this kind of analysis does not consider individuals that dispersed when juveniles and never nested in the Cayman Islands. Future studies could run models to approximate the dispersion of juveniles and breeders and combine modelling with genetic analysis to confirm the origin of other Caribbean breeders.

Ex-situ conservation could be the only option for several species in the near future. Species with low dispersal, damaged habitat and limited potential to colonise mew areas are potentially more vulnerable and more likely to become extinct. With this case study we could thus show how genetic analysis can contribute to the management of a conservation strategies, such as captive breeding, reintroduction and assisted colonisation, on many different levels. This methodology provides insight unexplored in the previous literature because such programs are fairly new in sea turtle conservation management and such monitoring program has never been carried out before on these species. The management of *ex-situ* conservation projects is extremely complicated and delicate, considering the welfare of the animals involved and also the economic cost of such programs. The consequences of *ex-situ* projects not only affect the captive and reintroduced population but also the environment in which it is being reintroduced and the communities sharing that habitat. For this reason, scientifically implemented genetic analysis studies must be put into place to understand the implications of captive breeding of a certain species in a certain area in order to inform conservation management before, during and after the captive program is run. These findings can be valuable not only for future management of sea turtle reintroductions, but can also be transferred to the study other complex species with similar challenging characteristics.

Genomic analysis for in-situ conservation planning

Conservation of endangered species should be carried out preferably *in-situ* so that animals are not taken from their natural habitat and the maintenance (i.e. food and artificial habitat) is less costly. *In-situ* conservation not only ensures the target species survival by directly interacting with it, but also involves the preservation of the habitat and surrounding ecosystem and the education of the public (Edwards *et al.,* 2020). Genetic markers are a powerful tool to inform conservation because can detect ongoing evolutionary processes not visible to mere observation and allows the assessment of a population status and fitness (Höglund, 2009). Both genetic and genomic analysis have been used in this thesis with the ultimate purpose of improving conservation management strategies by providing scientific evidence regarding the target population of study. Previous studies have proved that NGS methodologies and protocols can provide different outcomes when used with different species, and for this reason, they need to be optimised and adjusted. A foundation of the approach taken in this thesis is the use of the most effective and efficient analysis on the target populations of sea turtles. In order to achieve this, an optimised laboratory protocol was developed. Given the critical nature of conservation genetics, this thesis sought to optimise laboratory protocols for conservation genetics to achieve cost-effective analysis. Our pilot study (Chapter 3) showed the steps to follow when using 2b-RAD in non-model species. We decided to run several analyses on two IIB enzymes to optimise this technique tailoring it to our species of interest. In this way we were able to understand how loggerhead turtle DNA interacts with 2b-RAD reagents, and we could calibrate both the library building and sequencing protocol to our species needs.

First of all, we found that the quality of the initial DNA did not affect the number of raw reads nor the final number of loci. This proves already that the methodology used is well suited for our target species, the loggerhead turtle, since very often sea turtle samples come from dead highly degraded individuals with poor DNA quality. Second, we focused on the obtention of the best possible depth of coverage. A good depth coverage, in fact, is important to consider data reliable, since low mean depth per locus leads to less accurate genotype calling and thus higher percentage of missing data across loci (Casso, Turon & Pascual, 2019; Maruki & Lynch, 2017; Chow et al., 2019). Simulation analyses allowed relating the number of reads per sample and the resulting number of loci as well as the linear correlation between the mean depth per locus and the number of reads per individual resulting in 20x as a minimum depth for reliable genotyping. Combining these two functions the optimum number of individuals per one lane can be calculated easily, simplifying decision-making and analysis design at the lowest cost. Only AlfI enzyme should be used to build 2b-RAD libraries for loggerhead turtles, due to problems with digestion and library

building with other enzymes. Unfortunately, due to the big genome size of this species to achieve 20x-25x of coverage, only a few samples could be sequenced in a same lane, which would result in extremely expensive analyses. Simulating a selective-base ligation we found that we can reduce the number of loci being analysed, decreasing the costs of sequencing per sample while ensuring good loci coverage, without influencing the outcome. Finally, we used the results from this study to create a set of guidelines for future studies using this methodology to optimize effort, time, and sequencing cost, not only for our organism of interest but for non-model species in general. We therefore used our optimised protocol to study population structure and dynamics of loggerhead turtle rookeries of the Eastern Mediterranean Basin. The results of our genomic study (Chapter 4) indicated that the predictions we made with our pilot study were correct. In fact, we obtained between 4706 and 9821 polymorphic loci per population with a mean depth of 30.79 reads per locus by processing 48 samples per lane with 2bRAD and 5'-WN-3' base selection.

In-situ conservation is fundamental for the survival of the loggerhead turtle in the Eastern Mediterranean. For this reason, this thesis explored the application of In-situ genomics analysis in 9 populations of sea turtles in this region. This case study was chosen because although many years of beach patrolling and nest protection helped the recovery of several nesting populations, in this region the loggerhead turtle still needs conservation intervention to prevent extinction (Casale, 2015). Moreover, although previous studies already focused on this area, they left open important evolutionary questions that at that time could not be answered fully by using traditional markers. In addition, these same studies showed that the ability to detect differentiation relies on the number of markers used for the analyses (Carreras *et al.,* 2007; Clusa et al 2018), therefore using genomics we could increase the power and accuracy of the analyses. For this reason, we used genomic analysis to refine the population structure of the most important Mediterranean rookeries, assess their status and to identify environmental impact on genetic differentiation. In order to ensure validity and reliability we used the optimised protocol from chapter 3.

The understanding of the actual size of wild populations is critical to assess their status and to plan conservation measures. The size of a breeding population is in fact an indicator of their genetic health and a critical factor in the planning of conservation actions. However, this parameter is often difficult to obtain in marine turtles and mostly relying on nest counts and assumptions of certain population parameters (Casale and Heppel, 2016) that are not always accurate (Casale and Ceriani, 2020). Furthermore, the male segment of a population is challenging to study because they do not approach nesting beaches and remain at sea during all their life, so adult operational sex ratios (i.e. sex ratios of the breeding individuals) remain mostly

unknown. In this thesis we used parentage analysis on the Cayman green turtle populations to assess the breeding census size as well as the operational sex ratio that can be applied to monitor marine turtle populations. However, the risk of inbreeding and the vulnerability of a population to genetic drift are related to the effective population size (Ne) rather than the size of the breeding population (Figure 5). By computing the effective population size of Eastern Mediterranean loggerhead rookeries we were able to complement data on population size collected by counting the number of nests laid per nesting season by females of each population, and compare it with estimates of breeding adults projected from the nest count. Although the studied populations already had census data of population size, Ne allows to understand how certain evolutionary processes influence the population taking into account the number of mating individuals, sex ratio, variation in reproductive success, age structure, migration and other demographic factors (Waples *et al.,* 2016). The Turkish populations of Belek and Dalyan, followed by Kyparissia, had the highest effective population sizes within the Mediterranean, meaning that they are less vulnerable to detrimental genetic effects related to population sizes, such as the loss of alleles due to genetic drift.

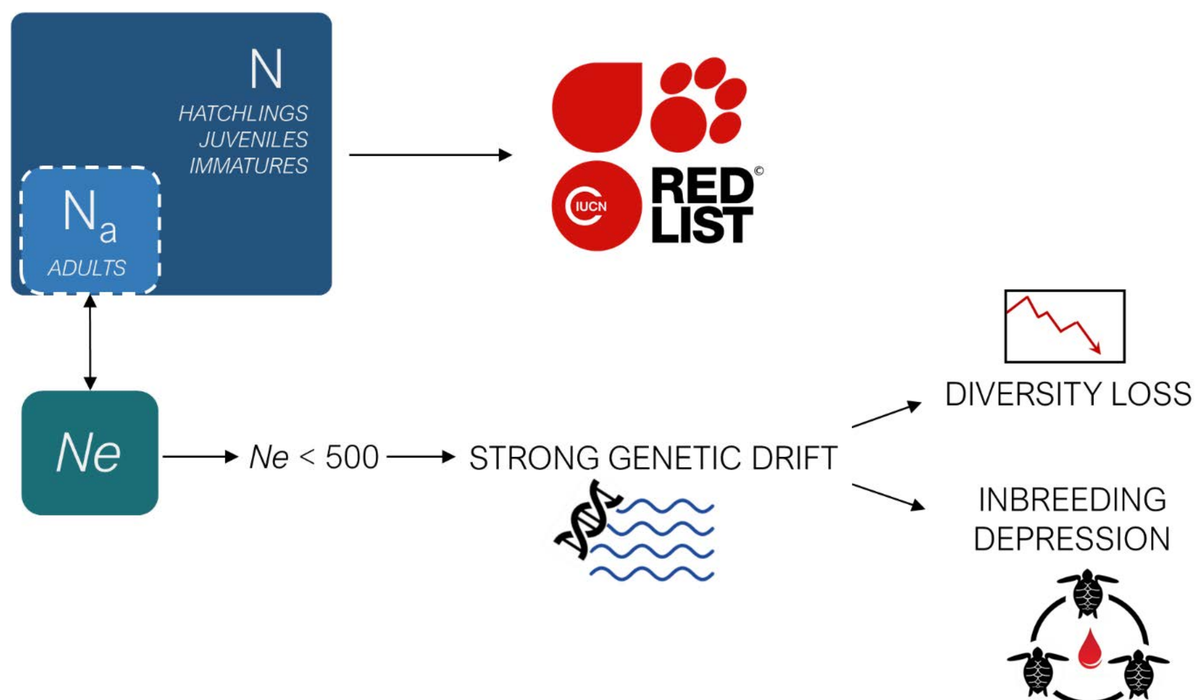


Figure 5. Conceptual map for estimates of population size for conservation. IUCN assessments rely on changes in census sizes (N). Within the total number of individuals, only adults reproduce (N_a = number of adults based on nest count), and thus produce gametes for the next generation. However, since not all breeders reproduce equally, the effective population size (*Ne*), provides a more accurate estimate of the effectively breeding population. *Ne* can therefore assess the actual vulnerability of a population to genetic drift, which could result in diversity loss or Inbreeding depression, and represents a more reliable source of information for conservation planning.

Populations with small Ne, in fact, can easily collapse in response to any drastic change in the environment or also just as a consequence of genetic drift. Alagadi, Messara and Rethymno showed the smallest effective sizes and are therefore more vulnerable to genetic drift. In the case of Alagadi and Messara. According to mtDNA haplotype composition, the populations of the two Crete and Cyprus have been proposed as colonised more recently in comparison of the populations of Greece, Libya and Turkey, and thus the lower effective population sizes can be the result of a more recent origin (Clusa *et al.,* 2013). Previous studies suggest that an effective population of 500 is necessary to maintain equilibrium between loss of adaptative genetic variation due to genetic drift and its replacement by mutation (Franklin, 1980, Franklin and Frankham, 1998). Lynch and Lande (1998) have also estimated that the target Ne for conservation programs of endangered species should range from 500 to 1000. Although these estimates likely depend on the studied species, previous studies on sea turtles considered populations with Ne of over a thousand breeders as healthy (Theissinger *et al.,* 2009) and Ne ranging between 90-220 individuals as sign of vulnerability (Rivalan *et al.,* 2006). Based on these different classes of extinction risk, our results support the management of Mediterranean rookeries as different units, given the range of Ne presented by the studied populations. Therefore, these results are extremely important to plan future conservation actions. Although we found a significant correlation between the effective population size and the number of breeders estimated from nest counts, the population by population analysis showed that the relationship between these two parameters can be very variable. Consequently, although the census size is globally related to the Ne, it cannot be used for population risk assessment. On the contrary, effective population size based on a genome wide panel of markers should be used as a complementary measure for conservation purposes, since it can be estimated without the need of temporal data information.

The analysis of population structure also points in the direction of redefining the Mediterranean management unit. Previous studies already detected a certain degree of population structure in Mediterranean rookeries (Clusa *et al.,* 2018), identifying five distinct units. In this study even the distribution of missing data alone indicated a strong difference in genetic markers among the studied populations. We found, in fact, a population-based pattern of missing loci where the frequency of our markers stood at either side of the spectrum, being fully present or fully absent in each population. This population barcoding can be used to interpret population structure, understand phylogenetic history and could be exceptionally useful in future analyses involving population assignation of individuals. Confirming this differentiation suggest by regional missing loci, we found no sign of overlapping populations and significant genetic differentiation between

all pairwise F_{ST} comparisons. These rookeries thus are more genetically isolated than previously found in studies based on microsatellites (Clusa *et al.,* 2018). We also could not find evidence of gene flow between these populations, which is probably caused by the strong philopatric behaviour of both adult females and males.

These results suggest that the degree of philopatry of these populations is very strong, reinforcing the genetic isolation of rookeries. This extreme philopatry, coupled with reduced effective population sizes, would favour differentiation among populations due to genetic drift. We therefore show that on top of female's strong nest fidelity, males also may present a degree of natal homing such that most of the mating would happen in the areas surrounding nesting beaches and only a very small portion would occur in feeding grounds or during migrations where individuals from several rookeries interact. Thus, even if turtles from different populations may share common foraging areas that define sub-regional management units (Casale and Mariani, 2014), individuals do not generally mingle genetically. This discovery is a key information for the conservation of loggerhead turtle populations, considering the tremendous damage that could be done to turtles mating in the vicinity of tourism-hotspot beaches. Bayesian clustering also found 6 genetic groups within our sample set that mirrored the pattern found in the Principal Coordinate Analysis (i.e. 1) Sirte, El Mansouri and Alagadi, 2) Akamas with Kyparissia, 3) Rethymno, 4) Belek, 5) Dalyan and 6) Messara). In addition, five individuals sampled at Alagadi showed to group with Messara in both PCoA and clustering analysis. These individuals could either come from another population and accidentally nest in Alagadi, or be hybrids resulting from different genetic groups. Both these hypotheses though remain to be confirmed. Groups found in this study don't fully match the units found in Clusa et al. (2018) as they found Akamas grouped with Sirte and identified Greek populations as one same unit. Based on our results we can therefore hypothesise the presence of a low level of gene flow between rookeries, which should definitely be managed as separate units.
At date, the whole Mediterranean basin is considered as one RMU for loggerhead turtles (Wallace *et al.,* 2010). In the past decade a few studies already pointed out the necessity of dividing this region in smaller units in order to improve and customise management strategies. In 2007, Carreras et al. identified 2 MUs in the Mediterranean using 7 nuclear markers; in 2018, Clusa et al. identified 5 MUs using 15 nuclear markers and in the present study, using 195 genomic markers, we found that all the main loggerhead rookeries of the Eastern Mediterranean have significant genetic differentiation, identifying 9 MUs. Considering that the sampled rookeries were not always the same, the number found in our study would be the minimum number of MUs to be considered in the Mediterranean. This highlights the importance of using markers with enhanced

resolution for management and conservation and specifically the potential of genomics studies for delineating conservation units. This also proves that increasing the number of markers improves the accuracy of detecting genetic differentiation. These results reinforce the importance of splitting the Mediterranean basin into sub-regional MUs and consider each rookery as a separate unit as previous studies have recommended (Laurent *et al.,* 2018).

Conservation management of endangered individuals not only focuses of the target species but should also consider its environment and the interaction between the two. In sea turtles for instance, sand temperature directly influences the development of sea turtle embryos and determines their sex during incubation, and atmospheric temperature can be used as a proxy for this parameter. Based on our results using atmospheric temperature, we could suggest that the studied populations might have adapted to the different temperatures of their nesting area, which results today in a strong genetic differentiation. Moreover, the fact that a higher number of markers was correlated with maximum than minimum temperature could be a sign that warmer climates have a stronger impact on sea turtles. In fact, the capacity to adapt to global warming could be crucial for the survival of these populations. We also found significant differentiation correlated to geographic locations of the rookeries. Several outlier markers candidate for adaptation were in fact correlated with latitude and longitude measurements of the sampling site. This can reflect how environmental factors linked with geographic gradients (either north-south or east-west) can have significant impact on local adaptation of each rookery to environmental conditions and therefore, enhance genetic differentiation. The effect of temperature on the genetic structuring is not an artifact of the geographic position of the nesting beaches, as shown by the partial mantel test and by the fact that the loci related to temperature were different from those related to either latitude or longitude. Unfortunately, the lack of a reference genome reduces the capacity of identifying the role of the candidate regions on adaptation. Further studies focusing on adaptation to temperature changes could help to understand even better the evolutionary dynamics of loggerhead turtle nests in this area. The outcome of these kinds of analyses will be extremely valuable to the management of nesting beaches, nest protection and also the monitoring of adults colonising new areas. Previous studies suggested, in fact, that marine turtle nesting behaviour can drive adaptive differentiation at remarkably fine spatial scales, and have important implications for how we define conservation units for protection (Weber *et al.,* 2012). The combination of environmental data and genomics represents a fundamental baseline for the studying and managing of sea turtles, since environmental conditions are critical factors for the survival of these species, particularly in the early stages of their life.

Final remarks

Genetic analysis can be complemented with and complementary to other kinds of analysis and data. In the previous literature it was common to use nest related data or success rates to estimate population parameters and fitness (Broderick *et al.,* 2003; Casale, 2010). In this thesis though, we have used data regarding nest fertility and success, inter-nesting intervals, geographical coordinates and environmental variables combined with genetic data to improve our understanding of nesting dynamics, population fitness and local adaptation. This is a much more reliable and valid approach as the combination of this data with genetics allows to crosscheck the validity of the analysis and the data collected, and it allows to increase the number of possible analysis to perform and their reliability. The combination of genetics and ecological or biological parameters can in fact be used to have a deeper understanding of the processes driving the genetic differentiation of populations. The combination of genetic and genomic data can thus provide more complete results to inform conservation management. Conservation planning must, in fact, consider several factors and potential variables (e.g. geographic location, genetic history, environmental factors etc.) at the time to ensure optimal planning. For this reason, interdisciplinary analyses are crucial to obtain holistic answers to conservation questions.

CONCLUSIONS

1. The Cayman Turtle Centre captive stock was related to 90% of the wild green turtle females nesting in Grand Cayman Island, suggesting a successful outcome of the reintroduction program.

2. The excess of heterozygotes in the captive F_1 is the result of the mixing of different genetic groups.

3. We show that continuous small replacements of the breeding stock using individuals from different cohorts is a better strategy to maintain diversity in captive populations.

4. We find high level of relatedness between hatchlings of both Little Cayman and Grand Cayman Islands and the CTC.

5. Significant genetic differentiation between Little Cayman and Grand Cayman Islands and the CTC suggests that founder effect can drive differentiation on a short time and geographic scale.

6. Heterozygosity of both islands was found lower in hatchlings related to the CTC. Nonetheless, neither of them showed lower female reproductive fitness due to this relationship.

7. High degree of nest-site fidelity in females nesting in Grand Cayman Island suggests low dispersal within the same season.

8. 2b-RAD genomic sequencing is a suitable methodology for non-model species with big genome sizes and potentially degraded DNA, such as the loggerhead turtle.

9. We found that a selective-base ligation is necessary to optimise the trade-off between samples per plate and number of reads per locus, in species with big genomes.

10. The reduction of loci using selective-base ligation did not affect the genetic distance among samples nor individual heterozygosity.

11. We found high population structure among Eastern Mediterranean rookeries.

12. Half of the rookeries show very low effective population size suggesting high vulnerability and indicating different conservation needs within the same Regional Management unit.

13. The temperature and geographic location are likely to have an important role on local adaptation of loggerhead turtle hatchlings.

14. Based on our results we suggest to consider at least 9 Management Units within the Eastern Mediterranean and to plan their conservation plans separately.

REFERENCES

Aiken, J. J., Godley, B. J., Broderick, A. C., Austin, T., Ebanks-Petrie, G. & Hays, G. C. (2001). Two hundred years after a commercial marine turtle fishery: the current status of marine turtles nesting in the Cayman Islands. Oryx, 35(2), 145-151.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature reviews genetics*, *11*(10), 697.

Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... & Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. Nature climate change, 1(6), 313-318.

Barlow, J., Lennox, G. D., Ferreira, J., Berenguer, E., Lees, A. C., Mac Nally, R., ... & Parry, L. (2016). Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature, 535(7610), 144-147.

Bell, C. D., Parsons, J., Austin, T. J., Broderick, A. C., Ebanks-Petrie, G., & Godley, B. J. (2005). Some of them came home: the Cayman Turtle Farm headstarting project for the green turtle Chelonia mydas. Oryx, 39(2), 137-148.

Bellard, C., Cassey, P., & Blackburn, T. M. (2016). Alien species as a driver of recent extinctions. Biology letters, 12(2), 20150623.

Blanco, M. B., Greene, L. K., Williams, R. C., Yoder, A. D., & Larsen, P. A. (2019). Next-generation in situ conservation and capacity building in Madagascar using a mobile genetics lab. BioRxiv, 650614.

Bjorndal, K. A., & Carr, A. (1989). Variation in clutch size and egg size in the green turtle nesting population at Tortuguero, Costa Rica. Herpetologica, 181-189.

Bjorndal, K. A., Wetherall, J. A., Bolten, A. B. & Mortimer, J. A. (1999). Twenty-six years of green turtle nesting at Tortuguero, Costa Rica: an encouraging trend. Conservation Biology, 13(1), 126-134.

Bolten, A. B., Bjorndal, K. A., Martins, H. R., Dellinger, T., Biscoito, M. J., Encalada, S. E., & Bowen, B. W. (1998). Transatlantic developmental migrations of loggerhead sea turtles demonstrated by mtDNA sequence analysis. Ecological Applications, 8(1), 1-7.

Bowen, B. W., & Karl, S. A. (2007). Population genetics and phylogeography of sea turtles. Molecular ecology, 16(23), 4886-4907.

Bowen, B. W., Abreu-Grobois, F. A., Balazs, G. H., Kamezaki, N., Limpus, C., & Ferl, R. J. (1995). Trans-Pacific migrations of the loggerhead turtle (*Caretta caretta*) demonstrated with mitochondrial DNA markers. Proceedings of the National Academy of Sciences, 92(9), 3731-3734.

Bradshaw, P. J., Broderick, A. C., Carreras, C., Fuller, W., Snape, R. T. E., Wright, L. I., & Godley, B. J. (2018). Defining conservation units with enhanced molecular tools to reveal fine scale structuring among Mediterranean green turtle rookeries. Biological Conservation, 222, 253-260.

Broderick, A. C., Glen, F., Godley, B. J. & Hays, G. C. (2002). Estimating the number of green and loggerhead turtles nesting annually in the Mediterranean. Oryx, 36(3), 227-235.

Broderick, A. C., Glen, F., Godley, B. J., & Hays, G. C. (2003). Variation in reproductive output of marine turtles. Journal of Experimental Marine Biology and Ecology, 288(1), 95-109.

Butchart, S. H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P., Almond, R. E., ... & Carpenter, K. E. (2010). Global biodiversity: indicators of recent declines. Science, 328(5982), 1164-1168.

Carlsson, J., Olsen, K. H., Nilsson, J., Øverli, Ø., & Stabell, O. B. (1999). Microsatellites reveal fine-scale genetic structure in stream-living brown trout. Journal of Fish Biology, 55(6), 1290-1303.

Carreras, C., Pascual, M., Cardona, L., Aguilar, A., Margaritoulis, D., Rees, A., ... & Khalil, M. (2007). The genetic structure of the loggerhead sea turtle (Caretta caretta) in the Mediterranean as revealed by nuclear and mitochondrial DNA and its conservation implications. Conservation Genetics, 8(4), 761-775.

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., ... & Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). Scientific reports, 8(1), 1435.

Carreras, C., García-Cisneros, A., Wangensteen, O. S., Ordóñez, V., Palacín, C., Pascual, M., & Turon, X. (2020). East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). Diversity and Distributions, 26(3), 382-398.

Carrión-Cortez, J. A., Zárate, P., & Seminoff, J. A. (2010). Feeding ecology of the green sea turtle (*Chelonia mydas*) in the Galapagos Islands. Journal of the Marine Biological Association of the United Kingdom, 90(5), 1005-1013.

Casale, P. (Ed.). (2010). Sea turtles in the Mediterranean: distribution, threats and conservation priorities. IUCN.

Casale, P. (2011). Sea turtle by-catch in the Mediterranean. Fish and Fisheries, 12(3), 299-316.

Casale, P. 2015. Caretta caretta(Mediterranean subpopulation). The IUCN Red List of Threatened Species 2015: e.T83644804A83646294. https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T83644804A83646294.en

Casale, P., & Mariani, P. (2014). The first 'lost year'of Mediterranean sea turtles: dispersal patterns indicate subregional management units for conservation. Marine Ecology Progress Series, 498, 263-274.

Casale, P., & Heppell, S. S. (2016). How much sea turtle bycatch is too much? A stationary age distribution model for simulating population abundance and potential biological removal in the Mediterranean. Endangered Species Research, 29(3), 239-254.

Casale, P., & Ceriani, S. A. (2020). Sea turtle populations are overestimated worldwide from remigration intervals: correction for bias. Endangered Species Research, 41, 141-151.

Casale, P., Affronte, M., Insacco, G., Freggi, D., Vallini, C., Pino d'Astore, P., ... & Argano, R. (2010). Sea turtle strandings reveal high anthropogenic mortality in Italian waters. Aquatic Conservation: Marine and Freshwater Ecosystems, 20(6), 611-620.

Casale, P., Broderick, A. C., Camiñas, J. A., Cardona, L., Carreras, C., Demetropoulos, A., ... & Lazar, B. (2018). Mediterranean sea turtles: current knowledge and priorities for conservation and research. Endangered species research, 36, 229-267.

Chassin-Noria, O., Macip-Ríos, R., Dutton, P. H. & Oyama, K. (2017). Multiple paternity in the East Pacific green turtle (*Chelonia mydas*) from the Pacific coast of Mexico. Journal of Experimental Marine Biology and Ecology, 495, 43-47.

Chesser, R. K. (1991). Gene diversity and female philopatry. Genetics, 127(2), 437-447.

Chow, J. C., Anderson, P. E., & Shedlock, A. M. (2019). Sea turtle population genomic discovery: Global and locus-specific signatures of polymorphism, selection, and adaptive potential. Genome biology and evolution, 11(10), 2797-2806.

Clusa, M., Carreras, C., Pascual, M., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... & Türkozan, O. (2013). Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). Journal of experimental marine biology and ecology, 439, 15-24.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Giacoma, C., ... & Maffucci, F. (2014). Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine biology, 161(3), 509-519.

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... & Aguilar, A. (2018). Philopatry in loggerhead turtles *Caretta caretta*: beyond the gender paradigm. Marine Ecology Progress Series, 588, 201-213.

Cury, P. (1994). Obstinate nature: an ecology of individuals. Thoughts on reproductive behavior and biodiversity. Canadian Journal of Fisheries and Aquatic Sciences, 51(7), 1664-1673.

de Carvalho, R. H., Lacerda, P. D., da Silva Mendes, S., Barbosa, B. C., Paschoalini, M., Prezoto, F., & de Sousa, B. M. (2015). Marine debris ingestion by sea turtles (Testudines) on the Brazilian coast: an underestimated threat? Marine pollution bulletin, 101(2), 746-749.

Denkinger, J., Parra, M., Muñoz, J. P., Carrasco, C., Murillo, J. C., Espinosa, E., ... & Koch, V. (2013). Are boat strikes a threat to sea turtles in the Galapagos Marine Reserve?. Ocean & coastal management, 80, 29-35.

DeSalle, R., & Amato, G. (2004). The expansion of conservation genetics. *Nature reviews genetics*, *5*(9), 702.

Dietz, S., & Adger, W. N. (2003). Economic growth, biodiversity loss and conservation effort. Journal of environmental management, 68(1), 23-35.

Doherty, T. S., Glen, A. S., Nimmo, D. G., Ritchie, E. G., & Dickman, C. R. (2016). Invasive predators and global biodiversity loss. Proceedings of the National Academy of Sciences, 113(40), 11261-11265.

Dutton, D. L., Dutton, P. H., Chaloupka, M. & Boulon, R. H. (2005). Increase of a Caribbean leatherback turtle *Dermochelys coriacea* nesting population linked to long-term nest protection. Biological conservation, 126(2), 186-194.

Edwards, R. C., Godley, B., & Nuno, A. (2020). Exploring connections among the multiple outputs and outcomes emerging from 25 years of sea turtle conservation in Northern Cyprus. Journal for Nature Conservation, 125816.

Encalada, S. E., Bjorndal, K. A., Bolten, A. B., Zurita, J. C., Schroeder, B., Possardt, E., ... & Bowen, B. W. (1998). Population structure of loggerhead turtle (*Caretta caretta*) nesting colonies in the Atlantic and Mediterranean as inferred from mitochondrial DNA control region sequences. Marine Biology, 130(4), 567-575.

Engeman, R. M., Martin, R. E., Constantin, B., Noel, R. & Woolard, J. (2003). Monitoring predators to optimize their management for marine turtle nest protection. Biological conservation, 113(2), 171-178.

Engeman, R. M., Martin, R. E., Smith, H. T., Woolard, J., Crady, C. K., Shwiff, S. A., ... Griner, J. (2005). Dramatic reduction in predation on marine turtle nests through improved predator monitoring and management. Oryx, 39(3), 318-326.

Fischer, J. & Lindenmayer, D. B. (2000). An assessment of the published results of animal relocations. Biological Conservation, 96(1), 1-11.

Flynn, D. F., Mirotchnick, N., Jain, M., Palmer, M. I., & Naeem, S. (2011). Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. Ecology, 92(8), 1573-1581.

Fontaine, C. (2005). Head-starting the Kemp's ridley sea turtle, Lepidochelys kempii, at the NMFS Galveston Laboratory, 1978-1992: A review. Chelonian Conservation Biology, 4, 838-845.

Franklin, I. R. (1980). Evolutionary changes in small populations. Conservation Biology: An Evolutionary-Ecological Prospective. M. Soulé, B.A. Wilcox (Eds.), Sinauer Associates, Sunderland, MA, pp. 135-150

Franklin, I. R., & Frankham, R. (1998). How large must populations be to retain evolutionary potential?. Animal conservation, 1(1), 69-70.

Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). Introduction to conservation genetics. Cambridge university press.

Frazer, N. B. (1986). Survival from egg to adulthood in a declining population of loggerhead turtles, *Caretta caretta*. Herpetologica, 47-55.

Fuentes, M. M. P. B., Maynard, J. A., Guinea, M., Bell, I. P., Werdell, P. J., & Hamann, M. (2009). Proxy indicators of sand temperature help project impacts of global warming on sea turtles in northern Australia. Endangered Species Research, 9(1), 33-40.

Fuentes, M. M. P. B., Limpus, C. J., & Hamann, M. (2011). Vulnerability of sea turtle nesting grounds to climate change. Global Change Biology, 17(1), 140-153.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. Trends in ecology & evolution, 27(9), 489-496.

Genome 10K Community of Scientists. (2009). Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. Journal of Heredity, 100(6), 659-674.

González, S., Maldonado, J. E., Leonard, J. A., Vilà, C., Barbanti Duarte, J. M., Merino, M., ... & Wayne, R. K. (1998). Conservation genetics of the endangered Pampas deer (*Ozotoceros bezoarticus*). *Molecular ecology*, *7*(1), 47-56.

Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. Animal Behaviour, 28(4), 1140-1162.

Grunwald, C., Stabile, J., Waldman, J. R., Gross, R., & Wirgin, I. (2002). Population genetics of shortnose sturgeon Acipenser brevirostrum based on mitochondrial DNA control region sequences. Molecular Ecology, 11(10), 1885-1898.

Guebert, F. M., Barletta, M., & da Costa, M. F. (2013). Threats to sea turtle populations in the Western Atlantic: poaching and mortality in small-scale fishery gears. Journal of Coastal Research, 65(sp1), 42-47.

Gurevitch, J., & Padilla, D. K. (2004). Are invasive species a major cause of extinctions?. Trends in ecology & evolution, 19(9), 470-474.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). Principles of population genetics (Vol. 116). Sunderland, MA: Sinauer associates.

Hawkes, L. A., Broderick, A. C., Godfrey, M. H. & Godley, B. J. (2005). Status of nesting loggerhead turtles *Caretta caretta* at Bald Head Island (North Carolina, USA) after 24 years of intensive monitoring and conservation. Oryx, 39(1), 65-72.

Hays, G. C., & Scott, R. (2013). Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. Functional Ecology, 27(3), 748-756.

Hedrick, P. W., & Miller, P. S. (1992). Conservation genetics: techniques and fundamentals. *Ecological Applications*, *2*(1), 30-46.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., ... & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular ecology resources, 11, 123-136.

Heppell, S. S., Crowder, L. B. & Crouse, D. T. (1996). Models to evaluate headstarting as a management tool for long-lived turtles. Ecological Applications, 6(2), 556-565.

Höglund, J. (2009). Evolutionary conservation genetics. Oxford University Press.

Hurtado, L. A., La Rosa-Reyna, D., Mateos, M., Shaver, D. J., Metz, R., Thiltges, J., ... & Johnson, C. (2016). Thousands of Single Nucleotide Polymorphisms in the Critically Endangered Kemp's Ridley Sea Turtle (*Lepidochelys kempii*) Revealed by Double-Digest Restriction-Associated DNA Sequencing: Opportunities for Previously Elusive Conservation Genetics Research. Gulf of Mexico Science, 33(2), 8.

IUCN 2020. The IUCN Red List of Threatened Species. Version 2020-1. https://www.iucnredlist.org. Downloaded on 19 March 2020.

James, M. C., Andrea Ottensmeyer, C., & Myers, R. A. (2005). Identification of high-use habitat and threats to leatherback sea turtles in northern waters: new directions for conservation. Ecology letters, 8(2), 195-201.

Janzen, F. J. & Paukstis, G. L. (1991). Environmental sex determination in reptiles: ecology, evolution, and experimental design. The Quarterly review of biology, 66(2), 149-179.

Jensen, M. P., FitzSimmons, N. N., Dutton, P. H., & Michael, P. (2013). Molecular genetics of sea turtles. The biology of sea turtles, 3, 135-161.

Jensen, M. P., Allen, C. D., Eguchi, T., Bell, I. P., LaCasella, E. L., Hilton, W. A., ... 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. e154.

Komoroske, L. M., Jensen, M. P., Stewart, K. R., Shamblin, B. M., & Dutton, P. H. (2017). Advances in the application of genetics in marine turtle biology and conservation. Frontiers in Marine Science, 4, 156.

Komoroske, L. M., Miller, M. R., O'Rourke, S. M., Stewart, K. R., Jensen, M. P., & Dutton, P. H. (2019). A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. Molecular ecology resources, 19(2), 497-511.

Kumar, V. P., Shrivastwa, A., Nigam, P., Kumar, D., & Goyal, S. P. (2017). Genetic characterization of wild swamp deer populations: ex situ conservation and forensics implications. Mitochondrial DNA Part A, 28(6), 965-970.

Laurent, L., Casale, P., Bradai, M. N., Godley, B. J., Gerosa, G., Broderick, A. C., ... & El-Mawla, E. A. (2018). Molecular Profile of Loggerhead Turtles Nesting in Northern Cyprus. Studies on the Ecology and Conservation of Marine Turtles, with Particular Reference to the Mediterranean., 181.

Lee, P. L., Luschi, P., & Hays, G. C. (2007). Detecting female precise natal philopatry in green turtles using assignment methods. Molecular Ecology, 16(1), 61–74.

Lemons, G., Lewison, R., Komoroske, L., Gaos, A., Lai, C. T., Dutton, P., ... & Seminoff, J. A. (2011). Trophic ecology of green sea turtles in a highly urbanized bay: insights from stable isotopes and mixing models. Journal of Experimental Marine Biology and Ecology, 405(1-2), 25-32.

Lewison, R. L., & Crowder, L. B. (2007). Putting longline bycatch of sea turtles into perspective. Conservation biology, 21(1), 79-86.

Lohmann, K. J., Lohmann, C. M., Brothers, J. R., & Putman, N. F. (2013). Natal homing and imprinting in sea turtles. The Biology of Sea Turtles, 3, 59–77.

Lynch, M., & Lande, R. (1998). The critical effective size for a genetically secure population. Animal Conservation, 1(1), 70-72.

Lolavar, A., & Wyneken, J. (2020). The impact of sand moisture on the temperature-sex ratio responses of developing loggerhead (*Caretta caretta*) sea turtles. Zoology, 138, 125739.

Mancini, A., & Koch, V. (2009). Sea turtle consumption and black market trade in Baja California Sur, Mexico. Endangered Species Research, 7(1), 1-10.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. Trends in ecology & evolution, 18(4), 189-197.

Martin, T. G., Kehoe, L., Mantyka-Pringle, C., Chades, I., Wilson, S., Bloom, R. G., ... & Diaz, B. P. (2018). Prioritizing recovery funding to maximize conservation of endangered species. Conservation Letters, 11(6), e12604.

Mayr, E. (1963). Animal Species and Evolution Belknap Press of Harvard University Press Cambridge

Mazaris, A. D., Schofield, G., Gkazinou, C., Almpanidou, V., & Hays, G. C. (2017). Global sea turtle conservation successes. Science advances, 3(9), e1600730.

McClenachan, L., Jackson, J. B. & Newman, M. J. (2006). Conservation implications of historic sea turtle nesting beach loss. Frontiers in Ecology and the Environment, 4(6), 290-296.

McLaughlin, J. F., Hellmann, J. J., Boggs, C. L., & Ehrlich, P. R. (2002). Climate change hastens population extinctions. Proceedings of the National Academy of Sciences, 99(9), 6070-6074.

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation?. Evolutionary Applications, 7(9), 999-1007.

Meffe, G. K. (1986). Conservation genetics and the management of endangered fishes. *Fisheries*, *11*(1), 14-23.

Michaels, C. J., Gini, B. F., & Preziosi, R. F. (2014). The importance of natural history and species-specific approaches in amphibian ex-situ conservation. The Herpetological Journal, 24(3), 135-145.

Midgley, G. F., Hannah, L., Millar, D., Rutherford, M. C., & Powrie, L. W. (2002). Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. Global Ecology and Biogeography, 11(6), 445-451.

Miller JD (1985) Embryology of marine turtles. In: Gans C, Billett F, Maderson PFA (eds) Biology of the Reptilia, Vol 14. Wiley Interscience, New York, p 271–328.

Milner-Gulland EJ, Bennett EL, SAMWM Group (2003) Wild meat: the bigger picture. Trends Ecol Evol 18:351–357

Mitrus, S. (2005). Headstarting in European pond turtles (*Emys orbicularis*): Does it work? Amphibia-Reptilia, 26(3), 333-341.

Mooney, A., Conde, D. A., Healy, K., & Buckley, Y. M. (2020). A system wide approach to managing zoo collections for visitor attendance and in situ conservation. Nature Communications, 11(1), 1-8.

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. Trends in ecology & evolution, 19(4), 208-216.

Moritz, C. (1994). Defining ‘evolutionarily significant units’ for conservation. Trends in ecology & evolution, 9(10), 373-375.

Mrosovsky, N. & Pieau, C. (1991). Transitional range of temperature, pivotal temperatures and thermosensitive stages for sex determination in reptiles. Amphibia-Reptilia, 12(2), 169-179.

Nada, M., & Casale, P. (2011). Sea turtle bycatch and consumption in Egypt threatens Mediterranean turtle populations. Oryx, 45(1), 143-149.

Naro-Maciel, E., Reid, B. N., Alter, S. E., Amato, G., Bjorndal, K. A., Bolten, A. B., ... & Pineda-Catalan, O. (2014). From refugia to rookeries: phylogeography of Atlantic green turtles. Journal of experimental marine biology and ecology, 461, 306-316.

Nichols, J. D. & Williams, B. K. (2006). Monitoring for conservation. Trends in Ecology & Evolution, 21(12), 668-673.

Paetkau, D., & Strobeck, C. (1994). Microsatellite analysis of genetic variation in black bear populations. Molecular ecology, 3(5), 489-495.

Pawar, P. R., Shirgaonkar, S. S., & Patil, R. B. (2016). Plastic marine debris: Sources, distribution and impacts on coastal and ocean biodiversity. PENCIL Publication of Biological Sciences, 3(1), 40-54.

Perzanowski, K., Bleyhl, B., Olech, W., & Kuemmerle, T. (2020). Connectivity or isolation? Identifying reintroduction sites for multiple conservation objectives for wisents in Poland. Animal Conservation, 23(2), 212-221.

Rivalan, P., Dutton, P. H., Baudry, E., Roden, S. E., & Girondot, M. (2006). Demographic scenario inferred from genetic data in leatherback turtles nesting in French Guiana and Suriname. Biological Conservation, 130(1), 1-9.

Roos, D., Pelletier, D., Ciccione, S., Taquet, M. & Hughes, G. (2005). Aerial and snorkelling census techniques for estimating green turtle abundance on foraging areas: a pilot study in Mayotte Island (Indian Ocean). Aquatic Living Resources, 18(2), 193-198.

Russello M. A., & Jensen E. L. (2018). Ex Situ Wildlife Conservation in the Age of Population Genomics. In: Population Genomics. Springer, Cham

Ryman, N. (1991). Conservation genetics considerations in fishery management. *Journal of Fish Biology*, *39*, 211-224.

Sanz, V. & Grajal, A. (1998). Successful Reintroduction of Captive-Raised Yellow-Shouldered Amazon Parrots on Margarita Island, Venezuela. Conservation Biology, 12(2), 430-441.

Seminoff, J. (2004). Chelonia mydas. IUCN Red List of Threatened Species. Version 2011.

Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., ... & Fawcett, K. D. (2015). Genomics and the challenging translation into conservation practice. Trends in ecology & evolution, 30(2), 78-87.

Shillinger, G. L., Palacios, D. M., Bailey, H., Bograd, S. J., Swithenbank, A. M., Gaspar, P., ... & Eckert, S. A. (2008). Persistent leatherback turtle migrations present opportunities for conservation. PLoS biology, 6(7).

Storfer, A. (1999). Gene flow and endangered species translocations: a topic revisited. Biological Conservation, 87(2), 173-180.

Taberlet, P. (1996). The use of mitochondrial DNA control region sequencing in conservation genetics. Molecular Genetic Approaches in Conservation. Oxford University Press, New York, 125-142.

Theissinger, K., FitzSimmons, N. N., Limpus, C. J., Parmenter, C. J., & Phillott, A. D. (2009). Mating system, multiple paternity and effective population size in the endemic flatback turtle (*Natator depressus*) in Australia. Conservation Genetics, 10(2), 329-346.

Troëng, S., & Chaloupka, M. (2007). Variation in adult annual survival probability and remigration intervals of sea turtles. Marine Biology, 151(5), 1721-1730.

Van Dierendonck, M., Bandi, N., Batdorj, D., Dügerlham, S. & Munkhtsog, B. (1996). Behavioural observations of reintroduced Takhi or Przewalski horses (*Equus ferus przewalskii*) in Mongolia. Applied Animal Behaviour Science, 50(2), 95-114.

Van Houtan, K. S., Gagne, T., Jenkins, C. N., & Joppa, L. (2020). Sentiment Analysis of Conservation Studies Captures Successes of Species Reintroductions. Patterns, 100005.

Vrijenhoek, R. C. (1998). Conservation genetics of freshwater fish. *Journal of fish Biology*, *53*, 394-412.

Wallace, B. P., DiMatteo, A. D., Hurley, B. J., Finkbeiner, E. M., Bolten, A. B., Chaloupka, M. Y., ... Bjorndal, K. A. (2010). Regional management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS One, 5(12), e15465.

Wallace, B. P., DiMatteo, A. D., Bolten, A. B., Chaloupka, M. Y., Hutchinson, B. J., Abreu-Grobois, F. A., ... Bjorndal, K. A. (2011). Global conservation priorities for marine turtles. PLoS One, 6(9), e24510.

Warwick, C., Arena, P. C. & Steedman, C. (2013). Health implications associated with exposure to farmed and wild sea turtles. JRSM Short Reports, 4(1), 1-7.

Weber, S. B., Broderick, A. C., Groothuis, T. G., Ellick, J., Godley, B. J., & Blount, J. D. (2012). Fine-scale thermal adaptation in a green turtle nesting population. Proceedings of the Royal Society B: Biological Sciences, 279(1731), 1077-1084.

Witzenberger, K. A., & Hochkirch, A. (2011). Ex situ conservation genetics: a review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. Biodiversity and Conservation, 20(9), 1843-1861.

Wright, L. I., Fuller, W. J., Godley, B. J., McGOWAN, A. N. D. R. E. W., Tregenza, T. O. M., & Broderick, A. C. (2012). Reconstruction of paternal genotypes over multiple breeding seasons reveals male green turtles do not breed annually. Molecular ecology, 21(14), 3625-3635.

Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., ... & Connors, S. (2018). Global Warming of 1.5 OC: An IPCC Special Report on the Impacts of Global Warming of 1.5° C Above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty (p. 32). V. Mass

SEA TURTLE CONSERVATION:
GENETICS AND GENOMICS FOR A BETTER MANAGEMENT

ANNA BARBANTI
2020