DOTTORATO DI RICERCA IN

BIOLOGIA EVOLUZIONISTICA ED ECOLOGIA

CICLO XXXVII - PON ex DM1061

Mitochondrial genome sequencing and applications for the conservation of endangered species: sea turtles as case studies

Sequenziamento di genomi mitocondriali e loro applicazioni nella conservazione di specie a rischio: casi di studio sulle tartarughe marine

Settore Scientifico Disciplinare

BIOS-05/A - Ecologia

Dottoranda

Dott.ssa Livia Tolve

Supervisore

Prof. Claudio Ciofi

Co-supervisore

Dott.ssa Sara Fratini

Coordinatore

Prof. Duccio Cavalieri

Index

Summary

Mitochondrial DNA (mtDNA) is widely used in evolutionary and population genetic studies thanks to its small size, conserved structure, high mutation rate, lack of recombination and maternal inheritance. These features make mtDNA one of the main markers of choice to define population structure and demography and reconstruct phylogeographic patterns of dispersal and migration, particularly in sea turtles where life cycles are based on female-philopatry to nesting sites. For decades, parts of single-genes or non-coding sequences of mtDNA have been used as genetic markers to address questions on the life history of organism. With the advent of next generation sequencing techniques, the characterization of whole mtDNA, as well as whole nuclear genome sequences, has become more accessible.

In this thesis, I developed wet and dry lab pipelines for whole mitogenome sequencing using biological samples collected from live or stranded individuals of three sea turtle species across different geographic areas. I compared mitogenome sequence data and information from single gene sequencing to study the population structure, demography and evolutionary history of loggerhead turtles (*Caretta caretta*), green turtles (*Chelonia mydas*) and hawksbill turtles (*Eretmochelys imbricata*).

The mitogenome pipeline proved effective in generating medium–to-high coverage whole mitochondrial genomes for *C. caretta* from the Mediterranean Sea and globally, as well as *C. mydas* and *E. imbricata* sampled in the Gulf of Guinea, West Africa. In all case studies, mitogenomes resolved matrilineages more clearly than traditional markers, revealing numerous mitogenomic variants for each single, widely distributed mtDNA control region haplotype.

Whole mitogenomes also helped clarify global-scale species phylogeography and provided insight into adaptive and non-neutral evolutionary processes that might have affected natural populations. Comparing the results of this thesis with previous studies based on single-gene markers, some discrepancies emerged, which helped explain divergent conclusions previously reported in the literature on the evolutionary history of sea turtle populations. These findings underscored the importance of using most informative markers (e.g. whole mitogenomes) for population and evolutionary genetic analyses.

To fully leverage mitogenomics, it is crucial to implement efficient methodological pipelines, as demonstrated in this thesis. Moreover, clearly defined research questions are

essential to optimize sampling strategies and ensure adequate sample sizes. At both local and global scales, whole mitogenomes, eventually integrated with broad single-gene datasets and nuclear sequences, hold significant potential to improve our understanding of population structure, ecology and evolution of sea turtles, while supporting the design of informed management and conservation strategies.

CHAPTER I

General Introduction

Mitochondrial DNA, characteristics and applications

Mitochondrial DNA (mtDNA) in animal species is a circular molecule present in multiple (1 - 10) copies within each mitochondrion. Each cell, in turn, harbours hundreds or thousands of mitochondria, resulting in a very high number of mtDNA molecules per organism (Cole 2016). Due to this high copy number, laboratory procedures have been streamlined, and mtDNA has been well characterized at a molecular level for nearly 50 years (Brown et al. 1979; Brown 1981; Avise 1986, and references therein).

Animal mtDNA typically ranges between 15 and 20 kilobases and consists of 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, 13 protein-coding genes for subunits of the electron transport chain and a control region for mtDNA replication and transcription, which includes short repetitive sequences and a displacement loop (D-loop) (Boore 1999, and references therein). Exceptions to this structure have been known for long in invertebrates such as nematodes, cnidarians and bivalves (e.g. Hoffmann et al. 1992; Okimoto et al. 1992; Keddie et al. 1998); vertebrates overall display a highly conserved gene content and genome architecture (Gissi et al. 2008), but recent technical advances in long read sequencing (Formenti et al. 2021) show that repeats and duplications might have been overlooked until now.

Mitochondrial genomes' conservation is likely due to the central role of mitochondria in cellular and organismal functions, namely metabolism, apoptosis, disease and aging (e.g. Brand 1997; Kroemer et al. 1998; Gorman et al. 2016). Despite the conserved nature of content and genome architecture, animal mtDNA evolves more rapidly than single-copy nuclear DNA (scnDNA; 5 to 10 times faster), according to estimates by Brown et al. (1979). Indeed, while structural rearrangements (e.g. duplications, deletions, translocations) are rare, base substitutions occur at a high rate (Brown 1981; Boore 1999), which can be explained by a smaller genome size than scnDNA and by a limited efficiency of mtDNA repair mechanisms (Alencar et al. 2019).

Within this generally high mtDNA substitution rate, there is marked heterogeneity among individual genes: some parts, such as the control region, evolve very rapidly, while others, such as the rRNA genes, evolve more slowly. Within the coding protein genes, the third codon position shows the highest substitution rate, 2 – 3 times faster than the first position and 5 – 10 times faster than the second one (Brown et al. 1979; Brown 1981; Kumar 1996).

Another well-known characteristic of mtDNA is the uniparental inheritance in major part of eukaryotes and specifically maternal in all the known vertebrates (Birky 1995), with individuals carrying mostly or uniquely one mitochondrial haplotype (homoplasmy). However, deviations from these rules occur, such as sperm-mediated transfer of mtDNA during the zygote formation (e.g. Sutovsky et al. 1999) and evidence of mtDNA heteroplasmy (e.g. Rensch et al. 2016), both of which are increasingly documented.

All these characteristics – small genome dimensions, conserved structure but rapid evolution rate, lack of recombination, maternal inheritance – have made mtDNA very suitable for evolutionary and ecological studies. Indeed, mtDNA has been extensively used in population biology, phylogenetics and biogeography, proving useful for tracing matrilineages, female-mediated gene flow and founder effects, for defining population units in female-philopatric species, reconstructing relationships among close taxa, and - in case sufficient polymorphisms are detected - revealing matrilineal relationships within populations (Harrison 1989). The mtDNA is also routinely employed in metazoan barcoding as a powerful tool for taxonomy and biodiversity monitoring (Hebert et al. 2003).

For years most of those studies have applied single mtDNA genes as markers. Phylogenetic studies have relied mainly on cytochrome oxidase subunit 1 (COI) and 16S rRNA genes for distant species and on control region for closely related ones (Bronstein et al. 2018). Due to its high evolution rate, the control region has also been the principal marker for genetics studies aiming at defining population structures (e.g. Di Rienzo and Wilson 1991; Norman et al. 1994; Rosel and Block 1996). Demographic history as well has used mainly the control region as a marker, due to its presumed evolutionary neutrality, making it a reliable marker for tracking changes in population size and structure unshaped by selective pressure (e.g. Merilä et al. 1997; Moum and Árnason 2001; Ersmark et al. 2016). Protein-coding genes are generally known to be under purifying selection in mtDNA (Galtier et al. 2009), but studies on several genes have challenged this assumption, focusing on positive selection and mitochondrial adaptation (e.g. Lamb et al. 2018; Baltazar-Soares et al. 2021). As far as barcoding is concerned, COI has been elected as the standardized marker in this field (Ratnasingham and Hebert 2007).

Although the traditional use of mtDNA's single genes has already proved to be efficient in a large plethora of applications, relying on whole mitogenomes means a substantial improvement (DeSalle and Hadrys 2017). Even if reference mitogenomes have been

produced since several years for a lot of species (Formenti et al. 2021), a massive use of complete mtDNA sequences in population genetics, phylogeography and evolutionary biology has been limited. The recent development of next generation sequencing techniques has lowered the costs and further streamlined the mitogenome availability, favouring the advent of mitogenomics.

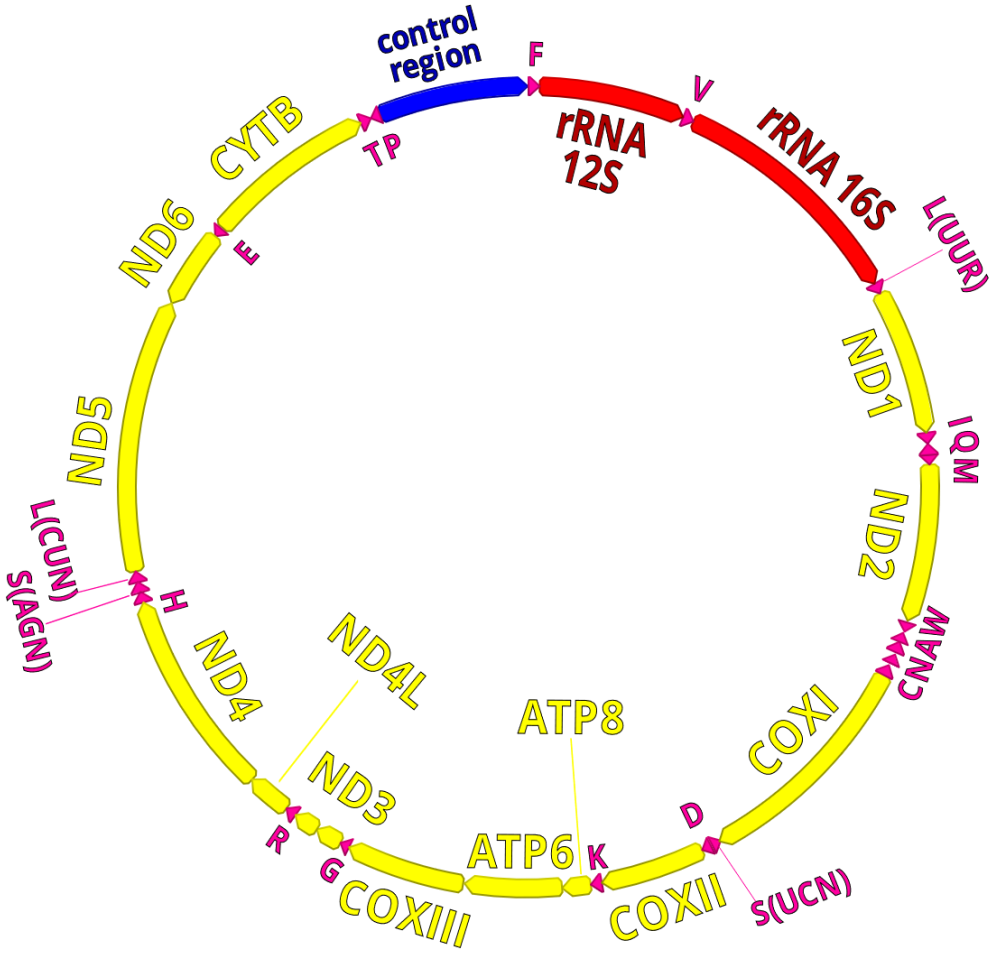


Fig. 1 Mitochondrial gene architecture most spread in vertebrate species (Boore 1999). In red, rRNA coding genes; in pink, tRNA coding genes; in yellow, protein coding genes; in blue, control region of the mtDNA replication and transcription.

Mitogenomics, an overview

The analysis of entire mitochondrial genomes is a step forward compared to the use of single mtDNA genes for a variety of applications.

In phylogenetic studies, whole mitogenomes allow for much deeper phylogenies than single gene markers, as for example in Bernt et al. (2013), where a comprehensive reconstruction of bilaterian evolutionary relationships has been developed based on 650 complete mitogenomes. Using complete mitochondrial genomes in phylogenetics also lead to analytical improvements, such as correcting the random rooting effect in star-like single gene trees of fast evolving species (Hirase et al. 2016). Moreover, whole mtDNA sequences often clarify the phylogenetic placement of species whose relationships remain uncertain when based solely on single-gene data. This is especially important for endangered and relict species (e.g. Zhang et al. 2017). In Morin et al. (2010a) whole mitogenome sequencing helped revealing that previously thought sympatric "ecotypes" of killer whales actually belong to divergent phylogenetic clades and should be considered different species.

Complete mitogenomes help resolving taxa definition also at subspecies level, with important outcomes for management and conservation. In species of commercial interest, subspecies identification allows to avoid undesired hybridization and damage to native taxa. For instance, thanks to whole mitochondrial genomes in Ruiz et al. (2021) *Bombus terrestris canariensis* was defined as a distinct, endemic subspecies of bumblebee in the Canary Islands, deserving appropriate planning of pollination networks. More examples in literature refer to endangered species, where appropriate taxa definition is fundamental to inform conservation measures. By means of mitogenomes the Northern Hemisphere fin whales *Balenoptera physalus* was reconsidered as belonging to two different subspecies – an Atlantic and a Pacific one (Archer et al. 2013), and the same occurs for the humpbacks *Megaptera novaeangliae* (Jackson et al. 2014).

Whole mitogenome analysis also improves the genetic population structure description respect to single gene markers, allowing a proper assessment of Management Units (MU) of conservation. In the critically endangered speartooth shark (*Glyphis glyphis*), the use of mitogenomes revealed barriers to gene flow which had not been detected using only the mtDNA control region, maybe due to genetic signal saturation (Feutry et al. 2014b). Similarly, in another endangered elasmobranch, the largetooth sawfish (*Pristis pristis*),

whole mtDNA sequences refined the population structure respect to the description previously done by means of single genes (Feutry et al. 2015).

In addition, mitogenomes are applied to phylogeography and demographic history. In arctic fox the human impact on demographic trends of the species was assessed by the use of mitogenomes (Larsson et al. 2019). In bats, mitochondrial genomes revealed different population-level responses to Pleistocenic climatic oscillations (Alberdi et al. 2015). By means of mitogenomes Bjork et al. (2010) revealed four subpopulations of chimpanzees with its own evolutionary history. Halley et al. (2016) specifically tested the discordance between complete mtDNA genomes and single mtDNA fragments (such as the control region) in drawing conclusions about the demographic history of northern bobwhites (*Colinus virginianus*), supporting the use of mitogenomes in such kind of investigation.

Access to complete mtDNA sequences also enables to scan all the protein coding partition, assessing the heterogeneity of selective pressure across genes, for example by means of sliding windows approaches. In widely distributed marine species, which are subjected to strong thermic or salinity gradients, evidence of environmental-induced selection has been unveiled in some genes thanks to the analysis of the whole mtDNA. This is the case for example of the Atlantic herrings, where ND2, ND4, ND5 genes were found to be under positive selection and to characterize distinct clades with different haplotypes (Teacher et al. 2012), or the Pacific Salmon, where the piston of ND5 subunit resulted to be involved in adaptive evolution of the species (Garvin et al. 2011).

Overall, mitogenomics is nowadays a useful tool for research in evolutionary biology and ecology of different species, with important applications for management and conservation as well.

Mitochondrial DNA in sea turtles research and conservation

Sea turtles are undoubtedly among the species to which mtDNA has been most thoroughly applied for decades (e.g. Allard et al. 1994; Bowen et al. 1994; Encalada et al. 1996). The main reason for this is female natal philopatry (also known as *natal homing*), which is the tendency of the females to come back to the beach of origin in order to reproduce (Bowen and Karl 2007). This causes different nesting sites (*rookeries*) to host specific aggregates of matrilines; therefore, they can be appropriately characterized by means of a maternally inherited marker such as the mtDNA (Jensen et al. 2013).

In the past decades, sea turtles' population genetics has been mainly carried out by means of fragments of the mtDNA control region. At first, haplotypes were defined based on relatively short mtDNA control region sequences, as for example those produced by the primers pair TCR5/TCR6 for *Chelonia mydas*, 380 base pairs long (Norman et al. 1994). As genetic studies spread, other primers pairs were attempted for a variety of sea turtles' species to obtain longer fragments – about 800 bp long - which included more polymorphic sites (Abreu et al. 2006).

The relative frequency of mtDNA control region haplotypes has been used to characterize turtles' rookeries and to assess their genetic diversity, which allowed for the definition of Management Units (MUs) and of Regional Management Units (RMUs) at oceanic scale, integrating behavioural, biological and ecological data (Wallace et al. 2010a; Wallace et al. 2025). The rookeries' genetic description is also fundamental as a baseline for Mixed Stock Analyses (MSA), which have been extensively used to characterize the aggregates of turtles at sea and to reconstruct turtles' migrations (Pella and Masuda 2001).

MSA is a Bayesian statistical approach to infer the most probable composition of a mixed stock in terms of percentage contributions from known sources. In the case of sea turtles, implementation of MSA involves the haplotypic frequencies in the mixed stock of turtles at sea (whose origin is unknown) and the haplotypic frequencies in the baseline of known sampled rookeries; throughout a MCMC computation, the MSA outputs relative contributions – with relative confidence intervals - from different rookeries to the stock at sea (Pella and Masuda 2001; Bolker et al. 2007; Komoroske et al. 2017).

The more informative the genetic marker, the finer the resolution of matrilines; the traditional mtDNA control region haplotypes of 800 bp have come to a threshold of resolution for several turtle species, with a few haplotypes largely shared among populations that hamper sensible improvements in genetic description of nesting sites and mixed stocks (e.g. Shamblin et al. 2012b; Tolve et al. 2018; Loisier et al. 2021). At this regard, upscaling turtle population genetics to mitogenomics might be promising, and only few studies have tried it by now (Shamblin et al. 2012a).

Mitochondrial DNA has been applied also to sea turtle phylogeography and helped assessing the distribution and the evolutionary history of the globally spread species (Dutton et al. 1999; Shamblin et al. 2014; Jensen et al. 2019; Arantes et al. 2020; Vilaça et al. 2022). Such studies relied on huge datasets and on single gene markers – sometimes in association with nuclear ones – but none has tempted a mitogenomic approach at population level.

The mtDNA single genes have been extensively used also for demographic history (e.g. Garofalo et al. 2009) and evolutionary biology (e.g. Naro-Maciel et al. 2008; Novelletto et al. 2016; Tikochinski et al. 2020), but few studies have been published which explore the potential of whole mtDNA sequencing in these research areas (Drosopoulou et al. 2012; Duchene et al. 2012; Hernández-Fernández and Delgado Cano 2018; Frandsen et al. 2020).

Even if the overall production of complete sea turtles mitogenomes have been limited by now, mitogenome analysis might have the potential to improve research on a variety of topics and conservation. In the present thesis I will focus on applications of mitogenomics to sea turtle species, most of which are nowadays at risk of extinction according to the IUCN Red List of Threatened Species and according to the re-assessment of global conservation status for marine turtles (Wallace et al. 2025).

The aim of the thesis

The main objectives of the present thesis are the application of whole mitogenome sequencing to the population genetics and phylogeography of three sea turtle species, and a critical evaluation of its methodological and biological implications.

In Chapter II, I give an overview on different approaches for mitogenome sequencing and assembly, ranging from traditional pipelines such as target-enrichment protocols to more advanced techniques like short-reads and long-reads sequencing. I compare these methods, highlighting their respective strengths and limitations.

In Chapter III, I focus on the use of whole mitogenome sequencing in loggerhead turtles (*Caretta caretta*) and explore its potential to improve the characterization of genetic diversity in the Central Mediterranean Sea, compared to traditional control region markers.

In Chapter IV, I apply mitogenomics to clarify the global phylogeography of *C. caretta,* which until now has been investigated using mtDNA control region or single-gene markers, and whose patterns remain debated. In addition, I use mitogenomic analysis to investigate selective pressure within *C. caretta,* aiming to explain a geographic pattern in haplogroup distribution across the Western Atlantic under the hypothesis of thermal adaptation.

In Chapter V, I introduce the case study of *Chelonia mydas* in the Gulf of Guinea, initially leveraging control region haplotypes data. I then apply mitogenome sequencing to assess its potential to improve genetic diversity resolution of matrilines in the study area, and to contribute to a more refined global phylogeographic reconstruction of the species.

In Chapter VI, I examine *Eretmochelys imbricata* in the Gulf of Guinea, presenting unprecedented control region haplotype data that provide the context for subsequent mitogenomic analyses. As with *C. caretta* and *C. mydas*, mitogenomic data are integrated into a global framework to offer new insights into the species phylogeography.

In Chapter VII, I synthesize and discuss the main methodological and biological contributions of the thesis and suggest future applications of mitogenomics along with more advanced approaches to the study of sea turtle populations.

References

Abreu A, Horrocks J, Formia A, LeRoux R, Velez-Zuazo X, Dutton P, Soares L, Meylan P, Browne D (2006) New mtDNA Dloop primers which work for a variety of marine turtle species may increase the resolution capacity of mixed stock analysis. Proceedings of the 26th Annual Symposium on Sea Turtle Biology and Conservation.

Alberdi A, Gilbert MTP, Razgour O, Aizpurua O, Aihartza J, Garin I (2015) Contrasting population-level responses to Pleistocene climatic oscillations in an alpine bat revealed by complete mitochondrial genomes and evolutionary history inference. Journal of Biogeography 42: 1689-1700.

Alencar RR, Batalha C, Freire TS, de Souza-Pinto NC (2019) Enzymology of mitochondrial DNA repair. Enzymes 45: 257-287.

Allard MW, Miyamoto MM, Bjorndal KA, Bolten AB, Bowen BW (1994) Support for natal homing in green turtles from mitochondrial DNA sequences. Copeia 1994: 34-41.

Arantes LS, Vargas SM, Santos FR (2020) Global phylogeography of the critically endangered hawksbill turtle (*Eretmochelys imbricata*). Genetics and molecular biology 43.

Archer FI, Morin PA, Hancock-Hanser BL, Robertson KM, Leslie MS, Bérubé M, Panigada S, Taylor BL (2013) Mitogenomic phylogenetics of fin whales (*Balaenoptera physalus* spp.): genetic evidence for revision of subspecies. PLOS ONE 8: e63396.

Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Philos Trans R Soc Lond B 312: 325-342.

Baltazar-Soares M, de Araújo Lima AR, Silva G (2021) Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. Genes 12: 91.

Bernt M, Bleidorn C, Braband A, Dambach J, Donath A, Fritzsch G, Golombek A, Hadrys H, Jühling F, Meusemann K, Middendorf M, Misof B, Perseke M, Podsiadlowski L, von Reumont B, Schierwater B, Schlegel M, Schrödl M, Simon S, Stadler PF, Stöger I, Struck TH (2013) A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Molecular Phylogenetics and Evolution 69: 352-364.

Birky C (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences 92: 11331-11338.

Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M (2010) Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. Molecular Biology and Evolution 28: 615-623.

Bolker BM, Okuyama T, Bjorndal KA, Bolten AB (2007) Incorporating multiple mixed stocks in mixed stock analysis: 'many-to-many' analyses. Molecular Ecology 16: 685-695.

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27: 1767-1780.

Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC (1994) Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. Evolution 48: 1820-1828.

Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. Mol Ecol 16.

Brand MD (1997) Regulation analysis of energy metabolism. Journal of Experimental Biology 200: 193-202.

Bronstein O, Kroh A, Haring E (2018) Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. BMC Evolutionary Biology 18: 80.

Brown WM (1981) Mechanisms of evolution in animal mitochondrial DNA.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences 76: 1967-1971.

Cole LW (2016) The evolution of per-cell organelle number. Frontiers in Cell and Developmental Biology Volume 4 - 2016.

DeSalle R, Hadrys H (2017) Evolutionary biology and mitochondrial genomics: 50 000 mitochondrial DNA genomes and counting eLS, pp 1-35

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proceedings of the National Academy of Sciences 88: 1597-1601.

Drosopoulou E, Tsiamis G, Mavropoulou M, Vittas S, Katselidis KA, Schofield G, Palaiologou D, Sartsidis T, Bourtzis K, Pantis J (2012) The complete mitochondrial genome of the loggerhead turtle *Caretta caretta* (Testudines:

Cheloniidae): Genome description and phylogenetic considerations. Mitochondrial DNA 23: 1-12.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK (1999) Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). J Zool 248: 397-409.

Encalada SE, Lahanas PN, Bjorndal KA, Bolten AB, Miyamoto MM, Bowen BW (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. Molecular Ecology 5: 473-483.

Ersmark E, Klütsch CFC, Chan YL, Sinding M-HS, Fain SR, Illarionova NA, Oskarsson M, Uhlén M, Zhang Y-p, Dalén L, Savolainen P (2016) From the past to the present: wolf phylogeography and demographic history based on the mitochondrial control region. Frontiers in Ecology and Evolution Volume 4 - 2016.

Feutry P, Kyne PM, Pillans RD, Chen X, Marthick JR, Morgan DL, Grewe PM (2015) Whole mitogenome sequencing refines population structure of the Critically Endangered sawfish *Pristis pristis*. Mar Ecol Prog Ser 533: 237-244.

Feutry P, Kyne PM, Pillans RD, Chen X, Naylor GJP, Grewe PM (2014) Mitogenomics of the Speartooth Shark challenges ten years of control region sequencing. BMC Evolutionary Biology 14: 232.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli FO, Ambrosini R, Houde P, Koren S, Oliver K, Smith M, Skelton J, Betteridge E, Dolucan J, Corton C, Bista I, Torrance J, Tracey A, Wood J, Uliano-Silva M, Howe K, McCarthy S, Winkler S, Kwak W, Korlach J, Fungtammasan A, Fordham D, Costa V, Mayes S, Chiara M, Horner DS, Myers E, Durbin R, Achilli A, Braun EL, Phillippy AM, Jarvis ED, Kirschel ANG, Digby A, Veale A, Bronikowski A, Murphy B, Robertson B, Baker C, Mazzoni C, Balakrishnan C, Lee C, Mead D, Teeling E, Aiden EL, Todd E, Eichler E, Naylor GJP, Zhang G, Smith J, Wolf J, Touchon J, Delmore K, Jakobsen K, Komoroske L, Wilkinson M, Genner M, Pšenička M, Fuxjager M, Stratton M, Liedvogel M, Gemmell N, Minias P, Dunn PO, Sudmant P, Morin P, Ayub Q, Kraus R, Vernes S,

Smith S, Lama T, Edwards T, Smith T, Gilbert T, Marques-Bonet T, Einfeldt T, Venkatesh B, Johnson W, Warren W, Bukhman Y, The Vertebrate Genomes Project C (2021) Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biology 22: 120.

Frandsen HR, Figueroa DF, George JA (2020) Mitochondrial genomes and genetic structure of the Kemp's ridley sea turtle (*Lepidochelys kempii*). Ecol Evol 10: 249-262.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18: 4541-4550.

Garofalo L, Mingozzi T, Mico A, Novelletto A (2009) Loggerhead turtle (*Caretta caretta*) matrilines in the Mediterranean: further evidence of genetic diversity and connectivity. Mar Biol 156: 2085-2095.

Garvin MR, Bielawski JP, Gharrett AJ (2011) Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific Salmon. PLOS ONE 6: e24127.

Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101: 301-320.

Gorman GS, Chinnery PF, DiMauro S, Hirano M, Koga Y, McFarland R, Suomalainen A, Thorburn DR, Zeviani M, Turnbull DM (2016) Mitochondrial diseases. Nature Reviews Disease Primers 2: 16080.

Halley YA, Oldeschulte DL, Bhattarai EK, Hill J, Metz RP, Johnson CD, Presley SM, Ruzicka RE, Rollins D, Peterson MJ, Murphy WJ, Seabury CM (2016) Northern bobwhite (*Colinus virginianus*) mitochondrial population genomics reveals structure, divergence, and evidence for heteroplasmy. PLOS ONE 10: e0144913.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology & Evolution 4: 6-11.

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences 270: 313-321.

Hernández-Fernández J, Delgado Cano DA (2018) Complete mitochondrial genome of the three Colombian Caribbean loggerhead sea turtles, using next-generation sequencing. Mitochondrial DNA B 3: 626-627.

Hirase S, Takeshima H, Nishida M, Iwasaki W (2016) Parallel mitogenome sequencing alleviates random rooting effect in phylogeography. Genome Biology and Evolution 8: 1267-1278.

Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the blue mussel, Mytilus edulis. Genetics 131: 397-412.

Jackson JA, Steel DJ, Beerli P, Congdon BC, Olavarría C, Leslie MS, Pomilla C, Rosenbaum H, Baker CS (2014) Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). Proceedings of the Royal Society B: Biological Sciences 281: 20133222.

Jensen MP, FitzSimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH (2019) The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). Journal of Biogeography 46: 860-870.

Jensen MP, FitzSimmons NN, Dutton PH, Michael P (2013) Molecular genetics of sea turtles. In: Wyneken J, Lohmann K, Musick JA (eds) The biology of sea turtles, Volume III. CRC Press, Boca Raton, FL, pp 135-161

Keddie EM, Higazi T, Unnasch TR (1998) The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol Biochem Parasitol 95: 111-127.

Komoroske LM, Jensen MP, Stewart KR, Shamblin BM, Dutton PH (2017) Advances in the application of genetics in marine turtle biology and conservation. Frontiers in Marine Science 4: 156.

Kroemer G, Dallaporta B, Resche-Rigon M (1998) The mitochondrial death/life regulator in apoptosis and necrosis. Annual Review of Physiology 60: 619-642.

Kumar S (1996) Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. Genetics 143: 537-548.

Lamb AM, Gan HM, Greening C, Joseph L, Lee Yin P, Morán-Ordóñez A, Sunnucks P, Pavlova A (2018) Climate-driven mitochondrial selection: a test in Australian songbirds. Molecular Ecology 27: 898-918.

Larsson P, von Seth J, Hagen IJ, Götherström A, Androsov S, Germonpré M, Bergfeldt N, Fedorov S, Eide NE, Sokolova N, Berteaux D, Angerbjörn A, Flagstad Ø, Plotnikov V, Norén K, Díez-del-Molino D, Dussex N, Stanton DWG, Dalén L (2019) Consequences of past climate change and recent human persecution on mitogenomic diversity in the arctic fox. Philosophical Transactions of the Royal Society B: Biological Sciences 374: 20190212.

Loisier A, Savelli M-P, Arnal V, Claro F, Gambaiani D, Sénégas JB, Cesarini C, Sacchi J, Miaud C, Montgelard C (2021) Genetic composition, origin and conservation of loggerhead sea turtles (*Caretta caretta*) frequenting the French Mediterranean coasts. Mar Biol 168: 52.

Merilä J, Björklund M, Baker AJ (1997) Historical demography and present day population structure of the greenfinch, *Cardeus chloris* - an analysis of mtDNA control-region sequences. Evolution 51: 946-956.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome research 20: 908-916.

Moum T, Árnason E (2001) Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. Molecular Ecology 10: 2463-2478.

Naro-Maciel E, Le M, FitzSimmons NN, Amato G (2008) Evolutionary relationships of marine turtles: A molecular phylogeny based on nuclear and mitochondrial genes. Molecular Phylogenetics and Evolution 49: 659-662.

Norman JA, Moritz C, Limpus CJ (1994) Mitochondrial DNA control region polymorphisms: genetic markers for ecological studies of marine turtles. Mol Ecol 3: 363-373.

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. Physiol Biochem Zool 89: 473-486.

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130: 471-498.

Pella J, Masuda M (2001) Bayesian methods for analysis of stock mixtures from genetic characters. Fishery Bulletin 99: 151-151.

Ratnasingham S, Hebert PD (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7: 355-364.

Rensch T, Villar D, Horvath J, Odom DT, Flicek P (2016) Mitochondrial heteroplasmy in vertebrates using ChIP-sequencing data. Genome Biology 17: 139.

Rosel PE, Block BA (1996) Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. Marine Biology 125: 11-22.

Ruiz C, Cejas D, Muñoz I, De la Rua P (2021) Characterizing the mitogenome of the endemic bumblebee subspecies from the Canary Islands for conservation purposes. Sociobiology 68: e5910.

Shamblin BM, Bjorndal KA, Bolten AB, Hillis-Starr ZM, Lundgren I, Naro-Maciel E, Nairn CJ (2012a) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol Ecol 21: 2330-2340.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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: e85956.

Shamblin BM, Bolten AB, Bjorndal KA, Dutton PH, Neilen JT, Abreu-Grobois FA, Reich KJ, Witherington BE, Bagley DA, Ehrhart LM, Tucker AD, Addison DS, Arenas A, Johnson C, Carthy RR, Lamont MM, Dodd MG, Gaines MS, LaCasella E, Nairn CJ (2012b) Expanded mitochondrial control region sequences increase resolution of stock structure among North Atlantic loggerhead turtle rookeries. Mar Ecol Prog Ser 469: 145-160.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. Nature 402: 371-372.

Teacher AGF, André C, Merilä J, Wheat CW (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC Evolutionary Biology 12: 248.

Tikochinski Y, Carreras C, Tikochinski G, Vilaça ST (2020) Population-specific signatures of intra-individual mitochondrial DNA heteroplasmy and their potential evolutionary advantages. Sci Rep 10: 211.

Tolve L, Casale P, Formia A, Garofalo L, Lazar B, Natali C, Novelletto A, Vallini C, Bužan E, Chelazzi G, Gaspari S, Fortuna C, Kocijan I, Marchiori E, Novarini N, Poppi L, Salvemini P, Ciofi C (2018) A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. Mar Biol 165: 68.

Vilaça ST, Hahn AT, Naro-Maciel E, Abreu-Grobois FA, Bowen BW, Castilhos JC, Ciofi C, FitzSimmons NN, Jensen MP, Formia A, Limpus CJ, Natali C, Soares LS, de Thoisy B, Whiting SD, Bonatto SL (2022) Global phylogeography of ridley sea turtles (Lepidochelys spp.): evolution, demography, connectivity, and conservation. Conservation Genetics 23: 995-1010.

Wallace BP, Bandimere AN, Abreu-Grobois FA, Acosta H, Akiti J, Akomedi M, Alfaro-Shigueto J, Allen CD, Angenda D, Ayissi I, Azanza Ricardo J, Barrientos-Muñoz KG, Barrios-Garrido H, Bjorndal KA, Bretón Vargas E, Broderick AC, Calderón Peña R, Carreras C, Ceriani SA, Colman LP, Cortés-Gómez AA, Crespo L, Cuevas E, Dah A, de Groene A, Delgado Trejo C, Demetropoulos S, Dias A, Diez C, Dos Santos NA, Dossou Bodjrenou JS, Early Capistrán MM, Eckert KL, Eizaguirre C, Ekanayake L, Escobedo Mondragón M, Esteban N, Feliciano D, Fernandes RS, Ferreira-Airaud B, Foley A, Fonseca LG, Forneiro Martín-Viaña Y, Fossette S, Fuentes M, Gaglo J, Gaos AR, Gidsicki D, Giffoni B, Girard A, Girondot M, Godfrey MH, Godley BJ, González Díaz Mirón RJ, Hamann M, Hancock JM, Hart CE, Hays GC, Herrera R, Hochscheid S, Hoekstra S, Huerta-Rodríguez P, Inteca G, Ishihara T, Jensen MP, Jribi I, Kale N, Kaska Y, Kelez S, Kelly IK, Köhnk S, Lara P, Lasfargue M, Lauritsen AM, Le Gouvello DZM, Liusamoa A, López M, López-Castro MC, Lopez-Mendilaharsu M, Louro CMM, Luna T, Madden CA, Mahabir D, Mancini A, Manoharakrishnan M, Marcovaldi MA, Martínez-Portugal RC, Mastrogiacomo A, Matilde E, Mawunyo Adzagba B, Mbungu S, Miranda C, Moncada F, Morales-Mérida BA, Mortimer JA, Murakawa SKK, Nalovic MA, Nel R, Ngafack R, Nishizawa H, Ogou M, Panagopoulou A, Patricio AR, Peralta Buendía E, Phillott AD, Pilcher NJ, Polyak MMR, Prince RIT, Raynus EH, Reina RD, Rguez-Baron JM, Robbins AE, Santos AS, Sarti-Martínez AL, Schofield G, Seminoff JA, Serrano I, Shamblin BM, Shanker K, Stacy BA, Stahelin G, Staman MK, Stelfox M, Stewart KR, Taxonera A, Tucker AD, Turkozan O, van Dam RP, van de Geer CH, Viera S, West L, Whiting AU, Whiting SD, Wienand L, Wijntuin SR, Wildermann N, Zárate PM, Casale P, DiMatteo A, Hurley BJ, Hutchinson BJ, Maxwell SM, Posnik ZA, Rodriguez I, Mast RB (2025) Updated global conservation status and priorities for marine turtles. Endangered Species Research 56: 247-276.

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen

BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5: e15465.

Zhang X, Unmack PJ, Kuchling G, Wang Y, Georges A (2017) Resolution of the enigmatic phylogenetic relationship of the critically endangered Western Swamp Tortoise *Pseudemydura umbrina* (Pleurodira: Chelidae) using a complete mitochondrial genome. Molecular Phylogenetics and Evolution 115: 58-61.

CHAPTER II

Mitogenomes production

Mitochondrial DNA (mtDNA) has been extensively studied, and a variety of methods have been developed to characterize this molecule. In this chapter, I will briefly review the main approaches described in literature, discussing respective advantages and disadvantages. A general distinction can be made between protocols based on i) target-enrichment before sequencing (reviewed in Calvignac et al. 2011), ii) whole genome sequencing using short reads, followed by mitogenome assembly, iii) a combination of long and short read next-generation sequencing - specifically, the *mitoVGP* method proposed by Formenti et al. (2021) as part of the Vertebrate Genome Project (VGP).

Target enrichment methods are considered direct sequencing methods, while the others are indirect, as mtDNA is recovered as a byproduct of broader sequencing experiments. Another key difference among these methods is how they deal with nuclear-encoded mitochondrial DNA (numts) (Richly and Leister 2004). Numts are nuclear paralogs of mtDNA - segments of mtDNA that have been transferred to the nuclear genome and have since evolved independently. Their presence can compromise phylogenetic and evolutionary analyses if mistaken for genuine mtDNA. Methods in the first category address the issue of numts at the molecular level, during the wet-lab phase before sequencing. In contrast, the other methods rely on bioinformatic tools to detect and manage numts after sequencing.

Target-enrichment

The mtDNA can be isolated by first separating mitochondria from other cellular organelles. One of the first methods developed for mitochondrial separation was described by Beckman et al. (1993). This method relies on the selective sedimentation of mitochondria. Tissues are homogenized and centrifuged at 1500 g for 10 minutes, then the supernatant is collected and centrifuged again at 10000 g for another 10 minutes. Mitochondria are recovered from the pellet of this second centrifugation, after which DNA extraction and sequencing are performed. Similar, more sophisticated protocols based on the physical separation of mitochondria have since been developed for medical applications. Examples include magnetic isolation using antibody-linked microbeads (Devall et al. 2015) and differential centrifugation followed by exonuclease digestion (Gould et al. 2015).

Instead of isolating mitochondria, it is possible to extract mtDNA directly from the sample using specialized kits, such as the mtDNA Isolation Kit (BioVision), or by using

plasmid DNA extraction kits, such as the QIAGEN Miniprep Kit. Quispe-Tintaya et al. (2013) showed that the use of mtDNA-specific kits resulted in higher mtDNA yield and quality than plasmid DNA extraction kits.

A simplified approach for relative mtDNA enrichment involves diluting the total DNA (e.g. Kidd and Friesen 1998; Ibarguchi et al. 2006; Calvignac et al. 2011). Since mtDNA exists in higher copy numbers than nuclear DNA, dilution causes a relatively faster decrease in nuclear DNA concentration compared to mtDNA. The optimal dilution is determined by performing parallel PCRs: one targeting nuclear markers and one targeting a mitochondrial marker. Dilution continues until the nuclear PCR fails to produce an amplicon, while the mitochondrial PCR still does.

Another way to study mtDNA, preventing numts affecting the analyses, is based on extracting RNA instead of DNA (Williams and Knowlton 2001; Calvignac et al. 2011), since numts are rarely transcribed. RNA extraction can be performed by means of various dedicated kits, extracted RNA is then reverse transcribed and cDNA is used in following PCR-based protocols to selectively amplify and sequencing mtDNA.

A widely used amplification protocol for mitogenome characterization is Long-Range PCR (Kee et al. 2023). This technique is typically applied to mtDNA fragments longer than 5 kbp, which are amplified using high-processivity polymerases with longer extension times than those used in traditional PCR (Hu et al. 2002). Long-Range PCR is advantageous for avoiding contamination from numts, as numts are generally small in size. Long-Range PCR strategies may use either two primer sets (Ye et al. 2014), generating partially overlapping fragments, or on a single primer set (e.g. Cui et al. 2013). Overlapping regions between amplicons must be removed from analyses to prevent bias in coverage depth across the molecule. A strategy based on a single primer set eliminates this issue. However, in this case, the polymerase must exhibit even higher processivity to successfully produce a single long amplicon.

Another PCR strategy used for mtDNA enrichment is Rolling Circle Amplification (RCA) (Dean et al. 2001; Dean et al. 2002; Marquis et al. 2017; Dhorne-Pollet et al. 2020; Long et al. 2020). This is an isothermal reaction that employs the highly processive phi29 DNA polymerase and preferentially replicates circular (plasmid) DNA. RCA uses multiple random hexamers as primers, enabling branched replication. Due to the polymerase's strand displacement activity, newly synthesized strands are displaced from the template, resulting in a concatemer composed of multiple copies of the mtDNA genome.

These methods are not mutually exclusive; on the contrary, they have often been combined in a wide variety of protocols to maximize their advantages (e.g. Wolff et al. 2012; Dhorne-Pollet et al. 2020). Overall, their common goal is to increase the specific enrichment of mtDNA molecules while minimizing the presence of numts in the sample prior to high-throughput sequencing.

Short reads approach

Thanks to the widespread adoption of next-generation sequencing technologies, a vast amount of genetic data has become available at low cost. In whole-genome shotgun sequencing experiments, mitochondrial genomes are also automatically sequenced. Since mtDNA is generally present in higher copy numbers than nuclear DNA in metabolically active tissues (Stier et al. 2013; Di Giacomo et al. 2015; Liu et al. 2018; Naue et al. 2024), the depth of coverage for mtDNA typically exceeds that of the nuclear genome. For this reason, genome skimming protocols with low target coverage for the nuclear DNA (Trevisan et al. 2019) have proven effective for reconstructing mitochondrial genomes in ecological studies.

Shotgun sequencing strategies allow the use of total DNA as input for the protocol, with no need for prior mtDNA enrichment. Typically, the input DNA is fragmented before library construction, which can be achieved through mechanical fragmentation (e.g. Ultrasonicator, Covaris), or enzymatic fragmentation (Ribarska et al. 2022) such as tagmentation (Adey et al. 2010). After fragmentation, sequencing adapters and indexes are incorporated, and an amplification step may follow before the shotgun sequencing (e.g. Tilak et al. 2015). The output reads (usually 150 bp) are then assembled into a mitogenome contig using a reference from the same species, if available, or from a closely related species in case of *de novo* assembly (e.g. Gibb et al. 2015).

When reconstructing mitogenomes of species whose reference mitogenome has already been produced, reads obtained by shotgun sequencing are usually mapped to the reference. Bowtie1 (Langmead et al. 2009), Bowtie2 (Langmead and Salzberg 2012) and BWA (Li and Durbin 2009) are among the most used aligner originally used also in genomic pipelines; both implement the Burrows-Wheeler Transform (BWT) algorithm and are optimal for short – medium reads (25 – 150 bp), with BWA outperforming Bowtie in indels detection. Alternatively, MITObim (Hahn et al. 2013) is a mapping-guided assembly approach which has been implemented for mtDNA. MITObim requires

at least a distantly related mitochondrial genome or mitochondrial barcode sequences of the species of interest.

In case of *de novo* assembly using short reads, GetOrganelle (Jin et al. 2020) is a specialized tool for assembling mitochondrial and chloroplast genomes. It relies on an iterative approach to extract organelle reads from the total read set by mapping them to a k-mer based reference database of organelle sequences. Another widely used software is NOVOPlasty (Dierckxsens et al. 2017), which employs a seed-and-extend algorithm to produce circular mitogenomes starting from a seed sequence, such as the COI of a related species. NOVOPlasty also includes a pipeline for detecting heteroplasmy (Dierckxsens et al. 2020).

Since in shotgun experiments mitogenomes are byproducts of whole genome sequencing, there is no way to physically separate mtDNA from nuclear DNA sequences. The filtering of mtDNA is achieved through a bioinformatic pipeline applied to the output reads, either by mapping to a reference or by *de novo* assembly using an algorithm. In both cases, numts may be mistaken for genuine mtDNA genes and could be mapped or assembled together. However, in a shotgun sequencing experiment the depth of coverage for numts is significantly lower than that for the mtDNA, resulting in numts being underrepresented compared to mtDNA. This can be visually inspected in bam files during the mapping process and measured computationally. If the goal is to reconstruct the consensus mitogenome, low frequency variant such as numts will be excluded by the assembler and will not affect the final result. Conversely, if the aim is to investigate mitochondrial heteroplasmy, specific tools for numts correction are required, such as MitoHiFi (Uliano-Silva et al. 2023), Numt Parser (de Flamingh et al. 2022), or the option implemented in NOVOPlasty pipeline.

Long reads approach

One limitation shared by all the previously described methods is their difficulty in accurately characterizing complex, highly repetitive regions - such as the microsatellites typically located between the control region and the Phe tRNA. This is a well-known shortcoming of both Sanger sequencing and short-read sequencing technologies, especially when repetitive elements exceed the read length (Collins 2018). Long-read sequencing can overcome this issue by generating reads 10 – 20 kbp in length, which are often sufficient to span nearly the entire mtDNA molecule in a single contig. However,

single-molecule sequencing technologies have their own limitations, particularly in terms of base-calling accuracy, which often leads to a higher error rate (Espinosa et al. 2024).

To take advantage of both the structural resolution and the base calling accuracy offered by different sequencing technologies, Formenti et al. (2021) proposed a mitogenome assembly pipeline called mitoVGP, which combines long and short read sequencing. They developed mitoVGP as a complementary tool to the nuclear genome assembly pipeline of the Vertebrate Genome Project (VGP), with the aim of generating complete and error free reference mitogenomes.

Specifically, WGS long reads produced by PacBio or Nanopore platforms are aligned to a reference mitogenome from the same or a closely related species, resulting in the production of several contigs. Low-quality contigs and those contaminated by numts are then filtered out using a combination of software tools.

In the next step, Illumina short reads are mapped to the remaining contigs and used to further polish the mitogenome assembly. Finally, repetitive regions and duplications are identified on the resulting contig, and the entire molecule is annotated.

The mitoVGP pipeline was applied by Formenti et al. (2021) to 125 vertebrate species, resulting in complete mitogenomes for 100 of them (80%). For 33 of these 100 species, no reference mitogenome had previously been available. In 25% of the cases where a reference mitogenome already existed, the assemblies generated by mitoVGP filled gaps, added missing repeats, genes or duplications. Indeed, Formenti et al. (2021) significantly advanced our understanding of vertebrate mitochondrial genomes, demonstrating that structural variations - such as repeats and duplications - have been historically underestimated and are far from exceptional.

Methodological comparison

Target-enrichment methods were developed before the widespread adoption of next-generation sequencing technologies, during a time when Sanger sequencing was still the most commonly used approach and the bioinformatic management of numts was particularly challenging. Under these conditions, obtaining the purest possible mtDNA input was crucial. Mitochondria separation proved to be less effective than other enrichment methods (Calvignac et al. 2011), especially because it relies on the excellent preservation of tissue samples – a condition that is often difficult to achieve in molecular ecology. Serial dilutions is a streamlined technique that yields good results (Wolff et al.

2012), provided that a large amount of high quantity total DNA is available. As with mitochondria separation, this requirement is not always met in ecological contexts. PCR is, by definition, the most efficient enrichment method, but it tends to introduce technical errors, which are then amplified and can bias downstream analyses.

In this context, genome skimming via shotgun sequencing has become a popular approach for mitogenome reconstruction (Trevisan et al. 2019). First, it does not require large amounts or high-quality tissues and performs well even with suboptimal samples. Second, it eliminates the need for PCR, thereby avoiding associated technical errors. Third, numts can be directly assessed through sequencing coverage, which reflects the relative abundance of mtDNA vs nDNA in the sample. Finally, since it relies on Illumina short reads sequencing, costs are currently affordable, and the method is scalable to relatively large numbers of samples – making it suitable for ecological studies.

As highlighted by Formenti et al. (2021), Illumina short-read sequencing also has its shortcomings, particularly when it comes to accurately reconstructing the structural features of mtDNA. In such cases, long-read sequencing should be employed to ensure a correct assembly—but this increases the cost of the experiment and is not always necessary. Indeed, for *de novo* assembly of a reference mitogenome, the mitoVGP pipeline is likely the most appropriate choice. On the other hand, when conducting mitogenome resequencing for population genetics or phylogeographic studies where a reliable reference mitogenome is already available, short-read sequencing is a suitable and cost-effective solution (Trevisan et al. 2019). If an accurate reconstruction of the mtDNA microsatellite region cannot be achieved, a reasonable compromise may be to trim the mitogenome sequence at the beginning of the repetition and to exclude it from the following analyses.

References

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biology 11: R119.

Beckman KB, Smith MF, Orrego C (1993) Purification of mitochondrial DNA with Wizard Minipreps DNA purification system. Promega Notes 43: 10-13.

Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). Mitochondrion 11: 246-254.

Collins A (2018) The challenge of genome sequence assembly. The Open Bioinformatics Journal 11.

Cui H, Li F, Chen D, Wang G, Truong CK, Enns GM, Graham B, Milone M, Landsverk ML, Wang J (2013) Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. Genetics in Medicine 15: 388-394.

de Flamingh A, Rivera-Colón AG, Gnoske TP, Kerbis Peterhans JC, Catchen J, Malhi RS, Roca AL (2022) Numt Parser: Automated identification and removal of nuclear mitochondrial pseudogenes (numts) for accurate mitochondrial genome reconstruction in Panthera. Journal of Heredity 114: 120-130.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J (2002) Comprehensive human genome amplification using multiple displacement amplification. Proceedings of the National Academy of Sciences 99: 5261-5266.

Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome research 11: 1095-1099.

Devall M, Burrage J, Caswell R, Johnson M, Troakes C, Al-Sarraj S, Jeffries AR, Mill J, Lunnon K (2015) A comparison of mitochondrial DNA isolation methods in frozen post-mortem human brain tissue—applications for studies of mitochondrial genetics in brain disorders. Biotechniques 59: 241-246.

Dhorne-Pollet S, Barrey E, Pollet N (2020) A new method for long-read sequencing of animal mitochondrial genomes: application to the identification of equine mitochondrial DNA variants. BMC genomics 21: 1-15.

Di Giacomo G, Campello S, Corrado M, Di Giambattista L, Cirotti C, Filomeni G, Gentile G (2015) Mature erythrocytes of *Iguana iguana* (Squamata, Iguanidae) possess functional mitochondria. PLOS ONE 10: e0136770.

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res 45: e18.

Dierckxsens N, Mardulyn P, Smits G (2020) Unraveling heteroplasmy patterns with NOVOPlasty. NAR Genom Bioinform 2: lqz011.

Espinosa E, Bautista R, Larrosa R, Plata O (2024) Advancements in long-read genome sequencing technologies and algorithms. Genomics 116: 110842.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli FO, Ambrosini R, Houde P, Koren S, Oliver K, Smith M, Skelton J, Betteridge E, Dolucan J, Corton C, Bista I, Torrance J, Tracey A, Wood J, Uliano-Silva M, Howe K, McCarthy S, Winkler S, Kwak W, Korlach J, Fungtammasan A, Fordham D, Costa V, Mayes S, Chiara M, Horner DS, Myers E, Durbin R, Achilli A, Braun EL, Phillippy AM, Jarvis ED, Kirschel ANG, Digby A, Veale A, Bronikowski A, Murphy B, Robertson B, Baker C, Mazzoni C, Balakrishnan C, Lee C, Mead D, Teeling E, Aiden EL, Todd E, Eichler E, Naylor GJP, Zhang G, Smith J, Wolf J, Touchon J, Delmore K, Jakobsen K, Komoroske L, Wilkinson M, Genner M, Pšenička M, Fuxjager M, Stratton M, Liedvogel M, Gemmell N, Minias P, Dunn PO, Sudmant P, Morin P, Ayub Q, Kraus R, Vernes S, Smith S, Lama T, Edwards T, Smith T, Gilbert T, Marques-Bonet T, Einfeldt T, Venkatesh B, Johnson W, Warren W, Bukhman Y, The Vertebrate Genomes Project C (2021) Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biology 22: 120.

Gibb GC, Condamine FL, Kuch M, Enk J, Moraes-Barros N, Superina M, Poinar HN, Delsuc F (2015) Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living Xenarthrans. Molecular Biology and Evolution 33: 621-642.

Gould MP, Bosworth CM, McMahon S, Grandhi S, Grimerg BT, LaFramboise T (2015) PCR-free enrichment of mitochondrial DNA from human blood and cell lines for high quality next-generation DNA sequencing. PLoS One 10: e0139253.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41: e129-e129.

Hu M, Chilton N, Gasser R (2002) Long PCR-based amplification of the entire mitochondrial genome from single parasitic nematodes. Molecular and cellular probes 16: 261-267.

Ibarguchi G, Friesen VL, Lougheed SC (2006) Defeating numts: Semi-pure mitochondrial DNA from eggs and simple purification methods for field-collected wildlife tissues. Genome 49: 1438-1450.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z (2020) GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. Genome Biology 21: 241.

Kee PS, Karunanathie H, Maggo SDS, Kennedy MA, Chua EW (2023) Long-Range Polymerase Chain Reaction. Methods Mol Biol 2967: 181-192.

Kidd MG, Friesen VL (1998) Sequence variation in the guillemot (Alcidae: Cepphus) mitochondrial control region and its nuclear homolog. Molecular Biology and Evolution 15: 61-70.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754-1760.

Liu R, Jin L, Long K, Tang Q, Ma J, Wang X, Zhu L, Jiang A, Tang G, Jiang Y, Li X, Li M (2018) Analysis of mitochondrial DNA sequence and copy number variation across five high-altitude species and their low-altitude relatives. Mitochondrial DNA B Resour 3: 847-851.

Long NY, Qiao Y, Xu ZY, Tu J, Lu ZH (2020) Recent advances and application in whole-genome multiple displacement amplification. Quant Biol 8: 279-294.

Marquis J, Lefebvre G, Kourmpetis YA, Kassam M, Ronga F, De Marchi U, Wiederkehr A, Descombes P (2017) MitoRS, a method for high throughput, sensitive, and accurate detection of mitochondrial DNA heteroplasmy. BMC genomics 18: 1-19.

Naue J, Xavier C, Hörer S, Parson W, Lutz-Bonengel S (2024) Assessment of mitochondrial DNA copy number variation relative to nuclear DNA quantity between different tissues. Mitochondrion 74: 101823.

Quispe-Tintaya W, White RR, Popov VN, Vijg J, Maslov AY (2013) Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. Biotechniques 55: 133-136.

Ribarska T, Bjørnstad PM, Sundaram AYM, Gilfillan GD (2022) Optimization of enzymatic fragmentation is crucial to maximize genome coverage: a comparison of library preparation methods for Illumina sequencing. BMC Genomics 23: 92.

Richly E, Leister D (2004) NUMTs in Sequenced Eukaryotic Genomes. Molecular Biology and Evolution 21: 1081-1084.

Stier A, Bize P, Schull Q, Zoll J, Singh F, Geny B, Gros F, Royer C, Massemin S, Criscuolo F (2013) Avian erythrocytes have functional mitochondria, opening novel perspectives for birds as animal models in the study of ageing. Frontiers in Zoology 10: 33.

Tilak M-K, Justy F, Debiais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP (2015) A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. Conservation Genetics Resources 7: 37-40.

Trevisan B, Alcantara DMC, Machado DJ, Marques FPL, Lahr DJG (2019) Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. PeerJ 7: e7543-e7543.

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter M, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, Kersey P, Hollingsworth P, Wilson W, Twyford A, Gaya E, Lawniczak M, Lewis O, Broad G, Martin F, Hart M, Barnes I, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, McCarthy SA, Darwin Tree of Life C (2023) MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinformatics 24: 288.

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. Molecular Biology and Evolution 18: 1484-1493.

Wolff JN, Shearman DCA, Brooks RC, Ballard JWO (2012) Selective enrichment and sequencing of Whole Mitochondrial Genomes in the presence of nuclear encoded mitochondrial pseudogenes (numts). PLOS ONE 7: e37142.

Ye F, Samuels DC, Clark T, Guo Y (2014) High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17: 157-163.

CHAPTER III

Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea

This chapter was published in:

Marine Biology, 171:19. DOI: 10.1007/s00227-023-04325-x

(APPENDIX I)

Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea

Livia Tolve[1,*], Alessio Iannucci[1,2,*], Luisa Garofalo[3], Andrea Ninni[4,5], Andrea Capobianco Dondona[6], Ilaria Ceciarini[7], Cristiano Cocumelli[3], Alessandra De Lucia[8], Mattia Falconi[4], Angela Formia[1,9], Federico Iacovelli[4], Cecilia Mancusi[7,10], Erica Marchiori[11], Letizia Marsili[7], Toni Mingozzi[12], Stefano Nannarelli[8], Chiara Natali[1], Giuliana Terracciano[13], Marco A. L. Zuffi[14], Andrea Novelletto[5,§], Claudio Ciofi[1]

[1] Department of Biology, University of Florence, 50019 Sesto Fiorentino (FI), Italy

[2] National Biodiversity Future Center, 90133 Palermo, Italy

[3] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", 00178 Rome, Italy

[4] Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy

[5] PhD Program in Evolutionary Biology and Ecology, Tor Vergata University of Rome, 00133 Rome, Italy

[6] Farm4Trade, 66041 Atessa (CH), Italy

[7] Department of Physical Sciences, Earth and Environment, University of Siena, 53100 Siena, Italy

[8] Hydrosphera onlus, Rome, Italy

[9] African Aquatic Conservation Fund, Chilmark, MA 02535, USA

[10] Agenzia Regionale per la Protezione Ambientale Toscana (ARPAT), 57125 Livorno, Italy

[11] Department of Comparative Biomedicine and Food Science, University of Padua, Agripolis, 35020 Legnaro (PD), Italy

[12] Department of Biology, Ecology and Earth Sciences, University of Calabria, 87030 Rende (CS), Italy

[13] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", 56123 Pisa, Italy

[14] Natural History Museum, University of Pisa, 56011 Calci (PI), Italy

*These authors have contributed equally to this work

[§] Co-senior author

Corresponding authors: Livia Tolve, livia.tolve@unifi.it; Alessio Iannucci, alessio.iannucci@unifi.it

Abstract

Population structure and phylogeography of the loggerhead sea turtle (*Caretta caretta*) have so far been assessed mainly by mitochondrial DNA (mtDNA) single gene sequencing studies. However, phylogenetic relationships among matrilines, genetic characterization of rookeries and mixed stock analyses have suffered from the limited resolution obtained by comparison of relatively short sequences such as from the mtDNA control region. Whole mitogenome sequencing can significantly improve population genetics, particularly in marine organisms showing female natal philopatry. Despite mitogenomics becoming increasingly common in biodiversity monitoring and conservation, only a few complete mitogenomes are available for *C. caretta*. In this study, we sequenced the complete mtDNA of 61 loggerhead turtles sampled between 2008 and 2021 along the Italian coastline and central Mediterranean Sea. We assigned complete mtDNA haplotypes to dead embryos and bycatch samples and introduced a first nomenclature for loggerhead mitogenomes. Analysis of mtDNA diversity, Maximum Parsimony and Bayesian phylogenetic reconstruction allowed improved resolution of lineages with respect to studies reporting on partial mtDNA control region sequence comparisons, and we were able to further inform previous analyses on loggerhead ancestry based on control region haplogroups. Overall, whole mitogenome analysis has potential for considerable improvement of evolutionary history and phylogeographic investigations as well as mixed-stock surveys of loggerhead turtles.

Keywords

Whole mitogenome sequencing; Mediterranean Sea; *Caretta caretta* phylogeny; sea turtle population genetics

Introduction

Female natal philopatry is the tendency for individual females to breed at or near their place of birth (Shields 1982). This behaviour can shape population structure in many marine species including whales, sharks and chelonids (Miller 1997; Sheridan et al. 2010; Baker et al. 2013; Feldheim et al. 2014). Each breeding site generally hosts specific aggregates of maternal descent; however, exceptions episodically occur which lead a female to colonize new sites and significantly increase the species' range (e.g. Carreras et al. 2018).

Population genetic approaches to study female philopatry have mostly relied upon mitochondrial DNA (mtDNA), which is maternally inherited and can therefore trace matrilines within populations (Moritz 1994). In particular, investigations on whole mtDNA genome sequences have significantly improved our understanding of intraspecific phylogeographic patterns in several marine taxa. For instance, mitogenome sequence analysis in killer whales helped clarifying phylogenetic relationships among sympatric ecotypes which were then classified as different species (Morin et al. 2010b). In speartooth shark and sawfish, mitogenome sequences comparison allowed to better resolve population structure and inform effective conservation measures for critically endangered populations (Feutry et al. 2014a; Feutry et al. 2015).

In sea turtles, female natal philopatry and population genetic structure have long been investigated by the analysis of partial mtDNA control region sequence diversity (Jensen et al. 2013). More recently, whole mtDNA analysis helped in the characterization of *Chelonia mydas* matrilines, which could not be resolved by comparison of control region haplotypes alone and allowed definition of fine-scale population structure in the critically endangered *Lepidochelys kempii* (Frandsen et al. 2020). For the loggerhead sea turtle (*Caretta caretta*), a few complete mitogenomes have been characterized worldwide and used to reconstruct interspecific phylogenies (Drosopoulou et al. 2012; Duchene et al. 2012; Hernández-Fernández and Delgado Cano 2018; Otálora et al. 2018). In particular, Drosopoulou et al. (2012) provided the first thorough description of the mtDNA genome of *C. caretta* using a sample collected from a rookery in western Greece, however, more detailed patterns of intraspecific mitogenomic variation were not investigated.

The loggerhead sea turtle can be found in tropical and subtropical seas as well as temperate waters of the Mediterranean basin, which hosts one of the 10 Regional Management Units defined for *C. caretta* to prioritize conservation efforts worldwide

(Wallace et al. 2010a). The most recent assessment of the IUCN Red List classifies the loggerhead turtles from the Mediterranean as a “least concern” Regional Management Unit. However, populations are conservation-dependent, severely fragmented and characterized by a continuous decline of mature individuals mainly because of habitat encroachment, pollution, fisheries bycatch, climate change and intentional take (Casale et al. 2018; Casale et al. 2020). A variation in the distribution of the species induced by global warming is already detectable (Maffucci et al. 2016; Carreras et al. 2018; Hochscheid et al. 2022) and advocated by projection modelling (Mancino et al. 2022). Conservation efforts can make use of a comprehensive understanding of several life history traits. In particular, the definition of maternal descent of individuals from foraging stocks and genetic profiling of nesting females and new nesting sites are fundamental to guide management programs for *C. caretta* in the Mediterranean.

Genetic characterization of Mediterranean rookeries and foraging stocks for the loggerhead has been so far conducted by comparison of partial mtDNA control region sequences (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). This information was used in conservation-oriented studies where, for instance, mixed stock analysis described the contributions from known rookeries to aggregations of turtles at sea (Garofalo et al. 2013; Clusa et al. 2014; Splendiani et al. 2017; Tolve et al. 2018; Loisier et al. 2021). To this regard, comparison of control region haplotypes has reached a resolution threshold which cannot allow any further sensible improvement in fine-scale population structure analysis since the main Mediterranean rookeries appear to all share the same control region haplotypes (e.g. Tolve et al. 2018). The very same molecular marker was also used to study patterns of loggerhead Mediterranean colonization and radiation (e.g. Clusa et al. 2013; Novelletto et al. 2016; Splendiani et al. 2017; Loisier et al. 2021). Nevertheless, limited information could be obtained on the evolution of different lineages because of the scarce resolution among matrilines defined by mtDNA control region sequences.

In this study, we produced a first whole mtDNA genome dataset for loggerhead turtles from the Central Mediterranean Sea in order to enhance resolution of matrilines and characterize rookeries which could be not distinguished by comparison of ubiquitous control region haplotypes. We obtained high coverage mitogenome sequences from dead embryos collected at nesting sites and from bycaught and stranded turtles. Eight out 27 nests were located in minor or occasional nesting areas, which were genetically described for the first time. We assessed intraspecific genetic diversity and performed phylogenetic

analysis to reconstruct evolutionary relationships among matrilines and the time frame of radiation of haplogroups and mtDNA haplotypes within haplogroups. Using additional, published loggerhead mtDNA genome sequences, we provided an overview of mitogenomic diversity for *C. caretta* in the Central Mediterranean. We argue that our findings will help supporting detailed inference in mixed stock analysis and foster thorough investigation of the phylogeographic history of loggerhead turtles.

Materials and methods

Sampling, DNA isolation and sequencing by synthesis

A total of 61 tissue biopsies were obtained from loggerhead turtles sampled along the coastline of the Italian peninsula and the island of Linosa and in the sea waters between Sicily and Tunisia from 2008 to 2021 (Fig. 1). Samples were collected from stranded turtles in Veneto and Emilia Romagna (N = 4), Tuscany (N = 1) and Latium (N = 2) and from bycaught turtles in the Strait of Sicily (N = 27). Samples were also collected from dead embryos from four nests in Tuscany, two nests in Latium, 19 nests in Calabria and two nests on the island of Linosa (Table S1). Numbered tags were applied to bycaught turtles before release.

Each sample was preserved in a cryovial with 95% ethanol and subsequently stored at -80 °C. Whole DNA was extracted following a standard phenol:chloroform:isoamyl alcohol procedure (Green and Sambrook 2012). Integrity of DNA was assessed by 1.5% agarose gel electrophoresis and DNA concentration was measured using a Qubit 4 fluorometer Broad Range Assay (Invitrogen). Mitochondrial DNA whole genome sequencing was performed by first generating dual-indexed (i5 and i7) genomic libraries using a PCR-free Tagmentation Kit and IDT for Illumina indexes set A (Illumina) according to the manufacturer's protocol. We envisaged a 1× target nuclear genome coverage for all samples in order to obtain a high depth coverage (> 200×) for the mitochondrial DNA, which is represented by a much higher number of copies per cell than nuclear DNA (Michaels et al. 1982; Robin and Wong 1988; Wiesner et al. 1992). Considering a nuclear DNA genome size of 2.13 Gb (Chang et al. 2023) and a mtDNA genome size of 16,737 bp for *C. caretta* (Drosopoulou et al. 2012), we sequenced the libraries paired-end on an Illumina NovaSeq 6000 System using a 300-cycle NovaSeq 6000 SP Reagent Kit v 1.5 for an expected yield of approximately 250 Gb.

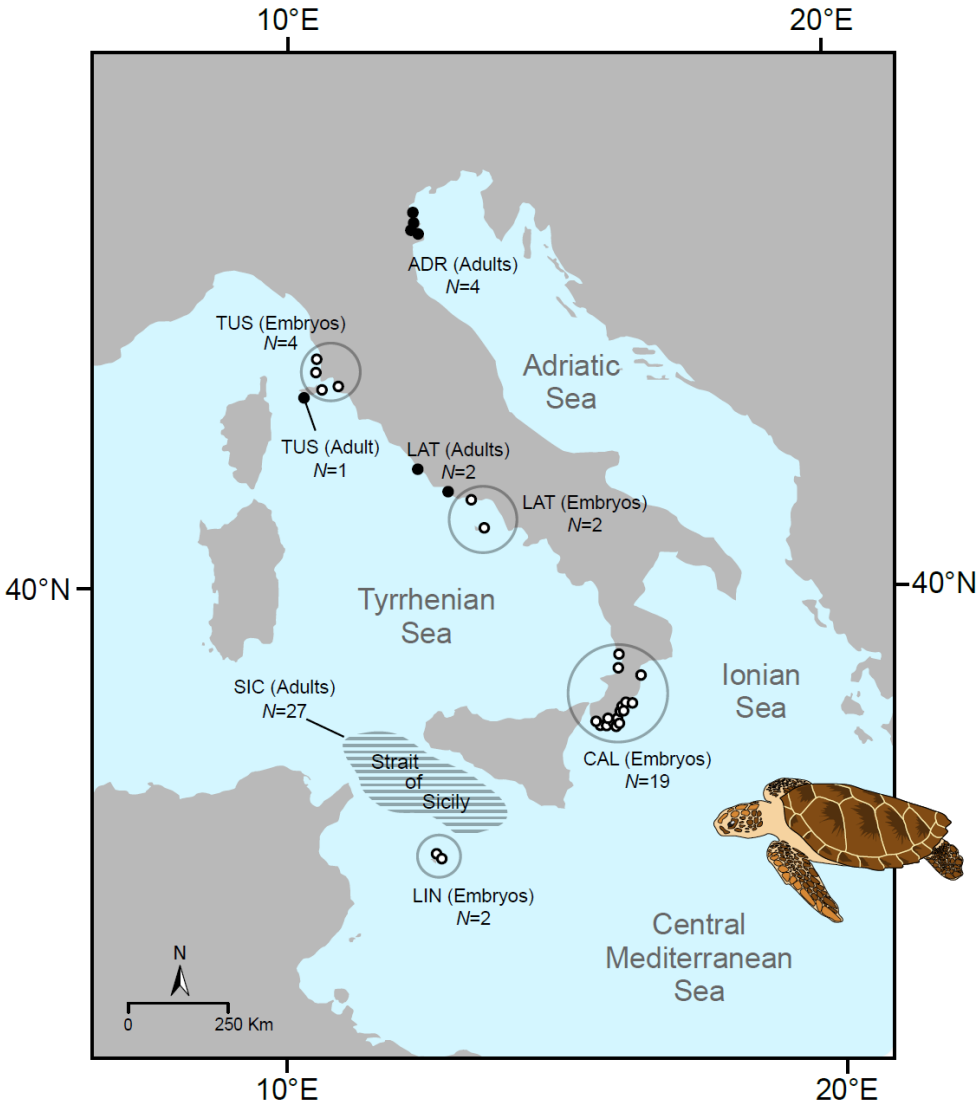


Fig. 1 Map of the study area. A total of 61 samples of *Caretta caretta* were collected between 2008 and 2021 along the coastline of the Italian peninsula and the Island of Linosa and in the sea waters of the Strait of Sicily. Nesting sites (circles) and stranded turtles (black dots) were sampled in Veneto and Emilia Romagna (ADR), Tuscany (TUS), Latium (LAT), Calabria (CAL) and the Island of Linosa (LIN). The horizontal shaded region shows the location of the bycatch area between Sicily and Tunisia.

Mitogenomes reconstruction and alignment

Demultiplexing and conversion of sequencing data from bcl to fastq format were performed using bcl2fastq version 2.20 (Illumina). FastQC 11.9 (Andrews 2010) was used to check for reads quality. Adaptors and low-quality reads were removed using Trimmomatic 0.39 (Bolger et al. 2014). Reads were mapped to the *C. caretta* complete annotated mitochondrial genome (GenBank accession number: FR694649.1) published by Drosopoulou et al. (2012) using Bowtie2-2.5.0 (Langmead and Salzberg 2012). This mitogenome was chosen as reference for it was the only one published for *C. caretta* from

the Mediterranean basin. Sequence Alignment Map (SAM) files were compressed in Binary Alignment Map (BAM) format, sorted and indexed using Samtools 1.16.1 (Danecek et al. 2021). Variant calling was performed with BCFtools 1.16 (Danecek et al. 2021) and Tablet alignment viewer (Milne et al. 2013) was used to visually check variants and relative coverage. Consensus sequences were extracted from Variant Calling Format (VCF) files using BCFtools by filtering out variants in the first 10 bp and those with a QUAL value < 30.

Mitogenomes alignment was performed by including five additional loggerhead mtDNA genome sequences available in GenBank (see also Table S2): FR694649.1 (Greece; used as reference mitogenome), JX454983.1 (Florida), JX454988.1 (Hawaii), MF554690.1 and MF579505.1 (Colombia) (Drosopoulou et al. 2012; Duchene et al. 2012; Hernández-Fernández and Delgado Cano 2018). The MUSCLE algorithm was used to produce a multiple sequence alignment, which was trimmed at position 16419, ahead of the (TATAT)n microsatellite repeat and annotated using Geneious Prime 2022.2. We then partitioned the alignment in i) 12S and 16S rRNAs, ii) tRNAs, iii) the control region and iv) coding genes. The final nucleotide dataset had no ambiguous positions and was numbered with reference to the FR694649.1 mitogenome (Drosopoulou et al. 2012). Genbank accession numbers are reported in Table S1 for each complete mtDNA haplotype sequence described in this study.

Mitogenomic haplotype nomenclature

Mitochondrial haplotypes were named following the criteria set by Shamblin et al. (2012a) and the guidelines of the Archie Carr Center for Sea Turtles Research (https://accstr.ufl.edu/resources/mtdna-sequences/). Complete mtDNA haplotypes were defined by three serial numbers. The first two numbers refer to the 815 bp long *C. caretta* control region haplotype code, while the third number refers to variants recorded in other genes of the mitogenome outside of the control region, so that complete mtDNA haplotype names are basically variants of the current control region haplotype nomenclature. For example, CC-A2.1.1, CC-A2.1.2 and CC-A2.1.3 all refer to complete mtDNA haplotype variants of the control region haplotype CC-A2.1.

Nomenclature for groups of complete mtDNA haplotypes also followed that of *C. caretta* control region variants, which identify three main haplotype clusters named haplogroups IA, IB, and II (Shamblin et al. 2014). Haplogroup IA includes haplotypes characterizing Pacific Ocean nests, haplogroup IB includes haplotypes which are found in

Atlantic rookeries and one single haplotype recorded in Indian Ocean nests (Shamblin et al. 2014). Finally, haplogroup II comprises haplotypes from both Atlantic and Mediterranean rookeries.

Mitogenomic diversity and phylogenetic analysis

Number of haplotypes, haplotypic diversity, number of variants, nucleotide diversity and GC content were computed using DnaSP (Librado and Rozas 2009) and the R package pegas v1.1 (Paradis 2010). Analysis was performed on whole mitogenome sequences and then single mitochondrial genes using the entire dataset, first and then on haplogroup II only. The McDonald-Kreitman test was implemented using DnaSP to check for neutrality of substitutions in the coding regions (McDonald and Kreitman 1991). The nonsynonymous (dN) to synonymous (dS) substitution rate ratio among haplogroup II sequences was then compared to the dN/dS ratio between haplogroup II sequences and mitogenomes of haplogroup IB (obtained by this study and from GenBank, accession number: MF579505.1). In order to rule out parallel patterns of nucleotide variations between haplogroup II and IB, the dN/dS ratio among haplogroup II variants was measured against the dN/dS between haplogroup II and the mitogenome of *Lepidochelys olivacea* as an outgroup (GenBank accession number: JX454987.1). Finally, differences in dN/dS among mitogenomes of the entire *C. caretta* dataset of our study were compared to dN/dS between our dataset and the mtDNA genome of *L. olivacea.*

Two haplotype networks were constructed using statistical parsimony (Templeton et al. 1992). The method, implemented in the TCS program (Clement et al. 2000), links haplotypes with the smaller number of differences as defined by a 95% confidence criterion. The first network was based on 815 bp control region haplotypes only, while the second one was a network of complete mitogenome haplotypes.

A Maximum Parsimony (MP) graph based on the entire dataset was also constructed using MEGA11 (Kumar et al. 2018) with the “use all sites” option in order to visualize similarities between sequences and polarize mutational changes. We then applied a Bayesian divergence time analysis to provide a time frame for the process of radiation of haplogroup II with respect to the other two major *C. caretta* mitochondrial haplogroups IA and IB. Phylogeny and divergence time analyses provided poor resolution when using *L. olivacea* as outgroup. We therefore conducted MP and Bayesian inference of phylogeny using haplogroup IB as reference in order to better resolve intraspecific topology of haplogroup II, which included the vast majority of haplotypes defined in our

study. We used BEAST suite v.1.10.4 (Suchard et al. 2018b) on four mitogenome partitions (12S and 16S rRNAs, tRNAs, the control region and coding genes). We performed 20 independent test-runs in order to optimize priors for chain convergence and defined, for each partition, the "unlink" option flagged for substitution model and for molecular clock and the "link" option flagged for generation of the tree. The HKY (Hasegawa et al. 1985) substitution model and a strict molecular clock were selected for each partition. The "unlink parameters on codon positions" option was flagged for the fourth partition. The HKY model was selected because the tests conducted using more complex substitution models gave a poor Effective Sample Size (ESS), possibly due to over-parametrization. Likewise, we chose a strict molecular clock because multiple tests conducted using more complex uncorrelated relaxed clocks resulted in poor ESS and zero values for the parameters uced.mean and meanRate, which indicate a *clock-like* behaviour of the data. No prior was set on the time to the most recent common ancestor (TMRCA) and no divergence time estimate was assessed for the root of the tree. We converted the substitutions value from (site × million year)$^{-1}$ to (site × generation)$^{-1}$, where the generation time is calculated as age of maturity plus half of reproductive longevity (Pianka 1973). We used 25 years as age of maturity and 22 years as reproductive longevity which resulted in a 36-year generation time (Margaritoulis et al. 2020; Baldi et al. 2023). We set a normal distribution (mean = 1.2E-7, S.D. = 4.0E-8) as prior on the substitution rates for each mitogenomic partition. The distribution was truncated with a lower bound equal to 0.0 and an upper bound of 2.41E-7 (corresponding to 6.7E-3 million year^{-1}), which is the highest value for mitochondrial substitution rates recorded in marine turtles (Dutton et al. 1999). Chain length was set to 10E8 steps with a 10% initial burn-in. We monitored the ESS for each output parameter after each run using Tracer v.1.7.1 (Rambaut et al. 2018). The overall BEAST output was summarized using TreeAnnotator (Drummond and Rambaut 2007) and the statistics related to the consensus tree nodes including node height, 95% Height Posterior Density (HPD) and posterior, were assessed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Mitogenomic diversity and complete mtDNA haplotype assignment

Whole mtDNA sequences were obtained for all 61 samples with an average genome coverage (×) of 272.4±37.1SE (see Table S3). Mitogenome sequences were trimmed at position 16419 so that the $(TATAT)_n$ microsatellite repeat positioned at the 3' end of the

control region was discarded from the analysis.

We recorded two mtDNA sequences in sample SIC04. We reported this pattern as a possible occurrence of heteroplasmy and assigned to SIC04 the mitogenomic haplotype name of the mtDNA variant (G>A at position 4671 of the genome) that was confirmed in 70% of the reads (Table S1).

All samples belonged to haplogroup II apart from the stranded adult TUS05, of unknown origin, which was assigned to haplogroup IB, control region haplotype CC-A1.1. We named this mitogenomic haplotype CC-A1.1.1. Downloaded mitogenome sequences FR694649.1 (our reference genome), JX454983.1 and MF554690.1 were assigned to haplogroup II, while MF579505.1 and JX454988.1 were assigned to haplogroups IB and IA, respectively (Table S2).

Haplogroup II sequence variants and their distribution across sampling sites are reported in Table 1 and Table 2, respectively. The loggerhead mitogenomes characterized in this study showed a total of 27 polymorphic sites with respect to the *C. caretta* reference mtDNA FR694649.1. Of these, two segregating sites were recorded across all 61 samples of our dataset and in particular an indel in position 1631 in the 16S rRNA gene and a T>C substitution in position 9974 in the ND4L gene (Table1). Mitogenomic haplotype CC-A2.1.1 was the most abundant both in embryos and adult turtles (found in 26 out of 61 samples). The second most abundant haplotype was CC-A2.1.8, found in six samples from the Calabrian nests, one sample from the Island of Linosa and three samples collected from bycaught adults in the Strait of Sicily. Seven embryos from different nests in Calabria carried mitogenomic haplotype CC-A20.1.1. This mitogenome was not found elsewhere in the study area. Haplotype CC-A2.1.3 was found in three specimens in Tuscany and southern Italy. All the other mitogenomic haplotypes were represented by either one or two individuals. The nests from Tuscany, Latium and Linosa Island were characterized by haplotypes CC-A2.1.1 (TUS01, TUS02, LAT01, LAT02), CC-A2.1.3 (TUS03), CC-A2.1.9 (TUS04), CC-A2.1.8 (LIN01) and CC-A2.9.2 (LIN02) (Table S1).

Overall, we characterized 286 variable sites with respect to 55 polymorphisms recorded by comparing control region sequences alone and assigned samples to 23 different mitogenomic haplotypes, while 10 haplotypes only were defined by control region sequence analysis (Table 3). Within haplogroup II, mitogenome sequences identified 27 variable sites and 20 haplotypes compared to six variable sites and seven haplotypes recorded by control region sequence comparison (Table 3; Fig. 2).

Genetic diversity analysis of whole mtDNA revealed a much higher number of haplotypes and higher haplotypic and nucleotide diversity than values recorded by control region sequences comparison (Table 3). Among the 16 gene partitions of the *C. caretta* mitogenome, the ND5 gene showed the highest number of segregating sites and haplotypic diversity (Tables S4 and S5). The same pattern was recovered by the analysis of haplogroup II mitogenomes for the high abundance (28%) of mitogenomic haplotypes CC-A2.1.8 and CC-A2.1.20 carrying the variant T at site 13373 of the ND5 gene. Overall, nucleotide diversity (π) showed a 3-fold variation along the mtDNA molecule, with peak values at positions 3000 to 5000, 9000 to 10000 and 13000 to 14000 (Fig. S1). When haplogroup II sequences only were considered, π values were highest at positions 9000 to 10000 and 13000 to 14000. The NADH dehydrogenase and ATP synthase subunit genes displayed the highest π values. Conversely, analysis restricted to haplogroup II mitogenomic haplotypes showed the highest π values for the ATP8 gene only (Tables S4 and S5).

The TCS haplotype networks clustered both the partial mtDNA control region sequences and the mitogenomes in three unlinked haplogroups with 95% probability (Fig. 2). Links between haplogroups were obtained with a lower probability threshold by fixing connection limits at 100 and 200 steps for control region and mitogenome haplotypes, respectively. Two internal ambiguities in the mitogenome cladogram were resolved by applying criteria derived from empirical validation of the coalescent theory (Crandall and Templeton 1993; Templeton and Sing 1993; Fetzner and Crandall 2003). Accordingly, a haplotype connection was maintained to high-frequency haplotypes with an interior position in the network (Castelloe and Templeton 1994) rather than to low-frequency ones, located in tip clades. Internal ambiguities could not be resolved for the mtDNA control region haplotypes network. We therefore maintained the three haplogroups unlinked. Haplogroup IA was represented by a single sequence in both reconstructions, while haplogroup IB included two haplotypes separated by 12 and 30 possible unobserved intermediate haplotypes in the partial mtDNA control region and mitogenomic network, respectively. The haplogroup II network based on control region sequences had a star-like configuration with a core CC-A2.1 haplotype and six equidistant derived haplotypes. A similar but more complex pattern was described by the mitogenomic analysis. Haplogroup II showed 10 equidistant haplotypes deriving from CC-A2.1.1, two haplotypes (CC-A2.1.10 and CC-A20.1.1) deriving from CC-A2.1.3 and

CC-A2.1.8, respectively and five additional branches with haplotypes linked to CC-A2.1.1 by unobserved intermediate haplotype states.

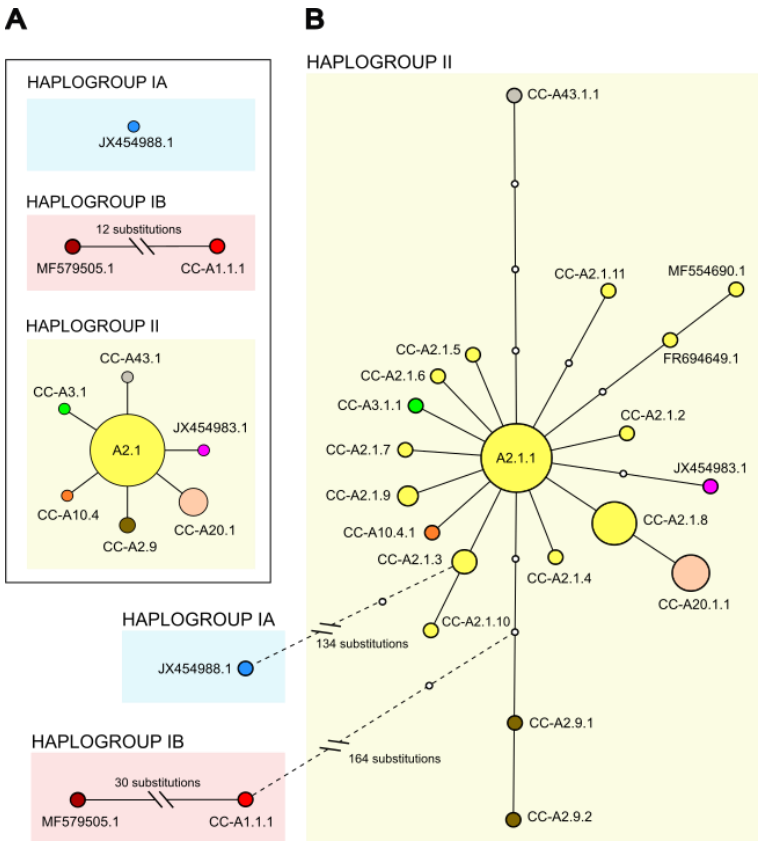


Fig. 2 TCS haplotype networks based on partial mtDNA control region (A) and complete mitogenome (B) sequences. Dashed lines represent links with probability lower than 95%.

Table 1 Mitochondrial genome polymorphic sites for *Caretta caretta* haplogroup II mitogenomic haplotypes. Numbers refer to nucleotide positions of the *C. caretta* reference mtDNA FR694649.1 (in grey).

mtgenomic haplotypes	mtDNA gene													
	16S	ND1		ND2		Cys	COI	COII	ATP8	ATP6		COIII		Gly
	1631	2936	3045	4277	4671	5303	5847	7433	7900	8398	8661	9102	9440	9482
FR694649.1	T	C	C	G	G	A	C	C	C	C	T	G	T	A
MF554690.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
JX454983.1	–	.	.	.	.	.	A	.	.	.	.	.	.	.
CC-A2.1.1	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.2	–	.	.	.	A	.	.	.	.	.	.	.	.	.
CC-A2.1.3	–	.	.	.	.	.	.	.	T	.	.	.	.	.
CC-A2.1.4	–	.	.	.	.	.	.	.	.	.	C	.	.	.
CC-A2.1.5	–	.	.	.	.	.	.	.	.	.	.	A	.	.
CC-A2.1.6	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.7	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.8	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.9	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.10	–	T	.	.	.	.	.	.	T	.	.	.	.	.
CC-A2.1.11	–	.	T	.	.	.	.	.	.	.	.	.	.	G
CC-A2.9.1	–	.	.	.	.	.	.	.	.	T	.	.	A	.
CC-A2.9.2	–	.	.	.	.	G	.	.	.	T	.	.	A	.
CC-A3.1.1	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A10.4.1	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A20.1.1	–	.	.	.	.	.	.	.	.	.	.	.	.	.
CC-A43.1.1	–	.	.	A	.	.	.	T	.	.	.	.	.	.

Table 1 Continued

mtgenomic haplotypes	mtDNA gene												
	ND3	ND4L	ND4	Leu	ND5		CYTB	Control region					
	9610	9974	10591	11800	13287	13373	14373	15562	15645	15731	15800	15970	15975
FR694649.1	T	T	T	G	G	C	G	T	C	A	G	A	A
MF554690.1	.	.	.	A	.	.	.	.	.	.	.	.	.
JX454983.1	.	C	.	.	.	.	.	.	.	.	.	.	–
CC-A2.1.1	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.2	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.3	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.4	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.5	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.6	.	C	C	.	.	.	.	.	.	.	.	.	.
CC-A2.1.7	.	C	.	.	A	.	.	.	.	.	.	.	.
CC-A2.1.8	.	C	.	.	.	T	.	.	.	.	.	.	.
CC-A2.1.9	.	C	.	.	.	.	A	.	.	.	.	.	.
CC-A2.1.10	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.1.11	.	C	.	.	.	.	.	.	.	.	.	.	.
CC-A2.9.1	.	C	.	.	.	.	.	C	.	.	.	.	.
CC-A2.9.2	.	C	.	.	.	.	.	C	.	.	.	.	.
CC-A3.1.1	.	C	.	.	.	.	.	.	.	.	A	.	.
CC-A10.4.1	.	C	.	.	.	.	.	.	.	.	.	G	.
CC-A20.1.1	.	C	.	.	.	T	.	.	T	.	.	.	.
CC-A43.1.1	C	C	.	.	.	.	.	.	.	G	.	.	.

Table 2 Number of whole mtDNA and control region haplotypes recorded for 61 loggerhead samples collected along the coastline of the Italian peninsula and the island of Linosa and in the sea waters of the Strait of Sicily.

mtDNA	Control region	ADR	TUS	LAT	CAL	LIN	SIC	Total
CC-A1.1.1	CC-A1.1	0	1	0	0	0	0	1
CC-A2.1.1	CC-A2.1	2	2	3	4	0	15	26
CC-A2.1.2		0	0	0	0	0	1	1
CC-A2.1.3		0	1	0	1	0	1	3
CC-A2.1.4		0	0	0	0	0	1	1
CC-A2.1.5		0	0	0	0	0	1	1
CC-A2.1.6		0	0	1	0	0	0	1
CC-A2.1.7		0	0	0	1	0	0	1
CC-A2.1.8		0	0	0	6	1	3	10
CC-A2.1.9		0	1	0	0	0	1	2
CC-A2.1.10		0	0	0	0	0	1	1
CC-A2.1.11		0	0	0	0	0	1	1
CC-A2.9.1	CC-A2.9	1	0	0	0	0	0	1
CC-A2.9.2		0	0	0	0	1	0	1
CC-A3.1.1	CC-A3.1	0	0	0	0	0	1	1
CC-A10.4.1	CC-A10.4	1	0	0	0	0	0	1
CC-A20.1.1	CC-A20.1	0	0	0	7	0	0	7
CC-A43.1.1	CC-A43.1	0	0	0	0	0	1	1
Total		4	4	2	19	2	27	61

ADR, Veneto and Emilia Romagna; TUS, Tuscany; LAT, Latium; CAL, Calabria; LIN, Island of Linosa; SIC, Strait of Sicily.

Table 3 Genetic diversity parameters for whole mitogenomes and control region sequences only. Values are reported for 61 samples plus five mitogenomes from published literature and a subset of haplogroup II mtDNA only (60 samples from this study and three sequences from published literature).

		N	Sequence length (bp)	*k*	*P*	*h*±SE	π±SE	GC
All samples	mtDNA	66	16,425	23	286	0.816±0.005	9.44E-4±5.82E-5	0.392
	Control region	66	919	10	55	0.419±0.009	3.26E-3±2.46E-4	0.340
Haplogroup II	mtDNA	63	16,425	20	27	0.798±0.006	9.58E-5±8.09E-10	0.392
	Control region	63	919	7	6	0.362±0.009	3.92E-4±2.29E-8	0.340

N, number of samples; *k*, number of haplotypes; *P*, number of polymorphic sites; *h*, haplotypic diversity ± Standard Error; π, nucleotide diversity ± Standard Error; GC, content of GC bases.

Variation in the protein-coding regions

The McDonald-Kreitman neutrality test returned no significant values when comparing dN/dS of haplogroup II with dN/dS between haplogroup II and either haplogroup IB or the mitogenome of *L. olivacea* (Table S4). We found an excess of synonymous substitutions in all genes when comparing haplogroup II with haplogroup IB sequences and a similar result was obtained when comparing haplogroup II variants to the mitogenome of *L. olivacea*. Ten variants in the coding regions of haplogroup II sequences were synonymous and six variants were non-synonymous. In particular, ND2, ND3 and CYTB genes displayed no synonymous substitutions and a non-synonymous one. Comparison of dN/dS of our complete *C. caretta* dataset with dN/dS between *C. caretta* mitogenomes and *L. olivacea* mtDNA revealed an excess of non-synonymous substitutions for ND1, ND2, ATP6, ND3, ND5 and CYTB genes. Nevertheless, a significant value was shown for the ND3 gene only.

Phylogenetic analyses

The phylogenetic tree inferred by Maximum Parsimony resolved equally parsimonious topologies whereby haplogroup II (Atlantic and Mediterranean) and haplogroup IA (Pacific) shared the same cluster, while haplogroup IB (Atlantic) included the most phylogenetically divergent sequences (Fig. 3). All haplogroup II sequences coalesced to a long branch defined by 55 substitutions. Of these, 18 substitutions occurred in the control region, three were found in rRNA genes, one in a tRNA and 33 in coding genes with 28 synonymous and five non-synonymous substitutions (Table S6).

Two main clades were defined within haplogroup II (Fig. 3). The clade characterized by the rare Mediterranean control region haplotype CC-A2.9 differed from the other haplogroup II sequences by the control region variant C at position 15562, the derived state C>T at position 8398 in the ATP6 gene and the ancestral state A at position 9440 in the COIII gene. Mitogenome analysis further distinguished mitogenomic haplotypes CC-A2.9.1 and CC-A2.9.2 as a result of a A/G variant at position 5303 in the Cys tRNA gene (Fig. 3, Table S7). The second and most conspicuous clade was defined by a basal A>T transversion at position 9440 and included all CC-A2.1 mitogenomic variants as well as haplotypes corresponding to control region haplotypes CC-A3.1, CC-A10.4, CC-A20.1 and CC-A43.1.

Of the CC-A2.1 mitogenomic variants, six of them (CC-A2.1.2 to CC-A2.1.7 and CC-A2.1.9) derived each by one private variant from the basal CC-A2.1.1 mitogenomic haplotype. Similarly, haplotypes CC-A3.1.1 and CC-A10.4.1 each differed from CC-A2.1.1 by a single control region variant. Mitogenomic haplotype CC-A2.1.10 stemmed from CC-A2.1.3 by a synonymous substitution in the ND1 gene, while haplotype CC-A2.1.11 differed from the basal CC-A2.1.1 by two variants. Mitogenomic haplotype CC-A43.1.1 was the most divergent sequence characterized by two non-synonymous, two synonymous substitutions and the control region variant. Haplotype CC-A20.1.1 derived from CC-A2.1.8, which in turn stemmed from the basal clade by a C>T variant in position 13373 in the ND5 gene (Table S7). All CC-A20.1 sequences were identical, denoting a recent common origin. Finally, the Atlantic mitogenomes JX454983.1 and MF554690.1 derived from CC-A2.1.1 by a single synonymous substitution and from the Mediterranean reference FR694649.1, respectively (Fig. 3, Tables S6 and S7).

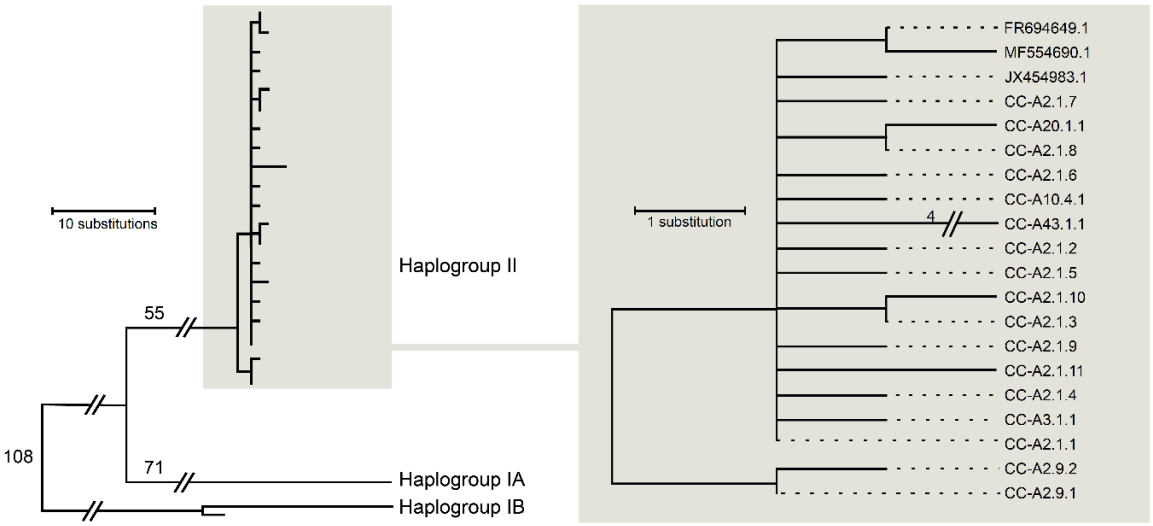


Fig. 3 Unrooted maximum parsimony (MP) tree of *Caretta caretta* obtained using 61 mitogenomes from this study and five sequences from published literature (left). Haplogroup IA (Pacific Ocean) corresponds to Genbank accession number JX454988.1. Haplogroup IB (Atlantic Ocean) includes Genbank accession number MF579505.1 and mitogenomic haplotype TUS05 from this study (see Tables S1 and S2). Details of the haplogroup II subtree (right) show phylogenetic relationships among mtDNA sequences of 60 samples from this study (Mediterranean Sea) and three sequences from published literature (Atlantic Ocean). Numbers on tree branches represent number of nucleotide substitutions.

The Bayesian tree confirmed the branching order of the haplogroups recovered by the MP tree (Fig. 4 and Table 4). The root of the tree (node I) separated haplogroup IB from haplogroups IA and II. The second node, positioned at 67% of the height of the total tree, divided haplogroup IA from haplogroup II mitogenomic haplotypes. According to Bayesian phylogeny, radiation of haplogroup IB (node III) occurred well before the branching of haplogroup II, which occupies only 4% of the total tree height. The Bayesian analysis also confirmed the basal position of the CC-A2.9 clade within haplogroup II. The only discrepancy between MP phylogeny and Bayesian reconstruction was the position of mitogenomic haplotype CC-A43.1.1, which resulted basal, instead of derived, with respect to CC-A2.1 variants. Other nodes that reached strong support (> 95% posterior probability) were those bifurcating CC-A20.1.1 and CC-A2.1.8, CC-A2.9.1 and CC-A2.9.2, CC-A2.1.3 and CC-A2.1.10, MF554690.1 and FR694649.1. All these nodes confirmed the phylogenetic clustering of the MP tree. The radiation of these mitogenomic haplotypes resulted to be a very recent event for they occupied just 1% of the total tree height. In particular, the most recently diverged mitogenomic sequences were FR694649.1 and MF554690.1, sampled in Greece and in the Atlantic Ocean, respectively (Table 4). Substitution rate estimates obtained for each mitogenome partition

are reported in Table S8. The highest substitution rate (number of substitutions × (site × million year)$^{-1}$) was recorded for the control region. The coding regions had a substitution rate three times lower than the control region while the rRNAs' partition showed the lowest values.

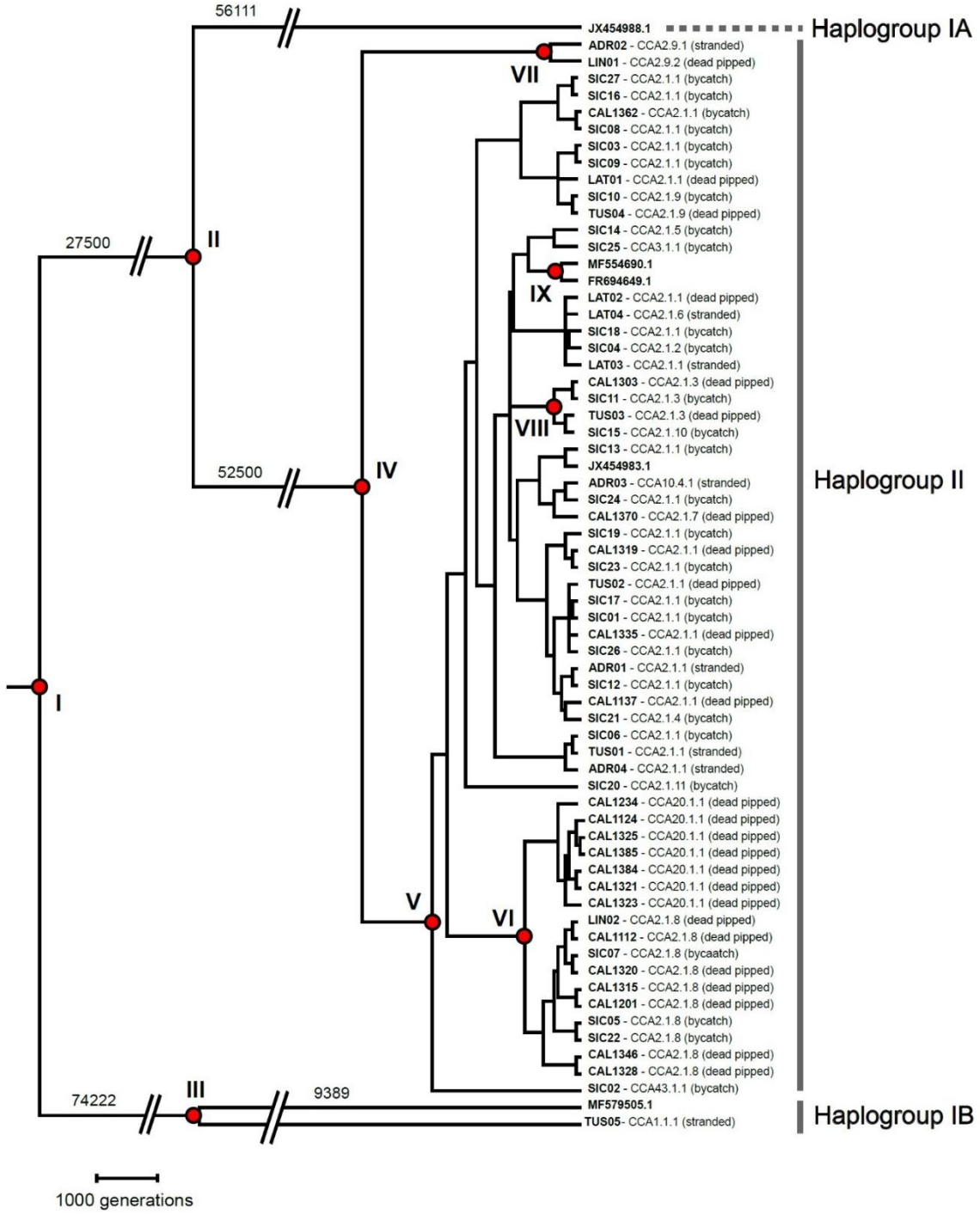


Fig. 4 Bayesian tree of *Caretta caretta* based on 61 mitogenomes from this study and five sequences from published literature. Nodes supporting > 95% posterior probability are shown in red. BEAST coerces multifurcations of identical sequences, so that nodes separating sequences assigned to the same

mitogenomic haplotype should not be considered. Numbers on tree branches represent number of generations. Sample names and haplotypes assignment as for Fig. 1 and Table S1, respectively.

Table 4 Bayesian tree's height estimates and 95% Height Posterior Density (HPD) for tree nodes with >90% support. Estimates are reported in number of *Caretta caretta* generations (1 generation = 36 years) and in thousands of years before present (KYBP).

Node	Node description	Median (generations)	Median (KYBP)	95% HPD (generations)	95% HPD (KYBP)
I	Haplogroup IB – Haplogroups IA and II	83611	3.01E3	49444 – 132500	1.78E3 – 4.77E3
II	Haplogroup IA – Haplogroup II	56111	2.02E3	32222 – 902777	1.16E3 – 3.25E3
III	CCA1.1.1 – MF579505.1	9389	338	4500 – 16555	162 – 596
IV	Haplotypes CCA2.9 – other mitotypes in Haplogroup II	3611	130	1555 – 6388	56 – 230
V	CCA43.1.1 – other mitotypes in Haplogroup II	2444	88	1166 – 4305	42 – 155
VI	CCA20.1.1 – CCA2.1.8	917	33	278 – 1889	10 – 68
VII	CCA2.9.1 – CCA2.9.2	500	18	8 – 1694	0.28 - 61
VIII	CCA2.1.3 – CCA2.1.10	389	14	3 – 1778	0.10 – 64
IX	MF554690.1 – FR694649.1	278	10	3 – 833	0.10 – 30

Discussion

Improved resolution in loggerhead mitochondrial genetic diversity

We characterized the mitochondrial genomes of 61 loggerhead turtles sampled between 2008 and 2021 along the coastline of Italy, on the Island of Linosa and in sea waters of the Strait of Sicily and provided a first dataset on whole mtDNA sequences for Mediterranean *C. caretta* populations. Mitogenome analysis allowed for an improved resolution of loggerhead matrilines with respect to previous studies based on partial mtDNA control region sequences comparison. The TCS haplotype networks based on statistical parsimony delineated similar star-like topologies in haplogroup II for both partial mtDNA control region and complete mtDNA analysis. However, statistical parsimony showed the increased resolution of the mitogenome haplotype network in respect to the 815 bp control region reconstruction, particularly for haplotype CC-A2.1.

The control region haplotype CC-A2.1 is the most abundant among the 13 main rookeries so far described in the Mediterranean Sea and it is also widespread in Atlantic nesting sites (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). The ubiquity of haplotype CC-A2.1 has often hampered the characterization of rookeries for turtles of unknown origin. In particular, mixed stock analyses appeared to suffer from large confidence intervals associated to estimates of contribution by rookeries to specific mixed stocks. Confidence intervals with a lower value equal to zero most probably underestimated contribution of minor but perhaps genetically important rookeries (Garofalo et al. 2013; Clusa et al. 2014; Splendiani et al. 2017; Tolve et al. 2018; Loisier et al. 2021). In our study, 52 individual turtles all shared the same control region haplotype CC-A2.1 while whole mitogenome analysis recovered 11 distinct mitogenomic haplotypes (CC-A2.1.1-CC-A2.1.11), which significantly improved genetic characterization of geographically distinct nesting sites. While the CC-A2.1.1 variant resulted the most abundant in our dataset with 28 records out of a total of 61 samples, characterization of the other 10 mitogenomic haplotypes enabled distinction of nesting sites from approximately 30 km down to only 5 km apart along the coastline of Tuscany, Latium and Calabria. These nests could not be distinguished by mtDNA control region sequences analysis alone.

Similarly, haplotype CC-A1.1, which was carried by TUS05 and never found in Mediterranean rookeries, is very abundant and widespread in the Atlantic and therefore poorly informative of turtles' provenance (Shamblin et al. 2012b). It is well known that CC-A1.1 turtles enter the western Mediterranean to forage and then migrate back to the Atlantic for reproduction (Bolten et al. 1998; Laurent et al. 1998; Carreras et al. 2006; Carreras et al. 2011). It may therefore be difficult to envisage a Mediterranean origin for TUS05. Nevertheless, future mtDNA characterization for Atlantic and Mediterranean loggerheads may unveil a number of mitogenomic variants of CC-A1.1 and allow description of differences in mtDNA genome frequencies across rookeries to improve resolution of turtles' origin.

Our mitogenomic approach also improved the resolution of matrilines based on mtDNA control region haplotype CC-A2.9. This is a rare Mediterranean haplotype so far recorded only in Libya and in a small Israelian rookery. The two mitogenomic haplotypes CC-A2.9.1 and CC-A2.9.2 were found in a stranded adult in the northern Adriatic Sea and most important in one nest on the Island of Linosa, respectively. This is also the first

evidence of a female carrying the rare haplotype nesting in a rookery other than Libya or Israel.

Mitogenomes comparison showed that in Calabrian nests there are at least five different mitogenomic haplotypes instead of only two variants (CC-A2.1 and CC-A20.1) detected by the analysis of mtDNA control region sequences. Of these, haplotype CC-A20.1.1 remains exclusive to the Calabrian rookery as for previous studies reporting on mtDNA control region haplotype CC-A20.1. However, we did not record mitogenomes characterized by control region haplotype CC-A31.1, also private to Calabrian nests (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Clusa et al. 2014).

Genetic characterization of occasional and minor nesting sites

Major Mediterranean loggerhead rookeries are found in the eastern and southern sections of the basin from Greece through Turkey, Cyprus, Lebanon, Israel to Libya (Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). The western and central parts of the Mediterranean have been so far home of occasional nests, and it was only in the last decade that nesting events were recorded on a more frequent and regular basis (Casale et al. 2018). According to Hochscheid et al. (2022), the number of turtle nests west of the main rookeries has increased from three nests per year in 2013 to 84 nests recorded in 2020, suggesting a 1300 km north-westward range expansion of loggerhead nesting distribution (Mancino et al. 2022). Samples collected from nests in Tuscany and Latium are evidence of such latitudinal dispersal, which seems now possible also as a result of a steady increase in water temperature (Casale et al. 2018; Hochscheid et al. 2022). All four nests recovered in Tuscany and the two nests from Latium had the same control region haplotype CC-A2.1, a highly shared sequence across both the Mediterranean basin and the Atlantic. On the other hand, analysis of mitogenome sequences recovered the presence of at least three distinct nesting females in this region of the northern Tyrrhenian Sea. Of these, a female that laid eggs in southern Tuscany (TUS03, see Table S1) carried mitogenomic haplotype CC-A2.1.3, which was also found in embryos from a nest on the Ionian coast of the Calabrian rookery (CAL1303). This pattern may suggest a southern Italian origin of the occasional nests recorded in Tuscany possibly as a result of the ongoing northward expansion of the distribution range of loggerheads in the Mediterranean (Mancino et al. 2022). As much as our evidence is based on a single observation, it nevertheless provides valuable information to further investigate dispersal of pioneer female loggerheads and the propagation of specific

matrilines for the establishment of new rookeries.

In the Pelagian archipelago, Mingozzi et al. (2006) recorded nesting events on the Islands of Lampedusa and Linosa between 1985 and 1999 and regular, but not annual, nesting activities from 2000 to 2004. Nevertheless, genetic data was not produced during these surveys, so that our represents the first study characterizing the mitochondrial signature of the region. The two nests sampled in 2006 and 2008 were assigned to different matrilines, suggesting a relatively high degree of genetic diversity considering the limited size (540 ha) of Linosa Island. This is nevertheless coherent with the mitogenomic analysis of samples collected in the nearby Strait of Sicily, where 11 mitogenomic haplotypes were recorded in 27 bycaught adult loggerheads. Additional surveys should assess the relevance of Linosa island as an additional nesting area for female loggerheads carrying a rare CC-A2.9.2 haplotype, given that the CC-A2.9 matriline had so far been recorded in Libyan and Israelian nests only and in several individuals foraging in Mediterranean waters (e.g. Garofalo et al. 2013, Splendiani et al. 2017; Tolve et al. 2018). Moreover, the higher resolution offered by a mitogenomic approach could allow analysis of the genetic diversity and genomic differentiation between Linosa and the Island of Lampedusa, for which genetic characterization of nesting sites is yet to be provided.

Heteroplasmy occurrence and patterns of substitutions in the coding regions

The deep sequencing coverage we obtained for *C. caretta* mitochondrial genomes revealed the presence of two mtDNA variants in one individual bycaught in the Strait of Sicily (SIC04). Heteroplasmy consisted in the coexistence of the CC-A2.1.1 mitogenomic haplotype with a mitogenome sequence carrying a truncating variant at mtDNA position 4671 on the ND2 gene, turning a tryptophan codon into a stop codon. The tryptophan codon located in the ND2 gene is a well conserved sequence across vertebrates (Avise 1986) and the presence of a termination codon instead of a coding sequence during translation would suggest a possible detrimental effect on gene functioning (Yan et al. 2019). Although a mitogenome carrying such variant may not persist in homoplasy, we assigned a distinct mitogenomic haplotype name (CC-A2.1.2) to specimen SIC04. Tikochinski et al. (2020) argued how intra-individual heteroplasmy may not be such a rare occurrence in sea turtles and could instead play a role in the evolution of population genetic diversity across generations. Unfortunately, the limitations of Sanger sequencing in the characterization of low-frequency variants have impeded the development of

studies on heteroplasmy in sea turtles. Nevertheless, regular use of high-throughput sequencing techniques should now lead to an increase in studies investigating this phenomenon and its evolutionary implications.

We found the highest nucleotide diversity in NADH dehydrogenase subunits genes. A similar pattern was previously recorded for genes encoding the ND1 and ND3 subunits (Novelletto et al. 2016). In particular, the ND5 subunit gene showed a relatively high proportion of polymorphism and the highest haplotype diversity. We also found a high nucleotide diversity in the genes encoding the subunits of ATP synthases across mitogenomes of haplogroup II, suggesting a pattern of sequence radiation within this clade. However, we did not record a significant excess of amino acid substitutions with respect to synonymous substitutions in haplogroup II mitogenomes. This would argue against adaptive selection acting on haplogroup II.

Mediterranean loggerhead phylogeography

The topologies of both Maximum Parsimony and Bayesian trees showed a marked star-like phylogeny of the main clade within haplogroup II, suggesting a fast and recent radiation of mitogenomic haplotypes from CC-A2.1.1 to CC-A2.1.10, CC-A3.1.1, CC-A10.4.1, CC-A20.1.1 and CC-A43.1.1. On the other hand, the sister branch defined by haplotypes CC-A2.9.1 and CC-A2.9.2 was basal in our phylogenetic reconstruction. A similar pattern was found for the CC-A2.9 mtDNA control region haplotype (Novelletto et al. 2016; Splendiani et al. 2017; Loisier et al. 2021), which was until now considered to be exclusive to Libyan and Israelian rookeries. This pattern supports the hypothesis of Clusa et al. (2013) by which the first loggerhead colonies of the Mediterranean settled in Libya during the Pleistocene, approximately 65 KYBP. The southern Mediterranean was probably a glacial refugium from where loggerhead turtles would have dispersed after the Last Glacial Maximum (LGM). Although our dataset only partially represents the Mediterranean mtDNA diversity, the time estimate we obtained for the divergence of the CC-A2.9 haplotype clade was consistent with a Pleistocenic colonization of Libya. Our phylogeny did not show additional matrilines deriving from mtDNA carrying the control region haplotype CC-A2.9, which was previously recorded to have a close phylogenetic relationship with haplotypes CC-A26.1 and CC-A66.1 (Novelletto et al. 2016). A more comprehensive sampling including Ionian and eastern Mediterranean rookeries may reveal additional mitogenomic variants of the CC-A2.9 clade, which has so far been reported to have a limited radiation with respect to the sister branch CC-A2.1 (Garofalo et

al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014).

A more recent, second colonization event of the Mediterranean from the Atlantic during the Holocene was suggested by (Clusa et al. 2013) to explain the origin and expansion of the Calabrian rookery in southern Italy. This hypothesis was based on the presence of mtDNA control region haplotype CC-A20.1 in both Calabria and three rookeries in Florida (Shamblin et al. 2012b). The mitogenomic haplotype CC-A20.1.1 described in our study shared a recent ancestor with CC-A2.1.8 and both were recorded only in the southern part of our study area including Calabria, Linosa Island and the Strait of Sicily. This may suggest a local origin of these matrilines after the end of the LGM. Comparison of whole mitogenomic sequences from both Mediterranean and Atlantic stocks will be crucial to help understand origin and dispersal patterns of the CC-A20.1 clade.

On a wider, trans-oceanic scale, a number of biogeographical scenarios have been suggested advocating a first colonization of the Mediterranean Sea by haplogroup II from Atlantic rookeries (Shamblin et al. 2014) or through a South African route (Bowen et al. 1994; Baltazar-Soares et al. 2020). Our topology recovered by both maximum parsimony and Bayesian analysis provided preliminary information on mitogenome shared ancestry. We envisage that further analysis based on e.g. an Approximate Bayesian Computation approach (see Baltazar-Soares et al. 2020) that makes use of whole mtDNA sequences from a wider sample set including individuals from the Atlantic, could help better elaborate current hypotheses and build a robust phylogeographic scenario of colonization and dispersal of loggerheads in the Mediterranean Sea.

Conclusions

Analysis of complete mitogenome sequences can improve population genetic and phylogeographic studies, particularly for marine species that exhibit female natal philopatry. In this study, we investigated patterns of mtDNA diversity in loggerhead sea turtles and presented a high-coverage, complete mitogenome dataset for *C. caretta* in the central Mediterranean Sea along with a first genetic description of the Italian nesting sites of Tuscany, Latium and the Island of Linosa. We also proposed the first nomenclature for *C. caretta* mitogenomic haplotypes and described a larger number of variants than those so far recovered by mtDNA control region sequence comparisons. Mitogenome analysis provided a more comprehensive definition of matriline phylogeny and improved genetic

characterization of rookeries. Our study represents a first example of the potential of whole mitogenome analysis for the study of the evolutionary history of Mediterranean loggerheads and fine-scale characterization of nesting sites to help mixed-stock analysis and the conservation of extant rookeries.

Acknowledgements

We are grateful to the following collaborators involved in sample collection: Nicola Novarini of the Natural History Museum of Venice, Lisa Poppi, Caterina Raso, Luca Marini and collaborators of the NGO TartaLazio and all voluntary workers and former students involved in the TartaCare Project of Calabria University including Sandra Berlingeri, Giulia Cambiè, Francesca Crispino, Concetta Denaro, Simona Fabiano, Elena Giardinazzo, Teresa Malito, Carmela Mancuso, Giampiero Masciari, Alessandro Massolo, Nunzia Micò, Gianni Parise, Brunella Pisani, Manuela Policastrese, Patrizia Rima, Manuela Russo, Salvatore Salice and Salvatore Urso. We also thank Luana Papetti and collaborators of the NGO tartAmare and Isa Tonso and collaborators of Legambiente Arcipelago Toscano for support during collection of pipped eggs from nesting sites in Tuscany. A special thanks to Enrico Negrisolo for valuable suggestions during the preparation of the manuscript.

Supplementary materials

Table S1 Details of biopsies obtained from 61 *Caretta caretta* individuals sampled along the coastline of the Italian peninsula and the island of Linosa and in the sea waters between Sicily and Tunisia from 2008 to 2021. Genbank accession numbers for each mitogenome haplotype are also reported.

Year	Region or Sea basin	Sample ID	Age class	Turtle condition	Location	Mitogenome haplotype	Genbank accession number
2013	Veneto	ADR01	Adult	Stranded	Chioggia (VE)	A2.1.1	OR166582
2013	Veneto	ADR02	Adult	Stranded	Chioggia (VE)	A2.9.1	OR166583
2013	Veneto	ADR03	Adult	Stranded	Rosolina (RO)	A10.4.1	OR166584
2013	Emilia Romagna	ADR04	Adult	Stranded	Lugo (RA)	A2.1.1	OR166585
2018	Tuscany	TUS01	Hatchling	Dead pipped	Santa Lucia (LI)	A2.1.1	OR166586
2020	Tuscany	TUS02	Hatchling	Dead pipped	Le Rocchette (GR)	A2.1.1	OR166588
2020	Tuscany	TUS03	Hatchling	Dead pipped	Baratti (GR)	A2.1.3	OR166590
2021	Tuscany	TUS04	Hatchling	Dead pipped	Morcone (LI)	A2.1.9	OR166592
2021	Tuscany	TUS05	Adult	Stranded	Portoferraio (LI)	A1.1.1	OR166594
2020	Latium	LAT01	Hatchling	Dead pipped	Fondi (LT)	A2.1.1	OR166595
2020	Latium	LAT02	Hatchling	Dead pipped	Ventotene (LT)	A2.1.1	OR166596
2021	Latium	LAT03	Adult	Stranded	Nettuno (RM)	A2.1.1	OR166597
2021	Latium	LAT04	Adult	Stranded	Fondi (LT)	A2.1.6	OR166598
2011	Calabria	CAL1112	Hatchling	Dead pipped	Brancaleone (RC)	A2.1.8	OR166599
2011	Calabria	CAL1124	Hatchling	Dead pipped	Palizzi (RC)	A20.1.1	OR166600
2011	Calabria	CAL1137	Hatchling	Dead pipped	Palizzi (RC)	A2.1.1	OR166601
2012	Calabria	CAL1201	Hatchling	Dead pipped	Cariati (CS)	A2.1.8	OR166602
2012	Calabria	CAL1234	Hatchling	Dead pipped	Bova Marina (RC)	A20.1.1	OR166603
2013	Calabria	CAL1303	Hatchling	Dead pipped	Bova Marina (RC)	A2.1.3	OR166604
2013	Calabria	CAL1315	Hatchling	Dead pipped	Brancaleone (RC)	A2.1.8	OR166605
2013	Calabria	CAL1319	Hatchling	Dead pipped	Bonifati (CS)	A2.1.1	OR166606
2013	Calabria	CAL1320	Hatchling	Dead pipped	Brancaleone (RC)	A2.1.8	OR166607

2013	Calabria	CAL1321	Hatchling	Dead pipped	Bova Marina (RC)	A20.1.1	OR166608
2013	Calabria	CAL1323	Hatchling	Dead pipped	Palizzi (RC)	A20.1.1	OR166609
2013	Calabria	CAL1325	Hatchling	Dead pipped	Palizzi (RC)	A20.1.1	OR166611
2013	Calabria	CAL1328	Hatchling	Dead pipped	Brancaleone (RC)	A2.1.8	OR166612
2013	Calabria	CAL1335	Hatchling	Dead pipped	Brancaleone (RC)	A2.1.1	OR166613
2013	Calabria	CAL1346	Hatchling	Dead pipped	Bova Marina (RC)	A2.1.8	OR166614
2013	Calabria	CAL1362	Hatchling	Dead pipped	Palizzi (RC)	A2.1.1	OR166615
2013	Calabria	CAL1370	Hatchling	Dead pipped	Bonifati (CS)	A2.1.7	OR166616
2013	Calabria	CAL1384	Hatchling	Dead pipped	Marina di Gioiosa Ionica	A20.1.1	OR166617
2013	Calabria	CAL1385	Hatchling	Dead pipped	Soverato (CZ)	A20.1.1	OR166618
2006	Linosa Island	LIN01	Hatchling	Dead pipped	Linosa Island	A2.9.2	OR166619
2008	Linosa Island	LIN02	Hatchling	Dead pipped	Linosa Island	A2.1.8	OR166620
2008	Strait of Sicily	SIC01	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166621
2008	Strait of Sicily	SIC02	Adult	Bycaught	Strait of Sicily	A43.1.1	OR166622
2008	Strait of Sicily	SIC03	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166623
2008	Strait of Sicily	SIC04	Adult	Bycaught	Strait of Sicily	A2.1.2	OR166647
2008	Strait of Sicily	SIC05	Adult	Bycaught	Strait of Sicily	A2.1.8	OR166624
2008	Strait of Sicily	SIC06	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166625
2008	Strait of Sicily	SIC07	Adult	Bycaught	Strait of Sicily	A2.1.8	OR166626
2008	Strait of Sicily	SIC08	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166627
2008	Strait of Sicily	SIC09	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166628
2008	Strait of Sicily	SIC10	Adult	Bycaught	Strait of Sicily	A2.1.9	OR166629
2008	Strait of Sicily	SIC11	Adult	Bycaught	Strait of Sicily	A2.1.3	OR166630
2008	Strait of Sicily	SIC12	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166631
2008	Strait of Sicily	SIC13	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166632
2008	Strait of Sicily	SIC14	Adult	Bycaught	Strait of Sicily	A2.1.5	OR166633
2008	Strait of Sicily	SIC15	Adult	Bycaught	Strait of Sicily	A2.1.10	OR166634
2008	Strait of Sicily	SIC16	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166635

2008	Strait of Sicily	SIC17	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166636
2008	Strait of Sicily	SIC18	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166637
2008	Strait of Sicily	SIC19	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166638
2008	Strait of Sicily	SIC20	Adult	Bycaught	Strait of Sicily	A2.1.11	OR166639
2008	Strait of Sicily	SIC21	Adult	Bycaught	Strait of Sicily	A2.1.4	OR166640
2008	Strait of Sicily	SIC22	Adult	Bycaught	Strait of Sicily	A2.1.8	OR166641
2008	Strait of Sicily	SIC23	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166642
2008	Strait of Sicily	SIC24	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166643
2008	Strait of Sicily	SIC25	Adult	Bycaught	Strait of Sicily	A3.1.1	OR166644
2008	Strait of Sicily	SIC26	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166645
2008	Strait of Sicily	SIC27	Adult	Bycaught	Strait of Sicily	A2.1.1	OR166646

Table S2 Mitogenome sequences of *Caretta caretta* included in this study.

GenBank accession number	Haplogroup	Provenance	Reference
FR694649.1	II	Mediterranean (Greece)	Drosopoulou et al. 2012
JX454988.1	IA	Pacific (Hawaii)	Duchene et al. 2012
JX454983.1	II	Atlantic (Florida)	Duchene et al. 2012
MF554690.1	II	Atlantic (Colombia)	Hernandez-Fernandez et al. 2018
MF579505.1	IB	Atlantic (Colombia)	Hernandez-Fernandez et al. 2018

Among available sequences in GeneBank, we excluded KP256531.1 (Otálora and Hernández-Fernández 2018) because it displayed a number of nucleotide differences with respect to the overall alignment we were not able to confirm. We also excluded JX454977.1 and JX454984.1 (Duchene et al. 2012), since they reported several missing base pairs (Ns) and NC016923.1 because after curation by the NCBI staff this record had the same nucleotide sequence of FR694649.1, which we used as reference mitogenome. MF579504.1 was also excluded from the analysis as it showed multiple poly-A stretches starting at positions 3763, 7009 and 9463. These poly-A stretches possibly resulted from non-independent, multiple substitutions at adjacent positions, which could have compromised diversity, branch length and divergence time estimates.

Table S3 Total number of reads obtained from *C. caretta* whole genome sequencing, number of reads mapping to the *C. caretta* reference mitogenome (GenBank accession number: FR694649.1) and mean mtDNA genome coverage obtained. Origin of samples as in Table S1.

Sample ID	Total no. of reads	Reads mapping to mtDNA reference genome sequence	mtDNA coverage (×)
ADR01	38,353,330	17,508	157
ADR02	18,593,330	51,135	458
ADR03	29,273,330	174,191	1561
ADR04	33,473,330	43,012	385
CAL1112	21,006,670	12,717	114
CAL1124	10,640,000	29,858	268
CAL1137	17,073,330	16,120	144
CAL1201	17,460,000	13,329	119
CAL1234	24,313,330	14,527	130
CAL1303	12,820,000	54,244	486
CAL1315	28,626,670	16,600	149
CAL1319	16,893,330	35,315	316
CAL1320	18,313,330	14,010	126
CAL1321	17,040,000	9,834	88
CAL1323	21,693,330	11,582	104
CAL1325	13,600,000	12,927	116
CAL1328	19,213,330	4,319	39
CAL1335	17,900,000	4,011	36
CAL1346	25,693,330	2,573	23
CAL1362	20,220,000	19,216	172
CAL1370	12,393,330	154,007	1380
CAL1384	10,400,000	10,148	91
CAL1385	17,333,330	77,430	694
LAT01	39,460,000	81,700	732
LAT02	33,706,670	17,402	156
LAT03	22,446,670	88,380	792
LAT04	11,260,000	10,673	96
LIN01	23,553,330	38,962	349
LIN02	21,266,670	14,364	129
SIC01	18,480,000	13,103	117
SIC02	28,540,000	29,728	266
SIC03	22,393,330	17,583	158
SIC04	31,520,000	32,216	289

SIC05	24,973,330	17,312	155
SIC06	30,813,330	22,896	205
SIC07	51,606,670	34,179	306
SIC08	20,640,000	25,182	226
SIC09	29,600,000	38,930	349
SIC10	19,206,670	18,299	164
SIC11	20,866,670	26,277	235
SIC12	35,493,330	42,543	381
SIC13	22,786,670	25,322	227
SIC14	30,106,670	22,085	198
SIC15	19,260,000	16,305	146
SIC16	27,946,670	28,450	255
SIC17	23,433,330	19,733	177
SIC18	17,546,670	13,653	122
SIC19	28,293,330	26,822	240
SIC20	18,640,000	13,862	124
SIC21	29,160,000	23,741	213
SIC22	28,626,670	32,407	290
SIC23	23,886,670	15,056	135
SIC24	21,586,670	14,605	131
SIC25	16,506,670	13,751	123
SIC26	12,553,330	9,954	89
SIC27	27,713,330	10,244	92
TUS01	42,140,000	7,226	65
TUS02	34,100,000	48,496	435
TUS03	30,266,670	23,937	215
TUS04	31,660,000	107,500	963
TUS05	19,220,000	12,789	115

Table S4 Loggerhead mitogenomic diversity and McDonald-Kreitman neutrality test estimated for single genes and concatenated tRNAs. Haplogroup II samples only are considered (60 samples from this study and three sequences from published literature). The McDonald-Kreitman test was conducted by comparing dN/dS of haplogroup II with dN/dS between haplogroup II and either haplogroup IB or the mitogenome of *Lepidochelys olivacea* (GenBank accession number: JX454987.1).

mtDNA gene	Sequence length (bp)	*k*	*P*	*h*±SE	π±SE	GC	McD-K test vs haplogroup IB	McD-K test vs *L. olivacea*
12S	970	1	0	0	0	0.416		
16S	1,615	2	1	0.063±0.005	0	0.374		
ATP6	682	3	2	0.093±0.000	1.38E-4±3.42E-5	0.377	$p = 0.49$	$p = 0.40$
ATP8	165	2	1	0.121±0.000	7.32E-4±1.62E-4	0.357	n.a.	$p = 1$
COXI	1548	2	1	0.032±0.004	2.05E-5±8.59E-6	0.416	n.a.	$p = 1$
COXII	687	2	1	0.032±0.004	4.62E-5±1.94E-5	0.371	n.a.	$p = 1$
COXIII	784	3	2	0.093±0.006	1.20E-4±2.99E-5	0.418	$p = 0.28$	$p = 0.26$
CYTB	1,144	2	1	0.063±0.005	5.46E-5±1.65E-5	0.434	$p = 0.36$	$p = 0.19$
ND1	967	3	2	0.063±0.005	6.56E-5±1.97E-5	0.399	$p = 0.53$	$p = 1$
ND2	1,042	3	2	0.063±0.005	6.09E-5±1.83E-5	0.388	$p = 0.58$	$p = 0.74$
ND3	350	2	1	0.032±0.004	9.07E-5±3.80E-5	0.383	$p = 1$	$p = 0.47$
ND4L	297	2	1	0.062±0.005	2.10E-4±6.35E-5	0.397	n.a.	$p = 1$
ND4	1,381	2	1	0.032±0.004	2.30E-5±9.64E-6	0.454	$p = 1$	$p = 1$
ND5	1,797	3	2	0.423±0.007	2.41E-4±3.06E-5	0.400	$p = 1$	$p = 1$
ND6	525	1	0	0	0	0.383	n.a.	n.a.
tRNAs	1,555	5	4	0.153±0.008	1.21E-4±2.18E-5	0.394		

k, number of haplotypes; *P*, number of polymorphic sites; *h*, haplotypic diversity ± Standard Error; π, nucleotide diversity ± Standard Error; GC, content of GC bases.

Table S5 Mitogenomic diversity and McDonald-Kreitman neutrality test estimated for single genes and concatenated tRNAs. All samples from our *Caretta caretta* dataset are considered (61 samples from this study plus five sequences from published literature). The McDonald-Kreitman test was conducted by comparing dN/dS of our *C. caretta* dataset with dN/dS between the *C. caretta* dataset and the mitogenome of *Lepidochelys olivacea*. Diversity indices as in Table S4.

mtDNA gene	Sequence length (bp)	*k*	*P*	*h*±SE	π±SE	GC	McD-K test vs *L. olivacea*
12S	971	3	6	0.089±0.006	2.77E-4±4.18E-5	0.416	
16S	1,612	4	22	0.146±0.007	6.67E-4±6.08E-5	0.374	
ATP6	682	5	12	0.174±0.008	1.05E-3±1.09E-4	0.377	$p = 1$
ATP8	165	3	2	0.197±0.008	1.53E-3±2.38E-4	0.356	$p = 1$
COXI	1,548	4	18	0.118±0.007	5.79E-4±5.58E-5	0.416	$p = 1$
COXII	687	4	11	0.118±0.007	7.82E-4±8.97E-5	0.372	$p = 0.57$
COXIII	784	5	10	0.174±0.008	7.17E-4±8.09E-5	0.418	$p = 1$
CYTB	1,144	4	18	0.146±0.007	8.86E-4±8.22E-5	0.433	$p = 0.20$
ND1	967	5	21	0.146±0.007	1.08E-3±9.93E-5	0.399	$p = 0.19$
ND2	1,042	6	20	0.147±0.007	1.11E-3±9.93E-5	0.388	$p = 0.56$
ND3	350	5	17	0.118±0.007	2.13E-3±2.18E-4	0.383	$p = 0.003$
ND4L	297	4	6	0.146±0.007	1.20E-3±1.61E-4	0.396	$p = 0.58$
ND4	1,381	4	15	0.118±0.007	6.04E-4±5.94E-5	0.367	$p = 0.19$
ND5	1,797	5	28	0.474±0.007	1.12E-3±8.74E-5	0.400	$p = 0.37$
ND6	525	4	7	0.089±0.006	5.72E-4±8.25E-5	0.383	$p = 0.27$
tRNAs	1,555	8	18	0.229±0.008	4.98E-4±6.13E-5	0.394	

Table S6 Substitutions in coding genes, rRNAs and tRNAs of the haplogroup II tree branch (see Fig. 2). Positions of nucleotide variants refer to nucleotide positions in the reference *Caretta caretta* mtDNA (GenBank accession number: FR694649.1). Codon position 1 where a synonymous (syn) or nonsynonymous (non-syn) nucleotide substitution was recorded is also reported along with the corresponding change in amino acid.

mtDNA position	Gene	Nucleotide substitution	Codon 1 position of nucleotide substitution	Change in amino acid
110	12S rRNA	T > C	-	-
1280	16S rRNA	A > C	-	-
1988	16S rRNA	A > G	-	-
3107	ND1	C > T	104	N > N (syn)
3191	ND1	T > C	132	Y > Y (syn)
3521	ND1	C > A	242	L > L (syn)
4223	ND2	A > G	85	N > D (non-syn)
4729	ND2	T > C	253	G > G (syn)
4909	ND2	G > A	313	Q > Q (syn)
4918	ND2	T > C	316	R > R (syn)
4997	ND2	C > T	343	L > L (syn)
5829	COXI	T > C	144	D > D (syn)
6150	COXI	T > C	251	F > F (syn)
7391	COX2	A > G	104	W > W (syn)
7628	COX2	C > A	183	T > T (syn)
8346	ATP6	G > A	117	G > G (syn)
8488	ATP6	T > C	165	L > L (syn)
9724	ND3	T > C	65	L > L (syn)
9843	ND3	C > T	104	I > I (syn)
9853	ND3	G > A	108	V > I (non-syn)
10016	ND4L	T > C	22	H > H (syn)
10193	ND4L	T > C	81	S > S (syn)
10735	ND4	C > A	165	L > L (syn)
11470	ND4	T > C	410	Y > Y (syn)
11549	ND4	C > T	437	L > L (syn)
11799	Leu tRNA	A > G	-	-
11880	ND5	C > T	18	L > F (non-syn)
11899	ND5	G > A	24	C > Y (non-syn)
11919	ND5	A > C	31	M > L (non-syn)
12116	ND5	T > C	96	S > S (syn)
12350	ND5	T > C	174	I > I (syn)
13076	ND5	C > T	416	T > T (syn)
13286	ND5	A > G	486	T > T (syn)
13499	ND5	T > C	557	G > G (syn)
13550	ND5	A > G	574	A > A (syn)
14340	CYTB	T > C	42	L > L (syn)
15212	CYTB	T > C	332	D > D (syn)

Table S7 Substitutions characterizing *Caretta caretta* mitotypes within haplogroup II. See table S6 for additional details.

Mitotype	mtDNA position	Gene	Nucleotide substitution	Codon 1 position of nucleotide substitution	Change in amino acid
CCA2.1.1	9440	COXIII	A > T	254	V > V (syn)
FR694649.1	9940	COXIII	A > T	254	V > V (syn)
	9974	ND4L	C > T	81	Y > Y (syn)
MF554690.1	9940	COXIII	A > T	254	V > V (syn)
	9974	ND4L	C > T	81	Y > Y (syn)
	11799	Leu tRNA	G > A	-	-
JX454983.1	9940	COXIII	A > T	254	V > V (syn)
	5847	COXI	C > A	150	L > L (syn)
CC-A2.1.7	9440	COXIII	A > T	254	V > V (syn)
	13287	ND5	G > A	487	A > T (non-syn)
CC-A2.1.8	9440	COXIII	A > T	254	V > V (syn)
	13373	ND5	C > T	515	H > H (syn)
CC-A20.1.1	9440	COXIII	A > T	254	V > V (syn)
	13373	ND5	C > T	515	H > H (syn)
	15645	Control region	C > T	-	-
CC-A2.1.6	9440	COXIII	A > T	254	V > V (syn)
	10591	ND4	T > C	117	I > I (syn)
CC-A10.4.1	9440	COXIII	A > T	254	V > V (syn)
	15970	Control region	A > G	-	-
CC-A43.1.1	9440	COXIII	A > T	254	V > V (syn)
	4277	ND2	G > A	103	A > T (non-syn)
	7433	COXII	C > T	118	F > F (syn)
	9610	ND3	T > C	27	L > S (non-syn)
	15731	Control region	A > G	-	-
CC-A2.1.2	9440	COXIII	A > T	254	V > V (syn)
	4671	ND2	G > A	234	W > X (non-syn)
CC-A2.1.5	9440	COXIII	A > T	254	V > V (syn)
	9102	COXIII	G > A	53	A > T (non-syn)
CC-A2.1.3	9440	COXIII	A > T	254	V > V (syn)
	7900	ATP8	C > T	20	Y > Y (syn)
CC-A2.1.10	9440	COXIII	A > T	254	V > V (syn)
	7900	ATP8	C > T	20	Y > Y (syn)
	2936	ND1	C > T	47	P > P (syn)
CC-A2.1.9	9440	COXIII	A > T	254	V > V (syn)
	14373	CYTB	G > A	53	A > T (non-syn)
CC-A2.1.11	9440	COXIII	A > T	254	V > V (syn)
	3045	ND1	C > T	84	L > L (syn)
	9482	Gly tRNA	A > G	-	-
CC-A2.1.4	9440	COXIII	A > T	254	V > V (syn)
	8661	ATP6	T > C	222	Y > Y
CC-A3.1.1	9440	COXIII	A > T	254	V > V (syn)
	15800	Control region	G > A	-	-
CC-A2.9.1	8398	ATP6	C > T	135	P > S (non-syn)
CC-A2.9.2	8398	ATP6	C > T	135	P > S (non-syn)
	5303	Cys tRNA	A > G	-	-

Table S8 Substitution rates and 95% Height Posterior Density (HPD) for *Caretta caretta* mitogenome partitions used in the Bayesian phylogenetic analysis. Time is estimated in number of *C. caretta* generations (1 generation = 36 years) and in million years.

Partition	Median no. substitutions × (site × generation)$^{-1}$	Median no. substitutions × (site × million year)$^{-1}$	95% HPD no. substitutions × (site × generation)$^{-1}$	95% HPD substitutions × (site × million year)$^{-1}$
12S rRNA+ 16S rRNA	3.85E-8	1.07E-3	1.73E-8 – 6.52E-8	0.48E-3 – 1.81E-3
tRNAs	4.75E-8	1.32E-3	2.09E-8 – 8.35E-8	0.58E-3 – 2.32E-3
Coding regions	6.3E-8	1.75E-3	3.46E-8 – 9.65E-8	0.96E-3 – 2.68E-3
Control region	1.98E-7	5.49E-3	1.31E-7 – 2.67E-7	3.63E-3 – 7.43E-3

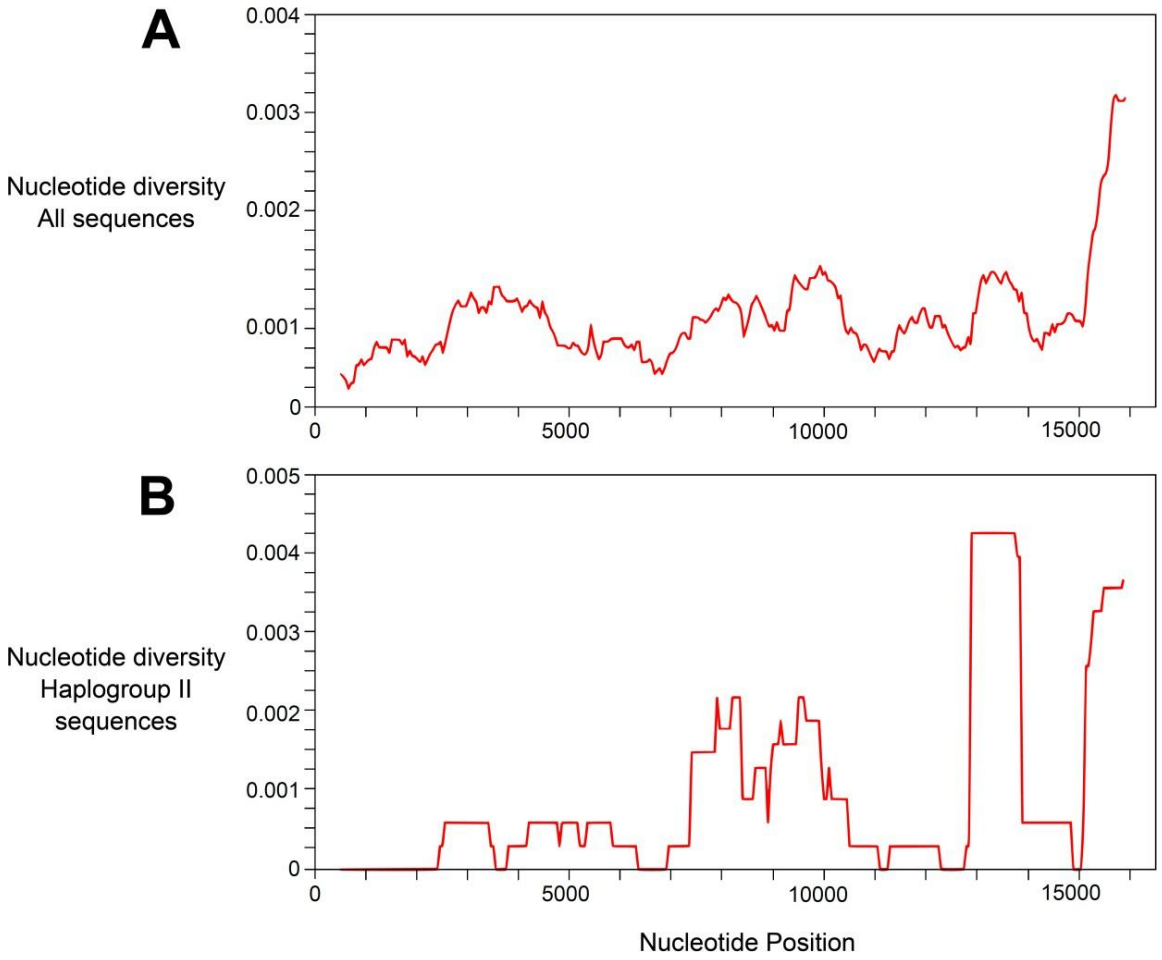


Figure S1 Nucleotide diversity (π) values across the 16425 sites of the *Caretta caretta* mitogenomes analysed in this study for A) the complete dataset (61 independent samples from this study plus five sequences from published literature) and B) haplogroup II mitogenomes only (60 independent samples from this study plus three sequences from published literature). Both profiles were obtained using π values calculated along a sliding window of 1000 bp at 50 bp intervals.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Philos Trans R Soc Lond B 312: 325-342.

Baker CS, Steel D, Calambokidis J, Falcone E, González-Peral U, Barlow J, Burdin AM, Clapham PJ, Ford JK, Gabriele CM (2013) Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. Mar Ecol Prog Ser 494: 291-306.

Baldi G, Furii G, Del Vecchio M, Salvemini P, Vallini C, Angelini V, Pari S, Lombardi Moraes K, Margaritoulis D, Çurri A, Lazar B, Casale P (2023) Growth rates and age at maturity of Mediterranean loggerhead sea turtles estimated from a single-population foraging ground. Mar Biol 170: 36.

Baltazar-Soares M, Klein JD, Correia SM, Reischig T, Taxonera A, Roque SM, Dos Passos L, Durão J, Lomba JP, Dinis H, Cameron SJK, Stiebens VA, Eizaguirre C (2020) Distribution of genetic diversity reveals colonization patterns and philopatry of the loggerhead sea turtles across geographic scales. Sci Rep 10: 18001.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bolten AB, Bjorndal KA, Martins HR, Dellinger T, Biscoito MJ, Encalada SE, Bowen BW (1998) Transatlantic developmental migrations of loggerhead sea turtles demonstrated by mtDNA sequence analysis. Ecol Appl 8: 1-7.

Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC (1994) Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. Evolution 48: 1820-1828.

Carreras C, Pascual M, Cardona L, Marco A, Bellido JJ, Castillo JJ, Tomás J, Raga JA, Sanfélix M, Fernández G, Aguilar A (2011) Living together but remaining apart: Atlantic and Mediterranean loggerhead sea turtles (*Caretta caretta*) in shared feeding grounds. J Hered 102: 666-677.

Carreras C, Pascual M, Tomas J (2018) Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). Sci Rep 8: 1435.

Carreras C, Pont S, Maffucci F, Pascual M, Barceló A, Bentivegna F, Cardona L, Alegre F, SanFélix M, Fernández G, Aguilar A (2006) Genetic structuring of immature loggerhead sea turtles (*Caretta caretta*) in the Mediterranean Sea reflects water circulation patterns. Mar Biol 149: 1269-1279.

Carreras C, Rees AF, Broderick AC, Godley BJ, 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. Sci Mar 78: 115-124.

Casale P, Broderick AC, Camiñas JA, Cardona L, Carreras C, Demetropoulos A, Fuller WJ, Godley BJ, Hochscheid S, Kaska Y (2018) Mediterranean sea turtles: current knowledge and priorities for conservation and research. Endanger Species Res 36: 229-267.

Casale P, Hochscheid S, Kaska Y, Panagopoulou A (2020) Sea Turtles in the Mediterranean Region: MTSG Annual Regional Report 2020. Report of the IUCN-SSC Marine Turtle Specialist Group

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol Phylogenet Evol 3: 102-113.

Chang G, Jones S, Leelakumari S, Ashkani J, Culibrk L, O'Neill K, Tse K, Cheng D, Chuah E, McDonald H, Kirk H, Pandoh P, Pari S, Angelini V, Kyle C, Bertorelle G, Zhao Y, Mungall A, Moore R, Vilaça ST, Jones S (2023) The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758 [version 2; peer review: 2 approved]. F1000Research 12.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1659.

Clusa M, Carreras C, Pascual M, Demetropoulos A, Margaritoulis D, Rees AF, Hamza AA, Khalil M, Aureggi M, Levy Y (2013) Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). J Exp Mar Biol Ecol 439: 15-24.

Clusa M, Carreras C, Pascual M, Gaughran SJ, Piovano S, Giacoma C, Fernández G, Levy Y, Tomás J, Raga JA, Maffucci F, Hochscheid S, Aguilar A, Cardona L (2014) Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Mar Biol 161: 509-519.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134: 959-969.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008.

Drosopoulou E, Tsiamis G, Mavropoulou M, Vittas S, Katselidis KA, Schofield G, Palaiologou D, Sartsidis T, Bourtzis K, Pantis J (2012) The complete mitochondrial genome of the loggerhead turtle *Caretta caretta* (Testudines: Cheloniidae): Genome description and phylogenetic considerations. Mitochondrial DNA 23: 1-12.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK (1999) Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). J Zool 248: 397-409.

Feldheim KA, Gruber SH, DiBattista JD, Babcock EA, Kessel ST, Hendry AP, Pikitch EK, Ashley MV, Chapman DD (2014) Two decades of genetic profiling yields first evidence of natal philopatry and long-term fidelity to parturition sites in sharks. Mol Ecol 23: 110-117.

Fetzner JW, Jr., Crandall KA (2003) Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). Evolution 57: 2101-2118.

Feutry P, Kyne PM, Pillans RD, Chen X, Marthick JR, Morgan DL, Grewe PM (2015) Whole mitogenome sequencing refines population structure of the Critically Endangered sawfish *Pristis pristis*. Mar Ecol Prog Ser 533: 237-244.

Feutry P, Kyne PM, Pillans RD, Chen X, Naylor GJ, Grewe PM (2014) Mitogenomics of the Speartooth Shark challenges ten years of control region sequencing. BMC Evol Biol 14: 1-9.

Frandsen HR, Figueroa DF, George JA (2020) Mitochondrial genomes and genetic structure of the Kemp's ridley sea turtle (*Lepidochelys kempii*). Ecol Evol 10: 249-262.

Garofalo L, Mastrogiacomo A, Casale P, Carlini R, Eleni C, Freggi D, Gelli D, Knittweis L, Mifsud CR, Mingozzi T, Novarini N, Scaravelli D, Scillitani G, Oliverio M, Novelletto A (2013) Genetic characterization of central Mediterranean stocks of the loggerhead turtle (*Caretta caretta*) using mitochondrial and nuclear markers, and conservation implications. Aquat Conserv 23: 868-884.

Garofalo L, Mingozzi T, Mico A, Novelletto A (2009) Loggerhead turtle (*Caretta caretta*) matrilines in the Mediterranean: further evidence of genetic diversity and connectivity. Mar Biol 156: 2085-2095.

Green M, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY

Hasegawa M, Kishino H, Yano T-a (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160-174.

Hernández-Fernández J, Delgado Cano DA (2018) Complete mitochondrial genome of the three Colombian Caribbean loggerhead sea turtles, using next-generation sequencing. Mitochondrial DNA B 3: 626-627.

Hochscheid S, Maffucci F, Abella E, Bradai MN, Camedda A, Carreras C, Claro F, de Lucia GA, Jribi I, Mancusi C, Marco A, Marrone N, Papetti L, Revuelta O, Urso S, Tomás J (2022) Nesting range expansion of loggerhead turtles in the Mediterranean: Phenology, spatial distribution, and conservation implications. Glob Ecol Conserv 38: e02194.

Jensen MP, FitzSimmons NN, Dutton PH, Michael P (2013) Molecular genetics of sea turtles. In: Wyneken J, Lohmann K, Musick JA (eds) The biology of sea turtles, Volume III. CRC Press, Boca Raton, FL, pp 135-161

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35: 1547-1549.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.

Laurent L, Casale P, Bradai MN, Godley BJ, Broderick G, Gerosa AC, Schroth W, Schierwater B, Levy AM, Freggi D, Abd EM, Hadoud DA, Gomati HE, Domingo M, Hadjichristophprou M, Kornaraky L, Demirayak F, Gautier CH (1998)

Molecular resolution of marine turtle stock composition in fishery bycatch: a case study in the Mediterranean. Mol Ecol 7: 1529-1542.

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

Loisier A, Savelli M-P, Arnal V, Claro F, Gambaiani D, Sénégas JB, Cesarini C, Sacchi J, Miaud C, Montgelard C (2021) Genetic composition, origin and conservation of loggerhead sea turtles (*Caretta caretta*) frequenting the French Mediterranean coasts. Mar Biol 168: 52.

Maffucci F, Corrado R, Palatella L, Borra M, Marullo S, Hochscheid S, Lacorata G, Iudicone D (2016) Seasonal heterogeneity of ocean warming: a mortality sink for ectotherm colonizers. Sci Rep 6: 23983.

Mancino C, Canestrelli D, Maiorano L (2022) Going west: range expansion for loggerhead sea turtles in the Mediterranean Sea under climate change. Glob Ecol Conserv 38: e02264.

Margaritoulis D, Dean CJ, Lourenço G, Rees AF, Riggall TE (2020) Reproductive longevity of loggerhead sea turtles nesting in Greece. Chelonian Conserv Biol 19: 133-136.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351: 652-654.

Michaels G, Hauswirth W, Laipis P (1982) Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 94: 246-251.

Miller J (1997) Reproduction in sea turtles. In: Lutz P, Musick JA (eds) The biology of sea turtles. CRC Press: Boca Raton, FL, pp 51–81

Milne I, Stephen G, Bayer M, Cock P, Pritchard L, Cardle L, Shaw P, Marshall D (2013) Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14: 193-202.

Mingozzi T, Masciari G, Paolillo G, Pisani B, Russo M, Massolo A (2006) Discovery of a regular nesting area of loggerhead turtle *Caretta caretta* in southern Italy: a new perspective for national conservation. In: Hawksworth D, Bull A (eds) Biodiversity conservation in Europe Topics in biodiversity and conservation, vol 7. Springer, Dordrecht, pp 277-299

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Abel Nielsen SC, Rasmussen M, Willerslev E, Gilbert MT, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis

of killer whales (*Orcinus orca*) indicates multiple species. Genome Res 20: 908-916.

Moritz C (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. Mol Ecol 3: 401-411.

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. Physiol Biochem Zool 89: 473-486.

Otálora K, Hernández-Fernández J, Carlos Salcedo-Reyes J (2018) Complete Colombian Caribbean loggerhead turtle mitochondrial genome: tRNA structure analysis and revisited marine turtle phylogeny. Universitas Scientiarum 23: 355-381.

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

Pianka ER (1973) Evolutionary ecology. Harper & Row, New York, NY

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol 67: 901-904.

Robin E, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136: 507-513.

Saied A, Maffucci F, Hochscheid S, Dryag S, Swayeb B, Borra M, Ouerghi A, Procaccini G, Bentivegna F (2012) Loggerhead turtles nesting in Libya: an important management unit for the Mediterranean stock. Mar Ecol Prog Ser 450: 207-218.

Shamblin BM, Bjorndal KA, Bolten AB, Hillis-Starr ZM, Lundgren I, Naro-Maciel E, Nairn CJ (2012a) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol Ecol 21: 2330-2340.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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: e85956.

Shamblin BM, Bolten AB, Bjorndal KA, Dutton PH, Neilen JT, Abreu-Grobois FA, Reich KJ, Witherington BE, Bagley DA, Ehrhart LM, Tucker AD, Addison DS, Arenas A, Johnson C, Carthy RR, Lamont MM, Dodd MG, Gaines MS, LaCasella E, Nairn CJ (2012b) Expanded mitochondrial control region sequences increase

resolution of stock structure among North Atlantic loggerhead turtle rookeries. Mar Ecol Prog Ser 469: 145-160.

Sheridan CM, Spotila JR, Bien WF, Avery HW (2010) Sex-biased dispersal and natal philopatry in the diamondback terrapin, *Malaclemys terrapin*. Mol Ecol 19: 5497-5510.

Shields WM (1982) Philopatry, inbreeding, and the evolution of sex. State University of New York Press, New York, NY

Splendiani A, Fioravanti T, Giovannotti M (2017) Mitochondrial DNA reveals the natal origin of *Caretta caretta* (Testudines: Cheloniidae) stranded or bycaught along the Southwestern Adriatic coasts. Eur Zool J 84: 566-574.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619-633.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics 134: 659-669.

Tikochinski Y, Carreras C, Tikochinski G, Vilaça ST (2020) Population-specific signatures of intra-individual mitochondrial DNA heteroplasmy and their potential evolutionary advantages. Sci Rep 10: 211.

Tolve L, Casale P, Formia A, Garofalo L, Lazar B, Natali C, Novelletto A, Vallini C, Bužan E, Chelazzi G, Gaspari S, Fortuna C, Kocijan I, Marchiori E, Novarini N, Poppi L, Salvemini P, Ciofi C (2018) A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. Mar Biol 165: 68.

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5: e15465.

Wiesner R, Rüegg J, Morano I (1992) Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. Biochem Biophys Res Commun 183: 553-559.

Yan K, An T, Zhai M, Huang Y, Wang Q, Wang Y, Zhang R, Wang T, Liu J, Zhang Y, Zhang J, Wang K (2019) Mitochondrial miR-762 regulates apoptosis and myocardial infarction by impairing ND2. Cell Death Dis 10: 500.

Yilmaz C, Turkozan O, Bardakci F (2011) Genetic structure of loggerhead turtle (*Caretta caretta*) populations in Turkey. Biochem Syst Ecol 39: 266-276.

Erratum to: Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea

Livia Tolve[1,*], Alessio Iannucci[1,2,*], Luisa Garofalo[3], Andrea Ninni[4,5], Andrea Capobianco Dondona[6], Ilaria Ceciarini[7], Cristiano Cocumelli[3], Alessandra De Lucia[8], Mattia Falconi[4], Angela Formia[1,9], Federico Iacovelli[4], Cecilia Mancusi[7,10], Erica Marchiori[11], Letizia Marsili[7], Toni Mingozzi[12], Stefano Nannarelli[8], Chiara Natali[1], Giuliana Terracciano[13], Marco A. L. Zuffi[14], Andrea Novelletto[5,§], Claudio Ciofi[1]

1 Department of Biology, University of Florence, 50019 Sesto Fiorentino (FI), Italy

2 National Biodiversity Future Center, 90133 Palermo, Italy

3 Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", 00178 Rome, Italy

4 Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy

5 PhD Program in Evolutionary Biology and Ecology, Tor Vergata University of Rome, 00133 Rome, Italy

6 Farm4Trade, 66041 Atessa (CH), Italy

7 Department of Physical Sciences, Earth and Environment, University of Siena, 53100 Siena, Italy

8 Hydrosphera onlus, Rome, Italy

9 African Aquatic Conservation Fund, Chilmark, MA 02535, USA

10 Agenzia Regionale per la Protezione Ambientale Toscana (ARPAT), 57125 Livorno, Italy

11 Department of Comparative Biomedicine and Food Science, University of Padua, Agripolis, 35020 Legnaro (PD), Italy

12 Department of Biology, Ecology and Earth Sciences, University of Calabria, 87030 Rende (CS), Italy

13 Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", 56123 Pisa, Italy

14 Natural History Museum, University of Pisa, 56011 Calci (PI), Italy

*These authors have contributed equally to this work

§ Co-senior author

Corresponding authors: Livia Tolve livia.tolve@unifi.it, Alessio Iannucci alessio.iannucci@unifi.it

Erratum to: Mar Biol doi.org/10.1007/s00227-023-04325-x

In the original manuscript, sample TUS05 (Genbank accession number OR166594) was assigned to haplogroup IB and identified as a mitogenomic variant of the control region haplotype CC-A1.1. Consequently, the corresponding mitogenomic haplotype was named CC-A1.1.1. The 800 bp control region haplotype of TUS05, however, does not correspond to CC-A1.1, nor the 380 bp control region haplotype of TUS05 corresponds to CC-A1. According to the Archie Carr Center for Sea Turtles Research database (https://accstr. ufl. edu/ resou rces/ mtdna- seque nces/), the control region haplotype described for TUS05 was yet to be described. We therefore renamed the mitogenomic haplotype of TUS05 as A79.1.1. Tables 2 and 4 and Table S1 should read CC-A79.1.1 and CC-A79.1 instead of CC-A1.1.1 and CC-A1.1, respectively. The corrected Figure 2 is produced here.

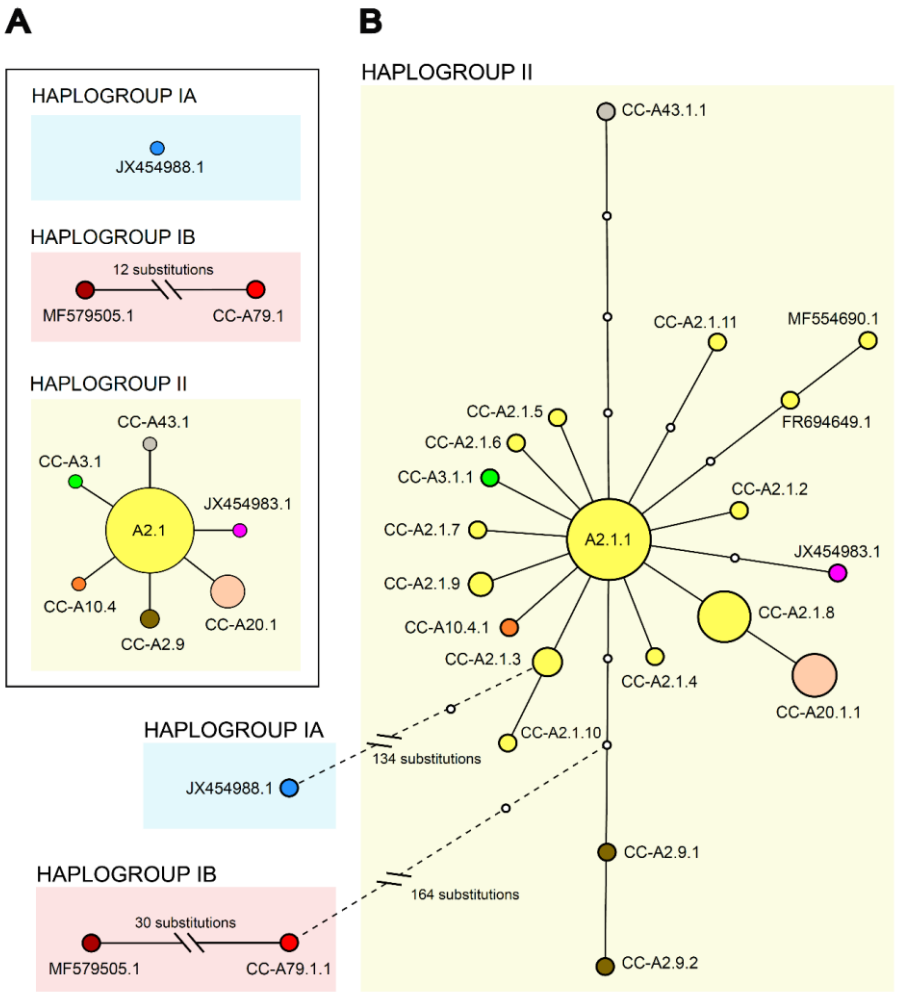


Fig. 2 TCS haplotype networks based on partial mtDNA control region (A) and complete mitogenome (B) sequences. Dashed lines represent links with probability lower than 95%.

CHAPTER IV

Mitogenomic analysis reveals the global phylogeography and adaptive signatures of the sea turtle *Caretta caretta*

Introduction

Unveiling patterns of dispersal and colonization dynamics over evolutionary history is a challenging task for conservation biology, especially in the case of widely distributed, highly migratory species (Eizaguirre and Baltazar-Soares 2014). Among marine vertebrates, sea turtles provide outstanding examples of globally spread species thanks to trans-oceanic movements, occurring in the first years of life via water currents and in adulthood through active swimming (Lutz and Musick 2017). Despite their wide distribution, sea turtles exhibit a remarkable genetic population structure, mainly due to natal philopatry, i.e. the tendency of females to reproduce at their natal beaches (*rookeries*) (Bowen and Karl 2007). In addition, adaptation to local environments contributes often to genetically structure populations or groups of populations under the same selective pressures (Barbanti et al., submitted). The characterization of the phylogenetic relationships among maternally inherited mitochondrial DNA (mtDNA) haplotypes, allows inferring some of the processes that have shaped the evolution of sea turtles' populations over the millions of years (Clusa et al. 2013). Furthermore, as this molecule does not recombine, the ancient phylogenetic signal is not erased in the event of a secondary contact of divergent lineages, in comparison to the nuclear genome (Galià-Camps et al. 2024).

Among the seven sea turtle species, the loggerhead (*Caretta caretta*) is one of the most widespread, inhabiting tropical and temperate waters worldwide, reproducing at the highest latitudes (Lutz and Musick 2017). Genetic studies based on sequences of mtDNA control region described three main extant mtDNA haplogroups for this species globally: the haplogroup (IA), being exclusively found in the Pacific Ocean, the haplogroup (IB), being mostly found in the Atlantic Ocean and in one rookery of the Northwest Indian Ocean, and the haplogroup (II), being widely spread in the Atlantic Ocean, in all rookeries of the Mediterranean Sea and in one rookery of the Southwest Indian Ocean (Shamblin et al. 2014). Over the decades there have been proposed several reconstructions of the evolution and colonization of the oceans by *Caretta caretta* based on the mtDNA control region regarding the phylogenetic relationship among the haplogroups IA, IB and II (e.g. Bowen et al. 1994; Bowen and Karl 2007; Carreras et al. 2007; Duchene et al. 2012; Shamblin et al. 2012b; Clusa et al. 2013; Shamblin et al. 2014), and the topic is still under debate.

Two main alternative scenarios can be summarized according to these studies, both taking the closure of the Isthmus of Panama (20-2.8 MYA) (O'Dea et al. 2016) as the first biogeographic event causing isolation and divergence of current mtDNA lineages. One scenario, supported by Shamblin et al. (2014), suggests that haplogroups I (ancestral state of IA and IB) and II separated respectively west and east of the Isthmus of Panama, with haplogroup II being the first one to settle in the Atlantic and later colonizing the Mediterranean Sea via North Atlantic Stream. After this first split, haplogroups IA and IB diverged within the Indo-Pacific Ocean, with IA settling in the Pacific and IB entering the Atlantic from the Indian Ocean via South Africa and reaching western Atlantic rookeries where already there was haplogroup II forming a secondary contact.

The alternative scenario, supported by Baltazar-Soares et al. (2020) by means of Approximate Bayesian Computation (ABC) method, envisages that haplogroup IB was the first that differentiated east of Panama and first settled in the Atlantic, while haplogroups IA and II were first settled west of Panama and diverged later in the Indo Pacific. In this scenario, haplogroup II entered the Atlantic via South Africa, heading north until the Mediterranean Sea, later colonizing the Atlantic rookeries where haplogroup IB was already present forming a secondary contact. However, only a fragment of the mtDNA control region was used in both studies, and differences between the studies could be explained using different sampling sets or the use of different methodological approaches. Alternatively, the whole mitogenome offers better resolution than traditional mtDNA control region markers, as recently demonstrated in Tolve et al. (2024). Consequently, whole mitogenome sequencing could be used to resolve the evolutionary processes leading to the global mitochondrial distribution of this species.

Interestingly, haplogroup IB and haplogroup II not only co-exist in the western Atlantic rookeries due to the known secondary contact but also present a parallel latitudinal gradient east and west of the Florida peninsula, ranging from nearly zero percent of haplogroup IB in Mexico and increasing with latitude until reaching almost one hundred percent in South Carolina (Encalada et al. 1998; Shamblin et al. 2012b). Given the global dispersal potential of these marine vertebrates, geographic distances and strong philopatry alone do not suffice to explain this symmetrical distribution at both sides of the peninsula. Mitochondrial adaptation to temperature gradients is known to occur in several marine species (e.g. Silva et al. 2014; Baltazar-Soares et al. 2021; Noll et al. 2022) and remains to be evaluated in loggerheads.

In the present study we aim to fulfil the gap of knowledge on the above-mentioned unresolved topics using whole mitogenome sequencing data. First, we seek to determine the most likely phylogeographic scenario explaining the origin and distribution of mitogenomic groups of *C. Caretta* worldwide. Second, we look for mitochondrial-scale evidence of selective pressure between haplogroups IB and II, supporting their geographic distribution in the Northwestern Atlantic under the hypothesis of adaptation to different temperature conditions, and at a global scale by evaluating the potential of the detected polymorphisms across the haplogroups to modify protein conformation and function.

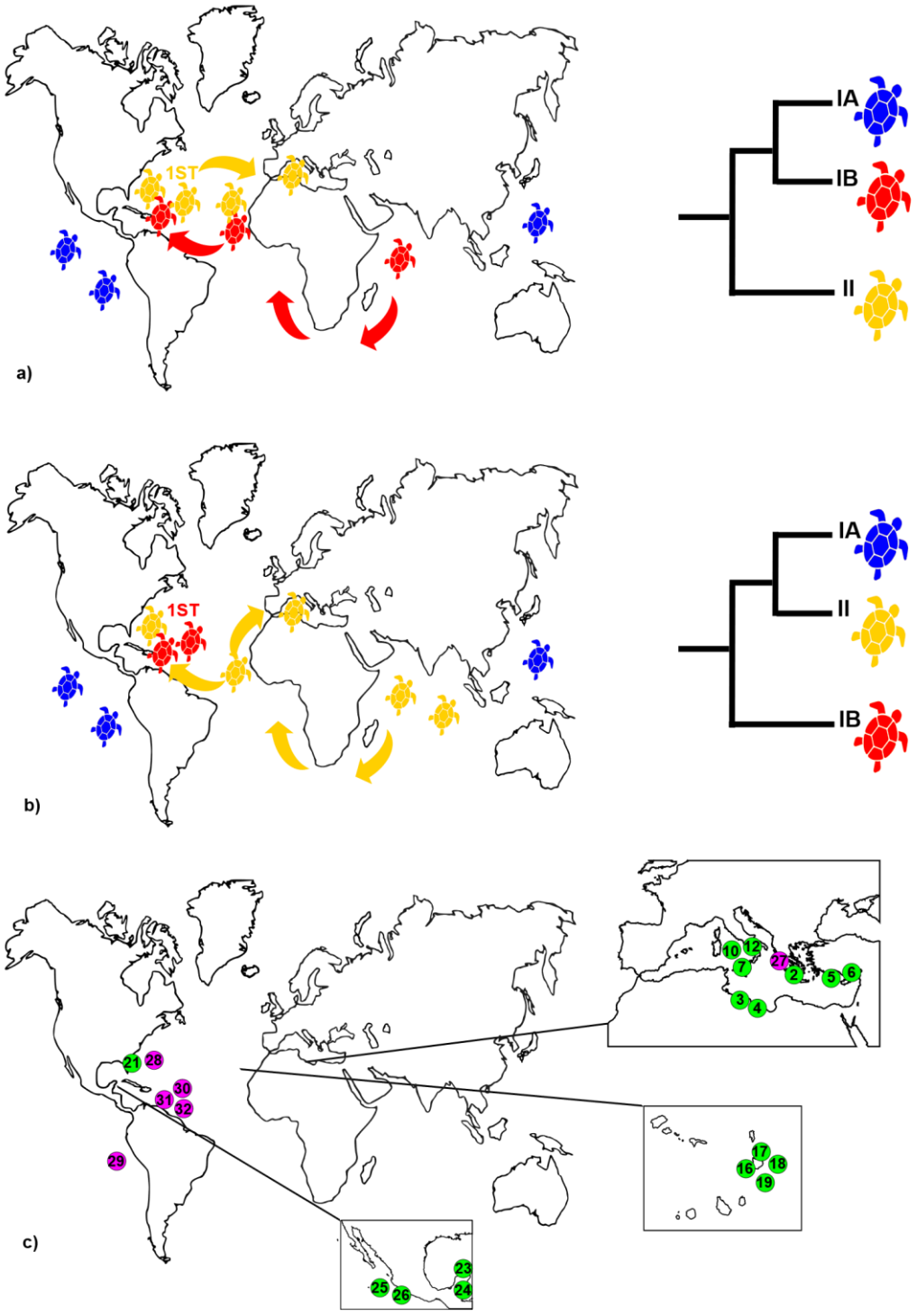


Fig. 1 In panels a) and b), phylogeographic hypotheses for oceans colonization by *Caretta caretta* and the corresponding control region haplogroup topologies supporting each scenario. In blue, haplogroup IA; in red, haplogroup IB, in yellow, haplogroup II. In panel c), samples used in genetic diversity and phylogenetic analysis of this study: in green, codes of the mitogenome sequences assembled in this study; in purple, codes of the mitogenome sequences downloaded from GenBank. Mitogenomes from GenBank are from Drosopoulou et al. (2012); Duchene et al. (2012); Hernandez-Fernandez et al. (2018); and Tolve et al. (2024). Code numbers for each mitogenome as in Table S1.

Materials and Methods

Sampling, sequencing and mitogenomes reconstruction

Tissue samples of 26 individuals preserved in ethanol 95% were analysed (Table S1). In detail: two samples from Mexico (Pacific Ocean), two samples from Quintana Roo, two samples from East Florida, five samples from Boa Vista (Atlantic Ocean), two samples from Libya, two samples from Turkey, two samples from Greece and nine samples from Italy (Mediterranean Sea) (Table S1). DNA extraction, DNA quality and concentration measures, library preparation and sequencing were performed as in Table S2. Sequencing data were demultiplexed and converted from bcl to fastq format using bcl2fastq version 2.20 (Illumina). Reads quality was checked with FastQC 11.9 (Andrews 2010) and low-quality reads were removed with Trimmomatic 0.39 (Bolger et al. 2014). Mitogenomes were assembled *de novo*, by a seed-and-extend algorithm implemented in NOVOPlasty (Dierckxsens et al. 2016). As a seed we used COI sequence of *C. caretta* (GenBank accession number NC016923) (Drosopoulou et al. 2012). Contigs assembled by NOVOPlasty were visually inspected and annotated using NC016923 (Drosopoulou et al. 2012) as reference mitogenome in Geneious Prime 2022.2.2. We considered a mitogenome to be properly reconstructed when it was circularized or when including the whole sequence of all genes. We trimmed the assembled mitogenomes at position 16425, to remove the microsatellite repeats (TATAT)n found in this region of the control region (Drosopoulou et al. 2012), since the polymorphism in this region is based on the number of AT rather than single point mutations.

Mitogenomes alignment and haplotypes assignment

We constructed a data set comprising the 26 mitogenomes generated in the present study and the 66 complete mitogenomes available in GenBank without N bases, resulting in 92 *C. caretta* sequences in total (Table S1). In order to compare our results with previous studies, we also generated a second data set including only the control regions from all mitogenomes. We aligned the two data sets using mafft v.7 (https://mafft.cbrc.jp/alignment/software/) and subsequently generated haplotype networks, using statistical parsimony (Templeton et al. 1992) as implemented in TCS (Clement et al. 2000). We classified the mitogenomic haplotypes following the criteria in Tolve et al. (2024) where the names were defined first by the control region haplotype and the last number was characteristic of a unique combination of all the variants outside

the control region. For the following analyses, we removed identical sequences and retained only a subset of haplotypes from haplogroup II (n = 13) to avoid its over-representation in relation to the other haplogroups, resulting in 23 total sequences from independent samples (Fig. 1c, Table S1).

Genetic diversity and phylogenetic analysis

Haplotype diversity, nucleotide diversity and dN/dS ratio were computed for single genes using DnaSP (Librado and Rozas 2009). For phylogenetic analyses, we included the mitogenomic sequences of *Lepidochelys olivacea* (NC028634.1) and *Lepidochelys kempii* (MN136061.1) as outgroups. Using Geneious Prime 2022.2.2, we built four mitogenomic partitions to be used in following analyses: rRNAs, tRNAs, control region and coding genes. We selected independent molecular evolution models for single genes, for each mitogenomic partition and for the complete mitogenome, based on AICc criterion. We then constructed Maximum Likelihood phylogenies using IqTree (Trifinopoulos et al. 2016) and constructed consensus trees with Astral-III (Zhang et al. 2018). Statistical support for the nodes was assessed by SH-aLRT (values > 80%) and ultrafast bootstrap (values > 95%).

To date the nodes of the consensus topology we used Bayesian phylogenetic analysis in BEAST suite v.1.10.4 (Suchard et al. 2018a). As before, we used the four mitogenomic partitions (rRNAs, tRNAs, control region and coding genes). For each partition, we used the 'unlink' option for substitution model and for molecular clock to estimate their parameters independently, the 'link' option to generate the tree, and we selected the options 'random starting tree' and 'coalescent constant size'. For priors setting, we adopted two different approaches: first, we set the prior on the mutation rates, each allowed to vary independently within a normal distribution of values ranging from zero to 6.7 E-3 million year−1, which is the highest value for mitochondrial substitution rates recorded in marine turtles (Dutton et al. (1999); second, we set the prior on the time of divergence between *Caretta* and *Lepidochelys*, with normal distribution truncated between 12 and 20 MYA, as in Duchene (2012). For each approach, we performed two independent runs, 10^8 steps each with a 10% burn-in. We used Tracer v.1.7.1 (Rambaut et al. 2018) to monitor ESS for each output parameter, Tree Annotator (Drummond and Rambaut 2007) to summarise the overall output, and FigTree v1.4.4 (http:// tree. bio. ed. ac. uk/ software/ figtr ee/) to visualize the tree.

Selection analysis

We tested the relationship between temperature and haplogroups in the Northwestern Atlantic, where haplogroups IB and II coexist in sympatry due to a secondary contact. We first downloaded data of sea surface water temperature and of air temperature from 1970 to 2020 (product CMEMS-INS-PUM-013-002-052) from Copernicus Marine Data Store (https://data.marine.copernicus.eu/products). For the same time period, data of soil temperature in layer 28–100 cm of depth were downloaded from Copernicus Climate Data Store (Muñoz Sabater 2019). Mean water, air and soil temperatures for the nesting season were then extrapolated for each rookery in western and eastern Atlantic, using the R packages 'ncdf4' (10.32614/CRAN.package.ncdf4) and 'raster' (10.32614/CRAN.package.raster). Frequencies of the control region haplotypes belonging to haplogroups IB and II in Northwestern Atlantic rookeries were taken from Shamblin et al. (2012b). The relationship between the relative frequency of haplogroups IB/II with temperatures was tested using linear and polynomial regressions with R packages 'ggplot2' (Wickham 2016) and 'stats' (R-core@r-project.org). Normality of the distribution of residuals was tested by Shapiro test. Best fitting models were selected based on AIC criterium.

To test for deviation from neutrality and for positive selection we used the 13 mitochondrial protein coding genes (Fig. S1) comparing the same whole mitogenome haplotypes retained for the phylogenetic analysis for haplogroup II and IB (Table S1). We performed the McDonald and Kreitman test and obtained the Neutrality Index (NI) with DnaSP (Librado and Rozas 2009). We considered 'polymorphisms' the non-synonymous (Pn) and synonymous (Ps) changes within haplogroups and 'divergence' the fixed non-synonymous (Dn) and synonymous (Ds) differences between haplogroups. The Neutrality Index (NI) is defined as Pn/Ps within groups divided by Dn/Ds between groups and indicates neutrality if = 1, purifying selection if > 1, positive selection if < 1.

Positive selection was also investigated using the 'codeml' package implemented in PAML (Yang 2007) which models ω (dN/dS) ratio with different methods and tests the statistical significance via Likelihood Ratio Tests (LRT). We used the branch method to model and estimate the significance of different ω ratio on a user-defined 'foreground' branch of the topology. The branch we selected as the 'foreground' branch was haplogroup IB, according to the phylogenetic results (see results section). In the branch method we compared the null model M0 (estimate of homogeneous ω across all the tree

and all the codons) with alternative model MBr1 (estimate of different ω between the background branch and the foreground branch IB). Then, we used the branch-site method to identify the codons implicated in values of ω > 1. This method therefore tests specifically the presence of positive selection across the sites of the foreground branch. We compared the null model M2a (fixed value of ω = 1 on the foreground branches across all sites) with alternative model MBr1_2S (estimate of ω > 1 across the sites of the foreground branches). We ran the analysis on each single gene, first using a dataset including only *C. caretta* sequences from our study and secondly a dataset including one mitogenomic sequence for each haplogroup of *C. caretta* and one mitogenomic sequence for each one of the other six sea turtle species (*Lepidochelys kempii* MN136053.1; *Lepidochelys olivacea* NC_028634.1; *Eretmochelys imbricata* NC_012398.1; *Chelonia mydas* NC_000886.1; *Natator depressus* NC_018550.1; *Dermochelys coriacea* CM019965.1, see also Table S1).

Then, we characterized the non-synonymous substitutions that are fixed between haplogroup II and haplogroup IB. To polarize these substitutions through the phylogeny, we reconstructed the ancestral sequence for the coding genes partitions at the first divergence within *C. caretta*, using BEAST. Each state of the non-synonymous position was therefore characterized as 'ancestral' or 'derived'. Those positions were also inspected and the states annotated across the other six sea turtle species, after downloading and aligning the complete mitogenomes available in GenBank (Table S1). From the comparison of the states at those positions across the sea turtle species we expected to find more similarities between *C. caretta* haplogroup II and the other species than between *C. caretta* haplogroup IB and the other species. This would be a consequence of the adaptation to cold temperatures in haplogroup IB, which is the 'coldest' group within the 'coldest' sea turtle species (indeed, *C. caretta*), the other sea turtles being more tropically distributed.

PROVEAN (Protein Variation Effect Analyzer) software (Choi et al. 2012) was used to predict whether those non-synonymous substitutions between haplogroup IB and II have an impact on the biological function of a protein (functional effects are associated with PROVEAN scores < -2.5). In the end, to visualize the non-synonymous substitutions in the 3D space, we built protein structures in SWISS-MODEL (https://swissmodel.expasy.org/) using turtle species proteins in AlphaFold as templates, and then RCSB PDB (https://www.rcsb.org/) to extract the images of the structures. The

physicochemical properties of amino acids corresponding to non-synonymous substitutions in were characterized using TreeSAAP (Woolley et al. 2003).

Results

Mitogenomes haplotypes assignment, genetic diversity and phylogenetic analysis

After mitogenome reconstruction with Novoplasty we obtained 26 high coverage mitogenomes (>100X, Table S1), 16,425 bp long. The organization of the mitogenomes is in agreement with the one described by Drosopoulou et al. (2012): 13 protein-coding genes, two rRNAs genes (12S and 16S rRNA), 22 tRNAs genes, and one non-coding region (control region) (Figure S1). The 54% of the assembled mitogenomes (13 out of 24) carried a total of 11 unpublished mitogenomic haplotypes (Table S1): in samples from Pacific Mexico, haplotypes P2.1.1 and P2.1.2 (haplogroup IA); in samples from Florida, haplotype A1.1.2; in samples from Cape Verde, haplotypes A1.3.1, A1.3.2, A17.2.1, A47.1.1 (haplogroup IB); in samples from Quintana Roo, haplotypes A2.1.15, A10.1.1 (haplogroup II); in samples from Mediterranean Sea, haplotypes A2.8.1, A2.1.14 (haplogroup II). The 46% of remaining samples carry mitogenomic haplotypes that have been described in literature (Table S1): A2.1.1 (Drosopoulou et al. 2012; Tolve et al. 2024), A2.1.3, A2.1.6, A2.1.8, A2.9.1, A20.1.1 (Tolve et al. 2024).

Haplotype diversity and nucleotide diversity are reported for complete mitogenomes and for single genes in Table S3. In summary, most genes have similar levels of haplotype and nucleotide diversity levels, with the control region having the higher levels for both haplotype and nucleotide diversity. ND3 and ND5 also stand out from average values, being the second highest values for nucleotide and haplotype diversity, relatively. To evaluate the relationship among the inferred haplotypes, we generated haplotype networks using TCS software (Figure S2). We observed differences in the topology of the networks built for each of the three haplogroups. We detected a starlike structure for haplogroup II, with a central abundant haplotype (A2.1, or A2.1.1) and several radiating variants, a more branched structure for haplogroup IB, and a linear relationship for haplogroup IA.

To investigate the phylogenetic relationships among haplogroups, we conducted ML phylogenetic analysis for each single gene and partition, after selecting the best molecular evolution models using AICc criterion (Table S4). We obtained topologies for all coding genes, rRNAs, control region and partitions. Single tRNAs did not contain sufficient

information to reconstruct reliable topologies, neither did the partition of all the tRNAs, therefore tRNAs trees were discarded in the following analyses (Table S5). Topologies obtained for ATP6, COXI, COXII, COXIII, CYTB, ND3, 12S rRNA, 16S rRNA all put haplogroup IB in a basal position, clustering together haplogroup II and haplogroup IA. The topology obtained for the partition of all the coding genes, behaved in the same way (Figure S3 and Figure S4). On the contrary, single gene trees obtained for ND1, ND2, ND4, ND4L, ND5, ND6 and control region tree, did not support this topology (Figure S4). The consensus trees built in Astral-III, the one with single gene trees and the one with partition trees as input, both clustered together haplogroup II and IA and put IB as basal. This output was also retrieved in the phylogeny of complete mitogenomes (Figure S3).

We also computed two types of Bayesian phylogenies that were constructed in BEAST, one with priors on mutation rates (i) and one with prior on the root (ii). The two Bayesian phylogenies were consistent with each other and in agreement with the most likely topology according to ML, with IB as basal and haplogroup II and IA clustering together (Figure 2 and Figure S5). Divergence of *C. caretta* from *Lepidochelys* was estimated to occur 13.86 (9.51–19.05) million years ago (MYA) in phylogeny (i) and 15.52 (13.68-18.13) MYA in phylogeny (ii). The divergence of haplogroup IB was dated 2.07 (1.37-2.85) MYA in phylogeny (i) and 2.31 (1.9-2.87) MYA in phylogeny (ii), then the separation between haplogroup IA and haplogroup II was 1.34 (0.17-0.45) MYA in phylogeny (i) and 1.5 (1.18-1.9) MYA in phylogeny (ii). The beginning of radiation of haplogroup II was dated 90 (40-150) thousands of years ago (KYA) in phylogeny (i) and 100 (50-170) KYA in phylogeny (ii). All these nodes were supported by bootstraps > 90%.

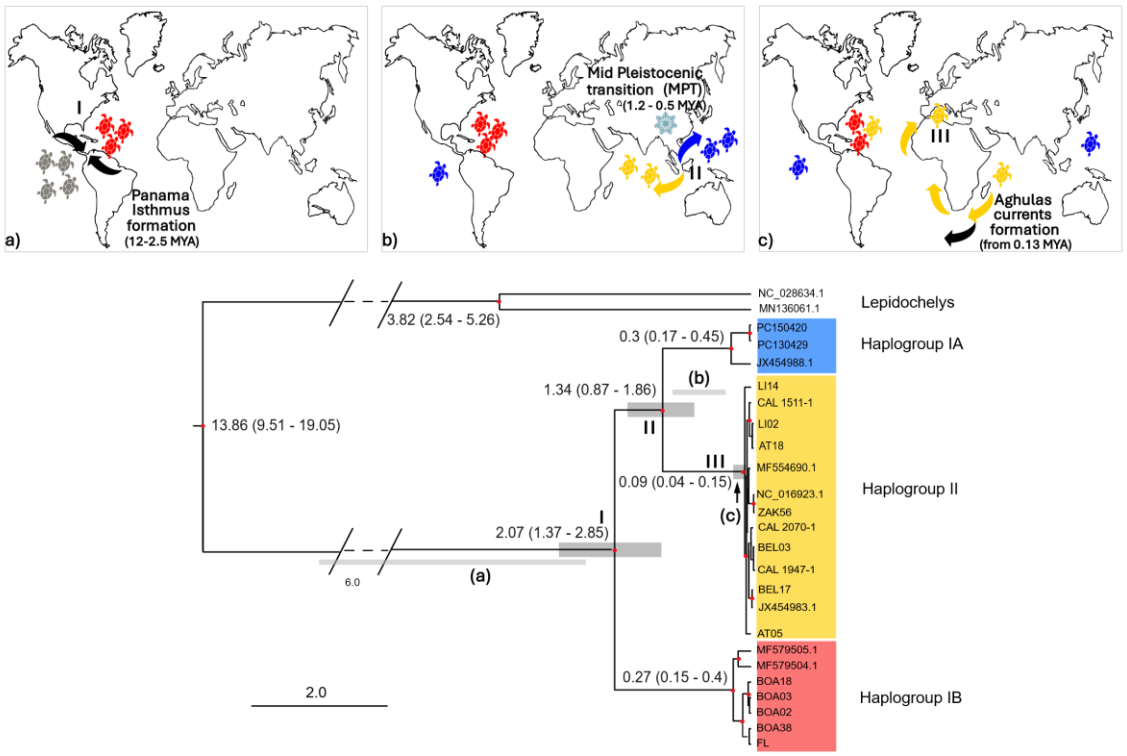


Fig. 2 Bayesian phylogeny with priors on the mutation rates of each mitogenomic partition. Red dots indicate the node with bootstrap > 90%. Distances are in millions of years. Height and 95% HDP are reported for the nodes I, II, III and associated geological and geographical events are shown in the panels above (a, b, c).

Selective pressure analyses

To test whether the relative frequency of haplogroups IB/II in Northwestern Atlantic rookeries is related with water, air or soil temperature we generated linear or polynomial regressions according to the best fitting models. We found a significant relationship among haplogroups frequency and each temperature tested during the nesting season (Figure 3). In particular, we detected an increase in the relative frequency of haplogroup II is associated with the increase in temperature. On the contrary, decreasing temperatures are associated with increasing frequency of haplogroup IB, which from now on we consider the 'cold' haplogroup.

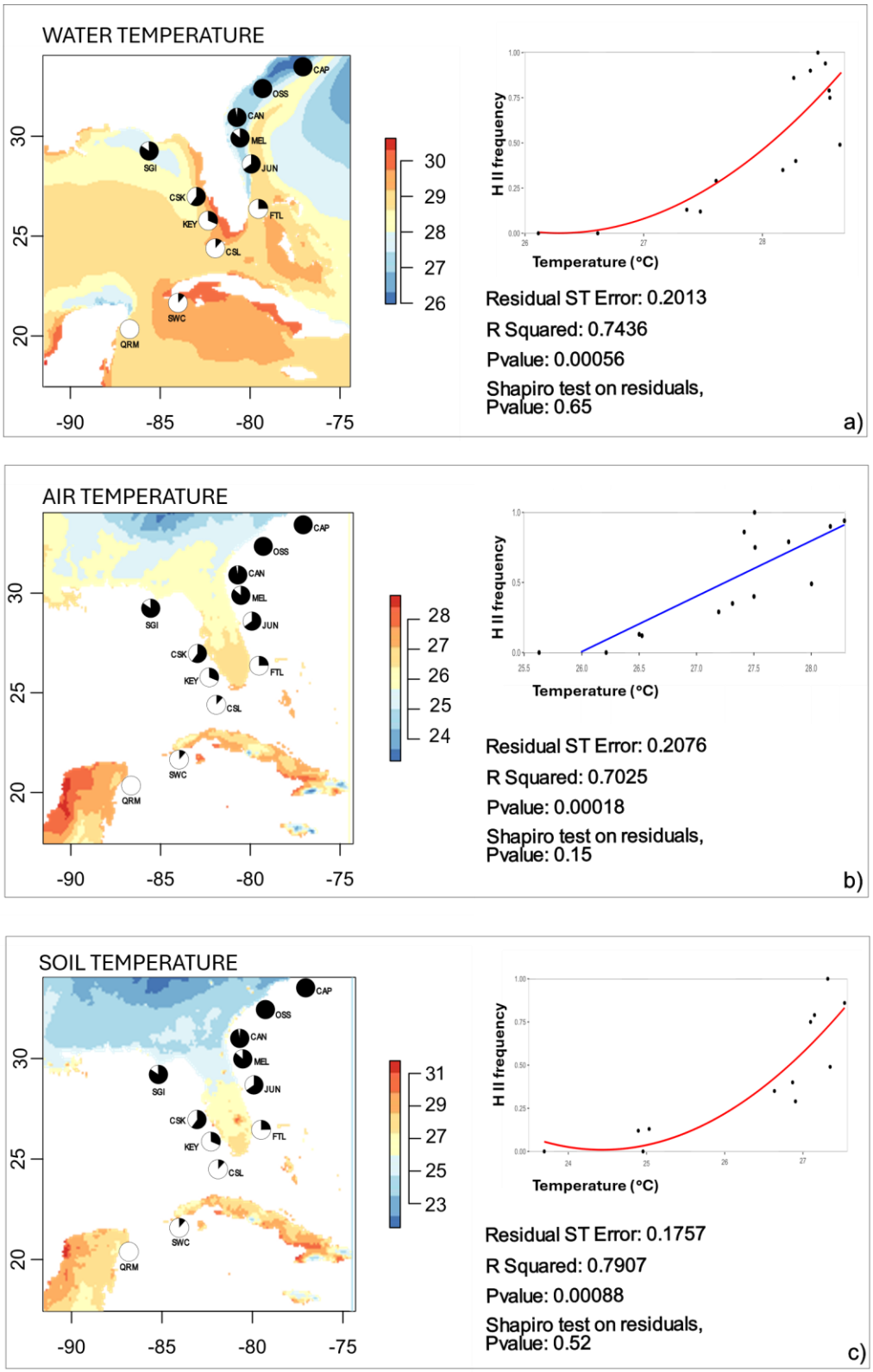


Fig. 3 On the left, sea water temperature (a), air temperature (b) and soil temperature (c) maps in Western Atlantic. In the pies, frequencies of haplogroups are reported (IB in black, II in white), data from Shamblin et al. 2012. On the right, best fitting models of linear or polynomial regressions and respective statistics**.**

To test for deviations from neutrality, we used different approaches. When computing McDonald – Kreitman test on haplogroup II vs haplogroup IB using complete mitogenomes, we obtained significant results (Neutrality Index (NI) = 3.65; p-value for Fisher's exact test = 0.0121). This result agrees with the higher proportion of nonsynonymous polymorphisms within haplogroups (Pn/Ps = 10/10) respect to nonsynonymous polymorphisms between haplogroups (Dn/Ds = 23/84). When single genes were tested, only ND3 resulted significant (p-value for Fisher's exact test = 0.0333), with six nonsynonymous and 0 synonymous polymorphisms within haplogroups against one nonsynonymous and three synonymous polymorphisms between haplogroups. We also tested deviations from neutrality using 'codeml' software (Tables S5 – S8). We obtained significant results in the branch method when targeting the IB branch, for ND3 gene when using the *C. caretta* dataset, and for CYTB, ND1 and ND5 when using the multiple species dataset. Although the values of dN/dS on the foreground branch IB were significantly different from the dN/dS on the background branches, they were never > 1 (Table S3), suggesting less stringent purifying selection rather than positive selection. This result is supported by the analysis on physicochemical properties of the amino acids in TreeSAAP, that failed to reveal signals of positive selection.

Interestingly, despite the lack of significant results in the branch site method (Tables S7 – S8), we found 24 nonsynonymous substitutions which are fixed between haplogroup IB and haplogroup II and they occur in ATP6, CYTB, ND1, ND2, ND3, ND4, ND5 genes. For all of them, we report the state carried by each *C. caretta* haplogroup and each sea turtle species (Fig. 4). In 15 out of these 24 cases of nonsynonymous substitutions, haplogroup II carries the ancestral state and haplogroup IB the derived state. Haplogroup IA carries the same state as haplogroup II in 14 out of those 15. In 10 out of these 24 cases, haplogroup II carries the same state as all the other sea turtle species and this state is the ancestral for *C. caretta.* On the contrary, only in three cases haplogroup IB carries the same state as all the other sea turtle species and it is the ancestral state also in this case.

According to PROVEAN scores (Table S9), four out of 24 substitutions fixed between IB and II are < -2.5 and thus are likely to affect protein functioning. In three cases out of these four affecting genes ATP6 (F151L), ND1 (G317S) and ND5 (S573L) haplogroup IB is carrying the derived state while all the other *C. caretta* haplogroups and species carry the other state. In ATP6 we found two amino acids affected by nonsynonymous

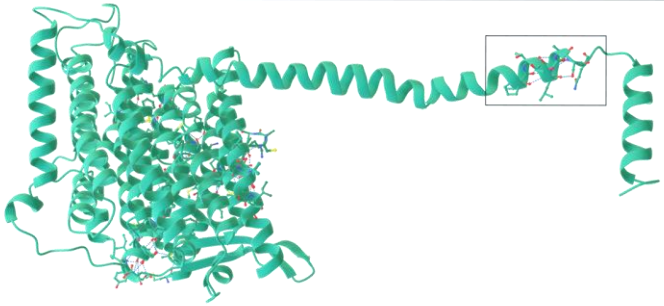
changes between IB and II (A78T and F151L) that are both in the intramembrane helix of the protein (Fig. S7). The total number of hydrogen bonds is maintained between haplogroups, but the Leu 151 carried by IB as a derived state is one of the likely affecting protein functions according to PROVEAN scores. In CYTB (Fig. S7) there are four nonsynonymous changes affecting intramembrane helix of the protein (I47T, F182L, I190V, L233F): overall, haplogroup IB has one more hydrogen bond respect to haplogroup II. In this gene as well, one amino acid change is likely to change protein functions according to PROVEAN; the derived Leu 182, which is shared by haplogroups II and IA, while haplogroup IB carries the ancestral Phe 182, which is shared by all other species. In ND1 (Fig. S7), three amino acids changes affect the intramembrane helix (I16V, I73T, T314A) and one affects the boundary of the membrane (G317S); in the last one, haplogroup IB carries the derived state Ser 317 and it is likely to affect protein functions according to PROVEAN score. Overall, in ND1 haplogroup IB has two hydrogen bonds less than haplogroup II. In ND5 (Figure 4b, c) 8 amino acids changes occur, four in intramembrane helix (L18F, A50T, V87A, I177V) and four at the membrane boundary (Y24C, M31L, T446M, S573L). Haplogroup IB carries the derived state 573Leu, which is likely to affect protein functions. In this case, haplogroup IB results in two more hydrogen bonds respect to haplogroup II.

Non-synonymous changes found in other genes are not likely to affect protein functions according to PROVEAN scores, since imply no changes or very few changes in the number of hydrogen bonds between haplogroups (Fig. 4). In ND2 (Fig. S7), one amino acid change occurs in the intramembrane helix (T31I), while two occur at the membrane boundary (H4Y, N85D). The total number of hydrogen bonds is maintained between haplogroups. In ND3 (Fig. S7) there are two amino acids changes occurring at the membrane boundary (M5T, V108I), with no differences in number of hydrogen bonds between haplogroups. In ND4 (Fig. S7) one amino acid change (A35S) occurs in the intramembrane helix; haplogroup IB has two hydrogen bonds less than haplogroup II.

a)

	ATP6		CYTB				ND1				ND2			ND3		ND 4	ND5							
	78	151	47	182	190	233	16	73	314	317	4	31	85	5	108	35	18	24	31	50	87	177	466	573
H IA	Thr (d)	Phe (a)	Ile (a)	**Leu (d)***	Ile (a)	Leu (a)	Val (d)	Ile (a)	Thr (a)	Gly (a)	Tyr (d)	Thr (a)	Asn (a)	Thr (d)	Val (a)	Ala (a)	Leu (a)	Cys (d)	Met (a)	Ala (a)	Val (a)	Ile (a)	Thr (a)	Ser (a)
C. caretta H II	Thr (d)	Phe (a)	Ile (a)	**Leu (d)***	Ile (a)	Leu (a)	Val (d)	Ile (a)	Thr (a)	Gly (a)	Tyr (d)	Thr (a)	Asp (d)	Thr (d)	Ile (d)	Ala (a)	Phe (d)	Tyr (a)	Leu (d)	Ala (a)	Val (a)	Ile (a)	Thr (a)	Ser (a)
H IB	Ala (a)	**Leu (d)***	Thr (d)	Phe (a)	Val (d)	Phe (d)	Ile (a)	Thr (d)	Ala (d)	**Ser (d)***	His (a)	Ile (d)	Asn (a)	Met (a)	Val (a)	Ser (d)	Leu (a)	Cys (d)	Met (a)	Thr (d)	Ala (d)	Val (d)	Met (d)	**Leu (d)***
L. kempii	Ala (a)	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile (a)	Thr (a)	Gly (a)	Gln	Thr (a)	Asn (a)	Met (a)	Val (a)	Ala (a)	Leu (a)	His	Met (a)	His	Met	Ile (a)	Thr (a)	Ser (a)
L. olivacea	Thr (d)	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile (a)	Thr (a)	Gly (a)	Gln	Thr (a)	Asn (a)	Met (a)	Val (a)	Ala (a)	Ile	His	Met (a)	Tyr	Met	Ile (a)	Thr (a)	Ser (a)
E. imbricata	Thr/ Ala	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile/ Thr/ Ser	Ala (d)	Gly (a)	His (a)	Thr (a)	Asn /Asp	Met (a)	Ile/ Leu	Ala (a)	Leu (a)	His/ Tyr	Met /Val	Ser	Val (a)	Ile (a)	Thr (a)	Ser (a)
C. mydas	Thr (d)	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile (a)	Thr (a)	Gly (a)	Tyr (d)	Thr (a)	Asp (d)	Met (a)	Ile (d)	Ala (a)	Leu (a)	Tyr (a)	Met /Thr	Val/I le	Val (a)	Ile (a)	Thr (a)	Ser (a)
N. depressus	Ala (a)	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile (a)	Ala (d)	Gly (a)	His (a)	Thr (a)	Asp (d)	Met (a)	Leu (a)	Ala (a)	Leu (a)	Tyr (a)	Met (a)	Ser	Val (a)	Ile (a)	Thr (a)	Ser (a)
D. coriacea	Ala (a)	Phe (a)	Ile (a)	Phe (a)	Ile (a)	Leu (a)	Ile (a)	Ile (a)	Ala/ Thr	Gly (a)	His (a)	Thr (a)	Asp (d)	Met (a)	Ile (d)	Ala (a)	Leu (a)	Tyr (a)	-	Ser	Val (a)	Ile (a)	Thr (a)	Ser (a)

b)



c)

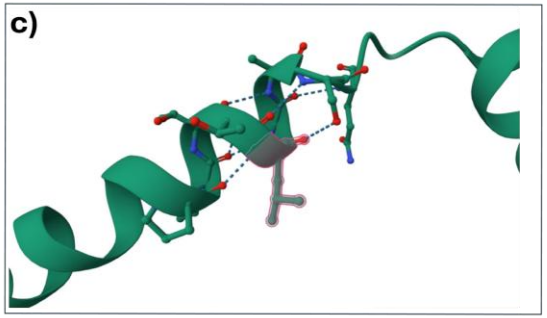


Fig. 4 In panel a), nonsynonymous substitutions, fixed between haplogroup II and haplogroup IB. Ancestral (a) and derived (d) states are considered in respect with the state at the divergence of haplogroups within *C. caretta.* In orange, the state that is shared by haplogroup II and other taxa; in red, the state that is shared by haplogroup IB and other taxa. In bold with *, the substitutions with a PROVEAN score < -2.5, likely influencing the protein function. In panel b), protein structure of ND5 gene as in haplogroup IB; amino acids differences respect to haplogroup II are shown. In panel c), detail of Leu 537 in ND5, discussed in the text.

Discussion

The species distribution is shaped by phylogeographic processes and adaptation. Disentangling their evolutionary history can help understanding the mechanisms responsible for the current patterns and thus providing insights on how the species could cope with the present threats by changing their distribution. Marine turtles have been colonising all temperate oceans and sea during the last million years, and their current distribution is tempered by the drastic biogeographic and climatic changes that shaped current earth's biodiversity. Our study shows that the mitogenome, as the printed legacy of this evolutionary history, can unveil the sequential steps that led to the present distribution and some of the adaptation drivers differentially affecting the present lineages.

Global phylogeography and past dynamics of oceans colonization

Whole mitogenome sequencing has proved to increase the resolution of maternal lineages in loggerhead turtles (Tolve et al. 2024) respect to traditional markers, such as single genes and control region fragments. Therefore, we used whole mtDNA sequences to disentangle the global phylogeography of this species, which until now has been studied mainly by means of control region haplotypes (e.g. Shamblin et al. 2014), without a worldwide spread sampling (Tolve et al. 2024) or by using only a very few numbers of mitogenomes (e.g. Duchene et al. 2012). The consensus trees and the mitogenomic Maximum Likelihood tree and most single gene trees supports that haplogroup IB diverges first, while haplogroups IA and II diverged later and cluster together. Of note, some single genes support the other scenario, which explains previous incongruent results found depending on the genetic marker used (Shamblin et al. 2014; Ludwig et al. 2024). For instance, in some genes selective pressure might be in action, and in the control region a high mutation rate can lead to saturation of the genetic signal (as shown in the TCS networks), both cases confounding the species phylogeny. The mitogenome is a single non-recombinant DNA molecule. Consequently, mutation rates and the effect of genetic drift are expected to be even across the whole molecule. However, selective pressures can be very variable along the mitogenome, and even within different positions within the genes, creating a landscape of differences in terms of genetic diversity and degree of persistence of the signal of the evolutionary trajectory. These differences likely

explain the different topologies found in our study and stress the importance of using the whole mitogenome rather than partial sequences, when performing phylogenetic analysis. Our data strongly supports that haplogroup IB diverged first, while haplogroups IA and II cluster together, in agreement with Baltazar-Soares et al. (2020), which reached the same conclusion without the inclusion of samples from the Pacific Ocean representative of haplogroup IA. This tree topology has already been found by other authors (e.g. Duchene et al. 2012; Tolve et al. 2024), but a reduced worldwide representation limited the interpretation of the tree to make inferences about the species' phylogeography. In our case, we obtained a solid support to solve the phylogeography of the species and we were able to link the major divergence events to known biogeographical events. According to our results, haplogroup IB diverged due to the closure of Panama Isthmus, as the timing of this node in our phylogeny (2.07; 1.37 – 2.85 MYA) and the geographic distribution of the corresponding branches is consistent with this geological event (12 – 2.8 MYA) (O'Dea et al. 2016). After this initial split, the haplogroup IB colonized the Atlantic rookeries, while haplogroups IA and II diverged later in the Pacific/Indo Pacific Oceans (1.37; 0.87 – 1.86 MYA), during the Mid Pleistocene Transition (MPT, 1.2 – 0.5 MYA). The MPT marks a change in the periodicity and intensity of the glaciations, from cycles of 41 KY to cycles of 100 KY (Head and Gibbard 2005). Therefore, the separation of haplogroups IA and II might have occurred as a consequence of a long and intense ice spreading in the MPT. The following ice melting and sea level rise would have limited haplogroups mixing, a possible cofactor being the minimum population size that sea turtle species reached in correspondence of MPT as suggested in other studies (Vilaça et al. 2021). When the Agulhas currents formed (around 130 KYA) (Turney and Jones 2010), haplogroup II had probably already spread in the Indo-Pacific at that time. Thus, the warm currents of Agulhas allowed the westward movements from South Africa for several species (e.g. Floeter et al. 2008), and likely for loggerheads as well. Finally. turtles from South Africa moved northward until haplogroup II entered the Atlantic and the Mediterranean.

The radiation of haplogroup II, according to our results, started 50 – 170 KYA, maybe along with the Mediterranean colonization, which has been dated around 65 KYA in previous studies (Clusa et al. 2013) and considered starting in Libya, as a climate refugia during the Last Glacial Maximum. In our tree, the first branch diverging within haplogroup II is haplotype A2.9.1 (so as in Tolve et al. 2024), whose correspondent control region haplotype (A2.9) that is characteristic of Libyan rookeries and

hypothesizes ad the first divergence within the Mediterranean (Clusa et al. 2013). Previous studies (Baltazar-Soares et al. (2020) hypothesized that turtles from the Mediterranean would have recolonized Atlantic rookeries where haplogroup IB had settled before, and they proposed Cape Verde as first would have been a stepping stone towards western Atlantic. However, analysis on a wider sampling set of the control region identified more ancestral splits of the haplogroup II (Shamblin et al. 2024), suggesting that the first arrival of this haplogroup in the Atlanto-Mediterranean region may be in some of the actual Atlantic nesting populations before entering the Mediterranean. Future work including more haplogroup II complete mitogenomes from Cape Verde and western Atlantic rookeries is needed in order to clarify the details and the timing of the spreading of this haplogroup in the region. Our study has a global perspective about loggerhead phylogeography; however, there are some regions not represented in our sampling, such as Indo Pacific or Southwestern Atlantic, where information would be crucial to refine the reconstruction of the species evolutionary history. For example, in Brazil, a new sub-haplogroup of the haplogroup IB has recently been described (Ludwig et al. 2024), as specific to the Southwestern Atlantic. This sub-haplogroup might include the most ancient lineages settling in the Atlantic, along with Reis et al. (2010), which would agree with the hypothesis of a first colonisation of the south America populations of this haplogroups and later colonisation of northwestern Atlantic and the Cape Verde (Soares et al. 2020). Complete mitogenomes for loggerhead turtles have been scarce to date, and we encourage a broader application of this technique to take full advantage of the enhanced information it can provide and to aid solving these few fine scale patterns of dispersal.

Selective pressure at mitochondrial level

One of the most interesting consequences of this global phylogeography pattern is the fact that individuals of very divergent lineages, haplogroups IB and II, are coexisting in the same nesting populations due to this secondary contact. These two haplogroups follow a clear cline in western Atlantic rookeries, and along two independent latitudinal gradients (e.g. at both sides of the Florida Peninsula) as described in the past years by other authors (Encalada et al. 1996; Shamblin et al. 2012b). However, no one has investigated to date the role of selective pressure in maintaining this pattern, as expected when independent clines are found. Our results indicate that the relative frequency of haplogroup IB is inversely related to temperature, suggesting that this haplogroup might be more

competitive, compared to haplogroup II, in tolerating a cold environment. This is not unprecedented in marine species, as mitochondrial evidence of adaptation to climatic/thermal gradients has been found at intraspecific and interspecific levels (e.g. Silva et al. 2014; Baltazar-Soares et al. 2021; Noll et al. 2022). Under this hypothesis, we looked for deviation from neutrality and signals of selection across the mitochondrial coding genes of our global mitogenome dataset, to obtain additional insights into the adaptation implications of the different haplogroups within the species, with a focus on haplogroups IB and II.

McDonald-Kreitman test between haplogroups IB and II revealed significant deviation from neutrality, mostly driven by ND3 gene. The excess of nonsynonymous changes that we found within haplogroups, compared to nonsynonymous changes between haplogroups, is usually explained by purifying selection preventing nonsynonymous changes to become fixed between groups. This goes along with the conservative nature of mitochondrion. Codeml analyses also gave significant results when applying the branch method. Our results supported less stringent purifying selection on branch IB, but without support for positive selection. Positive selection on mitochondrial DNA might be difficult to detect, because it often operates on few amino acids, and the signal might be masked. For this reason, some authors consider that the lack of positive selection should not be taken as reliable when sample size is small (Zhang et al. 2005; Deng et al. 2019). Consequently, even without statistical support for positive selection on haplogroup IB, we found evidence of a different selective force affecting IB and several fixed nonsynonymous changes between haplogroup IB and haplogroup II that were further investigated. For most of these substitutions, haplogroup IB carries the derived state and haplogroup II and IA the ancestral state, which is also shared by all the other sea turtle species. In other words, the other sea turtles, which are all more tropical and subtropical than *C. caretta*, are more like haplogroup II and IA than like haplogroup IB for those fixes nonsynonymous changes. This result is consistent with the hypothesis of adaptation to cold temperatures by haplogroup IB and suggests that haplogroup IB has been following a different evolutionary path respect to the other ‘warmer’ lineages. This also implies that in the event of coexistence of different mitochondrial lineages due to a secondary contact, selection may shape their relative frequencies and even can lead to the extinction of one of the two haplogroups at the edges of the cline, as observed in the western Atlantic populations. Looking in the details, these amino acids changes generally maintain the characteristics of the protein structures, but some of them may affect the

boundary of the membrane and can play a role in the homeoviscous adaptation, as other authors previously reported (Novelletto et al. 2016). These modifications of the membrane fluidity are likely to affect the adaptation to temperatures and support the potential role of the mitogenome in the adaptation to temperature variations.

Mitochondrial thermal adaptation has been studied in several marine species. In *Engraulis encrasicolus* (Silva et al. 2014) positive selection in one codon position of gene CYTB has been found to explain a major part of the two-clade mitochondrial structure in eastern Atlantic, with temperature as the main predictor. Similarly, in *Sardina pilchardus* (Baltazar-Soares et al. 2021) the frequencies of the most common CYTB and ATP6 haplotypes, which show signatures of positive selection, negatively correlate with minimum sea surface temperature and dissolved oxygen. In *Pygoscelis papua* (Noll et al. 2022) positive selection was found in ND genes, involving haplotypes that are fixed in a non-random manner among lineages and suggesting that mitogenome adaptation promoted penguins' survival in different geographic areas. Comparing our results for loggerheads with those for other species in literature, we did not find exact correspondence of the fixed aminoacidic changes under selection, even if the genes and the region of the encoded proteins sometimes are the same. For example, nonsynonymous changes in gene ND5, in correspondence of the piston of the encoded protein, such as 573 Leu in haplogroup IB, have been observed in genus *Oncorhynchus* to affect the function of ND5 subunit in response to temperatures (Garvin et al. 2011). These results are promising for further investigation in future studies, by increasing the number of mitogenomes and geographic coverage. A thorough understanding of such evolutionary processes is also crucial for comprehending the threats these animals will face in the near future, due to climate change.

Conclusions

In this study, we applied whole mitogenome sequencing to address global phylogeography and evolutionary biology of loggerhead turtles. We enhanced the resolution of matrilines and clarified the historical dynamics of oceans' colonization, assessing the first Atlantic settlements by haplogroup IB and the South African entrance in Atlantic by haplogroup II. We also used mitogenomic data to explore the role of selective pressure in mitochondrial differentiation and distribution. We found evidence

for differential selection on mitochondrial haplogroups and paved the way for future studies on thermal adaptation of loggerhead turtles.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Baltazar-Soares M, de Araújo Lima AR, Silva G (2021) Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. Genes 12: 91.

Baltazar-Soares M, Klein JD, Correia SM, Reischig T, Taxonera A, Roque SM, Dos Passos L, Durão J, Lomba JP, Dinis H, Cameron SJK, Stiebens VA, Eizaguirre C (2020) Distribution of genetic diversity reveals colonization patterns and philopatry of the loggerhead sea turtles across geographic scales. Sci Rep 10: 18001.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC (1994) Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. Evolution 48: 1820-1828.

Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. Mol Ecol 16.

Carreras C, Pascual M, Cardona L (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. Conserv Genet 8.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLOS ONE 7: e46688.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1659.

Clusa M, Carreras C, Pascual M, Demetropoulos A, Margaritoulis D, Rees AF, Hamza AA, Khalil M, Aureggi M, Levy Y (2013) Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). J Exp Mar Biol Ecol 439: 15-24.

Deng Z, Xu S, Gao T, Han Z (2019) Thermal adaptation of Japanese sand lance (*Ammodytes personatus*) in different ocean currents revealed by the cytochrome gene. Marine and Freshwater Research 70: 1484-1491.

Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18-e18.

Drosopoulou E, Tsiamis G, Mavropoulou M, Vittas S, Katselidis KA, Schofield G, Palaiologou D, Sartsidis T, Bourtzis K, Pantis J (2012) The complete mitochondrial genome of the loggerhead turtle *Caretta caretta* (Testudines: Cheloniidae): Genome description and phylogenetic considerations. Mitochondrial DNA 23: 1-12.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK (1999) Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). J Zool 248: 397-409.

Eizaguirre C, Baltazar-Soares M (2014) Evolutionary conservation—evaluating the adaptive potential of species. Evolutionary Applications 7: 963-967.

Encalada SE, Bjorndal KA, Bolten AB, Zurita JC, Schroeder B, Possardt E, Sears CJ, Bowen BW (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: 567-575.

Encalada SE, Lahanas PN, Bjorndal KA, Bolten AB, Miyamoto MM, Bowen BW (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. Molecular Ecology 5: 473-483.

Floeter SR, Rocha LA, Robertson DR, Joyeux J-C, Smith-Vaniz WF, Wirtz P, Edwards AJ, Barreiros JP, Ferreira C, Gasparini JL, Brito A, Falcón J, Bowen B, Bernardi G (2008) Atlantic reef fish biogeography and evolution. J Biogeogr 35: 22-47.

Galià-Camps C, Enguídanos A, Turon X, Pascual M, Carreras C (2024) The past, the recent, and the ongoing evolutionary processes of the worldwide invasive ascidian Styela plicata. Molecular Ecology 33: e17502.

Garvin MR, Bielawski JP, Gharrett AJ (2011) Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific Salmon. PLOS ONE 6: e24127.

Head MJ, Gibbard PL (2005) Early-Middle Pleistocene transitions: an overview and recommendation for the defining boundary. Geological Society, London, Special Publications 247: 1-18.

Hernández-Fernández J, Delgado Cano DA (2018) Complete mitochondrial genome of the three Colombian Caribbean loggerhead sea turtles, using next-generation sequencing. Mitochondrial DNA B 3: 626-627.

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

Ludwig S, Amorim L, Barcelos AC, Guimarães PR, Vargas SM (2024) Inferring the origin of new D-loop haplotypes of loggerhead sea turtles (Testudinata: Cheloniidae) from the Southwest Atlantic lineage. Mitochondrial DNA Part A: 1-8.

Lutz PL, Musick JA (2017) The Biology of Sea Turtles, Volume I. CRC Press

Muñoz Sabater J (2019) ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)

Noll D, Leon F, Brandt D, Pistorius P, Le Bohec C, Bonadonna F, Trathan PN, Barbosa A, Rey AR, Dantas GPM, Bowie RCK, Poulin E, Vianna JA (2022) Positive selection over the mitochondrial genome and its role in the diversification of gentoo penguins in response to adaptation in isolation. Scientific Reports 12: 3767.

'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD, Stallard RF, Woodburne MO, Aguilera O, Aubry M-P, Berggren WA, Budd AF, Cozzuol MA, Coppard SE, Duque-Caro H, Finnegan S, Gasparini GM, Grossman EL, Johnson KG, Keigwin LD, Knowlton N, Leigh EG, Leonard-Pingel JS, Marko PB, Pyenson ND, Rachello-Dolmen PG, Soibelzon E, Soibelzon L, Todd JA, Vermeij GJ, Jackson JBC (2016) Formation of the Isthmus of Panama. Science Advances 2: e1600883.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol 67: 901-904.

Reis EC, Soares LS, Vargas SM, Santos FR, Young RJ, Bjorndal KA, Bolten AB, Lôbo-Hajdu G (2010) Genetic composition, population structure and phylogeography of the loggerhead sea turtle: colonization hypothesis for the Brazilian rookeries. Conservation Genetics 11: 1467-1477.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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: e85956.

Shamblin BM, Bolten AB, Bjorndal KA, Dutton PH, Neilen JT, Abreu-Grobois FA, Reich KJ, Witherington BE, Bagley DA, Ehrhart LM, Tucker AD, Addison DS, Arenas A, Johnson C, Carthy RR, Lamont MM, Dodd MG, Gaines MS, LaCasella E, Nairn CJ (2012) Expanded mitochondrial control region sequences increase resolution of stock structure among North Atlantic loggerhead turtle rookeries. Mar Ecol Prog Ser 469: 145-160.

Silva G, Lima FP, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. Proceedings of the Royal Society B: Biological Sciences 281: 20141093.

Suchard MA, Lemey P, Baele G (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619-633.

Tolve L, Iannucci A, Garofalo L, Ninni A, Capobianco Dondona A, Ceciarini I, Cocumelli C, De Lucia A, Falconi M, Formia A, Iacovelli F, Mancusi C, Marchiori E, Marsili L, Mingozzi T, Nannarelli S, Natali C, Terracciano G, Zuffi MAL, Novelletto A, Ciofi C (2023) Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine Biology 171: 19.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44: W232-W235.

Turney CS, Jones RT (2010) Does the Agulhas Current amplify global temperatures during super-interglacials? J Quaternary Sci 25: 839-843.

Vilaça ST, Piccinno R, Rota-Stabelli O, Gabrielli M, Benazzo A, Matschiner M, Soares LS, Bolten AB, Bjorndal KA, Bertorelle G (2021) Divergence and hybridization in sea turtles: Inferences from genome data show evidence of ancient gene flow between species. Molecular Ecology 30: 6178-6192.

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. Bioinformatics 19: 671-672.

Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24: 1586-1591.

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19: 153.

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Molecular Biology and Evolution 22: 2472-2479.

Supplementary Materials

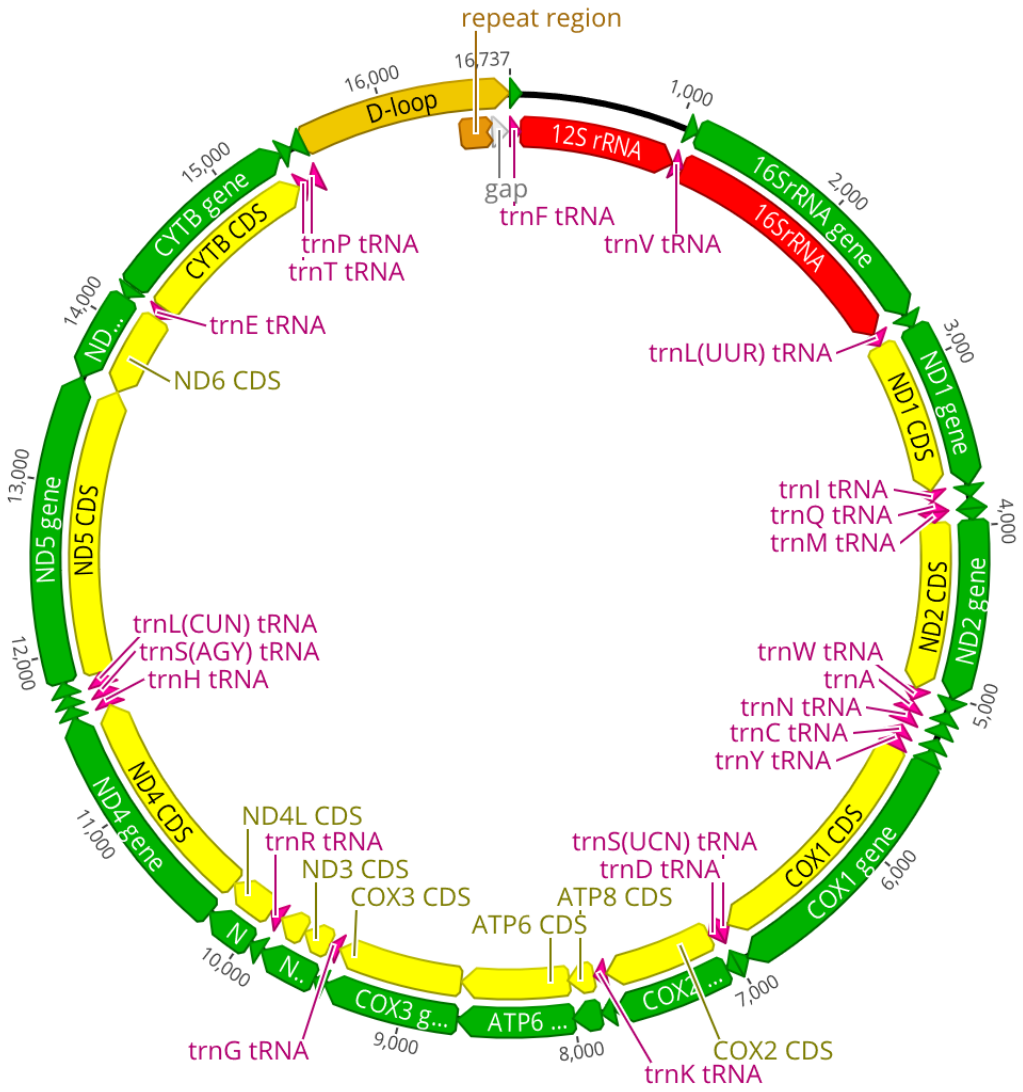


Fig. S1 *Caretta caretta*'s mitogenome structure; in red, rRNA genes 12S and 16S; in yellow, coding genes ND1, ND2, COXI, COXII, ATP8, ATP6, COXIII, ND3, ND4L, ND4, ND5, ND6, CYTB; in orange, control region sequence (D-loop) and its repeat region; in pink, tRNA genes, and in grey the gap corresponding to size changes according to the number of repeats.

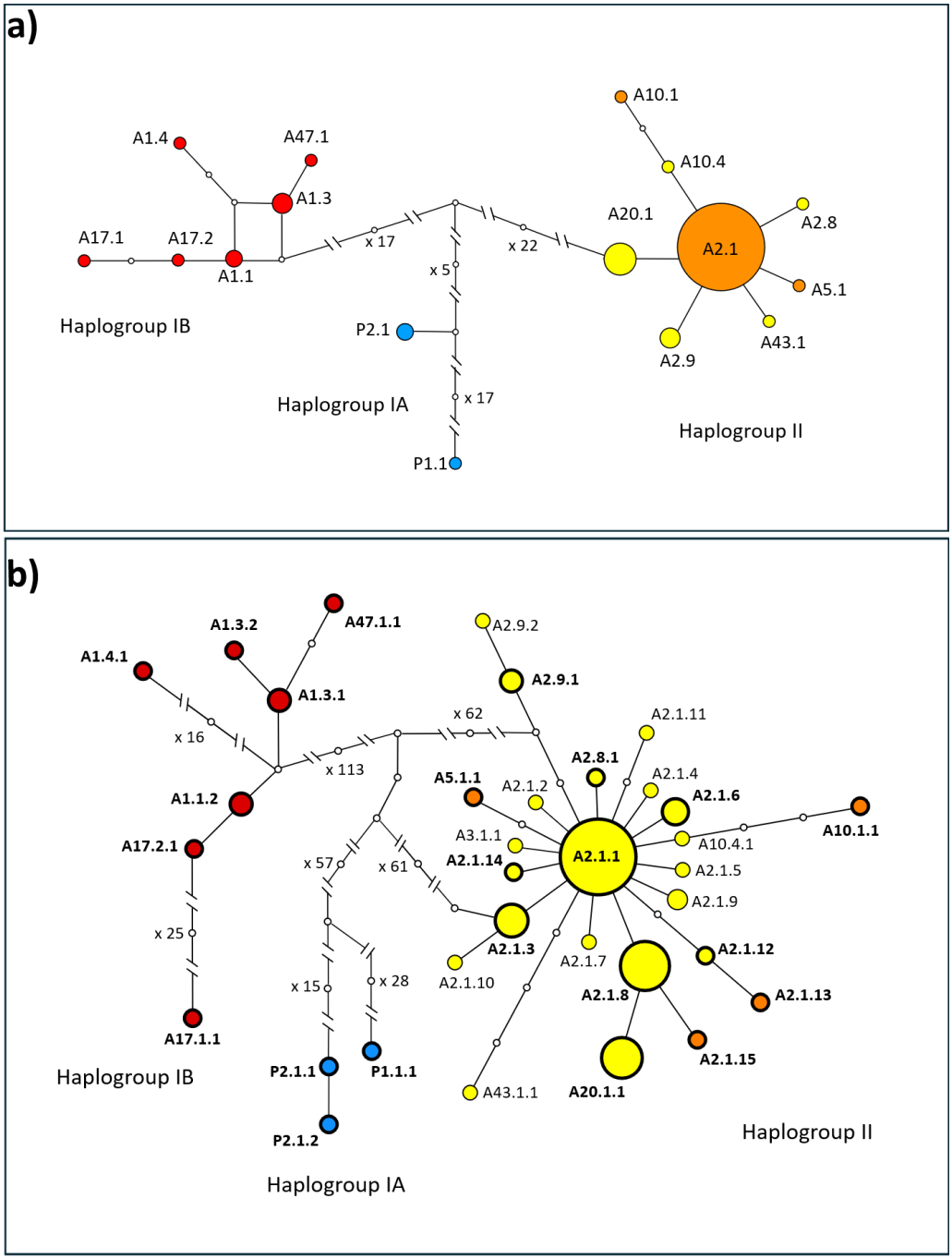


Fig. S2 TCS haplotype networks based on control region haplotypes (A) and on mitogenomic haplotypes (B), recorded in sequences from the present study and sequences downloaded from GenBank (n = 93, Supplementary Table S1). The size of each circle is proportional to the number of sequences carrying that haplotype. In red, haplotypes from haplogroup IB that have been sampled only in Atlantic; in blue, haplotypes from haplogroup IA that have been sampled only in Pacific; in yellow, haplotypes from haplogroup II that have been sampled only in Mediterranean; in orange, haplotypes from haplogroup II that have been sampled both in Mediterranean and Atlantic. Haplotypes in bold have been used in phylogenetic analyses.

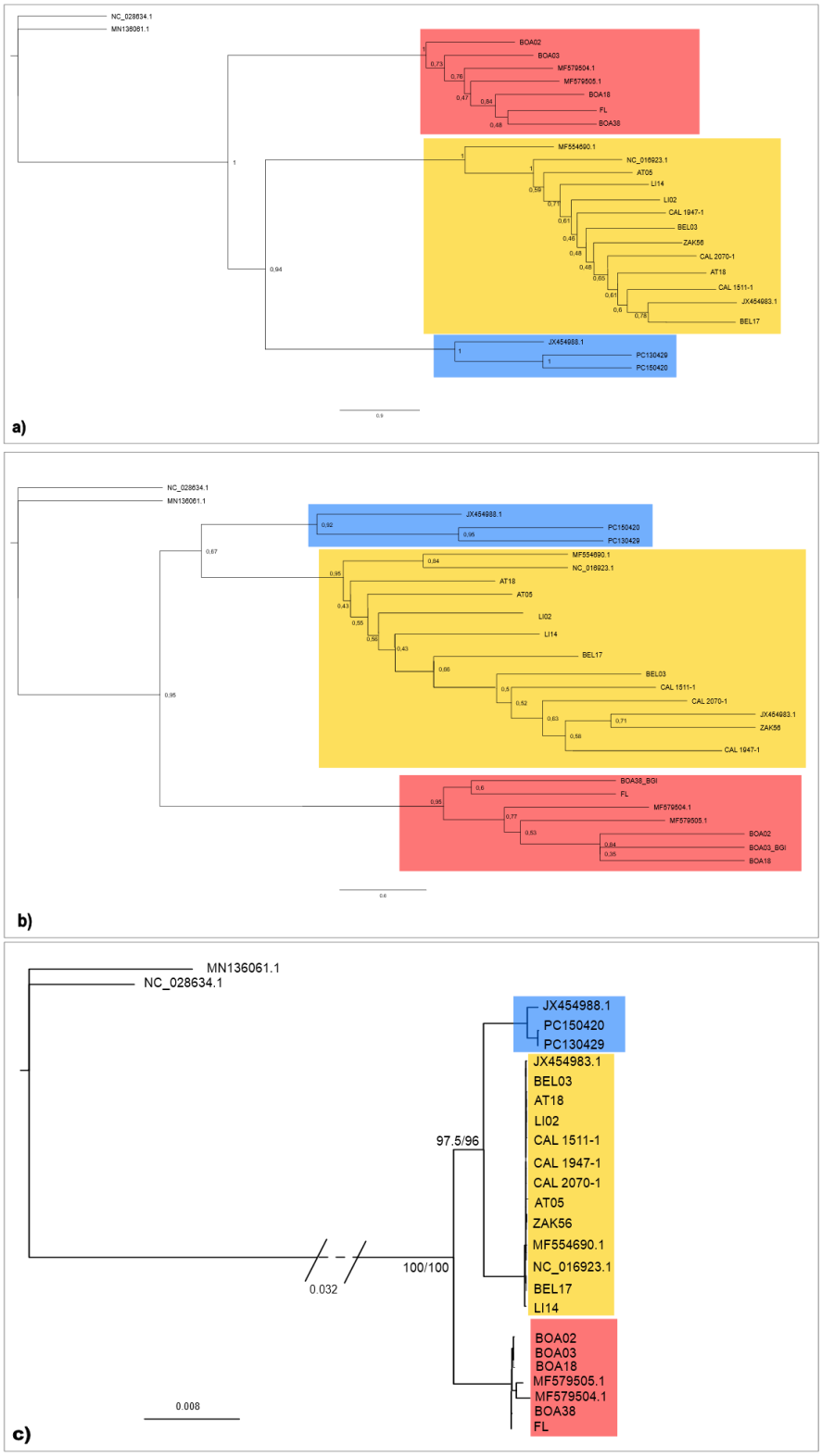


Fig. S3 Consensus trees built in Astral-III, with A) single gene trees, B) partitions trees. Support for the nodes is indicated in PP (posterior probability). Maximum Likelihood phylogeny of complete mitogenomes, C). Nodes support in SH-aLRT (%) /ultrafast bootstrap (%). Distances are in substitutions per site. In A, B, C haplogroup IA is shown in blue, haplogroup IB is shown in red, haplogroup II is shown in yellow.

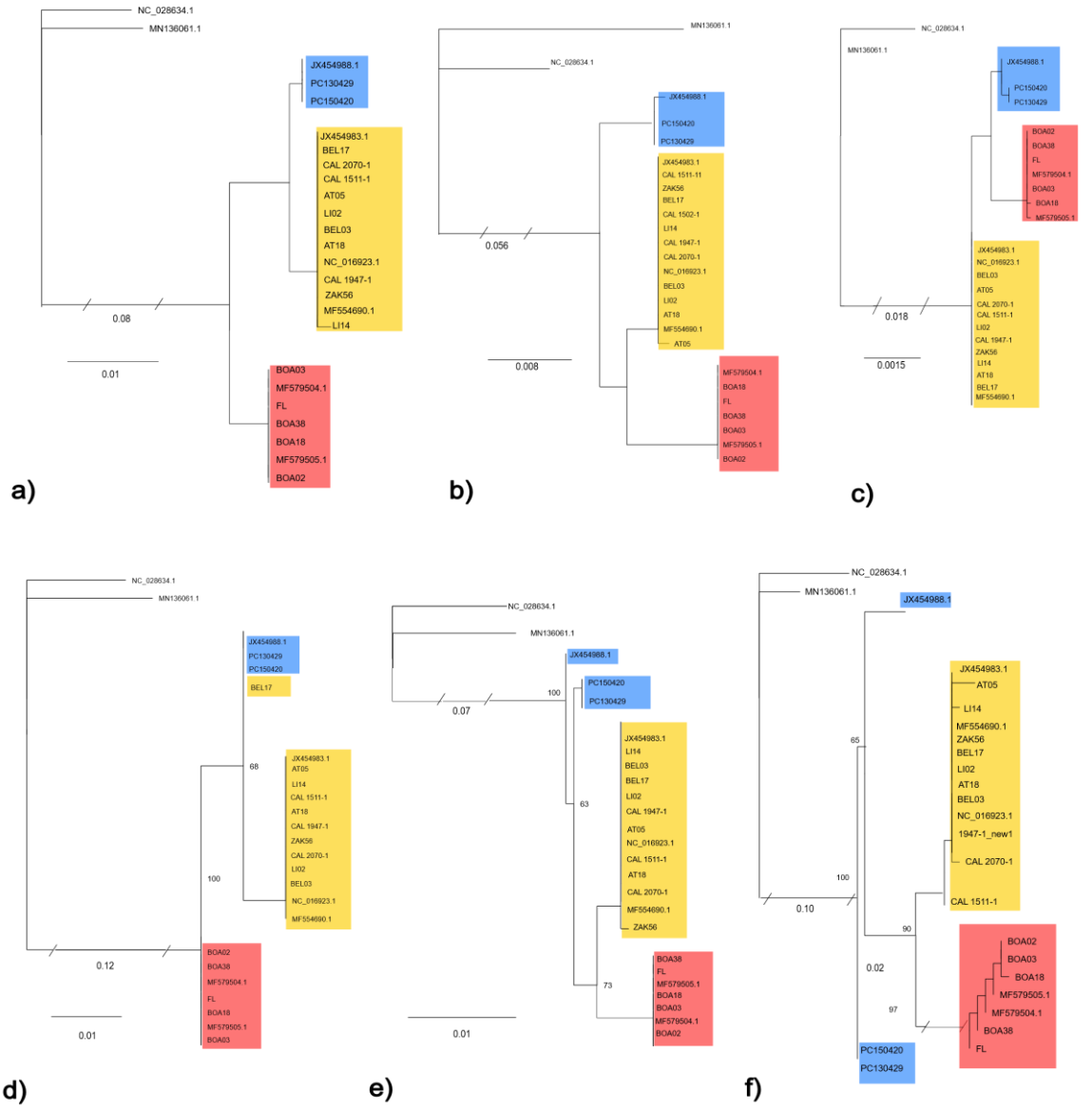


Fig. S4 Alternative topologies obtained by single gene phylogenetic analysis. a) is supported by ATP6, COXI, COXII, COXIII, CYTB, ND3, 12S rRNA, 16S rRNA; b) is supported by ND1, ND5, ND6; c) is supported by ND2, ND4L; d) is the one obtained using ATP8, showing paraphilia of haplogroup II; e) is the one obtained by ND4, showing paraphilia of haplogroup IA; f) is the one obtained by the control region, showing paraphilia of haplogroup IA. See also Table S3.

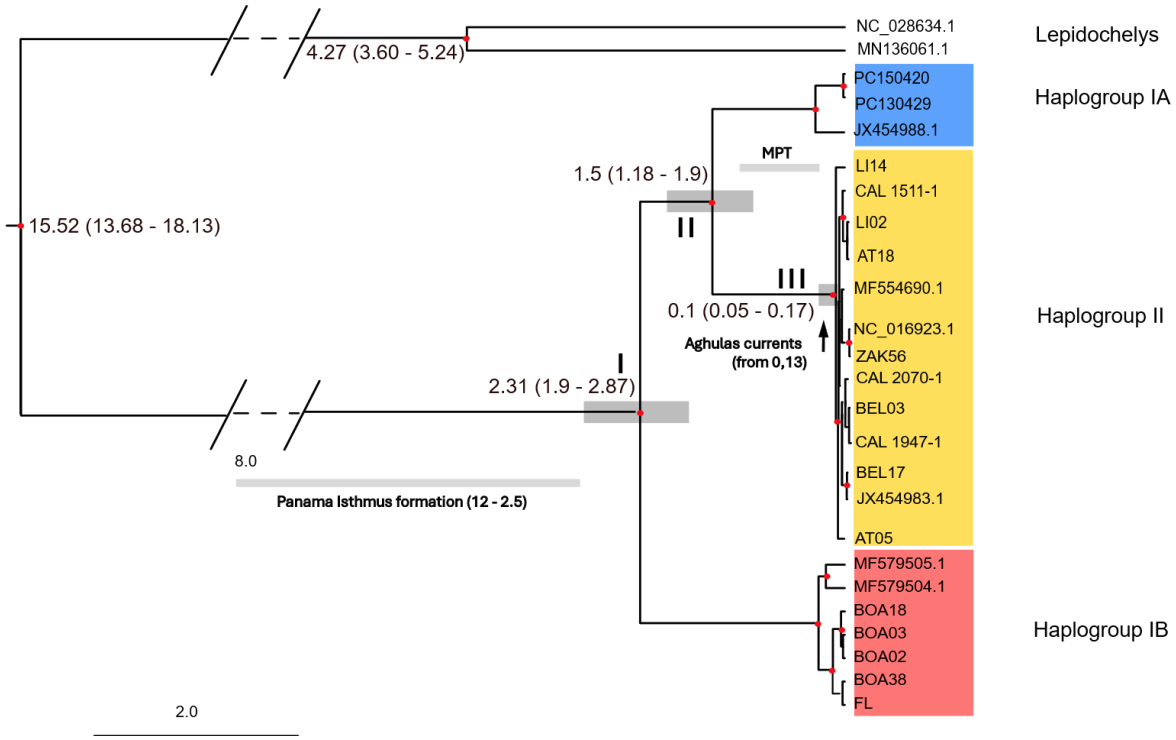


Fig. S5 Bayesian phylogeny built with prior on the time of the root (Duchene et al. 2012). Red dots indicate the node with bootstrap > 90%. Height and 95% HDP are shown for the nodes that are discussed in the text. Distances are in millions of years.

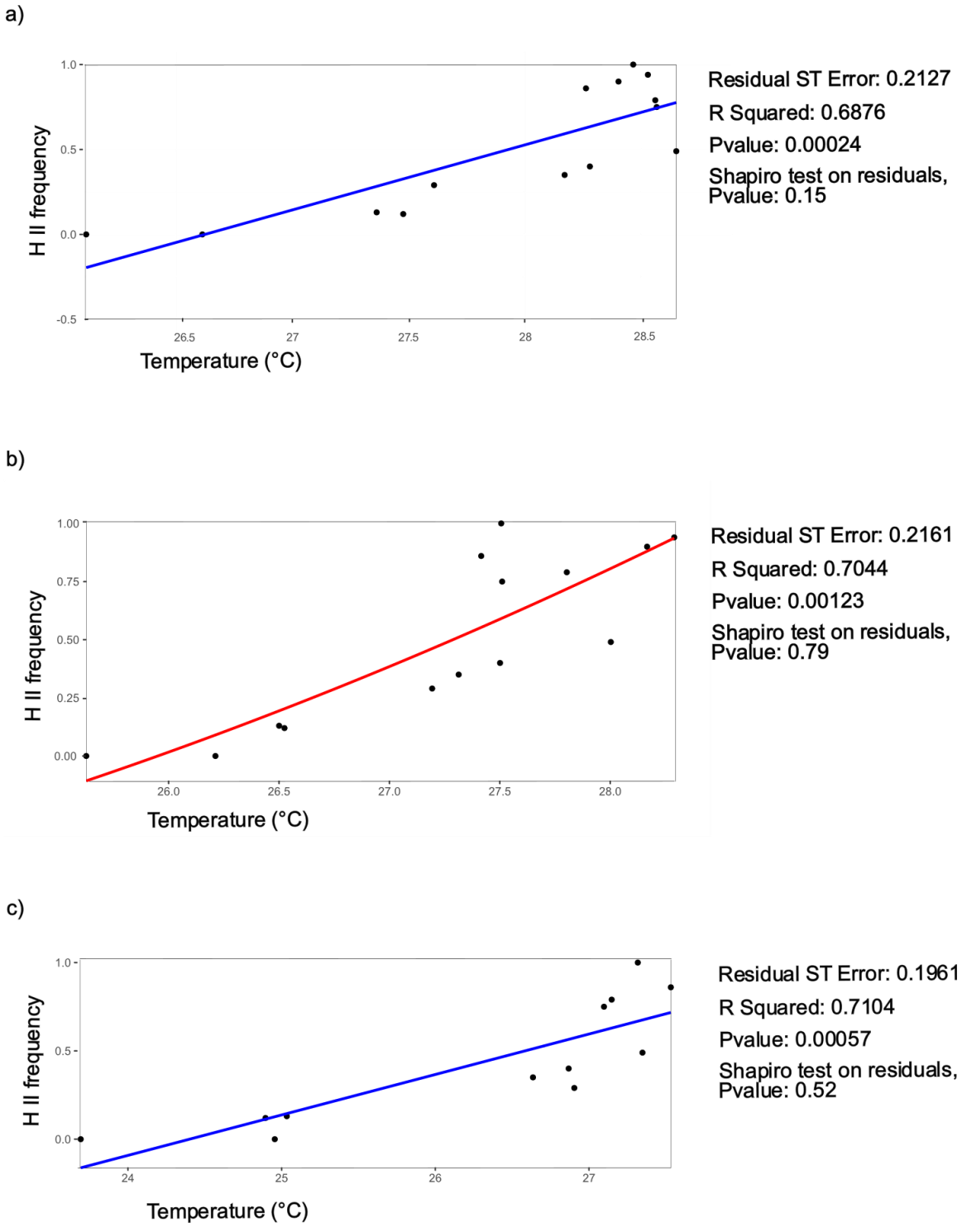


Fig. S6 Suboptimal models of regressions between temperatures and Haplogroup II's relative frequency in rookeries of Western Atlantic. Respectively, in a) sea water temperature, in b) air temperature, in c) soil temperature was used as independent variable for linear regression (a), (c) and for polynomial regression (b).

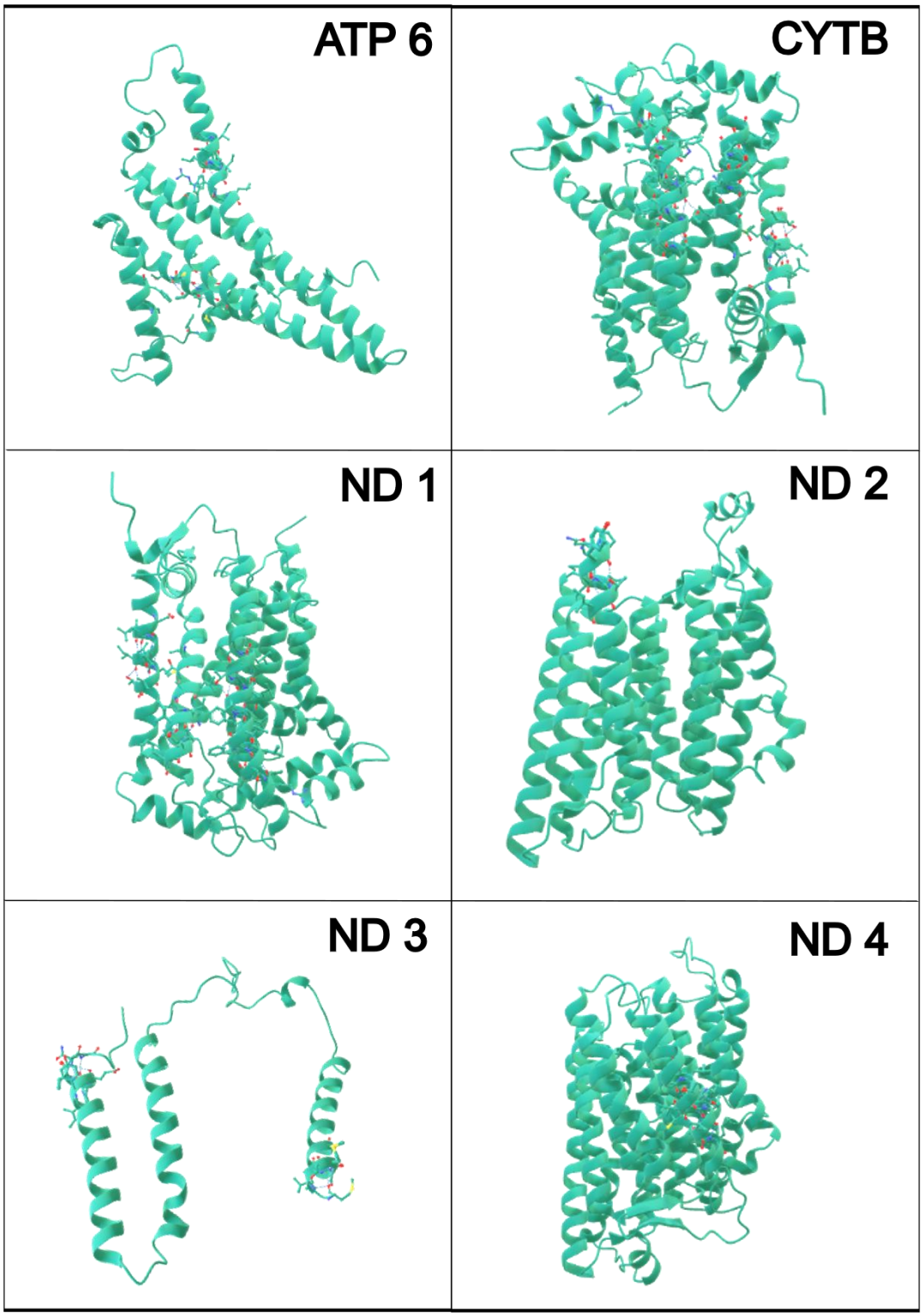


Fig. S7 Protein structure of coding genes as in Haplogroup IB. Only genes carrying nonsynonymous substitutions that are fixed between haplogroup IB and haplogroup II are shown (find ND5 in the main text). Aminoacidic changes are highlighted in each structure.

Supplementary Materials

Table S1 Mitogenomic sequences included in the present work. For each mitogenome we report the species, ID sample, accession number, reference, information on sampling and sequencing coverage. We provide corresponding control region haplotype, mitogenomic haplotype and haplogroup. (*) New mitogenomic haplotypes. In grey, sequences that have been used in genetic diversity, phylogenetic and selective pressure analyses.

n	Species	Sample ID	GenBank Accession Number	Reference	Sampling location	Geographic area	Sample type	Reads subsamp fraction	mtDNA sequencing coverage	*C. caretta* control region haplotype	*C. caretta* mtDNA haplotype	*C. caretta* Haplogroup
1	*C. caretta*	ZAK51		Present study	Zakinthos (Greece)	Mediterranean Sea	Nesting ground	98.97%	512	A2.1	A2.1.1	II
2	*C. caretta*	ZAK56		Present study	Zakinthos (Greece)	Mediterranean Sea	Nesting ground	99.1%	1061	A2.1	A2.1.6	II
3	*C. caretta*	LI02		Present study	Sirte (Libya)	Mediterranean Sea	Nesting ground	99.03%	1871	A2.1	A2.1.8	II
4	*C. caretta*	LI14		Present study	Sirte (Libya)	Mediterranean Sea	Nesting ground	99.56%	6558	A2.9	A2.9.1	II
5	*C. caretta*	BEL03		Present study	Belek (Turkey)	Mediterranean Sea	Nesting ground	97.47%	1763	A2.1	A2.1.1	II
6	*C. caretta*	BEL17		Present study	Belek (Turkey)	Mediterranean Sea	Nesting ground	98.91%	2766	A2.1	A2.1.3	II
7	*C. caretta*	CAL 1511-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	1220	A20.1	A20.1.1	II
8	*C. caretta*	CAL 1502-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	789	A2.1	A2.1.8	II

9	*C. caretta*	CAL 20105-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	1970	A2.1	A2.1.6	II
10	*C. caretta*	CAL 2070-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	806	A2.8	A2.8.1*	II
11	*C. caretta*	CAL 1942-2		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	5677	A2.1	A2.1.1	II
12	*C. caretta*	CAL 1947-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	545	A2.1	A2.1.14*	II
13	*C. caretta*	CAL 1653-1		Present study	Calabria (Italy)	Mediterranean Sea	Nesting ground	100%	346	A2.1	A2.1.3	II
14	*C. caretta*	RG Sic 6		Present study	Sicily (Italy)	Mediterranean Sea	Nesting ground	100%	209	A2.1	A2.1.8	II
15	*C. caretta*	SR Sic 7		Present study	Sicily (Italy)	Mediterranean Sea	Nesting ground	100%	109	A2.1	A2.1.1	II
16	*C. caretta*	BOA18		Present study	Boa Vista (Cape Verde)	Atlantic Ocean	Nesting ground	34.34%	195	A47.1	A47.1.1*	IB
17	*C. caretta*	BOA03		Present study	Boa Vista (Cape Verde)	Atlantic Ocean	Nesting ground	98.91%	932	A1.3	A1.3.1*	IB
18	*C. caretta*	BOA02		Present study	Boa Vista (Cape Verde)	Atlantic Ocean	Nesting ground	68.7%	323	A1.3	A1.3.2*	IB
19	*C. caretta*	BOA38		Present study	Boa Vista (Cape Verde)	Atlantic Ocean	Nesting ground	99.05%	883	A17.2	A17.2.1*	IB

20	*C. caretta*	BOA48		Present study	Boa Vista (Cape Verde)	Atlantic Ocean	Nesting ground	34.31%	301	A1.3	A1.3.1*	IB
21	*C. caretta*	FL		Present study	Northeast Florida (USA)	Atlantic Ocean	Nesting ground	3.2%	269	A1.1	A1.1.2*	IB
22	*C. caretta*	FL2		Present study	Northeast Florida (USA)	Atlantic Ocean	Nesting ground	23.2%	236	A1.1	A1.1.2*	IB
23	*C. caretta*	AT18		Present study	Quintana Roo (Mexico)	Atlantic Ocean	Nesting ground	98.99%	4273	A2.1	A2.1.15*	II
24	*C. caretta*	AT05		Present study	Quintana Roo (Mexico)	Atlantic Ocean	Nesting ground	98.97%	2129	A10.1	A10.1.1*	II
25	*C. caretta*	PC150420		Present study	(Mexico)	Pacific Ocean	At sea	98.99%	419	P2.1	P2.1.1*	IA
26	*C. caretta*	PC130429		Present study	(Mexico)	Pacific Ocean	At sea	98.97%	320	P2.1	P2.1.2*	IA
27	*C. caretta*	-	NC016923.1	(Drosopoulou et al. 2012)	Zakinthos (Greece)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.12	II
28	*C. caretta*	SWFSC-69599	JX454983.1	(Duchene et al. 2012)	Florida	Atlantic Ocean	At sea	-	-	A5.1	A5.1.1	II
29	*C. caretta*	SWFSC-46603	JX454988.1	(Duchene et al. 2012)	Perù	Pacific Ocean	At sea	-	-	P1.1	P1.1.1	IA
30	*C. caretta*	-	MF554690.1	(Hernández-Fernández and Delgado Cano	Colombia	Atlantic Ocean	At sea	-	-	A2.1	A2.1.13	II

				2018)								
31	*C. caretta*	-	MF579504.1	(Hernández-Fernández and Delgado Cano 2018)	Colombia	Atlantic Ocean	At sea	-	-	A17.1	A17.1.1	IB
32	*C. caretta*	-	MF579505.1	(Hernández-Fernández and Delgado Cano 2018)	Colombia	Atlantic Ocean	At sea	-	-	A1.4	A1.4.1	IB
33	*C. caretta*	ADR01	OR166582	(Tolve et al. 2024)	Veneto (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
34	*C. caretta*	ADR02	OR166583	(Tolve et al. 2024)	Veneto (Italy)	Mediterranean Sea	At sea	-	-	A2.9	A2.9.1	II
35	*C. caretta*	ADR03	OR166584	(Tolve et al. 2024)	Veneto (Italy)	Mediterranean Sea	At sea	-	-	A10.4	A10.4.1	II
36	*C. caretta*	ADR04	OR166585	(Tolve et al. 2024)	Emilia-Romagna (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
37	*C. caretta*	TUS01	OR166586	(Tolve et al. 2024)	Tuscany (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
38	*C. caretta*	TUS02	OR166588	(Tolve et al. 2024)	Tuscany (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
39	*C. caretta*	TUS03	OR166590	(Tolve et al. 2024)	Tuscany (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.3	II
40	*C. caretta*	TUS04	OR166592	(Tolve et al. 2024)	Tuscany (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.9	II
41	*C. caretta*	LAT01	OR166595	(Tolve et al. 2024)	Latium (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II

42	*C. caretta*	LAT02	OR166596	(Tolve et al. 2024)	Latium (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
43	*C. caretta*	LAT03	OR166597	(Tolve et al. 2024)	Latium (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
44	*C. caretta*	LAT04	OR166598	(Tolve et al. 2024)	Latium (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.6	II
45	*C. caretta*	CAL1112	OR166599	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
46	*C. caretta*	CAL1124	OR166600	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
47	*C. caretta*	CAL1137	OR166601	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
48	*C. caretta*	CAL1201	OR166602	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
49	*C. caretta*	CAL1234	OR166603	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
50	*C. caretta*	CAL1303	OR166604	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.3	II
51	*C. caretta*	CAL1315	OR166605	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
52	*C. caretta*	CAL1319	OR166606	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
53	*C. caretta*	CAL1320	OR166607	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
54	*C. caretta*	CAL1321	OR166608	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
55	*C. caretta*	CAL1323	OR166609	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
56	*C. caretta*	CAL1325	OR166611	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
57	*C. caretta*	CAL1328	OR166612	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
58	*C. caretta*	CAL1335	OR166613	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II

59	*C. caretta*	CAL1346	OR166614	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
60	*C. caretta*	CAL1362	OR166615	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.1	II
61	*C. caretta*	CAL1370	OR166616	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.7	II
62	*C. caretta*	CAL1384	OR166617	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
63	*C. caretta*	CAL1385	OR166618	(Tolve et al. 2024)	Calabria (Italy)	Mediterranean Sea	Nesting ground	-	-	A20.1	A20.1.1	II
64	*C. caretta*	LIN01	OR166619	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.9	A2.9.2	II
65	*C. caretta*	LIN02	OR166620	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	Nesting ground	-	-	A2.1	A2.1.8	II
66	*C. caretta*	SIC01	OR166621	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
67	*C. caretta*	SIC02	OR166622	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A43.1	A43.1.1	II
68	*C. caretta*	SIC03	OR166623	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
69	*C. caretta*	SIC04	OR166647	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.2	II
70	*C. caretta*	SIC05	OR166624	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.8	II
71	*C. caretta*	SIC06	OR166625	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
72	*C. caretta*	SIC07	OR166626	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.8	II
73	*C. caretta*	SIC08	OR166627	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
74	*C. caretta*	SIC09	OR166628	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
75	*C. caretta*	SIC10	OR166629	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.9	II

76	*C. caretta*	SIC11	OR166630	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.3	II
77	*C. caretta*	SIC12	OR166631	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
78	*C. caretta*	SIC13	OR166632	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
79	*C. caretta*	SIC14	OR166633	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.5	II
80	*C. caretta*	SIC15	OR166634	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.10	II
81	*C. caretta*	SIC16	OR166635	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
82	*C. caretta*	SIC17	OR166636	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
83	*C. caretta*	SIC18	OR166637	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
84	*C. caretta*	SIC19	OR166638	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
85	*C. caretta*	SIC20	OR166639	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.11	II
86	*C. caretta*	SIC21	OR166640	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.4	II
87	*C. caretta*	SIC22	OR166641	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.8	II
88	*C. caretta*	SIC23	OR166642	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
89	*C. caretta*	SIC24	OR166643	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
90	*C. caretta*	SIC25	OR166644	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A3.1	A3.1.1	II
91	*C. caretta*	SIC26	OR166645	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II
92	*C. caretta*	SIC27	OR166646	(Tolve et al. 2024)	Sicily (Italy)	Mediterranean Sea	At sea	-	-	A2.1	A2.1.1	II

93	*L.olivacea*	CGG-01	NC_028634.1	(Duchene et al. 2012)	Ostional (Costa Rica)	Atlantic Ocean	-	-	-	-	-	-
94	*L.olivacea*	SWFSC-55352	JX454979.1	(Duchene et al. 2012)	-	Pacific Ocean	At sea	-	-	-	-	-
95	*L.olivacea*	SWFSC-78920	JX454987.1	(Duchene et al. 2012)	Caribbean	Atlantic Ocean	At sea	-	-	-	-	-
96	*L. kempii*	SWFSC-68090	JX454981.1	(Duchene et al. 2012)	Texas (USA)	Atlantic Ocean	Nesting ground	-	-	-	-	-
97	*L. kempii*	SWFSC-68091	JX454982.1	(Duchene et al. 2012)	Texas (USA)	Atlantic Ocean	Nesting ground	-	-	-	-	-
98	*L. kempii*	-	MN136053.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
99	*L. kempii*	-	MN136054.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
100	*L. kempii*	-	MN136055.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
101	*L. kempii*	-	MN136056.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
102	*L. kempii*	-	MN136057.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
103	*L. kempii*	-	MN136059.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
104	*L. kempii*	-	MN136060.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
105	*L. kempii*	-	MN136061.1	(Frandsen et al. 2020)	Texas (USA)	Atlantic Ocean	-	-	-	-	-	-
106	*E. imbricata*	-	DQ533485.1	Unpublished	-	-	-	-	-	-	-	-
107	*E. imbricata*	SWFSC-5787	JX454970.1	(Duchene et al. 2012)	Hawaii (USA)	Pacific Ocean	-	-	-	-	-	-
108	*E. imbricata*	SWFSC-61392	JX454980.1	(Duchene et al. 2012)	Singapore	Indo Pacific Ocean	-	-	-	-	-	-
109	*E. imbricata*	SWFSC-72489	JX454986.1	(Duchene et al. 2012)	Tortugero (Costa Rica)	Atlantic Ocean	-	-	-	-	-	-

110	*E. imbricata*	-	KP221806.1	Unpublished	-	-	-	-	-	-	-	-
111	*E. imbricata*	-	MF571906.1	Unpublished	-	-	-	-	-	-	-	-
112	*C. mydas*	-	NC_000886.1	(Kumazawa and Nishida 1999)	-	-	-	-	-	-	-	-
113	*C. mydas*	SWFSC-8855	JX454971.1	(Duchene et al. 2012)	Hawaii (USA)	Pacific Ocean	-	-	-	-	-	-
114	*C. mydas*	SWFSC-9277	JX454972.1	(Duchene et al. 2012)	Karpaz (Cyprus)	Mediterranean Sea	-	-	-	-	-	-
115	*C. mydas*	SWFSC-13768	JX454974.1	(Duchene et al. 2012)	Revillagigedo (Mexico)	Pacific Ocean	-	-	-	-	-	-
116	*C. mydas*	SWFSC-71270	JX454976.1	(Duchene et al. 2012)	Yap (Micronesia)	Indo Pacific Ocean	-	-	-	-	-	-
117	*C. mydas*	CGG-01	JX454990.1	(Duchene et al. 2012)	Tortugero (Costa Rica)	Atlantic Ocean	-	-	-	-	-	-
118	*N. depressus*	SWFSC-21684	NC_018550.1	(Duchene et al. 2012)	Australia	Pacific Ocean	-	-	-	-	-	-
120	*D. coriacea*	-	CM019965.1	Unpublished	-	-	-	-	-	-	-	-
121	*D. coriacea*	SWFSC-5718	JX454969.1	(Duchene et al. 2012)	Guerrero (Mexico)	Pacific Ocean	Nesting ground	-	-	-	-	-
122	*D. coriacea*	-	JX454973.1	(Duchene et al. 2012)	-	-	-	-	-	-	-	-
123	*D. coriacea*	-	JX454989.1	(Duchene et al. 2012)	-	-	-	-	-	-	-	-
124	*D. coriacea*	-	MF460363.1	Unpublished	-	-	-	-	-	-	-	-

Table S2 Laboratory protocols that have been used to generate the mitogenomes for the present study.

Samples used	DNA extraction	Library preparation	Sequencing center	Sequencing platform	Target output per sample (Gb)
CAL 1511-1, CAL 1502-1, CAL 20105-1, CAL 2070-1, CAL 1942-2, CAL 1947-1, CAL 1653-1, RG Sic 6, SR Sic 7	phenol:chloroform:isoamyl alcohol procedure (Green and Sambrook 2012)	PCR-free Tagmentation protocol (Illumina)	Department of Biology, University of Florence	Novaseq 6000	2.13
AT05, AT18, LI02, LI14, ZAK51, ZAK56, PC150420, PC130429	Puregen Kit (Qiagen)	?	Novogene	Novaseq 6000	30
BEL03, BEL17	Puregen Kit (Qiagen)	?	CNAG	Novaseq 6000	30
BOA02, BOA03, BOA18, BOA38, BOA48	Puregen Kit (Qiagen)	?	BGI	Novaseq6000	30

Table S3 Information and diversity of the mitochondrial genes used. Sequence length in base pairs (bp). Number of haplotypes *k*, number of polymorphic sites *S*, haplotype (h) and nucleotide diversity (π) with their Standard Error (SE), and dN/dS ratios for coding genes. For ND6 dN/dS was not computable due to dS=0 (*).

mtDNA gene	Sequence length (bp)	*k*	*S*	*h*±SE	π±SE	dN/dS
12S	970	5	7	0.652±0.291	0.002±0.0001	-
16S	1612	5	21	0.609±0.037	0.005±0.0003	-
ATP6	682	4	11	0.644±0.036	0.006±0.0004	0.1336
ATP8	165	3	2	0.620±0.039	0.006±0.0003	0
COXI	1548	6	20	0.676±0.034	0.004±0.0002	0.0173
COXII	687	4	10	0.605±0.039	0.005±0.0004	0
COXIII	784	5	10	0.652±0.034	0.005±0.0003	0
CYTB	1144	5	19	0.652±0.042	0.006±0.0004	0.1525
ND1	967	5	20	0.652±0.034	0.007±0.0004	0.1913
ND2	1042	6	20	0.648±0.038	0.007±0.0004	0.1463
ND3	349	5	16	0.644±0.041	0.013±0.0016	0.4237
ND4L	297	4	6	0.684±0.033	0.007±0.0004	0
ND4	1381	5	16	0.652±0.034	0.004±0.0002	0.0228
ND5	1797	6	28	0.759±0.024	0.006±0.0002	0.1820
ND6	525	5	8	0.628±0.039	0.004±0.0003	*
tRNAs	1553	7	34	0.696±0.033	0.003±0.0003	-
Control region	919	13	51	0.850±0.02	0.021±0.0006	-

Table S4 Molecular evolutionary model, estimated for each gene and mitogenomic partition (considering all coding genes and all rRNA genes separately) used for phylogenetic analysis. The topology of the tree for each gene/partition, reported in newick format, was obtained by Maximum Likelihood with IqTree. Statistic support is reported as SH-aLRT (%) /ultrafast bootstrap (%). tRNAs were not used in phylogenetic analysis because of insufficient information content. The support of node I is based on using *Lepidochelys* as outgroup. Asterisks indicate paraphilia of haplogroup IA; double asterisks indicate paraphilia of haplogroup II.

Gene/partition	Molecular evolutionary model	Topology	Statistic support node I	Statistic support node II
ATP6	TIM+F+G4	(IB, (IA, II));	100/100	55.7/53
ATP8**	TPM3+F	(IB, (IA, II));	100/100	82/68
COXI	TIM2+F	(IB, (IA, II));	100/100	90.7/92
COXII	TN+F+G4	(IB, (IA, II));	100/100	62.9/55
COXIII	TIM2+F+G4	(IB, (IA, II));	100/100	0/35
CYTB	TIM3+F+G4	(IB, (IA, II));	100/100	78.5/76
ND1	TVM+F+I	(IA, (IB, II));	100/100	77.9 /61
ND2	HKY+F+I	(II, (IA, IB));	100/100	39.8/60
ND3	TIM3+F+I	(IB, (IA, II));	100/100	61.6/45
ND4*	TIM+F+G4	(IA, (IB, II));	79/3	90/73
ND4L	TN+F	(II, (IA, IB));	100/100	3.2/46
ND5	TIM3+F+G4	(IA, (IB, II));	100/100	57.2/59
ND6	HKY+F+I	(IA, (IB, II));	100/100	60.3/69
Coding genes partition	GTR+F+G4	(IB, (IA, II));	100/100	83.7/79
12S rRNA	TN+F	(IB, (IA, II));	100/100	89.4/84
16S rRNA	GTR+F	(IB, (IA, II));	100/100	92.3/86
rRNAs partition	TIM+F	(IB, (IA, II));	100/100	96.9/98
Control region*	TN+F+G4	(IA, (IB, II));	100/100	92/90

Table S5 Codeml results of branch method tests, on dataset with only *C. caretta* sequences. Null model M0 estimates one value of ω across all branches and sites. Alternative model MBr1 estimates one value of ω on the background branches and one value of ω on the foreground branch (here, haplogroup IB). lnL: logarithm of Likelihood, np: number of parameters, LRT: Likelihood Ratio Test, LRT = 2(lnLMBr1 – lnLM0), d.f.: degrees of freedom. The critical value for LRT statistic, for 1 degree of freedom and p = 0.005, is 3.841 and significant results are indicated by *. Values of ω = 999 means 'infinity' and may derive by a lack of synonymous substitutions.

C. caretta	**Null model M0**			**Alternative model MBr1**				**test**	
gene	**lnL**	**np**	**ω**	**lnL**	**np**	**ωb**	**ωf**	**LRT**	**d.f.**
ATP6	-868.870	54	0.1441	-868.857	55	0.1266	999	0.0248	1
ATP8	-205.229	54	0.0001	-205.229	55	0.0001	0.0001	0.0000	1
COXI	-2058.11	54	0.0148	-2057.00	55	0.0284	0.0001	1.2476	1
COXII	-912.868	54	0.0001	-912.868	55	0.0001	0.0001	0.0000	1
COXIII	-1043.37	54	0.0001	-1043.37	55	0.0001	0.0001	0.0000	1
CYTB	-1509.18	54	0.1137	-1507.46	55	0.0486	999	3.4438	1
ND1	-1299.53	54	0.1632	-1298.91	55	0.1017	0.3016	1.2548	1
ND2	-1380.86	54	0.1183	-1380.62	55	0.1050	0.3076	0.4915	1
ND3	-524.612	54	0.3891	-519.425	55	0.0972	999	10.3742*	1
ND4	-1776.37	54	0.0192	-1775.51	55	0.0001	999	1.7227	1
ND4L	-392.92	54	0.0001	-392.92	55	0.0001	999	0.0014	1
ND5	-2361.22	54	0.1552	-2360.52	55	0.1066	999	1.3977	1
ND6	-660.335	54	0.0001	-660.335	55	0.0001	0.0001	0.0000	1

Table S6 Codeml results of branch method tests, on dataset with multiple species (see text for accession numbers of sequences). Null model M0 estimates one value of ω across all branches and sites. Alternative model MBr1 estimates one value of ω on the background branches and one value of ω on the foreground branch (here, haplogroup IB). lnL: logarithm of Likelihood, np: number of parameters, LRT: Likelihood Ratio Test, LRT = 2(lnLMBr1 – lnLM0), d.f.: degrees of freedom. The critical value for LRT statistic, for 1 degree of freedom and p = 0.005, is 3.841; significant results are indicated by *. Values of ω = 999 means 'infinity' and may derive by a lack of synonymous substitutions.

Multiple sp	**Null model M0**			**Alternative model MBr1**				**test**	
gene	**lnL**	**np**	**ω**	**lnL**	**np**	**ωb**	**ωf**	**LRT**	**d.f.**
ATP6	-1846.63	18	0.0632	-1845.12	19	0.0613	999	3.03	1
ATP8	-507.77	18	0.1323	-507.77	19	0.1323	0.0001	0	1
COXI	-3702.01	18	0.0105	-3701.89	19	0.0106	0.0001	0.2533	1
COXII	-1642.82	18	0.0129	-1642.76	19	0.013	0.0001	0.1164	1
COXIII	-1873.82	18	0.0169	-1873.5	19	0.0172	0.0001	0.6444	1
CYTB	-2984.38	18	0.0258	-2981.34	19	0.0241	0.24	6.0652*	1
ND1	-2580.52	18	0.0269	-2578.08	19	0.0252	0.2281	4.8781*	1
ND2	-2804.47	18	0.0657	-2804.36	19	0.0663	0.4059	0.2171	1
ND3	-1030.78	18	0.0515	-1030.78	19	0.0515	0.0001	0.0001	1
ND4	-3762.65	18	0.0441	-3762.63	19	0.044	0.0539	0.0329	1
ND4L	-762.44	18	0.0333	-762.36	19	0.0336	0.0001	0.1659	1
ND5	-4945.65	18	0.0539	-4942.53	19	0.0521	0.2854	6.2362*	1
ND6	-1116	18	2.5217	-1114.9	19	2.4342	999	1.5613	1

Table S7 Codeml results of branch-site method tests, on dataset with only *C. caretta* sequences. Null model M2a fixes the value of ω = 1 on the foreground branch (here haplogroup IB). Alternative model MBr1_2S models 4 different site classes: in site class 0 all sites on all branches are under purifying selection ($0 < \omega_0 < 1$); in site class 1 all sites are under neutral evolution ($\omega_1 = 1$); in site class 2a, sites on foreground branch are under positive selection ($\omega_2 > 1$) and sites on background branches are under purifying selection ($0 < \omega_0 < 1$); in site class 2b, sites on foreground branch are under positive selection ($\omega_2 > 1$) and sites on background branches are under neutral evolution ($\omega_1 = 1$). Proportions of sites in classes are, respectively: p_0, proportion of sites in class 0; p_1, proportion of sites in class 1; p_{2a}, proportion of sites in class 2a; p_{2b}, proportion of sites in class 2b. lnL: logarithm of Likelihood, np: number of parameters, LRT: Likelihood Ratio Test, LRT = 2(lnLMBr1 – lnLM0), d.f.: degrees of freedom. The critical value for LRT statistic, for 1 degree of freedom and p = 0.005, is 3.841 and significant results are indicated by *. Values of ω = 999 means 'infinity' and may derive by a lack of synonymous substitutions.

C. caretta	**Null model M2a**			**Alternative model MBr1_2S**									**test**	
gene	**lnL**	**np**	**ω_0**	**lnL**	**np**	**ω_0**	**ω_1**	**ω_2**	**p_0**	**p_1**	**p_{2a}**	**p_{2b}**	**LRT**	**d.f.**
ATP6	-868.86	56	0,1269	-868.85	57	0.1245	1	110.90	0.9886	0	0.0114	0	0.0070	1
ATP8	-205.23	56	0	-205.23	57	0	1	2.7852	0.9999	0	0.0001	0	0	1
COXI	-2058.11	56	0,0148	-2058.11	57	0.0148	1	1	1	0	0	0	0	1
COXII	-912.86	56	0	-912.86	57	0	1	2.1701	1	0	0	0	0	1
COXIII	-1043.37	56	0	-1043.37	57	0	1	1.6361	1	0	0	0	0	1
CYTB	-1507.68	56	0,0425	-1507.46	57	0.0426	1	256.55	0	0	1	0	0.4462	1
ND1	-1298.91	56	0,1017	-1298.91	57	0.1017	1	1.5524	0.8554	0	0.1446	0	0.0004	1
ND2	-1380.46	56	0,6864	-1380.26	57	0	1	242.63	0.8847	0.1121	0.0028	0.0003	0.3936	1
ND3	-521.05	56	0,0961	-519.43	57	0.0972	1	999	0	0	1	0	3.2431	1
ND4	-1775.49	56	0	-1775.09	57	0	1	429.96	0.9972	0	0.0027	0	0.8032	1
ND4L	-392.92	56	0	-392.92	57	0	1	2.6909	1	0	0	0	0	1
ND5	-2360.52	56	0,1067	-2360.52	57	0.1066	1	38.410	0	0	1	0	0.0061	1
ND6	-660.332	56	0	-660.332	57	0	1	1	1	0	0	0	0	1

Table S8 Codeml results of branch-site method tests, on dataset with multiple species. Null model M2a fixes the value of ω = 1 on the foreground branch (here haplogroup IB). Alternative model MBr1_2S models 4 different site classes: in site class 0 all sites on all branches are under purifying selection ($0 < \omega_0 < 1$); in site class 1 all sites are under neutral evolution ($\omega_1 = 1$); in site class 2a, sites on foreground branch are under positive selection ($\omega_2 > 1$) and sites on background branches are under purifying selection ($0 < \omega_0 < 1$); in site class 2b, sites on foreground branch are under positive selection ($\omega_2 > 1$) and sites on background branches are under neutral evolution ($\omega_1 = 1$). Proportions of sites in classes are, respectively: p_0, proportion of sites in class 0; p_1, proportion of sites in class 1; p_{2a}, proportion of sites in class 2a; p_{2b}, proportion of sites in class 2b. lnL: logarithm of Likelihood, np: number of parameters, LRT: Likelihood Ratio Test, LRT = 2(lnLMBr1 – lnLM0), d.f.: degrees of freedom. The critical value for LRT statistic, for 1 degree of freedom and p = 0.005, is 3.841 and significant results are indicated by *. Values of ω = 999 means "infinity" and may derive by a lack of synonymous substitutions.

Multiple sp	**Null model M2a**			**Alternative model MBr1_2S**									**test**	
gene	**lnL**	**np**	**ω_0**	**lnL**	**np**	**ω_0**	**ω_1**	**ω_2**	**p_0**	**p_1**	**p_{2a}**	**p_{2b}**	**LRT**	**d.f.**
ATP6	-1834.3	20	0.0389	-1834.2	21	0.0390	1	261	0	0	0.95	0.048	0.2565	1
ATP8	-507	20	0.0932	-507	21	0.0932	1	1	0.9032	0.0968	0	0	0	1
COXI	-3693.1	20	0.0071	-3693.1	21	0.0071	1	1	0.993	0.007	0	0	0	1
COXII	-1632.4	20	0.0081	-1632.4	21	0.0081	1	1	0.9898	0.0102	0	0	0	1
COXIII	-1864.5	20	0.0054	-1864.5	21	0.0054	1	1	0.9659	0.0341	0	0	0	1
CYTB	-2961.7	20	0.0132	-2960.9	21	0.0132	1	75.256	0.9565	0.0321	0.0110	0.0004	1.6056	1
ND1	-2566.5	20	0.0141	-2566.3	21	0.0140	1	15.327	0.9543	0.0364	0.0089	0.0003	0.2563	1
ND2	-2781.4	20	0.0340	-2781.4	21	0.0340	1	1	0.9108	0.0742	0.0138	0.0011	0	1
ND3	-1020.2	20	0.0281	-1020.2	21	0.0281	1	1	0.9305	0.0695	0	0	0	1
ND4	-3738.0	20	0.0296	-3738.0	21	0.0296	1	1	0.9311	0.0351	0.0325	0.0012	0	1
ND4L	-762	20	0.0241	-762	21	0.0241	1	1	0.9697	0.0302	0	0	0	1
ND5	-4907	20	0.0330	-4907	21	0.0330	1	1.4140	0.8554	0.0417	0.0981	0.0048	0.0198	1
ND6	-1124	20	0	-1123	21	0	1	999	0	0	0.0745	0.9251	3.126	1

Table S9 (PROVEAN) Protein Variation Effect Analyzer scores for nonsynonymous substitutions that are fixed between haplogroup II and haplogroup IB. Shown with * significant values < - 2.5 indicating likelihood of a substitution to affect the protein function.

Gene	Position	Ancestral state	Derived state	PROVEAN score
ATP6	78	Ala	Thr	-0.382
	151	Phe	Leu	-3.999 *
CYTB	47	Ile	Thr	-2.088
	182	Phe	Leu	-2.838 *
	190	Ile	Val	-0.581
	233	Leu	Phe	2.136
ND1	16	Ile	Val	-0.292
	73	Ile	Thr	1.689
	314	Thr	Ala	0.266
	317	Gly	Ser	-3.683 *
ND2	4	His	Tyr	-0.038
	31	Thr	Ile	-1.678
	85	Asn	Asp	3.039
ND3	5	Met	Thr	-1.498
	108	Val	Ile	0.362
ND4	35	Ala	Ser	-1.030
ND5	18	Leu	Phe	-0.232
	24	Tyr	Cys	-1.206
	31	Met	Leu	-0.284
	50	Ala	Thr	0.346
	87	Val	Ala	-1.932
	177	Ile	Val	-0.896
	466	Thr	Met	-0.054
	573	Ser	Leu	-3.181 *

References

Drosopoulou E, Tsiamis G, Mavropoulou M, Vittas S, Katselidis KA, Schofield G, Palaiologou D, Sartsidis T, Bourtzis K, Pantis J (2012) The complete mitochondrial genome of the loggerhead turtle *Caretta caretta* (Testudines: Cheloniidae): Genome description and phylogenetic considerations. Mitochondrial DNA 23: 1-12.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

Frandsen HR, Figueroa DF, George JA (2020) Mitochondrial genomes and genetic structure of the Kemp's ridley sea turtle (*Lepidochelys kempii*). Ecol Evol 10: 249-262.

Green M, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY

Hernández-Fernández J, Delgado Cano DA (2018) Complete mitochondrial genome of the three Colombian Caribbean loggerhead sea turtles, using next-generation sequencing. Mitochondrial DNA B 3: 626-627.

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. Molecular Biology and Evolution 16: 784-792.

Tolve L, Iannucci A, Garofalo L, Ninni A, Capobianco Dondona A, Ceciarini I, Cocumelli C, De Lucia A, Falconi M, Formia A, Iacovelli F, Mancusi C, Marchiori E, Marsili L, Mingozzi T, Nannarelli S, Natali C, Terracciano G, Zuffi MAL, Novelletto A, Ciofi C (2023) Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine Biology 171: 19.

CHAPTER V

Mitochondrial and mitogenomic analyses of *Chelonia mydas* in the Gulf of Guinea

Introduction

The protection of marine environments and marine life has gained increasing attention in public awareness and international policies, such that the 14th Sustainable Development Goal (SDG) of the UN 2030 Agenda is dedicated to *'conserve and sustainably use the oceans, seas and marine resources for sustainable development'* (United Nations 2015). The specific targets of the 14th SDG address crucial issues such as marine pollution, ocean acidification, coastal urbanization and erosion, fishery bycatch and overexploitation, aiming to reduce these threats to oceans health and human well-being.

Sea turtles are among the many affected species, and despite a recent overall improvement in their risk status compared to previous assessments (Hamann et al. 2010; Wallace et al. 2025), fisheries bycatch continues to heavily impact all seven sea turtle species worldwide. Bycatch reduction is therefore a priority in sea turtle conservation, and its effective implementation relies on close collaboration with local communities (e.g. Proyecto MARES Comunidad 2024), based on extensive data collection about fishing effort, gears description, and of course about species presence in the area (Wallace et al. 2010b). Studying turtles at sea, both in terms of abundance and stock composition, is therefore crucial to inform appropriate conservation measures (Hamann et al. 2010).

Western Africa and the Gulf of Guinea are important regions for five sea turtle species (Wallace et al. 2010a), including the endangered green turtle *Chelonia mydas*. Major Eastern Atlantic rookeries for this species have been described in Ascension Island by Formia et al. (2007) and in Poilao (Guinea Bissau) by Patricio et al. (2017). Smaller rookeries were also described in Sao Tomé, Principe and Bioko (Formia et al. 2006). On the other hand, relatively scarce information has been collected and discussed regarding the stocks of turtles at sea (e.g. Patrício et al. 2017; Hancock et al. 2019), foraging or migrating and facing heavy industrial fishery bycatch (Riskas and Tiwari 2013; Casale et al. 2017).

Characterizing the stocks of turtles at sea in terms of rookeries of origin is always of major importance to understand which populations most rely on specific foraging areas, and which are threatened there. Genetic analyses have been applied extensively for this purpose, mainly based on mitochondrial DNA control region fragments which are used first for rookeries' profiling and then for assigning turtles of unknown origin in the mixed stocks (Pella and Masuda 2001). For *C. mydas*, the vast majority of the available data have been produced using control region fragments of ~ 490 bp (Allard et al. 1994),

which is still the reference marker for wide population genetic studies. Moreover, mtDNA control region has been used to reconstruct the species phylogeography, and to make inferences about the colonization dynamics of the oceans (Jensen et al. 2019). According to those studies, the Gulf of Guinea, and in specific Sao Tomé and Principe, would have been crucial as refugia for *C. mydas* during Pleistocenic glaciations, preserving genetic diversity which is still higher today than in recently colonized regions.

Although traditional control region markers are still used in population genetics and phylogenetic analyses, they have shortcomings. First, few common haplotypes, retrieved in all rookeries, may hamper the description of population and the assignment of turtles in the mixed stocks (e.g. Tolve et al. 2018); more powerful markers are needed, which can detect more polymorphisms within the mtDNA and resolve shared lineages. Second, single mtDNA genes may not be reliable in reconstructing the evolutionary history of a species, since mutation rate and selective constrains vary across the molecule (see chapter 4 of this thesis). Scaling up to whole mitogenome sequencing might help overcome these limitations.

In the present study, we focus on *C. mydas* in the Gulf of Guinea, providing genetic information about the mixed stocks of turtles at sea in such an important and understudied region. We take advantage of a broad sampling and of traditional genetic markers which allow a comparison with published literature. Moreover, we characterized a subsample of complete mitogenomes, in order to test their potential in genetic diversity resolution and phylogeographic application.

Materials and Methods

Control region analyses

Between 1998 and 2015, 1034 tissue biopsies were collected from nesting, captured, bycaught and stranded green turtles in the Gulf of Guinea. DNA was extracted by a standard protocol with phenol chloroform and a ‘short’ ~ 490 bp fragment of the mtDNA control region was characterized in our laboratory (Formia 2002, unpublished; Formia et al. 2006), using primers LTCM1 and HDCM1 (Allard et al. 1994). Then, a ‘long’ control region fragment ~ 780 bp in length was characterized for this dataset, using primers LCM15382 and H950 (Abreu et al. 2006) and following the same laboratory protocol as Vilaça et al. (2022). Haplotypes were assigned based on the datasets available at the Archie Carr Center for Sea Turtles Research website

(https://accstr.ufl.edu/resources/mtdna-sequences/). These prior data was produced before the present PhD project.

Number of haplotypes, haplotype and nucleotide diversity were computed in Arlequin 3.5.2.2 (Excoffier and Lischer 2010) for each control region dataset – the one of ~ 490 bp sequences (1034 samples) and the one with ~ 780 bp sequences (465 samples). Two haplotype networks were also constructed, using statistical parsimony (Templeton et al. 1992) implemented in TCS (Clement et al. 2000), allowing an infinite number of changes to link all haplogroups. The networks were visualized in tcsBU (Múrias dos Santos et al. 2015), exported as svg files and graphically edited in Inkscape (https://inkscape.org/).

To characterize the composition of the stocks at sea based on samples collected in the Gulf of Guinea in terms of contributing rookeries, we performed foraging ground – centric Mixed Stock Analyses (MSA) (Pella and Masuda 2001). We used the ~ 490 bp dataset, because it was the most numerous dataset that we had, and because the most comprehensive published baseline for green turtles has been constructed with the same marker. For the baseline we used data from Atlantic and Indo Pacific rookeries, with respective rookery size (number of nesting females per year) as prior (Supplementary Table S1). We ran separate analyses for the following mixed stocks of turtles sampled at sea (i.e. 'not nesting' turtles): i) Western Africa (samples from Liberia, Ivory Coast, Ghana, Togo, Benin: N = 64); ii) Continental Equatorial Guinea (N = 80); iii) Corisco Bay (N = 640); iv) Sao Tomé (N = 76), see Table 1. We used R mixstock package by Bolker (2007), setting 400,000 iterations and burn-in at 50% and checking chains convergence with Gelman and Rubin diagnostics (shrink factor < 1.2).

Mitogenomic sequencing and assembly

From the original dataset in Table 1, 13 individuals were subsampled for whole mitogenome sequencing; the choice of the samples was based on good DNA quality and quantity, assessed by 1% gel electrophoresis and Qubit 4 fluorometer Broad Range Assay (Invitrogen). Dual-indexed (i5 and i7) genomic libraries were built using a PCR-free Tagmentation Kit Illumina according to the manufacturer protocol. We envisaged a 1.8 X target nuclear genome coverage for 11 of these samples and 18X for two of these samples that were sequenced separately. Consequently, the target output was ~ 3.96 Gb and ~ 39.6 Gb per sample, respectively. We sequenced the libraries paired-end on an Illumina NovaSeq 6000 System using a 300-cycle NovaSeq 6000 SP Reagent Kit v 1.5.

Sequencing data were demultiplexed and converted to fastq format using bcl2fastq version 2.20 (Illumina). Reads quality and sequencing statistics were checked using FastQC 11.9 (Andrews 2010) and Trimmomatic 0.39 (Bolger et al. 2014) was used to remove reads of insufficient length and quality and to create paired reads to be used in following analyses. We applied two different pipelines in parallel to assemble complete mitogenomes, comparing their output. One pipeline consisted of reads mapping on a reference mitogenome, while the other pipeline assembled mitogenomes *de novo*, using a seed-and-extend algorithm.

For the mapping pipeline we chose the annotated sequence *C. mydas* NC_000886.1 as reference. Bowtie2-2.5.0 (Langmead and Salzberg 2012) was used for indexing and generating the SAM files, then converted in the human readable BAM files. Samtools 1.16.1 (Danecek et al. 2021) was used to sort and index the BAM files and to generate sorted.bam files thar can be visualized in IGV (Integrate Genomic Viewer) (Robinson et al. 2011). Variant calling and consensus sequence extrapolation were accomplished by BCFtools 1.16 (Danecek et al. 2021), filtering out variants with a quality score < 30.

For the *de novo* assembly pipeline, we used NOVOPlasty software (Dierckxsens et al. 2016). A COI sequence from *C. caretta* was used as a seed to start the assembly; k-mers option was set at 50 and eventually changed to higher values in case of failure of the assembly; genome range was set as 15000 – 18000 range. Successful assembly was achieved in case of circularized molecules or in case of overlapping contigs covering the entire length of the mitogenome.

For each sample, we compared the mitogenome sequence obtained by the mapping pipeline with the mitogenome sequence obtained by NOVOPlasty. In case of discrepancies, the visual inspection of vcf statistics and of sorted.bam files in IGV allowed us to check the reads coverage for each polymorphic variant, clarifying the correct reconstruction. The consensus sequences were aligned and trimmed at position 16443, removing the microsatellites from the following analyses. Mitogenomic haplotype names were assigned following the criterion of Shamblin et al. (2012a); the first two serial numbers refer to the respective ~ 780 bp control region haplotype and the third one refers to the variants outside the control region.

Mitogenomic diversity and phylogenetic analyses

To have a comprehensive picture of the information derived by using complete mitogenomes, we searched GenBank database to find all mitogenomic sequences

available for *Chelonia mydas* (Table 3). Sequences were downloaded, aligned with ours and partitioned in single genes alignments using Geneious Prime 2022.2.2; then, genetic diversity indices were computed for each gene with DnaSP (Librado and Rozas 2009). The whole mitogenomes alignment was used to build a mitogenomic TCS network.

To perform phylogenetic analyses on the mitogenomic dataset, *Natator depressus* mitogenomic sequence (GenBank Accession Number NC_018550.1) was selected as outgroup. The best molecular evolution model was assessed and then a Maximum Likelihood (ML) phylogeny was constructed using IqTree (Trifinopoulos et al. 2016). Statistical support for the nodes was assessed by ultrafast bootstrap (values > 95%). FigTree v1.4.4 (http:// tree. bio. ed. ac. uk/ software/ figtr ee/) was used to visualize the tree.

A Bayesian phylogenetic analysis using BEAST suite v.1.10.4 (Suchard et al. 2018b) was also performed to put dates on the nodes of the topology. The mitogenomic alignment was first partitioned in four: i) coding genes partition, ii) tRNAs partition, iii) rRNas partition; iv) control region partition. We used the four partitions with the 'unlink' option for substitution model and for molecular clock to estimate the parameters independently, the 'link tree' option, the 'random starting tree' option and 'coalescent constant size' option to generate the tree. As priors, we used the time of divergence between *N. depressus* and *C. mydas* suggested by Duchene et al. (2012), 36.43 MYBP (95% HPD 21.92 – 52.51) and we forced monophyly of *C. mydas*. We performed 5 independent runs, 10^8 steps each with a 10% burn-in. We used Tracer v.1.7.1 (Rambaut et al. 2018) to monitor ESS for each output parameter, Log Combiner to merge the runs, Tree Annotator (Drummond and Rambaut 2007) to summarise the overall output, and FigTree v1.4.4 (http:// tree. bio. ed. ac. uk/ software/ figtr ee/) to visualize the tree.

Selective pressure analyses

To test deviation from neutrality in *C. mydas* mitogenomic coding region, we performed a McDonald – Kreitman test on each gene, using DnaSP (Librado and Rozas 2009) and *N. depressus* NC_018550.1 as outgroup. The test computes the number of non-synonymous and synonymous polymorphic changes within a species (Pn and Ps, respectively) and the number of fixed non-synonymous and synonymous changes between species (Dn and Ds, respectively); the output is a Neutrality Index (NI), defined as Pn/Ps divided by Dn/Ds, and it should be interpreted as neutrality if = 1, purifying selection if > 1, positive selection if < 1.

To test more specifically the presence of different selective pressure within *C. mydas* mitogenomic coding partition, we used codeml package implemented in PAML (Yang 2007), using the ML phylogeny that we had previously generated and following the indications in Álvarez-Carretero et al. (2023). We first applied a Random Site Method to test the presence of sites under selection (values of ω) across the phylogeny. We used M0 as null model, assuming one value of ω across all sites and all branches. M0 was compared with M1a, a nearly-neutral model considering two site classes (one for $\omega = 1$ and one for $\omega < 1$); then, M1a was compared with M2a, which considers three site classes (one for $\omega = 1$, one for $\omega < 1$ and one for $\omega > 1$). Also, we applied the Random Site Method using M7 (null model) vs M8 (alternative model), which are more robust than M1a, M2a in case of weak positive selection (Álvarez-Carretero et al. 2023). This analysis indicates possible sites where selection might be acting without clustering on some lineages. Additionally, we tested specific hypotheses of selection in action on definite branches of the phylogeny, and we assessed i) whether selection on mtDNA affected the divergence of *C. mydas* from *N. depressus*, and ii) whether selection on mtDNA affected in specific the divergence of the Atlantic clade of *C. mydas* that we retrieved in our phylogeny (see Results). Therefore, a Branch Method was applied, to test the presence of a different value of ω on the foreground branch in comparison to the other branches (null model M0, alternative model MBr1); we first used the whole *C. mydas* clade, then the only Atlantic clade as 'foreground branch'. Finally, we used a Branch Site Method to identify specific loci under selection in the previously specified foreground branches. In this method the null model M0 is compared with an alternative model assuming four site classes; one class for sites under purifying selection on both branches, one class for sites under neutral selection on both branches, one class for sites under positive selection on the foreground branch and under purifying selection on the background branch, and one class for sites under positive selection on the foreground branch and under neutral selection on the background branch. For each test, the significance was assessed by a Likelihood Ratio Test (LRT).

We also used TreeSAAP software (Woolley et al. 2003) to test the presence of selective pressure with a complementary approach, not based on modelling of ω but on the physicochemical properties of amino acids corresponding to non-synonymous substitutions.

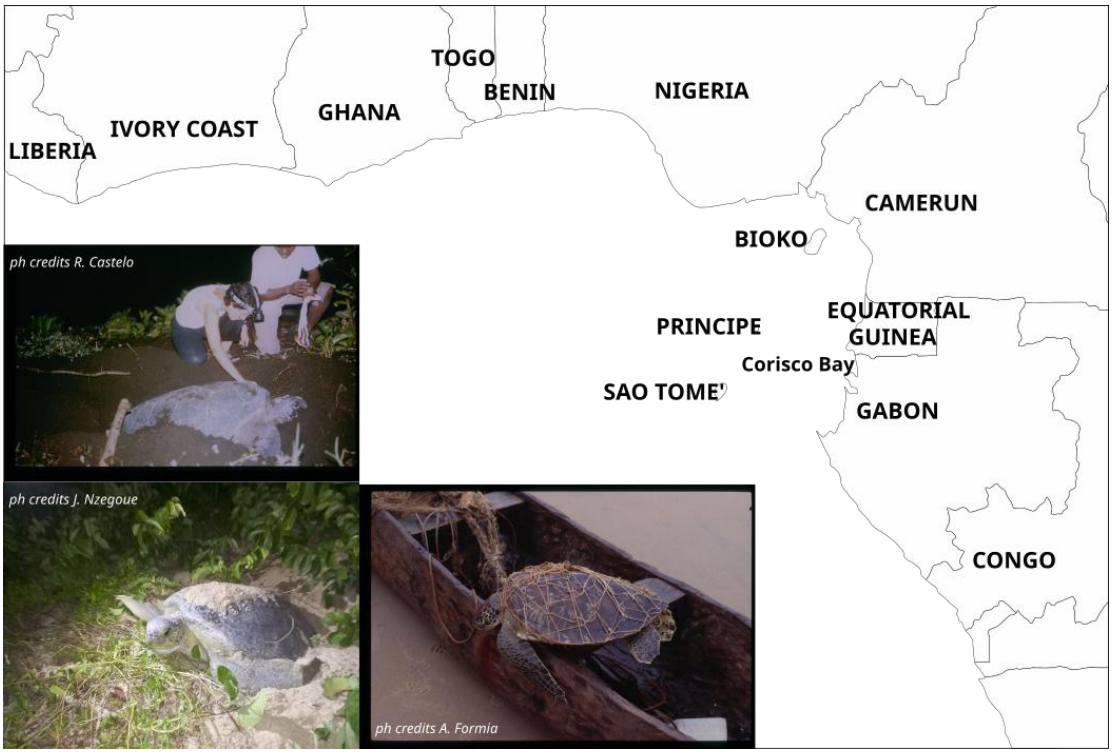


Fig. 1 Study area of the Gulf of Guinea and images from the field; on the left, green turtles nesting and researchers collecting data; on the right, bycaught green turtle in a fishing net.

Table 1. Control region haplotypes (~ 490 bp fragment) for samples collected in the Gulf of Nigeria. In parentheses, number of samples from nests. Samples collected specifically in Corisco Bay are reported separately from samples collected in other scattered locations in Gabon and Equatorial Guinea, because it is considered as a self-standing mixed stock in MSA.

Haplotype	Liberia	Ivory Coast	Ghana	Togo	Benin	Continental Equatorial Guinea	Gabon	Corisco Bay	Principe	Sao Tome	Bioko	TOT
CmA05	0	0	3	0	1	2	0	32	0	5 (1)	0	**43 (1)**
CmA06	0	0	2	1	1	6	3 (1)	53	1	4 (1)	6 (5)	**77 (7)**
CmA08	6	22	13	8	2	61	19 (3)	467	41 (39)	69 (12)	53 (46)	**761 (100)**
CmA09	0	0	0	0	0	2	0	7	0	0	0	**9**
CmA10	0	0	0	0	0	1	1	8	0	0	0	**10**
CmA23	0	0	0	0	0	0	0	1	0	0	0	**1**
CmA24	0	0	0	0	0	0	0	2	0	1	0	**3**
CmA25	0	0	0	0	0	0	0	2	0	0	0	**2**
CmA32	0	0	0	0	0	0	0	4	0	0	0	**4**
CmA35	0	0	0	0	0	2	0	10	0	1	0	**13**
CmA36	0	1	0	1	0	4	1	25	26 (25)	10 (1)	2	**70 (26)**
CmA37	0	0	0	0	0	0	0	4	0	2 (1)	0	**6 (1)**
CmA38	0	0	0	0	0	0	0	0	0	2 (2)	0	**2 (2)**
CmA39	0	0	0	0	0	0	0	1	0	0	0	**1**
CmA40	0	1	0	0	0	0	0	5	0	0	0	**6**
CmA41	0	0	0	0	0	0	0	2	0	0	0	**2**
CmA42	0	0	0	0	0	1	0	1	0	0	0	**2**
CmA43	1	0	0	0	0	1	1 (1)	4	0	0	0	**7 (1)**
CmA44	0	0	0	0	1	0	0	2	0	0	0	**3**
CmA72	0	0	0	0	0	0	0	1	0	0	0	**1**
CmA75	0	0	0	0	0	0	0	1	2 (2)	0	0	**3 (2)**
CmP47	0	0	0	0	0	0	0	1	0	0	0	**1**
CmP49	0	0	0	0	0	0	0	7	0	0	0	**7**
TOT	**7**	**24**	**18**	**10**	**5**	**80**	**25 (5)**	**640**	**70 (66)**	**94 (18)**	**61 (51)**	**1034 (140)**

Table 2. Control region haplotypes (~ 780 bp fragment) for a subset of samples collected in the Gulf of Nigeria. In parentheses, number of samples from nests. Samples collected specifically in Corisco Bay are reported separately from samples collected in other scattered locations in Gabon and Equatorial Guinea because it is considered as a self-standing mixed stock in MSA.

.

Haplotype	Ghana	Continental Equatorial Guinea	Gabon	Corisco Bay	Principe	Sao Tome	TOT
CmA5.1	0	0	0	20	0	0	**20**
CmA6.1	0	2	2 (1)	32	0	1	**37 (1)**
CmA8.1	1	12	16 (3)	270	37 (35)	0	**336 (38)**
CmA8.3	0	0	0	3	0	0	**3**
CmA9.1	0	0	0	3	0	0	**3**
CmA10.1	0	0	1	2	0	0	**3**
CmA24.1	0	0	0	1	0	0	**1**
CmA25.1	0	0	0	1	0	0	**1**
CmA32.1	0	0	0	3	0	0	**3**
CmA35.1	0	0	0	3	0	0	**3**
CmA36.1	0	0	1	15	22	0	**38**
CmA37.1	0	0	0	3	0	0	**3**
CmA40.1	0	0	0	1	0	0	**1**
CmA43.1	0	0	1	2	0	0	**3**
CmA44.1	0	0	0	1	0	0	**1**
CmA72.1	0	0	0	1	0	0	**1**
CmA75.1	0	0	0	1	2	0	**3**
CmP47.1	0	0	0	1	0	0	**1**
CmP49.1	0	0	0	3	0	0	**3**
CmP49.2	0	0	0	1	0	0	**1**
TOT	**1**	**14**	**21 (4)**	**367**	**61 (35)**	**1**	**465 (39)**

Table 3. *C. mydas* mitogenomic sequences that we used in this study.

ID sample / Accession Number	Reference	Target nuclear coverage (X)	Output nuclear coverage (X)	Output mitogenomic coverage (X)	Sample provenance	Ocean	Nesting vs at sea	Control region haplotype	Mitogenomic haplotype
G14	This study	1.8	0.59	22	Corisco	Atlantic	at sea	CMA8.1	CMA8.1.1
GE144	This study	1.8	1.80	143	Corisco	Atlantic	at sea	CMA8.1	CMA8.1.1
GE146	This study	1.8	1.85	153	Corisco	Atlantic	at sea	CMA8.1	CMA8.1.2
M036	This study	1.8	2.72	108	Corisco	Atlantic	at sea	CMA8.1	CMA8.1.1
M038	This study	1.8	2.30	400	Corisco	Atlantic	at sea	CMA8.1	CMA8.1.3
M090	This study	1.8	2.77	217	Principe	Atlantic	nesting	CMA8.1	CMA8.1.1
ST16	This study	1.8	3.89	289	Sao Tome	Atlantic	at sea	CMA8.1	CMA8.1.1
ST17	This study	1.8	2.01	179	Sao Tome	Atlantic	at sea	CMA8.1	CMA8.1.4
ST26	This study	1.8	3.38	236	Sao Tome	Atlantic	at sea	CMA8.1	CMA8.1.1
ST24	This study	18	15.9	1040	Sao Tome	Atlantic	at sea	CMA5.1	CMA5.1.4
ST61	This study	18	17.6	1362	Sao Tome	Atlantic	at sea	CMA5.1	CMA5.1.4
JQ026233	(Shamblin et al. 2012a)	-	-	-	Costa Rica	Atlantic	nesting	CMA3.1	CMA3.1.1
JQ034420	(Shamblin et al. 2012a)	-	-	-	Buck Island	Atlantic	nesting	CMA5.1	CMA5.1.1
JX454971	(Duchene et al. 2012)	-	-	-	Hawaii	Pacific	**?**	CMP1.1	CMP1.1.1
JX454972	(Duchene et al. 2012)	-	-	-	Cyprus	Mediterranean	**?**	CMA13.1	CMA13.1.1
JX454974	(Duchene et al. 2012)	-	-	-	Mexico	Pacific	**?**	CMP4.1	CMP4.1.1
JX454976	(Duchene et al. 2012)	-	-	-	Malaysia	Pacific	**?**	CMP20.1	CMP20.1.1
JX454978	(Duchene et al. 2012)	-	-	-	Galapagos	Pacific	**?**	CMP4.1	CMP4.1.2
JX454985	(Duchene et al. 2012)	-	-	-	Micronesia	Pacific	**?**	CMP20.1	CMP20.1.1
JX454990	(Duchene et al. 2012)	-	-	-	Costa Rica	Atlantic	**?**	CMA3.1	CMA3.1.2
NC000886	(Kumazawa and Nishida 1999)	-	-	-	?	Atlantic	**?**	CMA3.1	CMA3.1.3

Results

Control region analyses

Out of 1034 biopsies that had been genetically typed using the 'short' (~ 490 bp) fragment of the control region, 465 were successfully amplified and sequenced also for the 'long' (~ 780 bp) fragment.

Details about sampling location, sample type and control region haplotypes are reported in Tables 1 and 2 for the 'short' and 'long' fragment datasets, respectively.

From the dataset of samples analysed using the ~ 490 bp control region fragment (N = 1034), 23 haplotypes were retrieved (Tables 1 and 4). Haplotype CmA8 alone represents 74% of the total and the second most abundant haplotype is CmA6 which represents 7%; the other 21 haplotypes all together represent the remaining 19%. From the dataset of samples successfully characterized by the ~ 780 bp fragment (N = 465), 20 haplotypes were retrieved (Table 2 and 4). Haplotype CmA8.1 represents 72% of the total and CmA6.1 and CmA36.1 are the second most abundant ones, representing 8% each; the other 17 haplotypes all together represent the remaining 20%. Four 'long' control region haplotypes represent variants of two 'short' control region haplotypes; in detail, CmA8.1 and CmA8.3 split CmA8, while CmP49.1 and CmP49.2 split CmP49. In all the other cases, each 780 bp control region haplotype is the single, longer variant of one 490 bp control region haplotype. In both TCS control region networks (Fig. 1 and 2) haplotypes cluster in three groups. In both networks, the most represented group has a starlike shape and includes all the 'CmA' haplotypes, with CmA8 and CmA8.1 at the centre, respectively; the other two groups are represented by 'CmP' haplotypes.

Table 4 Genetic diversity statistics for datasets of ~ 490 and ~ 780 bp control region haplotypes. *N*, number of sequences; *k*, number of haplotypes; *h*, haplotypic diversity; π, nucleotide diversity; SE, Standard Error.

mtDNA control region fragment	***N***	***k***	***h*±SE**	**π±SE**
~ 490 bp	1034	23	0.4464±0.0007	0.0028±0.0001
~ 780 bp	465	20	0.4636±0.0015	0.0025±0.00005

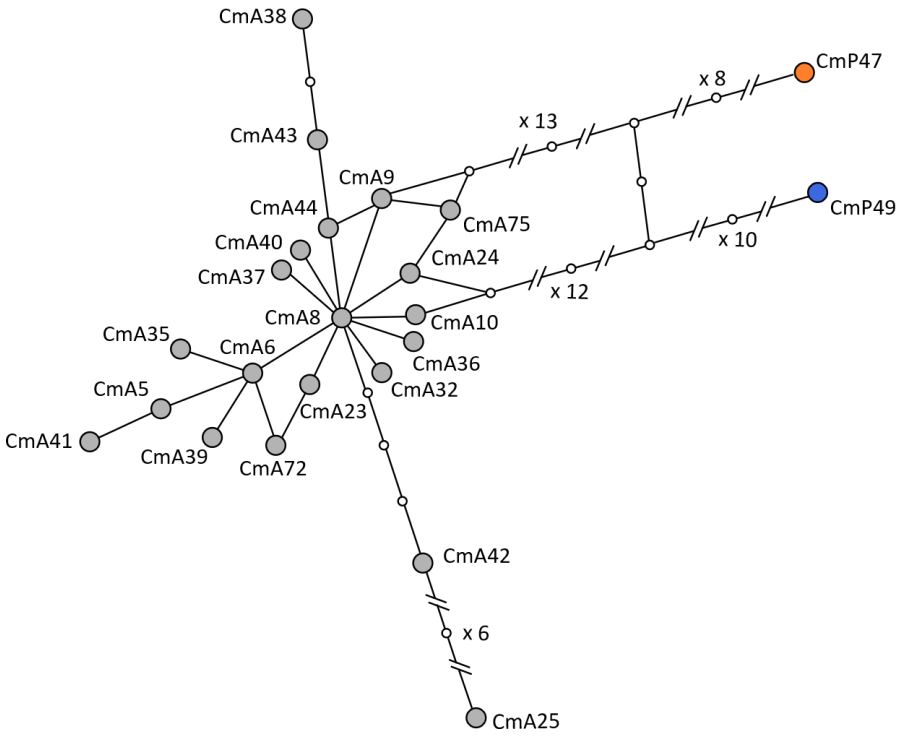


Fig. 1 TCS network for ~ 490 bp control region haplotypes retrieved in samples from dataset in Table 1. In grey, haplotypes from lineage II; in orange, haplotype from lineage IV; in blue, haplotype from lineage VIII, as defined in Jensen et al. (2019). Lineage II is the only one described in South Atlantic rookeries; lineages IV and VIII can be found in rookeries in Southwest Indian Ocean. White circles indicate unsampled haplotypes.

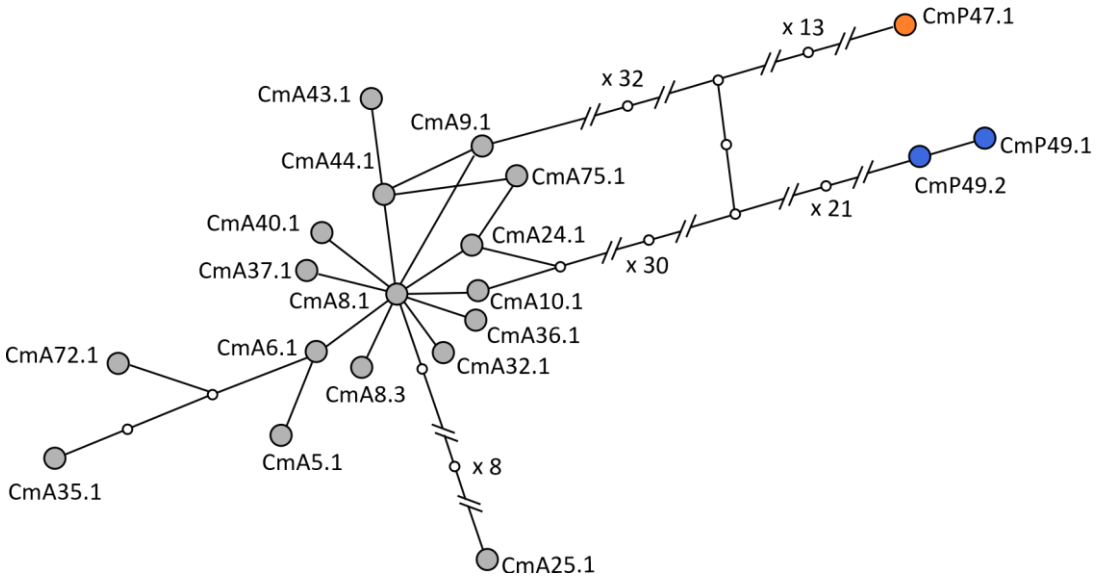


Fig. 2 TCS network for ~ 780 bp control region haplotypes retrieved in samples from dataset in Table 2. Colours and lineages as in Fig 1 and as defined in Jensen et al. (2019). White circles indicate unsampled haplotypes.

Mixed stock analysis

Contribution estimates from the Atlantic and Indo Pacific rookeries to the four mixed stock that we analysed generally had very wide confidence intervals, often crossing zero. For Western Africa mixed stock (Fig. S1), the highest contribution estimates were from Ascension Island and Bioko Island, 40% and 32%, but with confidence intervals 0 – 94% and 0 – 93%, respectively. For Equatorial Guinea mixed stock (Fig. S2), contribution estimates from Sao Tome, Ascension Island and Guinea Bissau were 38% (ci 14 - 72%), 32% (ci 0 – 69%) and 25% (ci 0 – 57%). For Corisco Bay mixed stock (Fig. S3), the highest contributions were from Sao Tomé (43%, ci 21 – 75%) and Ascension Island (37%, ci 20 – 58%), and in this case the estimates significantly differed from zero; in this case also a contribution from Europa (South Indian Ocean) was detectable (2%, ci 0 – 17%). For Sao Tomé mixed stock (Fig. S4) the main contribution estimate was from Sao Tomé itself (75%, ci 32 – 98%), followed by Ascension Island (14%, ci 0 – 51%).

Mitogenomic sequencing and assembly

From mitogenome sequencing we obtained an output nuclear coverage ≥ 1.8 X and optimal reads quality for 10 out of 13 samples, while three samples had an output nuclear coverage < 1.8 X. Two of these three samples also reported issues in FastQC statistics, due to bad per-sequence GC content, which impeded the reconstruction of the molecule by either NOVOPlasty or by mapping pipeline, therefore they were discarded from the analyses. For the remaining 11 samples, complete high coverage mitogenomes were obtained using NOVOPlasty and the mapping pipeline in parallel.

While comparing the contigs generated by NOVOPlasty and by the mapping pipeline for each sample, we found discrepancies on two occasions. For sample G14, NOVOPlasty reported three N positions, while the mapping pipeline assigned all the nucleotide bases; inspecting the bam files, we visually checked that the bases had been assigned correctly by the mapping. For sample M038 instead, the two contigs reported two different nucleotide bases at the same positions; inspecting the bam files, we visually checked that the correct assignment was achieved by NOVOPLasty, probably due to stringent per base quality filter of the mapping pipeline. Samples information, sequencing statistics, control region and mitogenomic haplotypes are reported for the mitogenomic dataset in Table 3.

Mitogenomic diversity and phylogenetic analyses

The dataset that we used for mitogenomic diversity analyses resulted by the alignment of our sequences (N = 11) with sequences downloaded from GenBank (N = 10) and counted 21 *C. mydas* mitogenomes (details in Table 3). We identified 14 mitogenomic haplotypes, four of which are mitogenomic variants of the 780 bp control region haplotype CmA8.1 and were retrieved from our samples collected in Corisco Bay, Principe and Sao Tomé. We also described a new mitogenomic variant of CmA5.1 (CmA5.1.4) in two samples from Sao Tomé. The other mitogenomic haplotypes came from published sequences. The most abundant mitogenomic haplotype is CmA8.1.1, carried by six sequences (29% of the total); two sequences carry CmA5.1.4 and CmP20.1.1 (10% each), one sequence carry CmA3.1.1, CmA3.1.2, CmA3.1.3, CmA5.1.1, CmA8.1.2, CmA8.1.3, CmA8.1.4, CmA13.1.1, CmP1.1.1, CmP4.1.1, CmP4.1.2 (5% each), see Table 3.

In the TCS network (Fig. 3) mitogenomic haplotypes cluster into 5 groups. One group includes the CmA8 and CmA5 variants, which had been sampled in both Eastern and Western Atlantic. Another group comprises CmA3 variants, also sampled in Western Atlantic, and it is located 15 substitutions away from the first group. The only Mediterranean haplotype stands apart, being 12 substitutions from the first group and 10 from the second. The greatest genetic distances are observed between these three groups and the two Pacific groups: one represented by CmP20.1.1 - sampled in Western Pacific - and one represented by CmP1.1.1, CmP4.1.1, CmP4.1.2 – sampled in Eastern Pacific. These Pacific groups are 138 and 154 substitutions away, respectively.

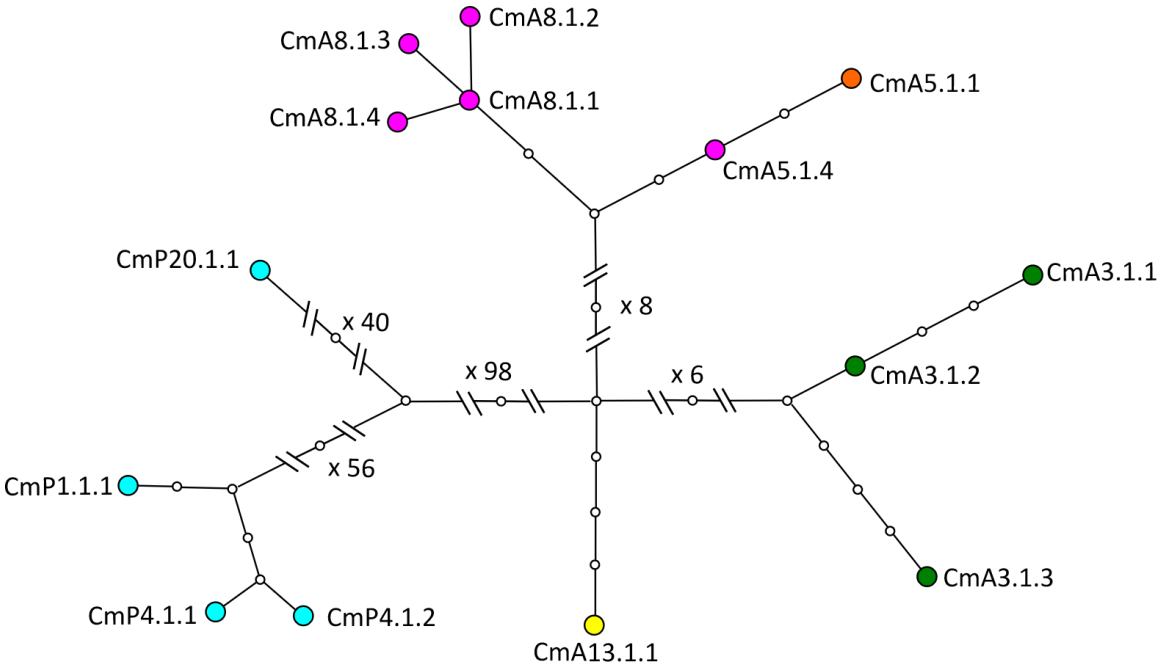


Fig. 3 TCS network for mitogenomic haplotypes retrieved in our dataset (Table 3). In light blue, haplotypes retrieved in samples from Pacific Ocean; in yellow, haplotype retrieved in samples from Mediterranea Sea; in purple, haplotypes retrieved in samples from the Gulf of Guinea; in orange, haplotype retrieved in samples from Buck Island; in green, haplotypes retrieved in samples from Atlantic Costa Rica. White circles indicate unsampled haplotypes.

Looking at the single gene diversity within the mitogenomic dataset (Table 5), the control region has the highest values of polymorphic sites, and its nucleotide diversity is an order of magnitude higher than for the other genes. Haplotypic diversity has the highest value in the control region, even if the same number of haplotypes (k = 7) is provided by ND2 gene. Among coding genes, CYTB and COXI have the highest haplotypic diversity, while ATP6 and ND4 have the highest nucleotide diversity.

Table 5 Information and diversity for single mitochondrial genes and tRNA partition of the mitogenomic alignment. Sequence length in base pairs (bp). Number of haplotypes (*k)*, number of polymorphic sites (*S)*, haplotype (h) and nucleotide diversity (π) with their Standard Error (SE), and dN/dS ratios for coding genes. * indicates that the test could not be computed, due to dS= 0.

mtDNA gene	Sequence length (bp)	*k*	*S*	*h*±SE	π±SE	dN/dS
12S	969	6	10	0.662±0.033	0.00259±0.00002	-
16S	1611	5	12	0.490±0.035	0.00193±0.00001	-
ATP6	683	6	14	0.662±0.033	0.00619±0.00006	0.097
ATP8	186	3	2	0.410±0.034	0.00343±0.00008	*
COXI	1546	5	12	0.767±0.029	0.00255±0.00001	0
COXII	687	3	6	0.410±0.034	0.00238±0.00003	0.063
COXIII	784	5	9	0.605±0.034	0.00238±0.00002	0
CYTB	1145	6	17	0.776±0.029	0.00457±0.00003	0.045
ND1	971	3	4	0.410±0.032	0.00131±0.00001	0.118
ND2	1039	7	15	0.671±0.032	0.00448±0.00003	0.172
ND3	350	2	1	0.257±0.030	0.00073±0.00002	*
ND4L	297	1	0	0	0	*
ND4	1381	4	25	0.595±0.034	0.00566±0.00004	0.107
ND5	1806	6	22	0.610±0.035	0.00363±0.00002	0.190
ND6	525	4	6	0.481±0.035	0.00319±0.00004	0.172
tRNAs	1550	3	5	0.410±0.034	0.00077±0.00006	-
Control region	927	7	58	0.790±0.028	0.01982±0.00014	-

The best-fitting molecular substitution model for the mitogenomic alignment of *C. mydas* and *N. depressus* was GTR + G4, which was used to construct the Maximum Likelihood (ML) phylogeny (Fig. 4). In the resulting topology, the first node with maximum statistical support separates haplotypes sampled in the Pacific Ocean from those sampled in the Atlantic Ocean and the Mediterranean Sea. The second most ancient and well-supported node corresponds to the divergence between CmP20.1.1 (sampled in the Western Pacific) and haplotypes sampled in the Eastern Pacific. Within the Atlantic clade, the ML topology supports a basal position of the Mediterranean haplotype CmA13.1.1, while the relationship between CmA3 variants and the other Atlantic haplotypes has low statistical support. The cluster of CmA8 and CmA5 variants is again well-supported, while more recent nodes are not.

Fig. 4 Maximum Likelihood phylogeny for complete mitogenomes dataset (Table 3). Colours as in Fig.3. Values of ultrafast bootstrap are reported for each node. The root does not have a value of statistical support because it was imposed by the user. Branch length is in number of substitutions per site.

The topology that we obtained using the Bayesian phylogenetic analysis (Fig. 5) mostly confirms the ML one, apart from the position of CmA13.1.1 which in this case clusters with CmA3 variants instead of being basal to the Atlantic clade; however, this node is not well-supported. In Table 6 we report the time estimates and respective 95% HPD for the seven most ancient nodes; the other nodes are not reported nor discussed since more recent events are less statistically supported.

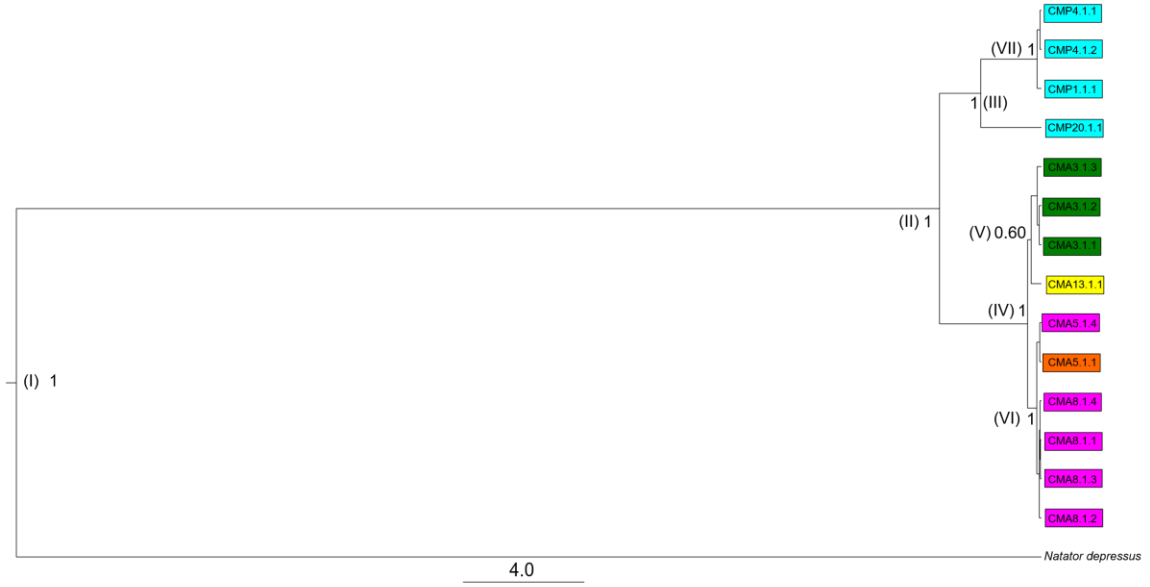


Fig. 5 Bayesian phylogeny for complete mitogenomes dataset (Table 3). Colours as in Fig. 3. Names of the nodes are in roman numbers (as in Table 6). Values of posterior probability are reported for each node. Branch length is in millions of years by present (MYBP).

Table 6 Bayesian tree's height estimates and 95% Height Posterior Density (HPD) for nodes shown in Fig 5. Estimates are reported in thousands of years by present (KYBP).

Node	Node description	Median (KYBP)	95% HPD
I	*N. depressus* – *C. mydas*	34,682	18,829 – 49,782
II	Pacific - Atlantic	3,464	1,879 – 5,176
III	CmP20.1.1 – Eastern Pacific	2,072	1,064 – 3,112
IV	CmA5/A8 – CmA3/13	476	215 – 750
V	CmA3 – CmA13	355	152 – 580
VI	CmA5 – CmA8	162	56 – 290
VII	CmP1.1.1 – CmP4	150	44 - 279

Selective pressure analyses

McDonald – Kreitman test, so as dN/dS, could not be computed for genes ATP8, ND3, ND4L due to the absence of synonymous substitutions. For all other genes where the test was performed, results were non-significant.

Tables 7 – 10 report the results of selective pressure tested with Codeml on the coding genes partition. The Random Site Method gave significant results for positive selection when comparing models M8 and M7; two sites, ATP6:AA3 and ND5:AA52, were detected by the bayes empirical bayes (BEB) approach as sites under positive selection with a probability > 80%. The Branch Method gave significant results for the best fitting of the alternative model than the null model, indicating different values of ω between the

foreground branch and the background branch, even if < 1. The tests were significant both in the case of *C. mydas* as foreground branch and in the case of the only Atlantic clade of *C. mydas* as foreground branch. Significant values of ω > 1 on the foreground branches were retrieved in both cases applying the Branch Site Method; in the case of Atlantic *C. mydas* as foreground branch, a site was detected by BEB as under positive selection with > 80% of probability (ND2:AA191). This site carries an Isoleucine in all Atlantic sequences and an Alanine in all Pacific sequences.

Those sites that resulted as potentially under positive selection for Codeml, namely ATP6:AA3, ND5:AA52, ND2:AA191, also scored significant in TreeSAAP respectively for 19, 16, 21 aminoacidic properties (of which 3, 3 and 1 belonging to a category of high impact ≥7).

Table 7 Codeml results for Random Site Method applied through nested models: null model, M0; alternative models, M1a, M2a. lnL: logarithm of Likelihood, np: number of parameters; ω_0 indicates purifying selection, ω_1 indicates neutral selection, ω_2 indicates positive selection; p_0: proportion of sites with ω_0, p_1: proportion of sites with ω_1, p_2: proportion of sites with ω_2. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. For M1a, LRT = 2(lnLM1a – lnLM0). For M2a, LRT = 2(lnLM2a – lnLM1a). The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.841. The critical value for LRT statistic, with 2 degrees of freedom and p = 0.005, is 5.99. Significant results are indicated by asterisks.

model	**lnL**	**np**	**ω_0**	**ω_1**	**ω_2**	**p_0**	**p_1**	**p_2**	**LRT**	**d.f.**
M0	-17467.1938	30	0.0378	-	-	1	-	-	-	-
M1a	-17435.4293	31	0.0161	1	-	0.9636	0.0364	-	63.52*	1
M2a	-17434.1561	33	0.0210	1	2.8423	0.9779	0.0116	0.0104	2.55	2

Table 8 Codeml results for Random Site Method applied through nested models: null model, M7 and alternative model M8. lnL: logarithm of Likelihood, np: number of parameters, p and q: parameters of beta distribution describing ω for sites in the range $0 \leq \omega \leq 1$. ω_2 indicates positive selection, p_0: proportion of sites with ω from beta (p, q) as in M7, p_1: proportion of sites with ω_2. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. Here, LRT = 2(lnLM8 – lnLM7). The critical value for LRT statistic, with 2 degrees of freedom and p = 0.005, is 5.99. Significant results are indicated by asterisks.

model	**lnL**	**np**	**p**	**q**	**ω_2**	**p_0**	**p_1**	**LRT**	**d.f.**
M7	-17438.2501	31	0.0157	0.2535	-	-	-	-	-
M8	-17434.0971	33	0.1581	4.6632	2.8501	0.9878	0.0122	8.31*	2

Table 9 Codeml results for Branch Method applied through models M0 and MBr1, first considering all *C. mydas* taxa as foreground branch, then considering only Atlantic *C. mydas* as foreground branch. lnL: logarithm of Likelihood, np: number of parameters; ω indicates purifying selection (homogeneous across all the sites and the tree); ω_b indicates the value for selection on the background branch; ω_f indicates the value for selection on the foreground branch. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. Here, LRT = 2(lnLMBr1 – lnLM0). The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.84. Significant results are indicated by asterisks.

Analysis setting	**model**	**lnL**	**np**	**ω**	**ω_b**	**ω_f**	**LRT**	**d.f.**
C. mydas as foreground branch	M0	-17467.1937	30	0.0378	-	-	-	-
	MBr1	-17455.8466	31	-	0.0234	0.0967	22.69*	1
Atlantic *C. mydas* as foreground branch	M0	-17467.1925	30	0.0378	-	-	-	-
	MBr1	-17455.8465	31	-	0.0347	0.1567	22.69*	1

Table 10 Codeml results for Branch Site Method applied through nested models M0, M2a_f and MBr1_2s, first considering all *C. mydas* taxa as foreground branch, then considering only Atlantic *C. mydas* as foreground branch. lnL: logarithm of Likelihood, np: number of parameters; ω_0 indicates purifying selection, ω_1 indicates neutral selection, ω_2 indicates positive selection; p_0: proportion of sites with ω_0, p_1: proportion of sites with ω_1, p_{2a}: proportion of sites with ω_0 on the background branch and ω_2 on the foreground branch, p_{2b}: proportion of sites with ω_1 on the background branch and ω_2 on the foreground branch. Here, LRT is computed comparing MBr1_2s first with M0: LRT = 2(lnLMBr1_2s – lnLM0); then, with M2a_f: LRT = 2(lnLMBr1 – lnLM2a_f). The critical value for LRT statistic, with 3 degrees of freedom and p = 0.005, is 7.81. The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.84. Significant results are indicated by asterisks.

Analysis setting	**model**	**lnL**	**np**	**ω_0**	**ω_1**	**ω_2**	**p_0**	**p_1**	**p_{2a}**	**p_{2b}**	**LRT**	**d.f.**
C. mydas as foreground branch	M0	-17467.1923	30	0.0377	-	-	-	-	-	-	-	-
	M2a_f	-17435.4303	32	0.0161	-	1	-	-	-	-	-	-
	MBr1_2s	-17435.4296	33	0.0162	1	2.0848	0.8871	0.0335	0.0765	0.0029	63.52*	3
											0.001	1
Atlantic *C. mydas* as foreground branch	M0	-17467.1939	30	0.0378	-	-	-	-	-	-	-	-
	M2a_f	-17431.1969	32	0.0156	-	1	-	-	-	-	-	-
	MBr1_2s	-17428.6939	33	0.0158	1	48.2178	0.9608	0.0333	0.0034	0.0001	77.00*	3
											5.01*	1

Discussion

The genetic analyses in this study aimed to enhance current knowledge about the diversity within foraging stocks and the movements of *C. mydas* reaching the Gulf of Guinea. We applied multiple genetic markers and analytical pipelines to characterize our samples, in specific we tested the potential of whole mitogenomes to improve genetic diversity resolution and to give insights into the species phylogeography. We also provide methodological considerations which might support future research planning. To the best of our knowledge, the present study is based on the largest dataset of green turtle samples ever collected in the Gulf of Guinea.

Control region analyses

Using 'short' control region haplotypes (490 bp), we characterized all the samples in the dataset. We confirmed some patterns already reported in literature; specifically, CmA8 is by far the most abundant haplotype at all sampling locations of the Gulf of Guinea, followed by CmA6 and CmA5, while CmA36 is characteristic of Sao Tomé and Principe (Formia et al. 2006; Hancock et al. 2019). However, we also identified some haplotypes that had never been reported in this study area (Formia et al. 2006; Hancock et al. 2019). Haplotypes CmA41 and CmA72, detected in two and one samples from Corisco Bay, respectively, have never been found in any rookery or foraging ground. The same applies to CmA43, which we found in one turtle caught in Liberia, one in Equatorial Guinea, and four in Corisco Bay; we also recorded CmA43 in a nesting turtle in Gabon, suggesting a possible origin for this rare haplotype. These findings indicate that some rookeries – whether nearby or distant – remain undersampled or even unknown. We also found haplotypes not previously documented in this area, but with records in known rookeries. Specifically, we found two turtles in Corisco Bay carrying haplotype CmA25, which has been recorded in the rookeries of Fernando Noronha (Brazil) (Encalada et al. 1996; Bjorndal et al. 2006) and in Ascension Island (Formia et al. 2007); we also found one turtle in Equatorial Guinea and one turtle in Corisco Bay carrying haplotype CmA42, which has been recorded in Guinea Bissau (Patrício et al. 2017). Finally, we found one turtle and seven turtles in Corisco Bay carrying haplotypes CmP47 and CmP49, respectively. As shown by our haplotypic networks and by phylogenetic reconstruction in Jensen et al. (2019), CmP47 and CmP49 belong to a distinct, Indo - Pacific phylogenetic clade compared to all the other haplotypes that we identified in our dataset. These

haplotypes are specific of Southern Indian Ocean rookeries (Bourjea et al. 2015), with no records from Atlantic nesting sites. This suggests that some individuals likely enter the Atlantic via South Africa, swimming northward until they reach Corisco Bay. The eight samples carrying these Indo Pacific haplotypes indicate that this migration is not a common behaviour, neither is it an isolated event. Moreover, according to Van der Zee et al. (2021), the crossing of the Cape of Good Hope has occurred more than once on an evolutionary timescale, both westward and eastward, allowing gene flow between Atlantic and Indo Pacific.

A more exhaustive description of the composition of the stocks at sea in terms of contributing rookeries was tempted through the Mixed Stock Analyses. Haplotype CmA8, widely shared by all Eastern Atlantic rookeries, severely hampers the resolution of the MSA. In all the four mixed stocks we analysed, the contributions with the narrowest confidence intervals were those from Western Atlantic, where haplotype CmA8 is not present, indicating 'zero contribution'. For most of the contribution estimates higher than zero, we had wide confidence intervals, the real value potentially being twice or half the estimate or even null, in case of confidence intervals including zero. This was the case for the whole Western Africa's MSA. For Western Africa we hypothesize an overestimation of contribution from Bioko and a contemporary underestimation of contribution from Poilao (Guinea Bissau), since Poilao is one of the major rookeries in Eastern Atlantic (Seminoff et al. 2015) and significant connections between Guinea Bissau and Western Africa have been already reported in literature (Patrício et al. 2017); on the other hand, Bioko is quite a small rookery (Tomás et al. 2010). For Sao Tomé, the main contribution was from local rookeries and in this case we consider it reliable since it departs from zero. Curiously, Hancock et al. (2019) found much higher contribution from Guinea Bissau in a similar MSA on Sao Tomé, while we estimated it to be quite low. The MSA for Equatorial Guinea took advantage of an exclusive haplotype of Sao Tomé (CmA36) and obtained contribution significantly higher than zero from this rookery. The best performing MSA was Corisco, where a very large sample size (ten times the sample size of Western Africa) allowed more reduced confidence intervals compared to the other MSA and main estimates significantly higher than zero, from Sao Tomé and Ascension Island.

Since the wide sharing of few common haplotypes has hampered the population structure description and the MSA for a long time, the characterization of the longer control region fragments (780 bp) was tempted for *C. mydas*, aimed to find more polymorphisms and to

increase the power of the genetic marker. In our study we faced the limitation of a low amplification and sequencing success for the 'long' control region fragment, reducing the dataset of more than half the total sample size. The number of 'long' haplotypes was lower than the number of 'short' haplotypes, therefore the resolution of the new marker did not compensate for the low yield. Overall, the genetic information obtained by long haplotypes did not provide additional insights to the previous results.

Mitogenomic analyses

In Tolve et al. (2024), we showed for *C. caretta* in the Mediterranean that complete mitogenomes can split the most shared control region haplotypes into numerous mitogenomic variants. In this study as well, we aimed to test the capability of mitogenomics to better resolve shared lineages in *C. mydas*, which long control region markers had almost failed to achieve. Additionally, we explored phylogeographic and adaptive signals in the mitochondrial DNA that could enhance our understanding of populations evolution, as we did for loggerheads in the previous chapter of this thesis.

Complete mitogenomes were retrieved as secondary outputs from whole-genome sequencing (WGS) data. Since genomic library preparation requires high DNA quality and quantity mitogenomics studies based on WGS are generally applied to a limited number of samples. Therefore, optimizing the sequencing output is crucial for obtaining datasets of sufficient size. In our protocol, adopting two assembly pipelines and comparing the results helped correct errors and uncertainties, thereby enabling the reconstruction of mitogenomic contigs from sub-optimally sequenced reads. We emphasize the importance of generating bam files through the mapping pipeline, as they are a key tool for visualizing polymorphisms and confirming or rejecting contig reconstruction. Overall, WGS with low - medium nuclear target coverage and a combination of mapping and *de novo* assembly allowed us to obtain medium – high coverage mitogenomes for 11 out of 13 samples that we had selected for this approach.

Mitogenomes revealed high haplotypic diversity, since we retrieved 5 mitogenomic haplotype among 11 samples and 4 of these haplotypes were variants of the control region haplotype CmA8.1, ubiquitous in the Eastern Atlantic rookeries; the other mitogenomic haplotype we retrieved is a variant of CmA5.1. This control region haplotype not only is present in Eastern Atlantic, but it is also the most widespread in Western Atlantic rookeries; Shamblin et al. (2012a) found three different mitogenomic variants, which differ from the one we found in our samples. This is promising for improving the genetic

assignment of turtles to different Regional Management Units. Since almost all our samples were collected at sea, we were not able to associate different mitogenomic variants to different nesting site, but thanks to these findings we characterized mitogenomic diversity in the Gulf of Guinea with a higher potential for resolution of lineages than long control region haplotypes. This paves the way for future, focused investigation on nesting sites.

The mitogenomic network and phylogenies largely align with the reconstructions based on the control region by Jensen et al. (2019), although some differences are present. Mitogenomic variants of CmA8 and CmA5, previously reported in the Eastern Atlantic and Caribbean (Shamblin et al. 2012a), cluster with mitogenomic variants of CmA3 (Western Atlantic), mirroring the grouping of control region clades II and I in Jensen et al. (2019). The Mediterranean mitogenome (CmA13.1.1) and CmA3 mitogenomic variants are more divergent from each other than their corresponding control region haplotypes (Jensen et al. 2019), and there is disagreement between our ML and Bayesian phylogenies regarding the placement of the Mediterranean haplotype. If the ML tree was correct (given its high statistical support, unlike Bayesian tree), the Mediterranean clade would represent the earliest divergence from the Atlantic haplotypes and not a recent one, as stated in Jensen et al. (2019). Of course, a single sample from the Mediterranean is not sufficient to draw firm conclusions, and broader sampling in the region is needed. The Pacific mitogenomic haplotypes included in our analyses, from both the Western and Eastern Pacific, cluster together; however, their corresponding control region clades, namely clade III and clade IX in Jensen et al. (2019), do not. Clade III is grouped with clades I and II, while clade IX clusters with other Pacific lineages. The phylogenetic reconstruction from our analyses is statistically robust, with strong support for all nodes involving these clades, and it is also consistent with geographic expectations; control region clade III and clades I – II are geographically disconnected, whereas clades III and IX likely diverged in the Pacific from a common ancestor. Regarding the dating of the nodes, the posterior estimate for the divergence between *C. mydas* and *N. depressus* is still consistent with the values in Duchene et al. (2012), that we used as prior. The divergence between Pacific and Atlantic turtles around 3.5 MYBP occurred in correspondence with the closure of the Isthmus of Panama (20 – 2.8 MYBP) (O'Dea et al. 2016), a pattern shared by other turtle species (e.g. Shamblin et al. 2014). The split of the mitogenomic clades of Western and Eastern Pacific date around 2 MYBP, at the beginning of the Pleistocene, while the divergence between our Atlantic mitogenomic

clades result more recent than what testified by the control region analyses (Jensen et al. 2019). Discrepancies between single-gene and whole-mitogenome phylogenies are not exceptional, since mutation rates and selective pressures vary considerably across different regions of the mitochondrial DNA. Therefore, different phylogeographic scenarios may emerge, depending on the mtDNA marker chosen for the analysis; for this reason, it is important to refer to the most informative one – in this case, the complete mitogenome - as discussed also in the previous chapter of this thesis in relation to the loggerhead turtle.

To understand the phylogeography of the species it is also necessary to investigate adaptive patterns. Indeed, the genetic structure and geographic distribution of mitochondrial lineages in many marine species result not only from neutral evolution but can also be shaped by selective pressure (e.g. Silva et al. 2014; Baltazar-Soares et al. 2021; Noll et al. 2022). Although mitochondrial DNA is a highly conserved molecule and primarily subject to purifying selection, signals of positive selection can emerge under appropriate biological hypotheses and analytical framework (e.g. Morales et al. 2015). From the tests we ran on green turtle mitogenomes, we found weak evidence of positive selection across the phylogeny, supported by the least stringent models within the Random Site Method (models M8 and M7). The selective pressure acting on *C. mydas* branch significantly differed from that on the outgroup *N. depressus*, according to both the Branch Method and Branch Site Method. However, the methods yielded different values, which were not always greater than 1, and thus did not consistently indicate positive selection. A similar pattern was observed on the Atlantic *C. mydas* branch, where we also identified a specific site under selection in the ND2 gene, with high probability, confirmed by the analyses of physicochemical aminoacidic properties. Upon inspecting our alignments, we also detected the point mutation in the ATP8 gene, previously reported and discussed by Duchene et al. (2012), which removes a stop codon and extends the overlapping region with ATP6. While those authors reported the mutation only in Mediterranean and Atlantic *C. mydas*, we also observed the mutation in two Pacific haplotypes, suggesting that it may have evolved independently more than once within the *C. mydas* phylogeny. The phenotypic effects of this mutation, as well as of the others we reported, remain to be fully understood and should be interpreted in the context of the species biology. A broader dataset of mitogenomes, combined with biochemical approaches, will be essential for drawing more robust results.

Conclusions

In the present study we meant to improve our knowledge of green turtles in the Gulf of Guinea by both leveraging data produced by traditional mtDNA markers from a broad sample dataset and employing newer techniques of whole mitogenome sequencing. The extensive characterization of control region haplotypes allowed us to compare our sequences with existing baseline information, thereby enabling inferences about the origin and movements of turtles within the Gulf. We confirmed the contribution of nearby rookeries to the mixed stocks and, notably, detected turtles from Indian Ocean reaching Corisco Bay, shedding light on an understudied migratory behaviour. On the other hand, whole mitogenome sequencing proved promising in resolving the ubiquitous control region mtDNA lineage, potentially improving phylogeographic reconstructions and studies on adaptation. While a full replacement of traditional markers by whole mitogenomes is not currently on the agenda for population genetics, we envisage complementarity between the two techniques, where mitogenomes can address well-defined questions within a broader context identified by traditional, widespread markers.

Acknowledgments

I am grateful to Angela Formia, African Aquatic Conservation Fund, and to the volunteers from many organizations in Central and West Africa who assisted with the collection of samples.

Table S1. Baseline data used in MSA: 490 bp haplotype frequencies in Atlantic and Indo Pacific green turtle rookeries, with respective rookery size estimates in number of nesting females per year. FL: Florida (Encalada et al. 1996; Bass and Wayne 2000); MX: Quintana Roo, Mexico (Encalada et al. 1996); CR: Tortugero, Costa Rica (Encalada et al. 1996; Bjorndal et al. 2005); CU: Cuba (Ruiz-Urquiola et al. 2010); BKI: Buck Island (Shamblin et al. 2012a); GU: Guadeloupe (Costa Jordao et al. 2017); AV: Aves Island, Venezuela (Encalada et al. 1996; Lahanas et al. 1998; Shamblin et al. 2015); SU: Matapica, Suriname (Encalada et al. 1996; Shamblin et al. 2012a); FG: French Guiana (Costa Jordao et al. 2017); FN: Fernando de Noronha, Brazil (Encalada et al. 1996; Bjorndal et al. 2006; Shamblin et al. 2015); TRI: Trinidade Island, Brazil (Encalada et al. 1996; Bjorndal et al. 2006; Shamblin et al. 2015); ASC: Ascension Island (Formia et al. 2007); GB: Guinea Bissau (Patrício et al. 2017); BIO: Bioko Island (Formia et al. 2006); SAO: Sao Tomé Island (Formia et al. 2006); PR: Principe Island (this study, unpublished data; Formia et al. 2006); MED: Lara bay, Cyprus (Encalada et al. 1996; Kaska 2000); EUR: Europa Island (Bourjea et al. 2015); JDN: Juan de Nova (Bourjea et al. 2015); NSWIO: Northern South West Indian Ocean (Bourjea et al. 2015); SEY: Seychelles (Bourjea et al. 2015).

Haplotype	FL	MEX	CR	CU	BKI	GU	AV	SU	FG	FN	TRI	ASC	GB	BIO	SAO	PR	MED	EUR	JDN	NSWIO	SEY
CMA1	224	7	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA2	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA3	219	5	395	16	0	1	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA5	6	1	32	0	45	35	69	68	43	0	0	0	0	0	1	0	0	0	0	0	0
CMA6	0	0	0	0	0	0	0	3	0	0	0	11	0	5	1	0	0	0	0	0	0
CMA7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA8	1	0	0	0	0	0	0	0	2	50	67	204	170	45	13	39	0	31	11	13	0
CMA9	0	0	0	0	0	0	0	0	0	7	19	9	0	0	0	0	0	0	0	0	0
CMA10	0	0	0	0	0	0	0	0	0	2	0	5	0	0	0	0	0	0	0	0	0
CMA11	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
CMA12	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
CMA13	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	241	0	0	0	0
CMA14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
CMA15	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA16	3	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA17	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA18	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA20	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA21	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA22	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CMA23	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	0	0	0	0	0
CMA24	0	0	0	0	0	0	0	0	0	0	1	7	0	0	0	0	0	0	0	0	0
CMA25	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0

CMA26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA27	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
CMA28	5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA32	0	0	0	0	0	0	0	0	0	1	5	1	0	0	0	0	0	0	0	0	0
CMA33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
CMA36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	25	0	0	0	0	0
CMA37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
CMA38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
CMA39	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CMA40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA42	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CMA43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA44	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CMA45	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CMA46	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
CMA47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA48	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA50	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CMA53	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA56	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA57	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
CMA62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
CMA63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
CMA68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMA75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
CMP40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CMP47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	37	24
CMP49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	240	54
CMP57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

CMP62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CMP75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	6
CMP76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1
CMP79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CMP87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
CMP114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CMP115	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CMP120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
TOT sequences	**485**	**20**	**433**	**28**	**49**	**36**	**75**	**73**	**46**	**69**	**99**	**245**	**171**	**50**	**20**	**66**	**248**	**33**	**20**	**307**	**89**
Rookery size	779	1,587	26,535	2,226	63	50	267	1,814	6,000	129	3,000	3,709	2,523	407	90	88	100	6,500	20	16,550	750

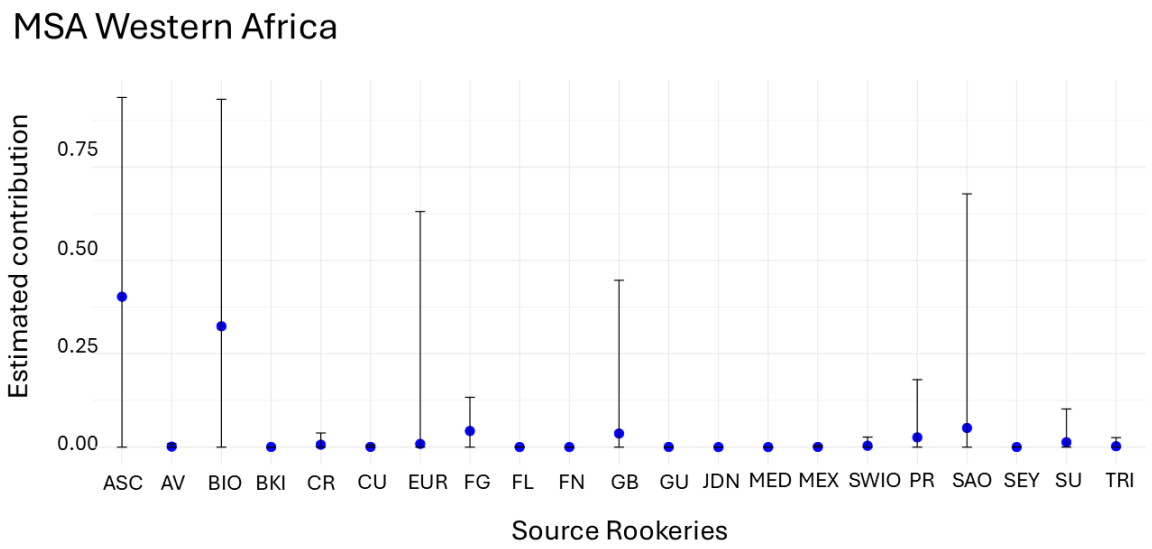


Fig. S1 Estimated contribution of Atlantic and Indo Pacific rookeries to Western Africa mixed stock. Acronyms as in Table S1.

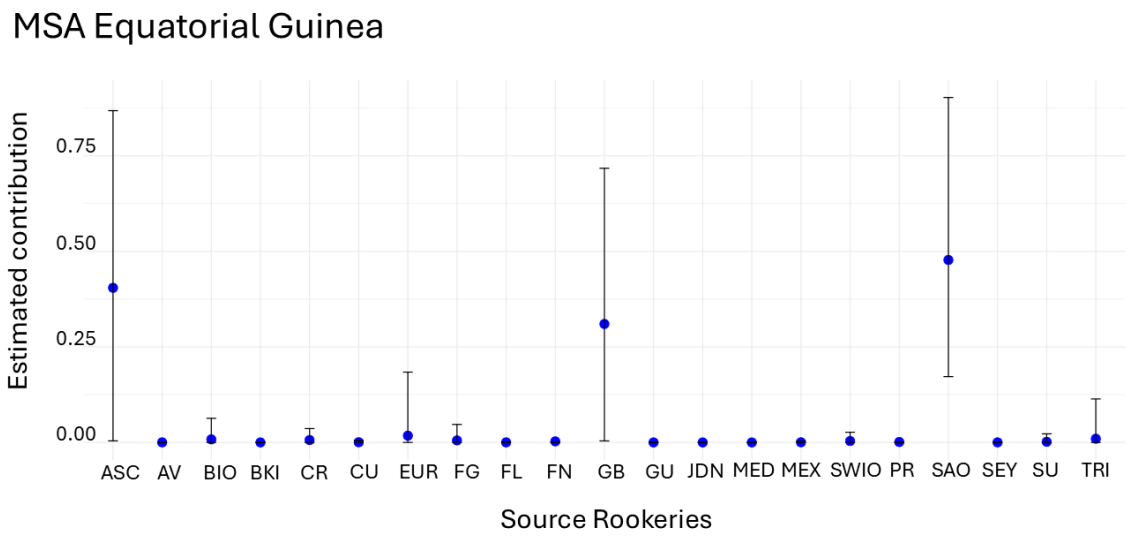


Fig. S2 Estimated contribution of Atlantic and Indo Pacific rookeries to Equatorial Guinea mixed stock. Acronyms as in Table S1.

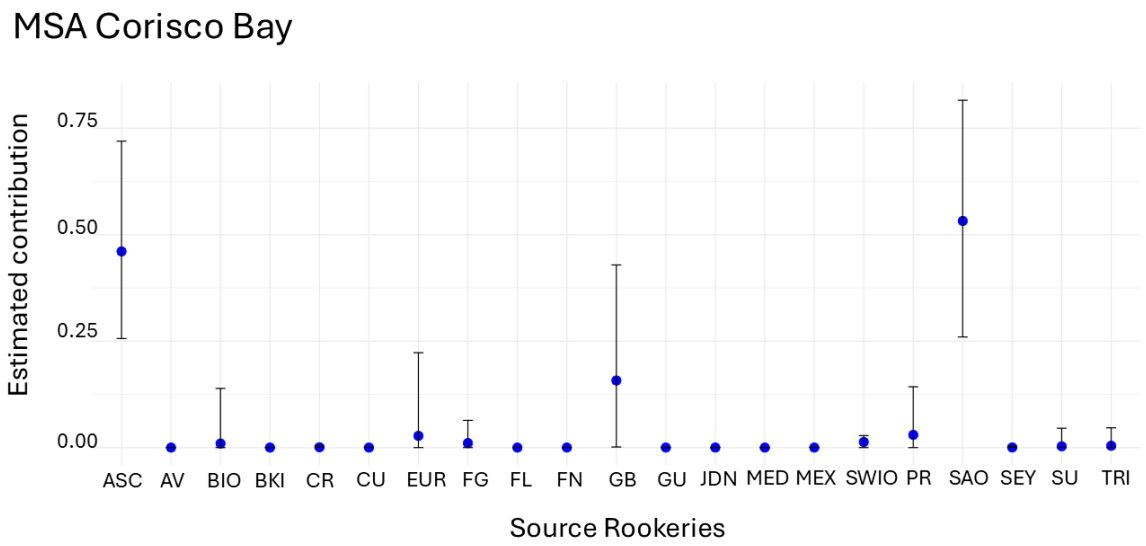


Fig. S3 Estimated contribution of Atlantic and Indo Pacific rookeries to Corisco mixed stock. Acronyms as in Table S1.

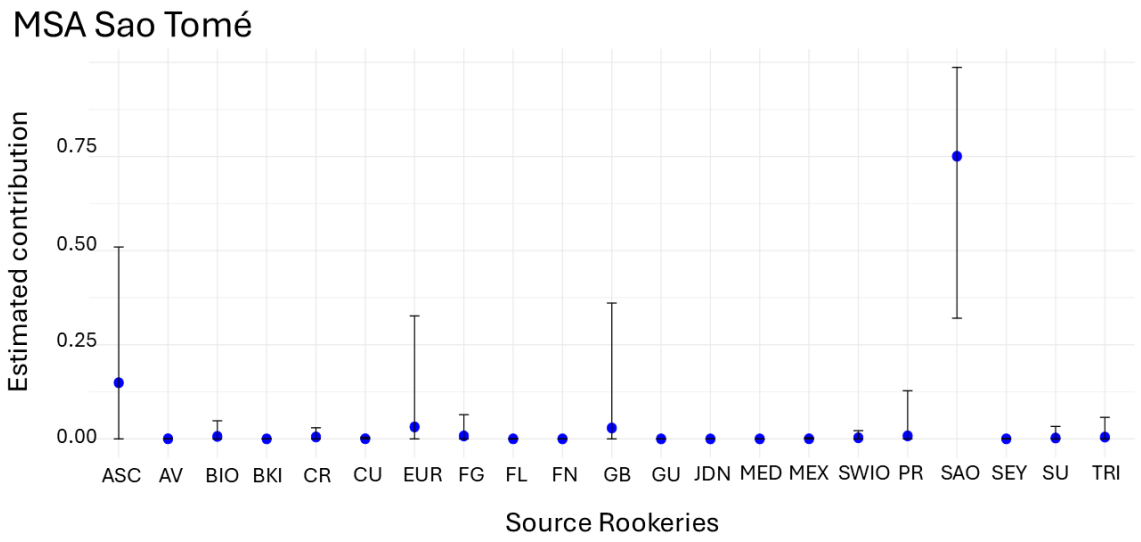


Fig. S4 Estimated contribution of Atlantic and Indo Pacific rookeries to Sao Tomé mixed stock. Acronyms as in Table S1.

References

Abreu A, Horrocks J, Formia A, LeRoux R, Velez-Zuazo X, Dutton P, Soares L, Meylan P, Browne D (2006) New mtDNA Dloop primers which work for a variety of marine turtle species may increase the resolution capacity of mixed stock analysis. Proceedings of the 26th Annual Symposium on Sea Turtle Biology and Conservation.

Allard MW, Miyamoto MM, Bjorndal KA, Bolten AB, Bowen BW (1994) Support for natal homing in green turtles from mitochondrial DNA sequences. Copeia 1994: 34-41.

Álvarez-Carretero S, Kapli P, Yang Z (2023) Beginner's Guide on the Use of PAML to Detect Positive Selection. Molecular Biology and Evolution 40.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Baltazar-Soares M, de Araújo Lima AR, Silva G (2021) Targeted sequencing of mitochondrial genes reveals signatures of molecular adaptation in a nearly panmictic small pelagic fish species. Genes 12: 91.

Bass AL, Wayne NW (2000) Demographic composition of immature green turtles (*Chelonia mydas*) from the East Central Florida coast: evidence from mtDNA markers. Herpetologica 56: 357-367.

Bjorndal KA, Bolten AB, Moreira L, Bellini C, Marcovaldi MÂ (2006) Population structure and diversity of Brazilian green turtle rookeries based on mitochondrial DNA sequences. Chelonian Conservation and Biology 5: 262-268, 267.

Bjorndal KA, Bolten AB, Troëng S (2005) Population structure and genetic diversity in green turtles nesting at Tortuguero, Costa Rica, based on mitochondrial DNA control region sequences. Marine Biology 147: 1449-1457.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Bolker BM, Okuyama T, Bjorndal KA, Bolten AB (2007) Incorporating multiple mixed stocks in mixed stock analysis: ‘many-to-many’ analyses. Molecular Ecology 16: 685-695.

Bourjea J, Mortimer JA, Garnier J, Okemwa G, Godley BJ, Hughes G, Dalleau M, Jean C, Ciccione S, Muths D (2015) Population structure enhances perspectives on regional management of the western Indian Ocean green turtle. Conservation Genetics 16: 1069-1083.

Casale P, Abitsi G, Aboro MP, Agamboue PD, Agbode L, Allela NL, Angueko D, Bibang Bi Nguema JN, Boussamba F, Cardiec F, Chartrain E, Ciofi C, Emane YA, Fay JM, Godley BJ, Kouerey Oliwiwina CK, de Dieu Lewembe J, Leyoko D, Mba Asseko G, Mengue M'adzaba P, Mve Beh JH, Natali C, Nyama-Mouketou C, Nzegoue J, Ogandagas C, Parnell RJ, Rerambyath GA, Gnandji MS, Sounguet G-P, Tiwari M, Verhage B, Vilela R, White L, Witt MJ, Formia A (2017) A first estimate of sea turtle bycatch in the industrial trawling fishery of Gabon. Biodiversity and Conservation 26: 2421-2433.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1659.

Costa Jordao J, Bondioli ACV, Almeida-Toledo LFd, Bilo K, Berzins R, Le Maho Y, Chevallier D, de Thoisy B (2017) Mixed-stock analysis in green turtles Chelonia mydas: mtDNA decipher current connections among west Atlantic populations. Mitochondrial DNA Part A 28: 197-207.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008.

Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18-e18.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

Encalada SE, Lahanas PN, Bjorndal KA, Bolten AB, Miyamoto MM, Bowen BW (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. Molecular Ecology 5: 473-483.

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

Formia A, Broderick AC, Glen F, Godley BJ, Hays GC, Bruford MW (2007) Genetic composition of the Ascension Island green turtle rookery based on mitochondrial DNA: implications for sampling and diversity. Endangered Species Research 3: 145-158.

Formia A, Godley BJ, Dontaine JF, Bruford MW (2006) Mitochondrial DNA diversity and phylogeography of endangered green turtle (*Chelonia mydas*) populations in Africa. Conservation Genetics 7: 353-369.

Hamann M, Godfrey MH, Seminoff JA, Arthur K, Barata PCR, Bjorndal KA, Bolten AB, Broderick AC, Campbell LM, Carreras C, Casale P, Chaloupka M, Chan SKF, Coyne MS, Crowder LB, Diez CE, Dutton PH, Epperly SP, FitzSimmons NN, Formia A, Girondot M, Hays GC, Cheng IS, Kaska Y, Lewison R, Mortimer JA, Nichols WJ, Reina RD, Shanker K, Spotila JR, Tomás J, Wallace BP, Work TM, Zbinden J, Godley BJ (2010) Global research priorities for sea turtles: informing management and conservation in the 21st century. Endangered Species Research 11: 245-269.

Hancock JM, Vieira S, Taraveira L, Santos A, Schmitt V, Semedo A, Patrício AR, Ferrand N, Gonçalves H, Sequeira F (2019) Genetic characterization of green turtles (*Chelonia mydas*) from São Tomé and Príncipe: Insights on species recruitment and dispersal in the Gulf of Guinea. Journal of Experimental Marine Biology and Ecology 518: 151181.

Jensen MP, FitzSimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH (2019) The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). Journal of Biogeography 46: 860-870.

Kaska Y (2000) Genetic structure of Mediterranean sea turtle populations. Turkish Journal of Zoology 24: 191 - 198.

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. Molecular Biology and Evolution 16: 784-792.

Lahanas PN, Bjorndal KA, Bolten AB, Encalada SE, Miyamoto MM, Valverde RA, Bowen BW (1998) Genetic composition of a green turtle (Chelonia mydas) feeding ground population: evidence for multiple origins. Marine Biology 130: 345-352.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359.

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

Morales HE, Pavlova A, Joseph L, Sunnucks P (2015) Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Molecular Ecology 24: 2820-2837.

Múrias dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M (2015) tcsBU: a tool to extend TCS network layout and visualization. Bioinformatics 32: 627-628.

Noll D, Leon F, Brandt D, Pistorius P, Le Bohec C, Bonadonna F, Trathan PN, Barbosa A, Rey AR, Dantas GPM, Bowie RCK, Poulin E, Vianna JA (2022) Positive selection over the mitochondrial genome and its role in the diversification of gentoo penguins in response to adaptation in isolation. Scientific Reports 12: 3767.

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD, Stallard RF, Woodburne MO, Aguilera O, Aubry M-P, Berggren WA, Budd AF, Cozzuol MA, Coppard SE, Duque-Caro H, Finnegan S, Gasparini GM, Grossman EL, Johnson KG, Keigwin LD, Knowlton N, Leigh EG, Leonard-Pingel JS, Marko PB, Pyenson ND, Rachello-Dolmen PG, Soibelzon E, Soibelzon L, Todd JA, Vermeij GJ, Jackson JBC (2016) Formation of the Isthmus of Panama. Science Advances 2: e1600883.

Patrício AR, Formia A, Barbosa C, Broderick AC, Bruford M, Carreras C, Catry P, Ciofi C, Regalla A, Godley BJ (2017) Dispersal of green turtles from Africa' s largest rookery assessed through genetic markers. Marine Ecology Progress Series 569: 215-225.

Pella J, Masuda M (2001) Bayesian methods for analysis of stock mixtures from genetic characters. Fishery Bulletin 99: 151-151.

Proyecto MARES Comunidad (2024) Identifying and supporting alternative economic opportunities to fishing.

Riskas KA, Tiwari M (2013) An overview of fisheries and sea turtle bycatch along the Atlantic coast of Africa. Munibe Monographs, Nature Series 1: 1-82.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29: 24-26.

Ruiz-Urquiola A, Riverón-Giró FB, Pérez-Bermúdez E, Abreu-Grobois FA, González-Pumariega M, James-Petric BL, Díaz-Fernández R, Álvarez-Castro JM, Jager M, Azanza-Ricardo J (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: 33-52.

Seminoff JA, Allen CD, Balazs GH, Dutton PH, Eguchi T, Haas H, Hargrove SA, Jensen M, Klemm DL, Lauritsen AM, MacPherson SL, Opay P, Possardt EE, Pultz S, Seney EE, Van Houtan KS, Waples RS (2015) Status review of the green turtle (*Chelonia mydas*) under the Engangered Species Act.

Shamblin BM, Bagley DA, Ehrhart LM, Desjardin NA, Martin RE, Hart KM, Naro-Maciel E, Rusenko K, Stiner JC, Sobel D, Johnson C, Wilmers TJ, Wright LJ, Nairn CJ (2015) Genetic structure of Florida green turtle rookeries as indicated by mitochondrial DNA control region sequences. Conservation Genetics 16: 673-685.

Shamblin BM, Bjorndal KA, Bolten AB, Hillis-Starr ZM, Lundgren I, Naro-Maciel E, Nairn CJ (2012) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol Ecol 21: 2330-2340.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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: e85956.

Silva G, Lima FP, Martel P, Castilho R (2014) Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. Proceedings of the Royal Society B: Biological Sciences 281: 20141093.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619-633.

Tolve L, Casale P, Formia A, Garofalo L, Lazar B, Natali C, Novelletto A, Vallini C, Bužan E, Chelazzi G, Gaspari S, Fortuna C, Kocijan I, Marchiori E, Novarini N, Poppi L, Salvemini P, Ciofi C (2018) A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. Mar Biol 165: 68.

Tolve L, Iannucci A, Garofalo L, Ninni A, Capobianco Dondona A, Ceciarini I, Cocumelli C, De Lucia A, Falconi M, Formia A, Iacovelli F, Mancusi C, Marchiori E, Marsili L, Mingozzi T, Nannarelli S, Natali C, Terracciano G, Zuffi MAL, Novelletto A, Ciofi C (2023) Whole mitochondrial genome sequencing

provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine Biology 171: 19.

Tomás J, Godley BJ, Castroviejo J, Raga JA (2010) Bioko: critically important nesting habitat for sea turtles of West Africa. Biodiversity and Conservation 19: 2699-2714.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44: W232-W235.

United Nations (2015) Sustainable Development Goal 14: Life Below Water

van der Zee JP, Christianen MJA, Bérubé M, Nava M, Schut K, Humber F, Alfaro-Núñez A, Becking LE, Palsbøll PJ (2021) The population genomic structure of green turtles (Chelonia mydas) suggests a warm-water corridor for tropical marine fauna between the Atlantic and Indian oceans during the last interglacial. Heredity 127: 510-521.

Vilaça ST, Hahn AT, Naro-Maciel E, Abreu-Grobois FA, Bowen BW, Castilhos JC, Ciofi C, FitzSimmons NN, Jensen MP, Formia A, Limpus CJ, Natali C, Soares LS, de Thoisy B, Whiting SD, Bonatto SL (2022) Global phylogeography of ridley sea turtles (Lepidochelys spp.): evolution, demography, connectivity, and conservation. Conservation Genetics 23: 995-1010.

Wallace BP, Bandimere AN, Abreu-Grobois FA, Acosta H, Akiti J, Akomedi M, Alfaro-Shigueto J, Allen CD, Angenda D, Ayissi I, Azanza Ricardo J, Barrientos-Muñoz KG, Barrios-Garrido H, Bjorndal KA, Bretón Vargas E, Broderick AC, Calderón Peña R, Carreras C, Ceriani SA, Colman LP, Cortés-Gómez AA, Crespo L, Cuevas E, Dah A, de Groene A, Delgado Trejo C, Demetropoulos S, Dias A, Diez C, Dos Santos NA, Dossou Bodjrenou JS, Early Capistrán MM, Eckert KL, Eizaguirre C, Ekanayake L, Escobedo Mondragón M, Esteban N, Feliciano D, Fernandes RS, Ferreira-Airaud B, Foley A, Fonseca LG, Forneiro Martín-Viaña Y, Fossette S, Fuentes M, Gaglo J, Gaos AR, Gidsicki D, Giffoni B, Girard A, Girondot M, Godfrey MH, Godley BJ, González Díaz Mirón RJ, Hamann M, Hancock JM, Hart CE, Hays GC, Herrera R, Hochscheid S, Hoekstra S, Huerta-Rodríguez P, Inteca G, Ishihara T, Jensen MP, Jribi I, Kale N, Kaska Y, Kelez S, Kelly IK, Köhnk S, Lara P, Lasfargue M, Lauritsen AM, Le Gouvello DZM, Liusamoa A, López M, López-Castro MC, Lopez-Mendilaharsu M, Louro CMM, Luna T, Madden CA, Mahabir D, Mancini A, Manoharakrishnan M, Marcovaldi

MA, Martínez-Portugal RC, Mastrogiacomo A, Matilde E, Mawunyo Adzagba B, Mbungu S, Miranda C, Moncada F, Morales-Mérida BA, Mortimer JA, Murakawa SKK, Nalovic MA, Nel R, Ngafack R, Nishizawa H, Ogou M, Panagopoulou A, Patricio AR, Peralta Buendía E, Phillott AD, Pilcher NJ, Polyak MMR, Prince RIT, Raynus EH, Reina RD, Rguez-Baron JM, Robbins AE, Santos AS, Sarti-Martínez AL, Schofield G, Seminoff JA, Serrano I, Shamblin BM, Shanker K, Stacy BA, Stahelin G, Staman MK, Stelfox M, Stewart KR, Taxonera A, Tucker AD, Turkozan O, van Dam RP, van de Geer CH, Viera S, West L, Whiting AU, Whiting SD, Wienand L, Wijntuin SR, Wildermann N, Zárate PM, Casale P, DiMatteo A, Hurley BJ, Hutchinson BJ, Maxwell SM, Posnik ZA, Rodriguez I, Mast RB (2025) Updated global conservation status and priorities for marine turtles. Endangered Species Research 56: 247-276.

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010a) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5: e15465.

Wallace BP, Lewison RL, McDonald SL, McDonald RK, Kot CY, Kelez S, Bjorkland RK, Finkbeiner EM, Helmbrecht Sr, Crowder LB (2010b) Global patterns of marine turtle bycatch. Conservation Letters 3: 131-142.

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. Bioinformatics 19: 671-672.

Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24: 1586-1591.

CHAPTER VI

Mitochondrial and mitogenomic analyses of *Eretmochelys imbricata* in the Gulf of Guinea

Introduction

According to the scientific community the *sixth mass extinction* is ongoing (Barnosky et al. 2011) and even relying on conservative estimates (Ceballos et al. 2015), the average rate of vertebrate species loss over the last century would be up to 100 times higher than the background rate of previous mass extinction. The magnitude of this phenomenon and its impact on ecosystems and human societies have nowadays been acknowledged as one of the most urgent challenges of our times (IPBES 2019).

To prevent species extinction through effective conservation actions, it is crucial to provide reliable assessments of risk and threats. This has become a collective effort, as in the context of the IUCN Red List of Threatened Species and in particular the Marine Turtle Specialist Group (MTSG). For appropriate sea turtles conservation strategies, the MTSG first developed the Regional Management Units (RMU) concept (Wallace et al. 2010a) and consequently a framework to evaluate the conservation status of all marine turtle RMUs using appropriate criteria and information: the 'conservation priorities portfolio' (CPP) (Wallace et al. 2011), which has been recently updated (Wallace et al. 2025) and shared as free online resource (https://www.seaturtlestatus.org/cpp-dashboard).

The hawksbill turtle (*Eretmochelys imbricata*) is one of the most historically impacted sea turtle species (Mortimer and Donnelly 2008), having suffered from intensive human exploitation. Nowadays, it is primarily affected by habitat degradation – particularly of coral reefs, which represent its preferred feeding ground. Although the overall conservation status of *E. imbricata* has improved according to the CPP (Wallace et al. 2025), assessments vary among different RMUs.

The Eastern Atlantic RMU is one of those in a critical condition, with low conservation capacity. Among the main risk factors are the low abundance of nesting females, the vulnerability of the rookeries and the reduced genetic diversity. All available population genetics data for the Eastern Atlantic have been generated using a 384 bp fragment of the mitochondrial DNA (mtDNA) control region (Monzón-Argüello et al. 2010; Monzón-Argüello et al. 2011). Sequence comparison revealed only a few haplotypes across two foraging grounds - Cape Verde and Principe - with just one fixed haplotype reported at the only described nesting site, Principe.

Longer mtDNA control region fragments (733 bp) have been used to characterize hawksbill populations globally (Arantes et al. 2020), offering higher resolution of genetic lineages than 384 bp fragments; however, the region of the Eastern Atlantic has remained

understudied and lacks information on genetic diversity and structure. A comprehensive investigation using the longer marker on a broad dataset was therefore needed to update our understanding of the genetic diversity in the Eastern Atlantic. This study aims to fill this knowledge gap for *E. imbricata* in the Gulf of Guinea.

Furthermore, since whole mitogenomic sequences proved promising in better resolving shared control region lineages in other sea turtle species (e.g. Tolve et al. 2024 and Chapter 5 of this thesis), we tested the potential of this technique in *E. imbricata*, exploring its usefulness for phylogeographic inference and the study of selective pressures. Overall, these data will help define future assessment for the species along the western African coast.

Materials and Methods

Control region analysis

Between 1998 and 2017, 293 tissue biopsies were collected from nesting, captured, bycaught and stranded hawksbill turtles in the Gulf of Guinea. DNA was extracted by a standard protocol using phenol chloroform. A ~ 733 bp fragment of the mtDNA control region was characterized using primers LTEi9 and H950 (Abreu et al. 2006) and following the protocol by Vilaça et al. (2022). Haplotypes were assigned based on the review by Arantes et al. (2020). These data were produced before this PhD project.

In order to visualize the haplotypes from our dataset along with the control region haplotypes described in the literature, a network was constructed, using the comprehensive haplotypes dataset in Arantes et al. (2020). We used statistical parsimony (Templeton et al. 1992) implemented in TCS (Clement et al. 2000). The network was visualized in tcsBU (Múrias dos Santos et al. 2015), exported as svg file and graphically edited in Inkscape (https://inkscape.org/).

We pooled all our samples collected from turtles ‘at sea’ (i.e. not nesting) into a single mixed stock. We used Arlequin 3.5.2.2 (Excoffier and Lischer 2010) to compute the number of haplotypes, as well as haplotype and nucleotide diversity for the mixed stock. We also compared the genetic diversity of our mixed stock with that of the other foraging grounds reported in Arantes et al. (2020), using both genetic diversity indices and pairwise F_{ST} values with 1000 permutations.

We conducted two foraging ground centric Mixed Stock Analyses (MSA) on our Gulf of Guinea mixed stock, using the R package mixstock (Bolker et al. 2007). In the first

analysis, we used data from Atlantic and Indo – Pacific rookeries reported in Arantes et al. (2020) as the baseline. In the second analysis, we grouped the samples from scattered nesting sites in our dataset into a single potential source population (Gulf of Guinea) and added this pool to the baseline. Our aim was to evaluate the effect of including or not our nesting samples from the Gulf of Guinea on the MSA estimates. As a prior, we used rookery size (i.e. number of clutches per year). For rookeries lacking available size data, we applied the average size estimated from those with available data (see Table S1). For the Gulf of Guinea, we used the estimate available for the Principe rookery (Monzón-Argüello et al. 2011), acknowledging that this likely underestimates the total number of clutches for the entire region. For both analyses we set the number of iterations to 400,000 with a 50% burn-in and assessed convergence of the MCMC chains using the Gelman and Rubin diagnostic (shrink factor < 1.2).

Mitogenome sequencing and assembly

We selected 10 samples from the dataset, based on good DNA quality and quantity, assessed by 1% gel electrophoresis and Qubit 4 fluorometer Broad Range Assay (Invitrogen). Dual-indexed (i5 and i7) genomic libraries were built using a PCR-free Tagmentation Kit Illumina according to the manufacturer protocol. We envisaged a 1.8 X target nuclear genome coverage for 5 of these samples and 18X for the other five samples that were sequenced separately. Consequently, the target output was ~ 3.96 Gb and ~ 39.6 Gb per sample, respectively. We sequenced the libraries paired-end on an Illumina NovaSeq 6000 System using a 300-cycle NovaSeq 6000 SP Reagent Kit v 1.5.

Sequencing data were handled in the same way as in chapter 5 of this thesis. In brief, after demultiplexing, reads quality check and trimming, mitogenome assembly was achieved through two different pipelines in parallel, one based on mapping on a reference mitogenome, and one based on *de novo* assembly using a seed-and-extend algorithm. As reference mitogenome we used the annotated sequence *E. imbricata* NC012398. For each sample we compared the mitogenome sequence obtained by the mapping with the mitogenome sequence obtained by NOVOPlasty, and in case of discrepancies, we clarified the correct reconstruction checking the coverage of the polymorphic sites using the sorted.bam. Mitogenomic haplotype names were assigned using the first code which refers to the respective ~ 733 bp control region haplotype and a serial number referring to the variants outside the control region (e.g. EiA70.1 would be a mitogenomic variant of control region haplotype EiA70).

We searched online databases for all the published mitogenomes of *E. imbricata*. We found a total of seven different samples for which the complete mitogenome had been reconstructed. However, two of these sequences (NC012398 and MF571906) lacked information on sampling location; for this reason, we considered them potentially misleading, and we excluded them from our analyses. The resulting mitogenome database is reported in Table 2.

Mitogenomic diversity and phylogenetic analyses

Downloaded mitogenomes were aligned with sequenced mitogenomes and trimmed at position 16,271, removing N bases and the microsatellite from the following analyses. The alignment of mitogenomic sequences was used to construct a mitogenomic haplotype network with TCS (Clement et al. 2000), allowing an infinite number of changes to link the haplogroups. As for the dloop, the network was visualized in tcsBU (Múrias dos Santos et al. 2015), exported as svg file and graphically edited in Inkscape (https://inkscape.org/). The mitogenomic alignment was then partitioned in single genes alignments using Geneious Prime 2022.2.2, and genetic diversity indices were computed for each gene using DnaSP (Librado and Rozas 2009).

For phylogenetic analyses, we included two *Caretta caretta* mitogenome sequences (NC016923 and JX454988) as outgroup in the alignment. First, we assessed the best molecular evolution model, and we constructed a Maximum Likelihood (ML) phylogeny using IqTree (Trifinopoulos et al. 2016). Statistical support for the nodes was assessed by ultrafast bootstrap (values > 95%). FigTree v1.4.4 (http:// tree. bio. ed. ac. uk/ software/ figtr ee/) was used to visualize the tree. Then, we performed a Bayesian phylogenetic analysis using BEAST suite v.1.10.4 (Suchard et al. 2018b) and four partitions as in chapter 5 of this thesis, setting the same options for the analysis. As priors, we forced monophyly of *E. imbricata* and we set the time of divergence between the branches *C. caretta* + *Lepidochelys* and *E. imbricata*, that we extrapolated by Duchene et al. (2012), around 32.5 MYBP (95% HPD, 23.5 – 43.5). We performed 5 independent runs, 10^8 steps each with a 10% burn-in. We used Tracer v.1.7.1 (Rambaut et al. 2018) to monitor ESS for each output parameter, Log Combiner to merge the runs, Tree Annotator (Drummond and Rambaut 2007) to summarise the overall output, and FigTree v1.4.4 (http:// tree. bio. ed. ac. uk/ software/ figtr ee/) to visualize the tree.

Selective pressure analyses

We tested deviation from neutrality comparing *E. imbricata* mitogenomic coding partition with the outgroup sequences of *C. caretta* (NC016923 and JX454988). We performed a McDonald-Kreitman test on each gene using DnaSP (Librado and Rozas 2009). For details about the computation of the Neutrality Index, see *Selective pressure analysis* in Chapter 5.

To test specifically the presence of selective pressure within *E. imbricata* mitogenomic coding partition, we used codeml package implemented in PAML (Yang 2007) with the ML phylogeny previously generated. As in Chapter 5, we applied the Random Site Method, then the Branch Method and the Branch Site Method, setting *E. imbricata* as foreground branch (see this section in Chapter 5 for details about the tests).

Results

Control region analyses

Within our dataset we found nine control region haplotypes (Table 1 and Fig 1). The most abundant one in our dataset was EiA48, also known in literature as EiIP16, which represents 62% of the total samples. The second most abundant haplotype was EiA70 (27% of the total). The other recorded haplotypes all represent minor percentages (3% or below). Among the rare haplotypes, we found two new haplotypes, that we called EiA50 and EiIP134, respectively. Haplotypes EiA48, EiA50, EiA70 were recorded not only in turtles at sea, but also in samples from nests. All haplotypes but EiA01, EiA32 and the new one EiIP134 belong to the II Indo Pacific clade, as defined in Arantes et al. (2020); EiA01 and EiA32 belong to the I Atlantic clade, while EiIP134 belongs to the I Indo Pacific clade (Fig 1).

Comparing our mixed stock from the Gulf of Guinea with the other foraging grounds reported in literature (Arantes et al. 2020), we found a higher value of haplotypic diversity for the Gulf of Guinea compared to the other Eastern Atlantic foraging ground for which this information is available (Ascension Island) (Table 2), although lower than other foraging grounds in the other oceanic basins. High and significant values of F_{ST} were recorded for each comparison involving the Gulf of Guinea with another foraging ground, with the only exception of Gulf of Guinea vs Principe and United States (USA) (not significant).

The results of the MSA differed consistently between the two analyses (Fig 2). The first run, the one with no samples from nesting in the Gulf of Guinea, gave very wide

confidence intervals spanning from zero to 100% for all the non-null contribution. Conversely, nearly all of the contribution to the stock (97%, ci 94 – 99%) was assigned to the nesting sites in the Gulf of Guinea, when our samples were included within the baseline, and null contribution was assigned to the other rookeries.

Table 1. Control region haplotypes (~ 733 bp fragment) for samples of *Eretmochelys imbricata* collected in the Gulf of Nigeria/Western Africa. In parentheses, number of samples from nesting. EiA48 is also reported as EiIP16. Haplotypes marked by § are previously undescribed control region haplotype.

Haplotype	Liberia	Benin	Continental Equatorial Guinea	Corisco Bay	Principe	Sao Tome	Gabon	TOT
EiA01	1	0	1	2	2	1	1	**8**
EiA32	0	0	1	1	0	0	0	**2**
EiA48*	1	0	27	86	22 (1)	35 (1)	10 (2)	**181 (4)**
EiA49	0	0	2	6	0	0	1	**9**
EiA50§	0	0	3	3	1	1 (1)	0	**8 (1)**
EiA70	0	1	11	35	26 (2)	6 (1)	1	**80 (3)**
EiA75	0	0	0	1	0	0	1	**2**
EiA87	0	0	0	1	0	0	0	**1**
EiIP134§	0	0	0	0	2	0	0	**2**
TOT	**2**	**1**	**45**	**135**	**53 (3)**	**43 (2)**	**14 (2)**	**293 (8)**

Fig 1. Control region haplotype network, comprehensive of all published haplotypes. Haplotypes recorded in samples sequenced in this study are shown in red. Haplotypes only recorded in samples from already published mitogenomes are shown in brown. Haplogroup names as in Arantes et al. (2020).

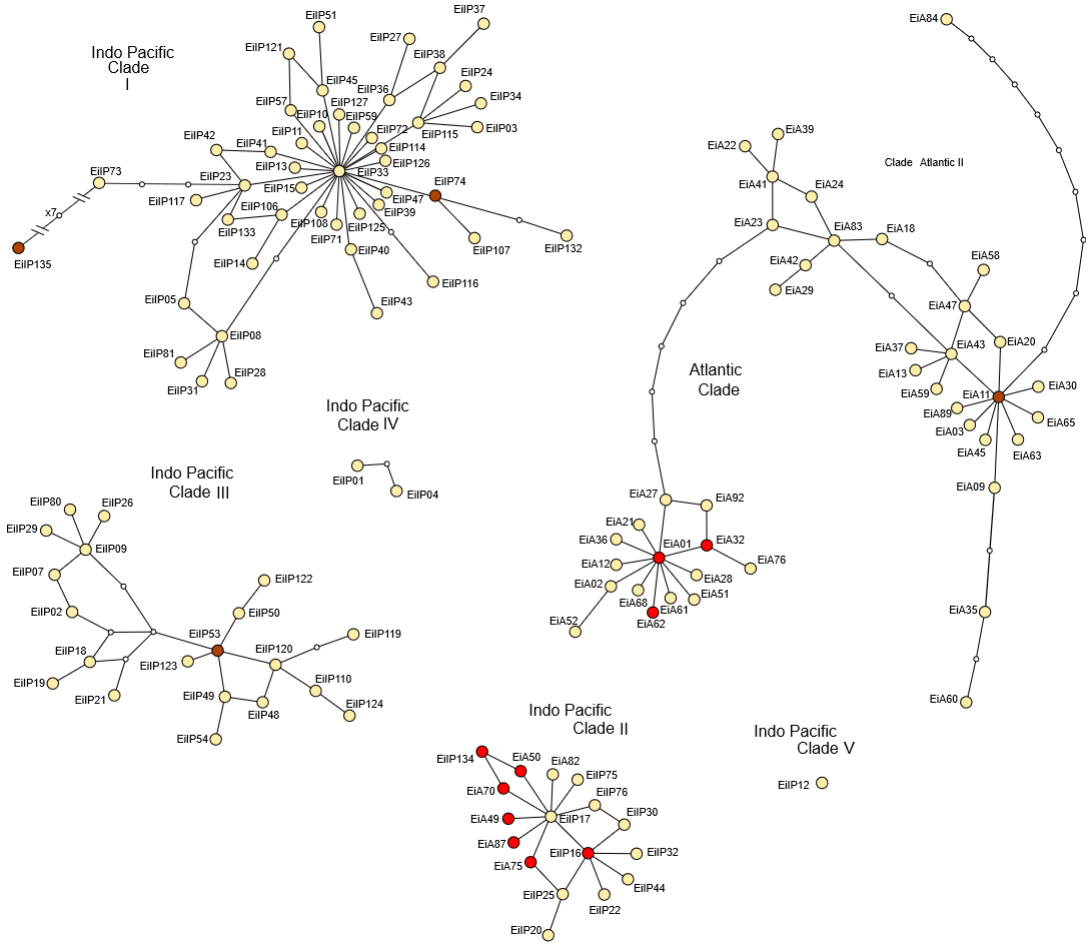


Table 2. Genetic diversity of *E. imbricata* foraging grounds reported in Arantes et al. (2020). *N*, number of samples; *K*, number of control region haplotypes (733 bp); *h*, haplotypic diversity; π, nucleotide diversity; SE, Standard Error.

Oceanic area	Locality	Abbreviation	*N*	*K*	*h* ± SE	π ± SE	Reference
Eastern Atlantic	Gulf of Guinea	GG	285	9	0.544 ± 0.002	0.0035 ± 0.0001	This study
	Ascension Island	AS	18	4	0.399 ± 0.040	0.0120 ± 0.0000	(Putman et al. 2014)
	Principe	PI	3	1	0.000 ± 0.000	0.0000 ± 0.0000	(Monzón-Argüello et al. 2011)
Northwest Atlantic	Florida, Key West	FL	50	12	0.770 ± 0.012	0.0061 ± 0.0002	(Gorham et al. 2014)
	Florida, Palm Beach County	PB	106	16	0.761 ± 0.006	0.0046 ± 0.0001	(Wood et al. 2013)
	Mexico, Punta Xen	PX	43	7	0.756 ± 0.014	0.0029 ± 0.0007	(Labastida-Estrada et al. 2018)
	Mexico, Isla Contoy	IC	7	3	0.667 ± 0.103	0.0085 ± 0.0001	(Labastida-Estrada et al. 2018)
	Mexico, Cozumel	CZ	37	9	0.634 ± 0.019	0.0069 ± 0.0000	(Labastida-Estrada et al. 2018)
	Mexico, Banco Chinchorro	MexBC	34	6	0.608 ± 0.021	0.0069 ± 0.0000	(Labastida-Estrada et al. 2018)
	Mexico, Xcalak	MexXC	2	2	1.000 ± 0.000	0.0179 ± 0.0002	(Labastida-Estrada et al. 2018)
	Colombia, Atlantic Coast	CO	6	5	0.933 ± 0.064	0.0091 ± 0.0002	(Trujillo-Arias et al. 2014)
	Jardines del Rey (Cuba)	CU	59	14	0.825 ± 0.009	0.0078 ± 0.0001	(Pérez-Bermúdez et al. 2017)
	Puerto Rico	PR	118	20	0.757 ± 0.005	0.0081 ± 0.0001	(Velez-Zuazo et al. 2008)
	Turks and Caicos	TC	15	4	0.371 ± 0.047	0.0026 ± 0.0000	(Richardson et al. 2009)
	Cayman Islands	CA	90	10	0.707 ± 0.007	0.0077 ±	(Blumenthal et al. 2009)

						0.0002	
	Tobago, Leeward Coast	ToLe	17	7	0.838 ± 0.032	0.0083 ± 0.0001	(Cazabon-Mannette et al. 2016)
	Tobago, Windward Coast	ToWi	47	8	0.588 ± 0.015	0.0065 ± 0.0001	(Cazabon-Mannette et al. 2016)
Southwest Atlantic	Brazil, Fernando de Noronha	NA	94	14	0.516 ± 0.008	0.0103 ± 0.0001	(Vilaca et al. 2012)
	Brazil, Sao Pedro Sao Paulo	SPSP	12	4	0.636 ± 0.059	0.0026 ± 0.0000	(Proietti et al. 2014)
	Brazil, Bahia	BA	32	5	0.434 ± 0.022	0.0022 ± 0.0000	(Proietti et al. 2014)
	Brazil, Cearà	CE	22	3	0.178 ± 0.025	0.0015 ± 0.0002	(Proietti et al. 2014)
	Brazil, Abrolhos Park	AB	65	5	0.203 ± 0.009	0.0010 ± 0.0001	(Proietti et al. 2014)
	Brazil, South	SO	22	2	0.368 ± 0.032	0.0005 ± 0.0001	(Proietti et al. 2014)
Indo-Pacific	Tun Sakaran Marine Park	TSMP	6	6	1.000 ± 0.000	0.0214 ± 0.0001	(Nishizawa et al. 2016)
	Peninsular Malaysia	PM	8	3	0.464 ± 0.094	0.0171 ± 0.0001	(Nishizawa et al. 2016)
	Howick Group	HG	91	15	0.618 ± 0.008	0.0062 ± 0.0002	(Bell and Jensen 2018)
	Palau Tiga	PT	2	2	1.000 ± 0.000	0.0041 ± 0.0000	(Nishizawa et al. 2016)
	Palau Sipadan	PS	5	5	1.000 ± 0.000	0.0027 ± 0.0001	(Nishizawa et al. 2016)
Eastern Pacific	Guatemala, Pacific Coast	GU	5	3	0.800 ± 0.126	0.0022 ± 0.0000	(Gaos et al. 2018)
	Panama, Pacific Coast	PA	68	6	0.580 ± 0.010	0.0012 ± 0.0000	(Gaos et al. 2018)
	El Salvador	ES	136	8	0.661 ± 0.005	0.0012 ± 0.0000	(Gaos et al. 2018)
	Nicaragua, Pacific Coast	NI	118	5	0.654 ± 0.006	0.0011 ±	(Gaos et al. 2018)

						0.0000	
	Mexico, Pacific Coast	MP	94	7	0.588 ± 0.007	0.0010 ± 0.0000	(Zuñiga-Marroquin and Espinosa de los Monteros 2017)
	Colombia, Pacific Coast	CP	9	3	0.556 ± 0.083	0.0010 ± 0.0000	(Trujillo-Arias et al. 2014)
	Ecuador	EC	64	5	0.407 ± 0.011	0.0006 ± 0.0000	(Gaos et al. 2018)
	Costa Rica, Pacific Coast	CR	62	4	0.333 ± 0.011	0.0006 ± 0.0000	(Gaos et al. 2018)
	USA, Pacific Coast	USA	1	1	1.000 ± 0.000	0.0000 ± 0.0000	(Gaos et al. 2018)
	Perù	PE	4	1	0.000 ± 0.000	0.0000 ± 0.0000	(Gaos et al. 2018)

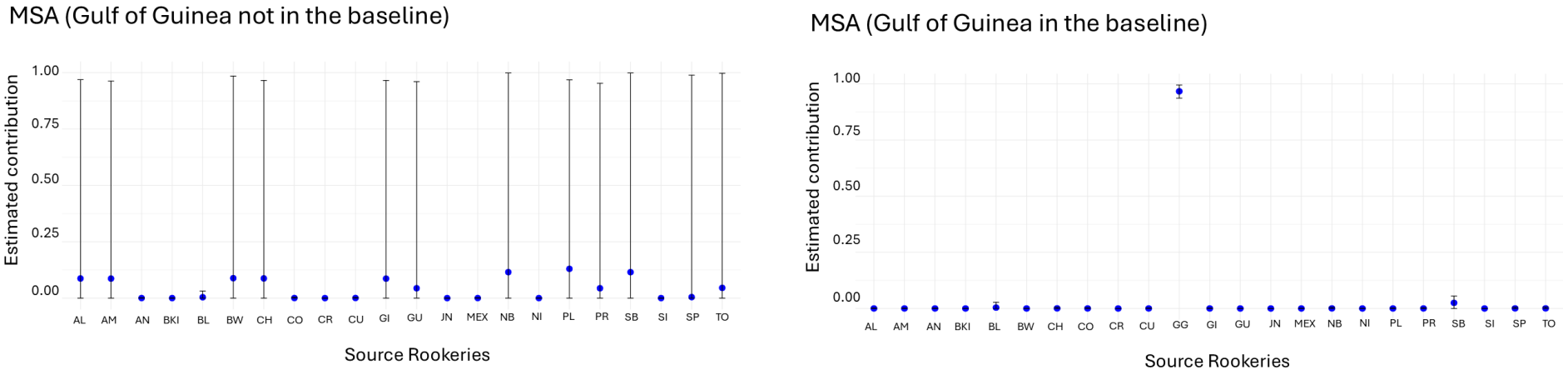


Fig 2. Estimated contributions of Atlantic and Indo Pacific rookeries to the Gulf of Guinea mixed stock. On the left, analysis excluding Gulf of Guinea from the baseline. On the right, analysis including Gulf of Guinea in the baseline. Acronyms as in Table S1.

Table 3. *E. imbricata* mitogenomic sequences that we used in this study. EiA48 is also known as EiIP16. In §, a previously undescribed control region haplotype.

ID sample/ Accession Number	Reference	Target nuclear coverage (X)	Output nuclear coverage (X)	Output mitogenomic coverage (X)	Sample provenance	Ocean	Nesting vs at sea	Control region haplotype	MtGenome haplotype
8025733	This study	1.8	1.43	73	Principe	Atlantic	Nesting	EiA70	EiA70.1
8025738	This study	1.8	8.47	1,309	Principe	Atlantic	Nesting	EiA70	EiA70.1
8026680	This study	1.8	12.85	568	Principe	Atlantic	At sea	EiA48*	EiA48.3
8028257	This study	1.8	2.30	248	Principe	Atlantic	At sea	EiA48*	EiA48.1
8029500	This study	1.8	1.75	180	Principe	Atlantic	At sea	EiA48*	EiA48.2
M123	This study	18	15.31	774	Sao Tomé	Atlantic	Nesting	EiA50	EiA50.1
M136	This study	18	9.19	380	Sao Tomé	Atlantic	At sea	EiA48*	EiA48.1
M167	This study	18	16.83	2,503	Sao Tomé	Atlantic	At sea	EiA48*	EiA48.1
M168	This study	18	17.15	2,141	Sao Tomé	Atlantic	At sea	EiA48*	EiA48.1
M169	This study	18	20.03	2,692	Sao Tomé	Atlantic	At sea	EiA48*	EiA48.1
JX454970	(Duchene et al. 2012)	-	-	-	Hawaii	Pacific	?	EiIP74	EiIP74.1
JX454980	(Duchene et al. 2012)	-	-	-	Singapore	Pacific	?	EiIP53	EiIP53.1
JX454986	(Duchene et al. 2012)	-	-	-	Costa Rica	Atlantic	?	EiA11	EiA11.1
KP221806	Unpublished	-	-	-	Colombia	Atlantic	?	EiIP135§	EiIP135.1
OZ223910	(Rhie et al. 2021)	-	-	-	Hawaii	Pacific	?	EiIP74	EiIP74.2

Mitogenomic analysis and genetic diversity

For some samples, the output nuclear coverage differed from the target nuclear coverage; however, in all cases it ensured a high mitogenomic coverage, with minimum and maximum values of 73X and 2,692X, respectively (Table 3). For all assembled sequences, a few discrepancies were observed at polymorphic sites reported by NOVOPlasty and the mapping pipeline. In all such cases, visual inspection of the bam files confirmed the most supported variant (see also Chapter 5 of this thesis).

Among 10 assembled mitogenomes, we identified five different mitogenomic haplotypes; three of these were variants of the most common control region haplotype, EiA48 (Table 3). Additionally, we assigned corresponding control region and mitogenomic haplotypes to the sequences retrieved from databases. Among these, we found two mitogenomic variants of haplotype EiIP74, one mitogenomic variant of EiIP53 and one mitogenomic variant of EiA11. The sequence KP221806 from Atlantic Colombia carried a previously unreported control region haplotype belonging to I Indo Pacific Clade, that we designated EiIP135, along with its mitogenomic variant, EiIP135.1. Unexpectedly, this sample was collected in Atlantic Colombia.

The mitogenomic haplotype network (Fig3) mostly resembles the control region haplotype network (Fig1). Five clusters can be identified: one corresponding to the haplotypes recorded in our samples (EiA48.1, EiA48.2, EiA48.3, EiA50.1, EiA70.1), which fall within the control region II Indo Pacific Clade; a second cluster for EiIP53.1, corresponding to the control region III Indo Pacific Clade; a third cluster including the two mitogenomic haplotypes EiIP74.1 and EiIP74.2, corresponding to the control region I Indo Pacific Clade. EiIP135.1 forms a separate cluster, located 65 substitutions away from the EiIP74 variants. In this case, the distances between mitogenomic clusters are greater than those between the corresponding control region clusters; however, the closest haplogroup to EiIP135.1 remains that including EiIP74.1 and EiIP74.2. The fifth cluster includes EiA11.1, which corresponds to the control region I Atlantic Clade and is the most divergent from all other clusters.

Looking at the single gene diversity (Table 4), the control region is confirmed to be the most variable one, even if the same number of haplotype (k = 7) was recorded within COXI gene, along with similar values of haplotypic diversity. All dN/dS ratios were below 1, consistent with purifying selection, although values varied considerably among genes, with ATP8 and ND3 showing the highest ratios -indicating a relatively elevated rate of non-synonymous substitutions in these regions.

Table 4 Information and diversity for single mitochondrial genes and tRNA partition of the mitogenomic alignment. Sequence length in base pairs (bp). Number of haplotypes (*k*), number of polymorphic sites (*S*), haplotype (*h*) and nucleotide diversity (π) with their Standard Error (SE), and dN/dS ratios for coding genes.

mtDNA gene	Sequence length (bp)	*k*	*S*	*h* ± SE	π ± SE	dN/dS
12S	971	4	8	0.467 ± 0.049	0.0015 ± 0.0000	-
16S	1620	5	13	0.562 ± 0.048	0.0017 ± 0.0000	-
ATP6	683	4	12	0.543 ± 0.048	0.0035 ± 0.0000	0.0714
ATP8	186	2	3	0.113 ± 0.033	0.0021 ± 0.0001	0.5670
COXI	1548	7	27	0.724 ± 0.044	0.0036 ± 0.0000	0.1260
COXII	690	3	5	0.448 ± 0.049	0.0016 ± 0.0000	0
COXIII	784	2	8	0.133 ± 0.033	0.0014 ± 0.0003	0.0672
CYTB	1144	6	19	0.705 ± 0.044	0.0034 ± 0.0000	0.1224
ND1	974	4	15	0.553 ± 0.048	0.0050 ± 0.0000	0.1279
ND2	1048	4	26	0.552 ± 0.048	0.0068 ± 0.0000	0.1032
ND3	350	3	8	0.448 ± 0.049	0.0037 ± 0.0002	0.5058
ND4L	297	3	5	0.457 ± 0.049	0.0047 ± 0.0000	0
ND4	1384	6	30	0.714 ± 0.044	0.0040 ± 0.0000	0.1726
ND5	1803	5	34	0.562 ± 0.048	0.0045 ± 0.0000	0.2851
ND6	526	5	6	0.562 ± 0.048	0.0028 ± 0.0000	0.1527
tRNAs	1562	5	19	0.476 ± 0.049	0.0026 ± 0.0000	-
Control region	733	7	49	0.781 ± 0.040	0.0194 ± 0.0002	-

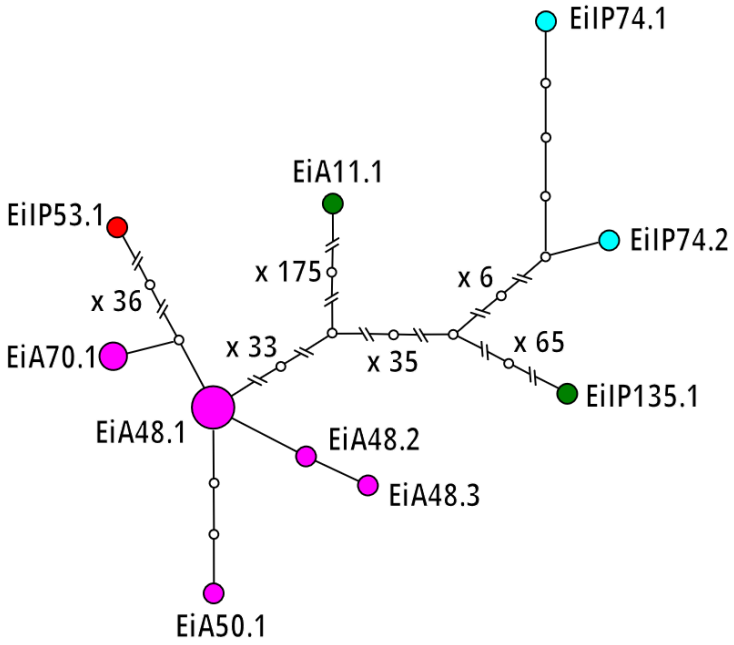


Fig 3. TCS network for mitogenomic haplotypes retrieved in our dataset (Table 3). Circle size is proportional to the number of sequences carrying the haplotype. In purple, haplotypes sampled in the Gulf

of Guinea; in red, haplotype sampled in Indo Pacific Ocean; in green, haplotype sampled in Western Atlantic; in light blue, haplotypes sampled in Eastern Pacific.

Phylogenetic analyses

The best-fitting molecular substitution model for the mitogenomic alignment of *E.imbricata* and the outgroup sequences of *C. caretta* was GTR + G4, which was used to construct the Maximum Likelihood (ML) phylogeny (Fig. 4). In the resulting topology, the deepest and well supported divergence within *E.imbricata* is between EiIA11.1 and other mitogenomic haplotypes. This confirms a primary split between the Atlantic Clade (represented here by a single haplotype) and the Indo Pacific Clades (which include all the remaining haplotypes). The second major node within *E.imbricata* separates EiIP74 variants and EiIP135.1 (corresponding to the control region I Indo Pacific Clade) from the other haplotypes (control region II Indo Pacific Clade). Therefore, the ML tree confirms the placement of the novel haplotype EiIP135.1 within the EiIP74 cluster. All these nodes are highly supported, while the branching pattern within the cluster of haplotypes from our samples (EiA48.1, EiA48.2, EiA48.3, EiA50, EiA70) is less clearly resolved.

The Bayesian phylogeny (Fig5) confirms the ML topology; all the main nodes are strongly supported, and the clustering of the haplotypes is the same as previously described. In Table 5 we report the time estimates and respective 95% HPD for the supported nodes.

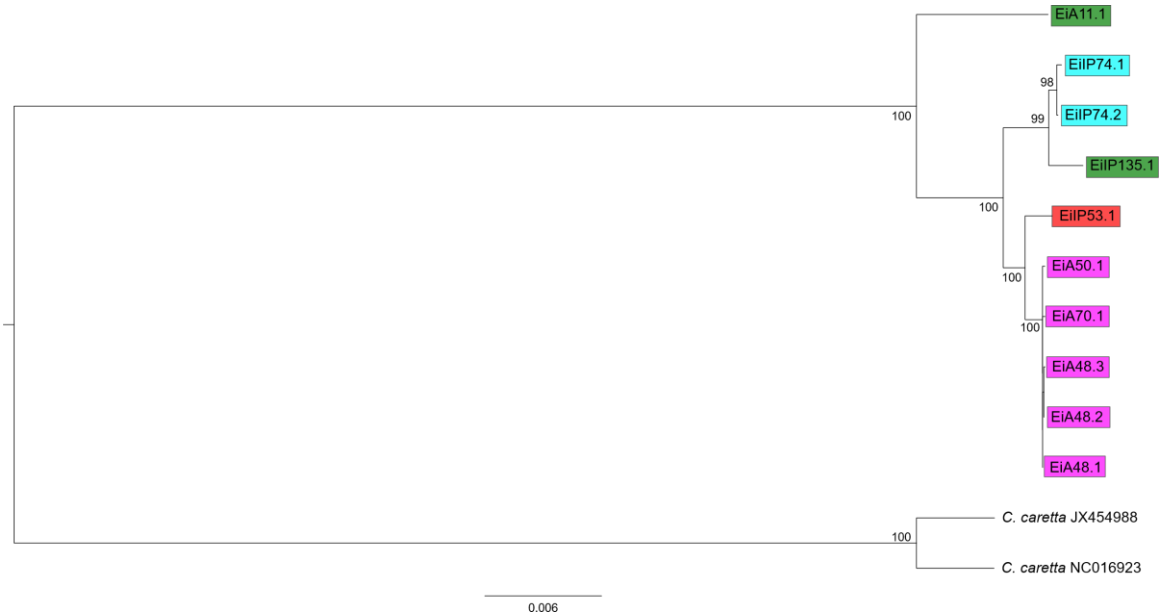


Fig 4. Maximum Likelihood phylogeny for complete mitogenomes dataset (Table 3). Values of ultrafast bootstrap are reported for each node. The root does not have a value of statistical support because it was imposed by the user. Branch length is in number of substitutions per site.

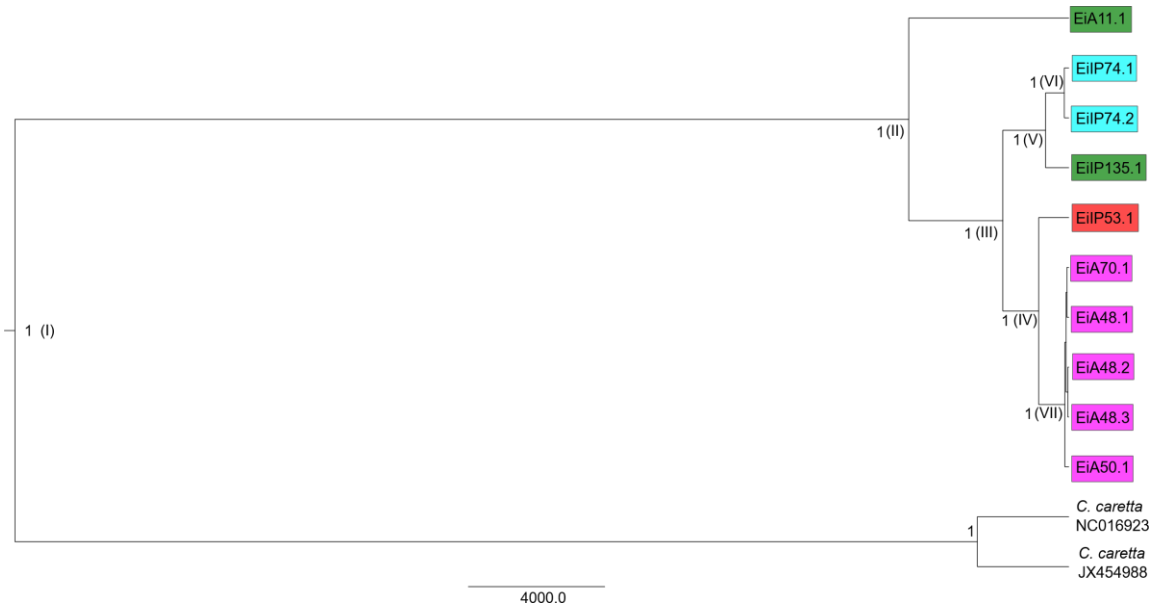


Fig 5. Bayesian phylogeny for complete mitogenomes dataset (Table 3). Names of the nodes are in roman numbers (as in Table 5). Values of posterior probability are reported for each node. Branch length is in thousands of years by present (KYBP).

Table 5 Bayesian tree's height estimates and 95% Height Posterior Density (HPD) for nodes shown in Fig 5. Estimates are reported in thousands of years by present (KYBP). Names of the clades follow Arantes et al. (2020) as in the text.

Node	Node description	Median (KYBP)	95% HPD
I	*C. caretta* – *E. imbricata*	30,884	16,908 – 44,679
II	I Atlantic – Indo Pacific clades	4,703	2,552 – 7,000
III	I Indo Pacific – other Indo Pacific clades	1,947	997 – 2,900
IV	II Indo Pacific – III Indo Pacific	895	435 – 1,396
V	EiIP135.1 – other I Indo Pacific	691	329 – 1,081
VI	EiIP74.1 – EiIP74.2	141	30 – 273
VII	EiA50.1 – other II Indo Pacific	121	35 - 222

Selective pressure analyses

The McDonald – Kreitman test yielded non-significant results for all genes except ND5, which showed a p-value of 0.01 and a Neutrality Index of 2.458, indicating purifying selection.

Tables XXX present the results of selective pressure analyses performed with Codeml on the coding partition. The Random Site Method produced significant results when comparing nested models M0, M1a, M2a, but not for M8 versus M7. Despite the significancy of the likelihood ratio test (LRT), no candidate sites under positive selection were identified by the Bayes Empirical Bayes (BEB). In both the Branch Method and the Branch Site Method, the value of omega on the foreground branch (i.e. *E. imbricata*) was significantly different from that on the background branch (*C. caretta*), although lower than 1 (i.e. purifying selection). Consistently, no positively selected sites were identified in either approach.

Table 6. Codeml results for Random Site Method applied through nested models: null model, M0; alternative models, M1a, M2a. lnL: logarithm of Likelihood, np: number of parameters; ω_0 indicates purifying selection, ω_1 indicates neutral selection, ω_2 indicates positive selection; p_0: proportion of sites with ω_0, p_1: proportion of sites with ω_1, p_2: proportion of sites with ω_2. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. For M1a, LRT = 2(lnLM1a – lnLM0). For M2a, LRT = 2(lnLM2a – lnLM1a). The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.841. The critical value for LRT statistic, with 2 degrees of freedom and p = 0.005, is 5.99. Significant results are indicated by asterisks.

model	lnL	np	ω_0	ω_1	ω_2	p_0	p_1	p_2	LRT	d.f.
M0	-19,352.7695	28	0.3256	-	-	1	-	-	-	-
M1a	-19,310.4196	29	0	1	-	0.6393	0.3607	-	84.70*	1
M2a	-19,308.3010	31	0.1443	1	2.4872	0.8543	0.0661	0.0796	88.94*	2

Table 7. Codeml results for Random Site Method applied through nested models: null model, M7 and alternative model M8. lnL: logarithm of Likelihood, np: number of parameters, p and q: parameters of beta distribution describing ω for sites in the range $0 \leq \omega \leq 1$. ω_2 indicates positive selection, p_0: proportion of sites with ω from beta (p, q) as in M7, p_1: proportion of sites with ω_2. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. Here, LRT = 2(lnLM8 – lnLM7). The critical value for LRT statistic, with 2 degrees of freedom and p = 0.005, is 5.99. Significant results are indicated by asterisks.

model	lnL	np	p	q	ω_2	p_0	p_1	LRT	d.f.
M7	-19,310.6745	29	0.0055	0.0103	-	-	-	-	-
M8	-19,308.2978	31	0.9595	3.9796	2.4550	0.9130	0.0870	4.7533	2

Table 8. Codeml results for Branch Method applied through models M0 and MBr1, considering *E. imbricata* as foreground branch. lnL: logarithm of Likelihood, np: number of parameters; ω indicates purifying selection (homogeneous across all the sites and the tree); ω_b indicates the value for selection on the background branch; ω_f indicates the value for selection on the foreground branch. LRT: Likelihood Ratio Test, d.f.: degrees of freedom. Here, LRT = 2(lnLMBr1 – lnLM0). The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.84. Significant results are indicated by asterisks.

model	lnL	np	ω	ω_b	ω_f	LRT	d.f.
M0	-19,352.7693	28	0.3255	-	-	-	-
MBr1	-19,342.8170	29	-	0.4618	0.0855	19.905*	1

Table 9. Codeml results for Branch Site Method applied through nested models M0, M2a_f and MBr1_2s, considering *E. imbricata* as foreground branch. lnL: logarithm of Likelihood, np: number of parameters; ω_0 indicates purifying selection, ω_1 indicates neutral selection, ω_2 indicates positive selection; p_0: proportion of sites with ω_0, p_1: proportion of sites with ω_1, p_{2a}: proportion of sites with ω_0 on the background branch and ω_2 on the foreground branch, p_{2b}: proportion of sites with ω_1 on the background branch and ω_2 on the foreground branch. Here, LRT is computed comparing MBr1_2s first with M0: LRT = 2(lnLMBr1_2s – lnLM0); then, with M2a_f: LRT = 2(lnLMBr1 – lnLM2a_f). The critical value for LRT statistic, with 3 degrees of freedom and p = 0.005, is 7.81. The critical value for LRT statistic, with 1 degree of freedom and p = 0.005, is 3.84. Significant results are indicated by asterisks.

model	lnL	np	ω_0	ω_1	ω_2	p_0	p_1	p_{2a}	p_{2b}	LRT	d.f.
M0	-19,352.769	28	0.325	-	-	-	-	-	-	-	-
M2a_f	-19,310.419	30	0	-	1	-	-	-	-	-	-
MBr1_2s	-19,310.419	31	0	1	1	0.616	0.348	0.022	0.012	84.70*	3
										0.00	1

Discussion

This study provides an unprecedentedly broad dataset of 733 bp-long control region haplotypes from samples collected in the Gulf of Guinea, including both at-sea turtles and nesting individuals from previously undocumented locations. These data fill a significant knowledge gap for *E. imbricata*, as the lack of genetic information generated using this longer mtDNA marker has so far hindered the integration of the Eastern Atlantic into a comprehensive phylogeographic framework for the species. Additionally, we applied mitogenomic sequencing to a subsample of turtles to evaluate its potential for resolving widespread haplotypes and to gain further phylogeographic insights into this species at global level.

Control region analyses

Among the turtles sampled at sea in the Gulf of Guinea, we recorded haplotypes belonging to both the I Atlantic clade and the II Indo Pacific clade, although the latter was largely predominant. The presence of this phylogenetic clade in the Eastern Atlantic was already observed in the foraging grounds of Cape Verde (Monzón-Argüello et al. 2010) and Principe Island (Monzón-Argüello et al. 2011), where it was mostly represented by the EATL haplotype (384 bp in length). According to those studies, haplotype EATL was fixed in the Cape Verde foraging ground while only a few individuals in the Principe Island foraging ground carried other haplotypes (EiA87, II Indo Pacific clade, and A, I Atlantic clade). Haplotypes EiA70 and EiA48 (733 bp in length), which constitute the majority of samples analysed in this study for the Gulf of Guinea, actually include the shorter EATL sequence. Therefore, our findings support previous results from these foraging grounds, while also providing greater resolution of the EATL genetic lineage. Among the haplotypes that we found in turtles at sea, EiA48 was previously found both in foraging grounds and in rookeries of the Indo-Pacific Ocean (Arantes et al. 2020). However, additional haplotypes were previously recorded in other foraging grounds of the Indian Ocean, but not in any known rookery. These were EiA70, very abundant in our stock, and EiA49, EiA75, EiA87, which we found in minor percentage. Moreover, we identified two completely new haplotypes in the stock of the Gulf of Guinea, EiA50 and EiIP134, both belonging to II Indo Pacific clade. This indicates that many rookeries have been under-sampled so far, or even not discovered.

Our dataset also included samples from nesting turtles in Principe and Sao Tomé Islands and in Gabon. While Monzón-Arguello et al. (2011) described some nesting females in Principe with the 384 bp haplotypes (EATL was fixed), we provided the first genetic dataset from nesting in Sao Tomé and Gabon. Among these samples from nesting turtles, we reported exclusively the II Indo Pacific clade, represented by EiA48 and EiA70 in Principe Island, EiA48, EiA50 and EiA70 in Sao Tomé, EiA48 in Gabon. All these findings are somehow new; EiA48 has never been recorded in Atlantic nests, and EiA50 and EiA70 have never been recorded in any nest at all. Until other nesting sites will be characterized by those haplotypes, EiA50 and EiA70 are potentially exclusive haplotypes of Sao Tomé and of Sao Tomé – Principe, respectively, and this is of crucial importance for genetic assignment of turtles at sea.

Thanks to our data, we performed the first MSA in the Gulf of Guinea using a 733 bp-long control region haplotype. Although the Gulf of Guinea cannot be considered a true rookery, such as those usually included in MSA baselines, we refer to the Gulf of Guinea as a broad source region. Including or excluding these nesting data in our MSA significantly altered the results: without the Gulf of Guinea in the baseline, our analysis was uninformative, indicating many potential source populations with contributions ranging from zero to 100%. In contrast, when the Gulf of Guinea was included, nearly all individuals in the mixed stock were assigned to it, with a narrow confidence interval. Of course, comprehensive sampling of rookeries in the region is still needed and will improve the resolution of future assignments. Through this MSA, we confirmed that hawksbills found at sea in the Gulf of Guinea are predominantly of local origin, rather than arriving from the Indian Ocean, as previously suggested by Monzón-Arguello et al. (2011) using 384 bp haplotypes. Connectivity between the Atlantic and the Indian Oceans has been documented in other sea turtle species, such as green turtles (see Chapter 5 of this thesis) and likely occurs in hawksbill as well (Putman et al. 2014). However, massive interoceanic migration is unlikely to constitute a major source for the foraging stock in the Gulf of Guinea. On the other hand, westward movements of turtles from the Indian Ocean on an evolutionary time-scale likely led to the colonization of Eastern Atlantic rookeries by Indo Pacific haplotypes.

Mitogenomic analyses

Whole mitogenome sequencing of a subsample from our dataset revealed that the most common control region haplotype, EiA48, can be divided into at least three distinct

mitogenomic lineages. The genetic diversity uncovered by this technique is promising for future studies: applying mitogenomics more extensively to the characterization of rookeries in the region could clarify the distribution of these mitogenomic lineages and potentially identify unique mitogenomic haplotypes specific to certain nesting areas. In parallel with control region haplotype-based MSA, this approach could contribute to study population structure across both rookeries and foraging grounds. Unfortunately, we could only sequence whole mitogenomes for a few samples, due to protocol restrictions regarding DNA quality and quantity. For highly migratory species such as sea turtles, identifying the natal beach is crucial for conducting population genetics studies. While the sampling location alone may not be sufficient for this purpose, it can still serve as a basis for formulating hypotheses and drawing inferences. When even the sampling location is unknown, genetic data become difficult to interpret and offer poor insights, especially for elusive and understudied species like the hawksbill turtle. In our analyses, we included sequence KP221806, which is listed in GenBank as originating from Atlantic Colombia. This sequence carries an unknown haplotype that clusters, in both control region and mitogenomic networks, with I Indo Pacific clade haplotypes from the Eastern Pacific. Phylogenetic analyses further support this placement. This finding is unexpected, as these genetic lineages have never been observed in Caribbean stocks and, a circumglobal migration from the Eastern Pacific to the Western Atlantic appears highly unlikely. At the moment, we are more prone to hypothesize a data entry error in the database, rather than looking for a phylogeographic explanation, but we highlight this case as an example of how poor sampling metadata can hinder phylogeographic inference.

Overall, our mitogenomic phylogeny supports previous reconstructions, with a deep divergence between Atlantic and Indo Pacific clades in correspondence of the closure of the Isthmus of Panama (12 – 2.5 MYBP) (O'Dea et al. 2016), as observed in other sea turtle species (e.g. Jensen et al. 2019; Baltazar-Soares et al. 2020). The I Indo Pacific clade diverged from the other Indo Pacific clades, which consequently separated in the Pleistocene during recurring events of ice formation and melting (Head and Gibbard 2005). The entrance of Indo Pacific haplotypes in the Atlantic would have occurred as a consequence of the formation of the Agulhas currents (Turney and Jones 2010) since 130 KYBP; indeed, when the phylogenetic spread of II Indo Pacific Clade begun (see Results, Table 5), these warm westward currents had already formed. In future analyses a broader sampling including more Atlantic and Pacific sequences will enable further details and

even more robust dating of the nodes. Similarly, adding more mitogenomic sequences to the database would allow to test selective pressure in more detail, eventually to enforce evidence of different evolutionary patterns within the branches of the phylogeny, identifying sites under selection.

Conclusions

This study provides the first genetic data based on 733 bp control region haplotypes and whole mitogenomic haplotypes for hawksbill turtles foraging and nesting in the Gulf of Guinea. We report a predominantly local origin for most turtles at sea and the exclusive presence of Indo Pacific haplotypes at sampled nesting sites in the area. These findings are consistent with previous studies using shorter control region fragments but offer increased resolution in the genetic characterization of the Gulf of Guinea population. We advocate that future assessments of hawksbill in Eastern Atlantic Regional Management Unit will benefit from this data. In addition, we explored the potential of whole mitogenome sequencing to further resolve maternal lineages of hawksbill in the region and to provide deeper insights into their phylogeography and adaptative history of the species. We envisage that expanding local mitogenome datasets would help clarify the fine-scale population structure of hawksbills in the Gulf of Guinea, and that broader global mitogenomics efforts would allow us to take full advantage of this technique to enhance our understanding of this critically endangered species.

Acknowledgments

I am grateful to Angela Formia, African Aquatic Conservation Fund, and to the volunteers from many organizations in Central and West Africa who assisted with the collection of samples.

Supplementary materials

Table S1. Baseline data used in MSA: 733 bp haplotype frequencies in Atlantic and Indo Pacific hawksbill turtle rookeries, with respective rookery size estimates in number of clutches per year; in bold, average size estimated from available rookeries in case of missing data for the specific rookery. CU, Cuba (Leroux et al. 2012); PR, Puerto Rico (Velez-Zuazo et al. 2008); SI, Saona Island (Carreras et al. 2013); BW, Barbados Windward (Browne et al. 2010); BKI, Buck Island (Leroux et al. 2012); JN, Jaragua National Park (Carreras et al. 2013); AN, Antigua (Leroux et al. 2012); CR, Costa Rica (Leroux et al. 2012); GU, Guadalupe (Leroux et al. 2012); MEX, Atlantic Mexico (Labastida-Estrada et al. 2018); NI, Nicaragua (Leroux et al. 2012); SP, Sandy Point (Hill et al. 2018); CO, Atlantic Colombia (Trujillo-Arias et al. 2014); TO, Tobago (Cazabon-Mannette et al. 2016); BL, Barbados Leeward (Browne et al. 2010); NB, Northern Brazil (Vilaca et al. 2012); SB, Sergipe Bahia Brazil (Marcovaldi et al. 2007; Vilaca et al. 2012); AM, Amirantes Island, Seychelles (Mortimer et al. 2011; Vargas et al. 2015); PL, Platte Island, Seychelles (Vargas et al. 2015); GI, Granitics Island, Seychelles (Vargas et al. 2015); CH, Chagos Archipelago (Vargas et al. 2015; Mortimer et al. 2020); AL, Aldabra, Seychelles (Vargas et al. 2015). Nest GG, Gulf of Guinea, present study and estimate for the rookery size from Monzón-Argüello et al. (2011).

Haplotype	CU	PR	SI	BW	BKI	JN	AN	CR	GU	MEX	NI	SP	CO	TO	BL	NB	SB	AM	PL	GI	CH	AL	Nest GG	MS GG
EiA01	62	1	3	3	8	1	42	0	2	0	0	33	12	24	54	22	54	0	0	0	0	0	0	8
EiA02	0	0	0	0	0	0	0	11	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA03	0	0	0	0	2	0	29	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
EiA09	0	1	0	6	2	2	0	3	69	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0
EiA11	1	59	22	21	50	3	1	33	2	0	54	7	13	9	0	0	0	0	0	0	0	0	0	0
EiA12	0	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA13	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA18	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA20	0	25	6	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA21	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA22	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA23	0	0	0	0	1	4	0	0	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA28	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
EiA29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA30	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

EiA32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	2
EiA39	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA41	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA43	0	2	0	0	0	4	0	0	0	1	16	0	0	1	0	0	0	0	0	0	0	0	0	0
EiA47	0	0	0	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
EiA50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7
EiA52	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
EiA62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	0	0	0
EiA63	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
EiA65	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	77
EiA75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
EiA84	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
EiIP02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	4	2	0	4	177
EiIP17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	14	18	10	0	0	0
EiIP18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	6	13	4	2	0	0
EiIP19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0	0	0	0
EiIP20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
EiA52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
EiA61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
EiA62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

EiA63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
EiA70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiA87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
EiIP03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
EiIP05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	1	0	0
EiIP07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
EiIP16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
EiIP26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
EiIP27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

EiIP34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
EiIP40	62	1	3	3	8	1	42	0	2	0	0	33	12	24	54	22	54	0	0	0	0	0	0	8
EiIP41	0	0	0	0	0	0	0	11	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP42	0	0	0	0	2	0	29	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
EiIP43	0	1	0	6	2	2	0	3	69	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0
EiIP44	1	59	22	21	50	3	1	33	2	0	54	7	13	9	0	0	0	0	0	0	0	0	0	0
EiIP47	0	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP48	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP49	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP50	0	25	6	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP51	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP53	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP54	0	0	0	0	1	4	0	0	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP57	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP74	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
EiIP75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP76	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	2
EiIP81	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP106	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP107	0	2	0	0	0	4	0	0	0	1	16	0	0	1	0	0	0	0	0	0	0	0	0	0
EiIP108	0	0	0	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
EiIP114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
EiIP115	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EiIP124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
EiIP125	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	0	0	0
New Haplo	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Rookery size	130	740	100	150	158	15	203	25	151	311	205	300	**410**	577	1504	304	1345	300	**410**	**410**	1200	**410**	50	-

References

Abreu A, Horrocks J, Formia A, LeRoux R, Velez-Zuazo X, Dutton P, Soares L, Meylan P, Browne D (2006) New mtDNA Dloop primers which work for a variety of marine turtle species may increase the resolution capacity of mixed stock analysis. Proceedings of the 26th Annual Symposium on Sea Turtle Biology and Conservation.

Arantes LS, Vargas SM, Santos FR (2020) Global phylogeography of the critically endangered hawksbill turtle (*Eretmochelys imbricata*). Genetics and molecular biology 43.

Baltazar-Soares M, Klein JD, Correia SM, Reischig T, Taxonera A, Roque SM, Dos Passos L, Durão J, Lomba JP, Dinis H, Cameron SJK, Stiebens VA, Eizaguirre C (2020) Distribution of genetic diversity reveals colonization patterns and philopatry of the loggerhead sea turtles across geographic scales. Sci Rep 10: 18001.

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA (2011) Has the Earth's sixth mass extinction already arrived? Nature 471: 51-57.

Bell I, Jensen MP (2018) Multinational genetic connectivity identified in western Pacific hawksbill turtles, *Eretmochelys imbricata*. Wildlife Research 45: 307-315.

Blumenthal J, Abreu-Grobois A, Austin T, Broderick A, Bruford M, Coyne M, Ebanks-Petrie G, Formia A, Meylan P, Meylan A, Godley B (2009) Turtle groups or turtle soup: dispersal patterns of hawksbill turtles in the Caribbean. Molecular Ecology 18: 4841-4853.

Bolker BM, Okuyama T, Bjorndal KA, Bolten AB (2007) Incorporating multiple mixed stocks in mixed stock analysis: 'many-to-many' analyses. Molecular Ecology 16: 685-695.

Browne DC, Horrocks JA, Abreu-Grobois FA (2010) Population subdivision in hawksbill turtles nesting on Barbados, West Indies, determined from mitochondrial DNA control region sequences. Conservation Genetics 11: 1541-1546.

Carreras C, Godley BJ, León YM, Hawkes LA, Revuelta O, Raga JA, Tomás J (2013) Contextualising the last survivors: population structure of marine turtles in the Dominican Republic. PLOS ONE 8: e66037.

Cazabon-Mannette M, Browne D, Austin N, Hailey A, Horrocks J (2016) Genetic structure of the hawksbill turtle rookery and foraging aggregation in Tobago, West Indies. Journal of Experimental Marine Biology and Ecology 485: 94-101.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM (2015) Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1: e1400253.

Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci U S A 114: E6089-e6096.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9: 1657-1659.

del Monte-Luna P, Nakamura M, Vicente A, Pérez-Sosa LB, Yáñez-Arenas A, Trites AW, Lluch-Cota SE (2023) A review of recent and future marine extinctions. Cambridge Prisms: Extinction 1: e13.

Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB, Collen B (2014) Defaunation in the Anthropocene. Science 345: 401-406.

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65: 241-250.

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

Gaos A, Lewison R, Jensen M, Liles M, Henriquez A, Chavarria S, Pacheco C, Valle M, Melero D, Gadea V, Altamirano E, Torres P, Vallejo F, Miranda C, LeMarie C, Lucero J, Oceguera K, Chácon D, Fonseca L, Abrego M, Seminoff J, Flores E, Llamas I, Donadi R, Peña B, Muñoz J, Ruales D, Chaves J, Otterstrom S, Zavala A, Hart C, Brittain R, Alfaro-Shigueto J, Mangel J, Yañez I, Dutton P (2018) Rookery contributions, movements and conservation needs of hawksbill turtles at foraging grounds in the eastern Pacific Ocean. Mar Ecol Prog Ser 586: 203-216.

Gorham JC, Clark DR, Bresette MJ, Bagley DA, Keske CL, Traxler SL, Witherington BE, Shamblin BM, Nairn CJ (2014) Characterization of a subtropical hawksbill sea turtle (*Eretmocheyles imbricata*) assemblage utilizing shallow water natural and artificial habitats in the Florida Keys. PLOS ONE 9: e114171.

Head MJ, Gibbard PL (2005) Early-Middle Pleistocene transitions: an overview and recommendation for the defining boundary. Geological Society, London, Special Publications 247: 1-18.

Hill JE, King CM, Stewart KR, Paladino FV, Dutton PH (2018) Genetic differentiation of hawksbill turtle rookeries on St. Croix, US Virgin Islands. Chelonian Conservation and Biology 17: 303-308.

IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services (summary for policy makers), IPBES 7, Paris.

Jensen MP, FitzSimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH (2019) The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). Journal of Biogeography 46: 860-870.

Labastida-Estrada E, Machkour-M'Rabet S, Díaz-Jaimes P, Cedeño-Vázquez JR, Hénaut Y (2018) Genetic structure, origin, and connectivity between nesting and foraging areas of hawksbill turtles of the Yucatan Peninsula: a study for conservation and management. Aquatic Conservation: Marine and Freshwater Ecosystems 29: 211-222.

Leroux RA, Dutton PH, Abreu-Grobois FA, Lagueux CJ, Campbell CL, Delcroix E, Chevalier J, Horrocks JA, Hillis-Starr Z, Troëng S, Harrison E, Stapleton S (2012) Re-examination of population structure and phylogeography of hawksbill turtles in the wider Caribbean using longer mtDNA sequences. Journal of Heredity 103: 806-820.

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

Marcovaldi M, Lopez G, Soares L, Santos A, Bellini C, Barata P (2007) Fifteen years of hawksbill sea turtle (*Eretmochelys imbricata*) nesting in Northern Brazil. Chelonian Conservation and Biology 6: 223-228.

Monzón-Argüello C, Loureiro NS, Delgado C, Marco A, Lopes JM, Gomes MG, Abreu-Grobois FA (2011) Príncipe island hawksbills: Genetic isolation of an eastern Atlantic stock. Journal of Experimental Marine Biology and Ecology 407: 345-354.

Monzón-Argüello C, Rico C, Marco A, López P, López-Jurado LF (2010) Genetic characterization of eastern Atlantic hawksbill turtles at a foraging group indicates major undiscovered nesting populations in the region. Journal of Experimental Marine Biology and Ecology 387: 9-14.

Mortimer J, Camille J, Boniface N (2011) Seasonality and status of nesting hawksbill (*Eretmochelys imbricata*) and green turtles (*Chelonia mydas*) at D'Arros Island, Amirantes Group, Seychelles. Chelonian Conservation and Biology 10: 26-33.

Mortimer JA, Donnelly M (2008) Eretmochelys imbricata. The IUCN Red List of Threatened Species. Version 2014.2. www.iucnredlist.org

Mortimer JA, Esteban N, Guzman AN, Hays GC (2020) Estimates of marine turtle nesting populations in the south-west Indian Ocean indicate the importance of the Chagos Archipelago. Oryx 54: 332-343.

Múrias dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M (2015) tcsBU: a tool to extend TCS network layout and visualization. Bioinformatics 32: 627-628.

Nishizawa H, Joseph J, Chong YK (2016) Spatio-temporal patterns of mitochondrial DNA variation in hawksbill turtles (*Eretmochelys imbricata*) in Southeast Asia. Journal of Experimental Marine Biology and Ecology 474: 164-170.

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD, Stallard RF, Woodburne MO, Aguilera O, Aubry M-P, Berggren WA, Budd AF, Cozzuol MA, Coppard SE, Duque-Caro H, Finnegan S, Gasparini GM, Grossman EL, Johnson KG, Keigwin LD, Knowlton N, Leigh EG, Leonard-Pingel JS, Marko PB, Pyenson ND, Rachello-Dolmen PG, Soibelzon E, Soibelzon L, Todd JA, Vermeij GJ, Jackson JBC (2016) Formation of the Isthmus of Panama. Science Advances 2: e1600883.

Pérez-Bermúdez E, Urquiola A, Rebordinos-González L, Espinosa-López G, Lee-González I, Petric B (2017) A new approach to mixed stock analysis that informs on the management and conservation of hawksbill in the wider Caribbean: the case of the legal fishery in Jardines del Rey, Cuba. International Journal of Marine Science 7: 316 - 343.

Proietti MC, Reisser J, Marins LF, Rodriguez-Zarate C, Marcovaldi MA, Monteiro DS, Pattiaratchi C, Secchi ER (2014) Genetic structure and natal origins of immature hawksbill turtles (*Eretmochelys imbricata*) in Brazilian waters. PLOS ONE 9: e88746.

Putman NF, Abreu-Grobois FA, Broderick AC, Ciofi C, Formia A, Godley BJ, Stroud S, Pelembe T, Verley P, Williams N (2014) Numerical dispersal simulations and genetics help explain the origin of hawksbill sea turtles in Ascension Island. Journal of Experimental Marine Biology and Ecology 450: 98-108.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C, Ko BJ, Chaisson M, Gedman GL, Cantin LJ, Thibaud-Nissen F, Haggerty L, Bista I, Smith M, Haase B, Mountcastle J, Winkler S, Paez S, Howard J, Vernes SC, Lama TM, Grutzner F, Warren WC, Balakrishnan CN, Burt D, George JM, Biegler MT, Iorns D, Digby A, Eason D, Robertson B, Edwards T, Wilkinson M, Turner G, Meyer A, Kautt AF, Franchini P, Detrich HW, 3rd, Svardal H, Wagner M, Naylor GJP, Pippel M, Malinsky M, Mooney M, Simbirsky M, Hannigan BT, Pesout T, Houck M, Misuraca A, Kingan SB, Hall R, Kronenberg Z, Sović I, Dunn C, Ning Z, Hastie A, Lee J, Selvaraj S, Green RE, Putnam NH, Gut I, Ghurye J, Garrison E, Sims Y, Collins J, Pelan S, Torrance J, Tracey A, Wood J, Dagnew RE, Guan D, London SE, Clayton DF, Mello CV, Friedrich SR, Lovell PV, Osipova E, Al-Ajli FO, Secomandi S, Kim H, Theofanopoulou C, Hiller M, Zhou Y, Harris RS, Makova KD, Medvedev P, Hoffman J, Masterson P, Clark K, Martin F, Howe K, Flicek P, Walenz BP, Kwak W, Clawson H, Diekhans M, Nassar L, Paten B, Kraus RHS, Crawford AJ, Gilbert MTP, Zhang G, Venkatesh B, Murphy RW, Koepfli KP, Shapiro B, Johnson WE, Di Palma F, Marques-Bonet T, Teeling EC, Warnow T, Graves JM, Ryder OA, Haussler D, O'Brien SJ, Korlach J, Lewin HA, Howe K, Myers EW, Durbin R, Phillippy AM, Jarvis ED (2021) Towards complete and error-free genome assemblies of all vertebrate species. Nature 592: 737-746.

Richardson PB, Bruford MW, Calosso MC, Campbell LM, Clerveaux W, Formia A, Godley BJ, Henderson AC, McClellan K, Newman S, Parsons K, Pepper M, Ranger S, Silver JJ, Slade L, Broderick AC (2009) Marine Turtles in the Turks and Caicos Islands: Remnant Rookeries, Regionally Significant Foraging Stocks, and a Major Turtle Fishery. Chelonian Conservation and Biology 8: 192-207.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619-633.

Tolve L, Iannucci A, Garofalo L, Ninni A, Capobianco Dondona A, Ceciarini I, Cocumelli C, De Lucia A, Falconi M, Formia A, Iacovelli F, Mancusi C, Marchiori E, Marsili L, Mingozzi T, Nannarelli S, Natali C, Terracciano G, Zuffi MAL, Novelletto A, Ciofi C (2023) Whole mitochondrial genome sequencing

provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Marine Biology 171: 19.

Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44: W232-W235.

Trujillo-Arias N, Amorocho DF, López-Álvarez D, Mejía-Ladino LM (2014) Relaciones filogeográficas de algunas colonias de alimentación y anidacion de la tortuga carey (*Eretmochelys imbricata*) en el Pacífico y Caribe Colombianos. Boletín de Investigaciones Marinas y Costeras - INVEMAR 43: 159-182.

Turney CS, Jones RT (2010) Does the Agulhas Current amplify global temperatures during super-interglacials? J Quaternary Sci 25: 839-843.

Vargas SM, Jensen MP, Ho SYW, Mobaraki A, Broderick D, Mortimer JA, Whiting SD, Miller J, Prince RIT, Bell IP, Hoenner X, Limpus CJ, Santos FR, FitzSimmons NN (2015) Phylogeography, genetic diversity, and Management Units of hawksbill turtles in the Indo-Pacific. Journal of Heredity 107: 199-213.

Velez-Zuazo X, Ramos W, Van Dam R, Diez C, Abreu-Grobois A, McMillan W (2008) Dispersal, recruitment and migratory behaviour in a hawksbill sea turtle aggregation. Molecular Ecology 17: 839-853.

Vilaça ST, Hahn AT, Naro-Maciel E, Abreu-Grobois FA, Bowen BW, Castilhos JC, Ciofi C, FitzSimmons NN, Jensen MP, Formia A, Limpus CJ, Natali C, Soares LS, de Thoisy B, Whiting SD, Bonatto SL (2022) Global phylogeography of ridley sea turtles (Lepidochelys spp.): evolution, demography, connectivity, and conservation. Conservation Genetics 23: 995-1010.

Vilaca ST, Vargas SM, Lara-Ruiz P, Molfetti E, Reis EC, Lobo-Hajdu G, Soares LS, Santos FR (2012) Nuclear markers reveal a complex introgression pattern among marine turtle species on the Brazilian coast. Molecular Ecology 21: 4300-4312.

Wallace BP, Bandimere AN, Abreu-Grobois FA, Acosta H, Akiti J, Akomedi M, Alfaro-Shigueto J, Allen CD, Angenda D, Ayissi I, Azanza Ricardo J, Barrientos-Muñoz KG, Barrios-Garrido H, Bjorndal KA, Bretón Vargas E, Broderick AC, Calderón Peña R, Carreras C, Ceriani SA, Colman LP, Cortés-Gómez AA, Crespo L, Cuevas E, Dah A, de Groene A, Delgado Trejo C, Demetropoulos S, Dias A, Diez C, Dos Santos NA, Dossou Bodjrenou JS, Early Capistrán MM, Eckert KL, Eizaguirre C, Ekanayake L, Escobedo Mondragón M, Esteban N, Feliciano D, Fernandes RS, Ferreira-Airaud B, Foley A, Fonseca LG, Forneiro Martín-Viaña Y,

Fossette S, Fuentes M, Gaglo J, Gaos AR, Gidsicki D, Giffoni B, Girard A, Girondot M, Godfrey MH, Godley BJ, González Díaz Mirón RJ, Hamann M, Hancock JM, Hart CE, Hays GC, Herrera R, Hochscheid S, Hoekstra S, Huerta-Rodríguez P, Inteca G, Ishihara T, Jensen MP, Jribi I, Kale N, Kaska Y, Kelez S, Kelly IK, Köhnk S, Lara P, Lasfargue M, Lauritsen AM, Le Gouvello DZM, Liusamoa A, López M, López-Castro MC, Lopez-Mendilaharsu M, Louro CMM, Luna T, Madden CA, Mahabir D, Mancini A, Manoharakrishnan M, Marcovaldi MA, Martínez-Portugal RC, Mastrogiacomo A, Matilde E, Mawunyo Adzagba B, Mbungu S, Miranda C, Moncada F, Morales-Mérida BA, Mortimer JA, Murakawa SKK, Nalovic MA, Nel R, Ngafack R, Nishizawa H, Ogou M, Panagopoulou A, Patricio AR, Peralta Buendía E, Phillott AD, Pilcher NJ, Polyak MMR, Prince RIT, Raynus EH, Reina RD, Rguez-Baron JM, Robbins AE, Santos AS, Sarti-Martínez AL, Schofield G, Seminoff JA, Serrano I, Shamblin BM, Shanker K, Stacy BA, Stahelin G, Staman MK, Stelfox M, Stewart KR, Taxonera A, Tucker AD, Turkozan O, van Dam RP, van de Geer CH, Viera S, West L, Whiting AU, Whiting SD, Wienand L, Wijntuin SR, Wildermann N, Zárate PM, Casale P, DiMatteo A, Hurley BJ, Hutchinson BJ, Maxwell SM, Posnik ZA, Rodriguez I, Mast RB (2025) Updated global conservation status and priorities for marine turtles. Endangered Species Research 56: 247-276.

Wallace BP, DiMatteo AD, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Mortimer JA, Seminoff JA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Briseño Dueñas R, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Finkbeiner EM, Girard A, Girondot M, Hamann M, Hurley BJ, López-Mendilaharsu M, Marcovaldi MA, Musick JA, Nel R, Pilcher NJ, Troëng S, Witherington B, Mast RB (2011) Global conservation priorities for marine turtles. PLOS ONE 6: e24510.

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5: e15465.

Wood LD, Hardy R, Meylan PA, Meylan AB (2013) Characterization of a hawksbill turtle (*Eretmochelys imbricata*) foraging aggregation in a high-latitude reef community in southeastern Florida, USA. Herpetological Conservation and Biology 8: 258-275.

Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24: 1586-1591.

Zuñiga-Marroquin T, Espinosa de los Monteros A (2017) Genetic characterization of the Critically Endangered hawksbill turtle (*Eretmochelys imbricata*) from the Mexican Pacific region. Latin american journal of aquatic research 45: 555-562.

CHAPTER VII

General Conclusions

The characterization of mitochondrial DNA (mtDNA) has long been a widely used technique in molecular ecology and conservation genetics (Harrison 1989). For decades, it has primarily relied on the analysis of individual genes or fragments of the molecule (e.g. Di Rienzo and Wilson 1991; Norman et al. 1994; Bronstein et al. 2018). However, in more recent years, the use of whole mitogenome sequencing and of genomics has significantly increased, driven by the growing accessibility of high-throughput sequencing technologies (Formenti et al. 2021; Rhie et al. 2021).

The use of mitochondrial DNA is particularly suitable for population genetic studies of some species; marine turtles are undoubtedly among them, due to a pronounced female philopatry to natal beaches, which leads to strong population structuring (Bowen and Karl 2007). Mitochondrial DNA – maternally inherited – allows researchers to trace female lineages and thus to describe populations associated with specific nesting beaches (Jensen et al. 2013). This type of information, integrated with behavioural and biological data, has been fundamental in defining Regional Management Units (RMUs) (Wallace et al. 2010) for conservation purposes, and in reconstructing both present and historical patterns of global distribution (e.g. Shamblin et al. 2014).

Even in female-philopatric animals, not only the analysis of mtDNA is important, but also of nuclear data, as it allows for the reconstruction of male-mediated gene flow among oceanic stocks and thus contributes to a more comprehensive understanding of genetic diversity and movements from rookeries to foraging grounds (e.g. Carreras et al. 2007). Historically, nuclear data have primarily derived from microsatellite characterization (e.g. Aggarwal et al. 2004) or single-gene amplification and sequencing (e.g. Brito et al. 2020). Today, the increasing availability of reference genomes (Chang et al. 2023) is paving the way for population genomics (Chow et al. 2019) and for evolutionary studies on a genomic scale (e.g. Bentley et al. 2023), in sea turtles as well as in many other species (e.g. Dussex et al. 2018; Morin et al. 2021). Overall, whole mitogenome and whole-genome sequencing represent the new frontier in molecular ecology; it is therefore crucial to assess the potential of these techniques for species of recognized conservation concern, such as sea turtles.

In this PhD thesis, I set up a pipeline for the sequencing, assembly and analysis of mitogenomes in three sea turtle species. A PCR-free genomic library protocol based on tagmentation allowed to avoid amplification errors. However, an inherent limitation of this approach was the requirement for high-quality and high-quantity DNA, which is often difficult to obtain from tissue biopsies of deceased animals - such as stranded

turtles. This limited the number of samples that could be sequenced. Whole genome sequencing with low nuclear coverage was performed to obtain medium–to-high coverage mitogenomes, although coverage heterogeneity among samples occurred, due to differences in input library quality. In all my case studies, a reference mitogenome was already available in online databases; I therefore used it to map the reads, or to support a *de novo* assembly pipeline. I also assembled the mitogenomes using both approaches in parallel, which ensured robust reconstruction even in case of suboptimal read quality - the only exception being the microsatellite at the end of the control region. As discussed in Chapter 2, reliable assembly of repetitive regions is not achievable with short-reads sequencing. I therefore chose to exclude the microsatellite from the bioinformatic analyses, as it was not required for our objectives. Overall, optimizing the laboratory procedure was crucial to fully leverage the available samples. The methodological choices made throughout this PhD thesis were driven by a balanced approach prioritizing quality, efficiency and cost-effectiveness.

This pipeline was applied to a range of biological questions involving three sea turtle species in different geographic areas. In Chapter 3, I first assessed the potential of whole mitogenomes to improve the resolution of maternal lineages of loggerhead turtle (*Caretta caretta*) that could not be distinguished using control region fragments. I focused on the Central Mediterranean, where the widely shared CcA2.1 haplotype had historically limited the characterization of rookeries and the assignment of individuals of mixed stocks to their natal populations (Tolve et al. 2018). I identified 11 mitogenomic variants associated with the same control region haplotype CcA2.1, and I also provided new genetic data for previously undescribed or recently colonized nesting sites. These preliminary findings advocate for future re-analyses of Mediterranean rookeries, where distinct mitogenomic haplotypes could subdivide the widespread CcA2.1 control region haplotype, potentially contributing to a refined mitogenomic baseline for Mixed Stock Analyses (MSAs).

Not only complete mitogenomes can be a useful tool for conservation genetics, but also for improving ecological studies and evolutionary history reconstruction. In Chapter 4, I used complete mitogenomes to clarify the global phylogeography of *C. caretta*, still debated by researchers supporting different colonization patterns for the three mtDNA haplogroups (e.g. Shamblin et al. 2014; Baltazar-Soares et al. 2020). I compared single genes topologies and mitogenomic topology of the same dataset and found divergent results; I concluded that conflicting phylogeographic hypotheses arose from the use of

single mtDNA genes in previous studies, and I highlighted the importance of using the most informative marker for phylogenetic inferences. In the same study I also investigated potential selective pressure in action on the Atlantic haplogroup (IB), under the hypothesis of adaptation to cold temperatures, which could help explain the peculiar latitudinal gradient in the relative distribution of this haplogroup and Atlantic-Mediterranean haplogroup (II) in Western Atlantic rookeries. I found significant correlation between temperature and haplogroup frequencies in that region; moreover, comparing the complete coding genes partitions of *C. caretta* haplogroups with the complete coding genes partitions of the other sea turtle species, I identified a pattern of nonsynonymous substitutions which suggests a different evolutionary pathway for haplogroup IB. Detailed analysis of mitogenomic alignment allowed detection of weak signals of selection which remained unnoticed using statistical approaches, due to the very conserved nature of the mitochondrial DNA.

Chapters 5 and 6 of this thesis were dedicated to *Chelonia mydas* and *Eretmochelys imbricata* in the Gulf of Guinea. In those studies, I leveraged both data produced by traditional mtDNA markers from broad sample datasets and whole mitogenome sequences produced from subsamples. For *C. mydas*, control region haplotypes 384 bp long allowed to perform four MSAs using published baselines, on Western Africa, Sao Tomé, Equatorial Guinea, and Corisco Bay. I recorded foraging turtles in the Gulf of Guinea coming from Indo Pacific rookeries, testifying interoceanic migrating behaviour. Results for the different MSA had very wide confidence intervals, unless a very numerous mixed stocks was used for the MSA, such as Corisco Bay. The main reason for this was the common haplotype CmA8. Interestingly, in the case of *C. mydas*, shifting to longer control region markers did not provide a better resolution of the shared lineages. On the other hand, whole mtDNA analysis recovered 5 mitogenomic haplotypes among 11 sequenced samples, and CmA8 control region haplotype was split into four mitogenomic lineages. Mitogenomic sequences also gave insights into the phylogeny of the species, giving different results compared to the control region phylogenies (Jensen et al. 2019) regarding the clustering of Indo Pacific groups and the placement of the Mediterranean haplotype. As in previous Chapters, I acknowledged the importance of using the most informative mtDNA sequences to make phylogeographic inferences.

For *E. imbricata*, I presented the first description of Eastern Atlantic foraging grounds and nesting using 733 bp control region haplotypes, which allowed to fill the gap of knowledge about this geographic area that had remained understudied until now and only

described by 384 bp fragments. Thanks to investigations on longer control region haplotypes, I confirmed that most turtles at sea in the Gulf of Guinea are of local origin, and not migrating from the Indo Pacific; also, I retrieved higher genetic diversity than previous studies, and such information will be useful to update regional assessment for the species. We also identified mitogenomic variants of abundant control region haplotypes were identified, although the limited sample size with very few individuals from nesting did not allow to make comprehensive inferences about the origin of the maternal lineages. This study highlighted the scarcity of mitogenomic data available now globally and how this may hinder the possibility of phylogeographic analyses, advocating for future increase in the use of whole mtDNA genome sequences. Overall, in Chapters 5 and 6 I showed how – even in the mitogenomics era - control region sequences can still be useful to population genetics, answering some broad ecological questions, mainly due to the huge amount of data that have been published over the decades and thanks to the generation of new sequences.

The potential of mitogenomics to improve population genetics and studies on evolutionary biology of sea turtles emerges across all the thesis. However, it is important to delineate the cases and requirements for efficient implementation of this technique. Although whole mitogenomes provided a higher resolution of matrilines than traditional markers, they are not yet a massive substitute of control region haplotypes. To assess mitogenomic diversity at population level and to perform mitogenomic Mixed Stock Analyses a huge effort of re-sequencing needs to be done, for all the main rookeries in the baselines and the main foraging grounds. The feasibility of such a transition relies on the research community's commitment to adopt mitogenomics, which is still more expensive than traditional genetic methods. Additionally, re-analysing all major Mediterranean rookeries and foraging grounds using mitogenomes might require more time than fully implementing more advanced population genomics tools. The outcome will depend on the relative pace at which the cost and feasibility of mitogenomics decline compared to the rate at which genomic tools become affordable and are adopted.

Multidimensional scaling (MDS) based on genome-wide data (Tzeng et al. 2008) looks promising as a method to reconstruct population structure, therefore, to define Management Units of turtle rookeries. Moreover, Discriminant Analysis of Principal Components (DAPC) using SNP data is likely to be used in the next future to make assignments of turtles to the populations of origin, so as for other animal species (Iannucci et al. 2021). The strength of these techniques, and in general of genomics

approaches, is to take advantage of a huge amount of genetic information from a relatively lower number of sampled individuals than what required by MSA with single gene markers. This is of main interest for endangered and elusive species, such as all extant sea turtles, whose sampling is mostly occasional and not always straightforward; the possibility to leverage limited sample size datasets will certainly empower sea turtle conservation genetics.

Until population genomics spreads for sea turtle studies, mitogenomics can be useful to address specific questions about fine-scale description of genetic diversity. Within defined geographic areas and a well-planned experimental design, a mitogenomic assessment could be done to refine rookeries description and track turtles' movements to local foraging grounds, not implying MSA. This is possible whenever a high number of mitogenomic lineages is described in limited areas, where mostly local turtles are likely to move. Moreover, in this thesis we demonstrated that whole mitogenomes have an underestimated potential to shed light on phylogeographic questions and adaptive studies. In conclusion, for future research on sea turtle species we envisage a combination of traditional and more advanced techniques, from single gene markers, across mitogenomes, until genomics. All approaches have strengths and weaknesses, and should be used as complementary tools, under appropriate experimental design and well-defined biological questions.

References

Aggarwal RK, Velavan TP, Udaykumar D, Hendre PS, Shanker K, Choudhury BC, Singh L (2004) Development and characterization of novel microsatellite markers from the olive ridley sea turtle (*Lepidochelys olivacea*). Molecular Ecology Notes 4: 77-79.

Baltazar-Soares M, Klein JD, Correia SM, Reischig T, Taxonera A, Roque SM, Dos Passos L, Durão J, Lomba JP, Dinis H, Cameron SJK, Stiebens VA, Eizaguirre C (2020) Distribution of genetic diversity reveals colonization patterns and philopatry of the loggerhead sea turtles across geographic scales. Sci Rep 10: 18001.

Bentley BP, Carrasco-Valenzuela T, Ramos EKS, Pawar H, Souza Arantes L, Alexander A, Banerjee SM, Masterson P, Kuhlwilm M, Pippel M, Mountcastle J, Haase B, Uliano-Silva M, Formenti G, Howe K, Chow W, Tracey A, Sims Y, Pelan S, Wood J, Yetsko K, Perrault JR, Stewart K, Benson SR, Levy Y, Todd EV, Shaffer HB,

Scott P, Henen BT, Murphy RW, Mohr DW, Scott AF, Duffy DJ, Gemmell NJ, Suh A, Winkler S, Thibaud-Nissen F, Nery MF, Marques-Bonet T, Antunes A, Tikochinski Y, Dutton PH, Fedrigo O, Myers EW, Jarvis ED, Mazzoni CJ, Komoroske LM (2023) Divergent sensory and immune gene evolution in sea turtles with contrasting demographic and life histories. Proceedings of the National Academy of Sciences 120: e2201076120.

Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. Mol Ecol 16.

Brito C, Vilaça ST, Lacerda AL, Maggioni R, Marcovaldi MÂ, Vélez-Rubio G, Proietti MC (2020) Combined use of mitochondrial and nuclear genetic markers further reveal immature marine turtle hybrids along the South Western Atlantic. Genetics and molecular biology 43.

Bronstein O, Kroh A, Haring E (2018) Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. BMC Evolutionary Biology 18: 80.

Carreras C, Pascual M, Cardona L (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. Conserv Genet 8.

Chang G, Jones S, Leelakumari S, Ashkani J, Culibrk L, O'Neill K, Tse K, Cheng D, Chuah E, McDonald H, Kirk H, Pandoh P, Pari S, Angelini V, Kyle C, Bertorelle G, Zhao Y, Mungall A, Moore R, Vilaça ST, Jones S (2023) The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758 [version 2; peer review: 2 approved]. F1000Research 12.

Chow JC, Anderson PE, Shedlock AM (2019) Sea turtle population genomic discovery: global and locus-specific signatures of polymorphism, selection, and adaptive potential. Genome biology and evolution 11: 2797-2806.

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proceedings of the National Academy of Sciences 88: 1597-1601.

Dussex N, Von Seth J, Robertson BC, Dalén L (2018) Full mitogenomes in the critically endangered kākāpō reveal major post-glacial and anthropogenic effects on neutral genetic diversity. Genes 9: 220.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli FO, Ambrosini R, Houde P, Koren S, Oliver K, Smith

M, Skelton J, Betteridge E, Dolucan J, Corton C, Bista I, Torrance J, Tracey A, Wood J, Uliano-Silva M, Howe K, McCarthy S, Winkler S, Kwak W, Korlach J, Fungtammasan A, Fordham D, Costa V, Mayes S, Chiara M, Horner DS, Myers E, Durbin R, Achilli A, Braun EL, Phillippy AM, Jarvis ED, Kirschel ANG, Digby A, Veale A, Bronikowski A, Murphy B, Robertson B, Baker C, Mazzoni C, Balakrishnan C, Lee C, Mead D, Teeling E, Aiden EL, Todd E, Eichler E, Naylor GJP, Zhang G, Smith J, Wolf J, Touchon J, Delmore K, Jakobsen K, Komoroske L, Wilkinson M, Genner M, Pšenička M, Fuxjager M, Stratton M, Liedvogel M, Gemmell N, Minias P, Dunn PO, Sudmant P, Morin P, Ayub Q, Kraus R, Vernes S, Smith S, Lama T, Edwards T, Smith T, Gilbert T, Marques-Bonet T, Einfeldt T, Venkatesh B, Johnson W, Warren W, Bukhman Y, The Vertebrate Genomes Project C (2021) Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. Genome Biology 22: 120.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology & Evolution 4: 6-11.

Iannucci A, Benazzo A, Natali C, Arida EA, Zein MSA, Jessop TS, Bertorelle G, Ciofi C (2021) Population structure, genomic diversity and demographic history of Komodo dragons inferred from whole-genome sequencing. Molecular Ecology 30: 6309-6324.

Jensen MP, FitzSimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH (2019) The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). Journal of Biogeography 46: 860-870.

Jensen MP, FitzSimmons NN, Dutton PH, Michael P (2013) Molecular genetics of sea turtles. In: Wyneken J, Lohmann K, Musick JA (eds) The biology of sea turtles, Volume III. CRC Press, Boca Raton, FL, pp 135-161

Morin PA, Archer FI, Avila CD, Balacco JR, Bukhman YV, Chow W, Fedrigo O, Formenti G, Fronczek JA, Fungtammasan A, Gulland FMD, Haase B, Peter Heide-Jorgensen M, Houck ML, Howe K, Misuraca AC, Mountcastle J, Musser W, Paez S, Pelan S, Phillippy A, Rhie A, Robinson J, Rojas-Bracho L, Rowles TK, Ryder OA, Smith CR, Stevenson S, Taylor BL, Teilmann J, Torrance J, Wells RS, Westgate AJ, Jarvis ED (2021) Reference genome and demographic history of the most endangered marine mammal, the vaquita. Molecular Ecology Resources 21: 1008-1020.

Norman JA, Moritz C, Limpus CJ (1994) Mitochondrial DNA control region polymorphisms: genetic markers for ecological studies of marine turtles. Mol Ecol 3: 363-373.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C, Ko BJ, Chaisson M, Gedman GL, Cantin LJ, Thibaud-Nissen F, Haggerty L, Bista I, Smith M, Haase B, Mountcastle J, Winkler S, Paez S, Howard J, Vernes SC, Lama TM, Grutzner F, Warren WC, Balakrishnan CN, Burt D, George JM, Biegler MT, Iorns D, Digby A, Eason D, Robertson B, Edwards T, Wilkinson M, Turner G, Meyer A, Kautt AF, Franchini P, Detrich HW, Svardal H, Wagner M, Naylor GJP, Pippel M, Malinsky M, Mooney M, Simbirsky M, Hannigan BT, Pesout T, Houck M, Misuraca A, Kingan SB, Hall R, Kronenberg Z, Sović I, Dunn C, Ning Z, Hastie A, Lee J, Selvaraj S, Green RE, Putnam NH, Gut I, Ghurye J, Garrison E, Sims Y, Collins J, Pelan S, Torrance J, Tracey A, Wood J, Dagnew RE, Guan D, London SE, Clayton DF, Mello CV, Friedrich SR, Lovell PV, Osipova E, Al-Ajli FO, Secomandi S, Kim H, Theofanopoulou C, Hiller M, Zhou Y, Harris RS, Makova KD, Medvedev P, Hoffman J, Masterson P, Clark K, Martin F, Howe K, Flicek P, Walenz BP, Kwak W, Clawson H, Diekhans M, Nassar L, Paten B, Kraus RHS, Crawford AJ, Gilbert MTP, Zhang G, Venkatesh B, Murphy RW, Koepfli K-P, Shapiro B, Johnson WE, Di Palma F, Marques-Bonet T, Teeling EC, Warnow T, Graves JM, Ryder OA, Haussler D, O'Brien SJ, Korlach J, Lewin HA, Howe K, Myers EW, Durbin R, Phillippy AM, Jarvis ED (2021) Towards complete and error-free genome assemblies of all vertebrate species. Nature 592: 737-746.

Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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: e85956.

Tolve L, Casale P, Formia A, Garofalo L, Lazar B, Natali C, Novelletto A, Vallini C, Bužan E, Chelazzi G, Gaspari S, Fortuna C, Kocijan I, Marchiori E, Novarini N, Poppi L, Salvemini P, Ciofi C (2018) A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. Mar Biol 165: 68.

Tzeng J, Lu HH-S, Li W-H (2008) Multidimensional scaling for large genomic data sets. BMC Bioinformatics 9: 179.

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5: e15465.

APPENDIX I

Marine Biology (2024) 171:19
https://doi.org/10.1007/s00227-023-04325-x

ORIGINAL PAPER

Whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea

Livia Tolve[1] · Alessio Iannucci[1,2] · Luisa Garofalo[3] · Andrea Ninni[4,5] · Andrea Capobianco Dondona[6] · Ilaria Ceciarini[7] · Cristiano Cocumelli[3] · Alessandra De Lucia[8] · Mattia Falconi[4] · Angela Formia[1,9] · Federico Iacovelli[4] · Cecilia Mancusi[7,10] · Erica Marchiori[11] · Letizia Marsili[7] · Toni Mingozzi[12] · Stefano Nannarelli[8] · Chiara Natali[1] · Giuliana Terracciano[13] · Marco A. L. Zuffi[14] · Andrea Novelletto[5] · Claudio Ciofi[1]

Received: 7 April 2023 / Accepted: 13 October 2023 / Published online: 23 November 2023

Abstract
Population structure and phylogeography of the loggerhead sea turtle (*Caretta caretta*) have so far been assessed mainly by mitochondrial DNA (mtDNA) single-gene sequencing studies. However, phylogenetic relationships amongst matrilines, genetic characterisation of rookeries and mixed-stock analyses have suffered from the limited resolution obtained by comparison of relatively short sequences such as from the mtDNA control region. Whole mitogenome sequencing can significantly improve population genetics, particularly in marine organisms showing female natal philopatry. Despite mitogenomics becoming increasingly common in biodiversity monitoring and conservation, only a few complete mitogenomes are available for *C. caretta*. In this study, we sequenced the complete mtDNA of 61 loggerhead turtles sampled between 2008 and 2021 along the Italian coastline and central Mediterranean Sea. We assigned complete mtDNA haplotypes to dead embryos and bycatch samples, and introduced a first nomenclature for loggerhead mitogenomes. Analysis of mtDNA diversity, Maximum Parsimony and Bayesian phylogenetic reconstruction allowed improved resolution of lineages with respect to studies reporting on partial mtDNA control region sequence comparisons, and we were able to further inform previous analyses on loggerhead ancestry based on control region haplogroups. Overall, whole mitogenome analysis has potential for considerable improvement of evolutionary history and phylogeographic investigations as well as mixed-stock surveys of loggerhead turtles.

Keywords Whole mitogenome sequencing · Mediterranean Sea · *Caretta caretta* phylogeny · Sea turtle population genetics

Introduction

Female natal philopatry is the tendency for individual females to breed at or near their place of birth (Shields 1982). This behaviour can shape population structure in many marine species including whales, sharks and chelonids (Miller 1997; Sheridan et al. 2010; Baker et al. 2013; Feldheim et al. 2014). Each breeding site generally hosts specific aggregates of maternal descent. However, exceptions episodically occur which lead a female to colonise new sites and significantly increase the species' range (e.g. Carreras et al. 2018).

Population genetic approaches to study female philopatry have mostly relied upon mitochondrial DNA (mtDNA), which is maternally inherited and can therefore trace matrilines within populations (Moritz 1994). In particular, investigations on whole mtDNA genome sequences have significantly improved our understanding of intraspecific phylogeographic patterns in several marine taxa. For instance, mitogenome sequence analysis in killer whales helped clarifying phylogenetic relationships amongst sympatric ecotypes which were then classified as different species (Morin et al. 2010). In speartooth shark and sawfish, mitogenome sequences comparison allowed to better resolve population structure and inform effective conservation measures

Responsible Editor: C. Eizaguirre.

Livia Tolve and Alessio Iannucci have contributed equally to this work.

Andrea Novelletto Co-senior author.

Extended author information available on the last page of the article

for critically endangered populations (Feutry et al. 2014, 2015).

In sea turtles, female natal philopatry and population genetic structure have long been investigated through the analysis of partial mtDNA control region sequence diversity (Jensen et al. 2013). More recently, whole mtDNA investigations helped characterising *Chelonia mydas* matrilines which could not be resolved by comparison of control region haplotypes alone and allowed definition of fine-scale population structure in the critically endangered *Lepidochelys kempii* (Frandsen et al. 2020). For the loggerhead sea turtle (*Caretta caretta*), a few complete mitogenomes have been characterised worldwide and used to reconstruct interspecific phylogenies (Drosopoulou et al. 2012; Duchene et al. 2012; Hernández-Fernández and Delgado Cano 2018; Otálora et al. 2018). In particular, Drosopoulou et al. (2012) provided the first thorough description of the mtDNA genome of *C. caretta* using a sample collected from a rookery in western Greece, however, more detailed patterns of intraspecific mitogenomic variation were not investigated.

The loggerhead sea turtle can be found in tropical and subtropical seas as well as temperate waters of the Mediterranean basin, which hosts one of the 10 Regional Management Units defined for *C. caretta* to prioritise conservation efforts worldwide (Wallace et al. 2010). The most recent assessment of the IUCN Red List classifies the loggerhead turtles from the Mediterranean as a "least concern" Regional Management Unit. However, populations are conservation-dependent, severely fragmented and characterised by a continuous decline of mature individuals mainly because of habitat encroachment, pollution, fisheries bycatch, climate change and intentional take (Casale et al. 2018, 2020). A variation in the distribution of the species induced by global warming is already detectable (Maffucci et al. 2016; Carreras et al. 2018; Hochscheid et al. 2022) and advocated by projection modelling (Mancino et al. 2022). Conservation efforts can make use of a comprehensive understanding of several life history traits. In particular, the definition of maternal descent of individuals from foraging stocks and the genetic profiling of nesting females and new nesting sites are fundamental to guide management programmes for *C. caretta* in the Mediterranean Sea.

Genetic characterisation of Mediterranean rookeries and foraging stocks for the loggerhead has been so far conducted by comparison of partial mtDNA control region sequences (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). This information was used in conservation-oriented studies where, for instance, mixed-stock analysis described the contributions from known rookeries to aggregations of turtles at sea (Garofalo et al. 2013; Clusa et al. 2014; Splendiani et al. 2017; Tolve et al. 2018; Loisier et al. 2021). To this regard, comparison of control region haplotypes has reached a resolution threshold which cannot allow any further sensible improvement in fine-scale population structure analysis as the main Mediterranean rookeries appear to all share the same control region haplotypes (e.g. Tolve et al. 2018). The very same molecular marker was also used to study patterns of loggerhead Mediterranean colonisation and radiation (e.g. Clusa et al. 2013; Novelletto et al. 2016; Splendiani et al. 2017; Loisier et al. 2021). Nevertheless, limited information could be obtained on the evolution of different lineages because of the scarce resolution amongst matrilines defined by mtDNA control region sequences.

In this study, we produced a first whole mtDNA genome dataset for loggerhead turtles from the Central Mediterranean Sea in order to enhance resolution of matrilines and characterise rookeries which could be not distinguished by comparison of ubiquitous control region haplotypes. We obtained high-coverage mitogenome sequences from dead embryos collected at nesting sites and from bycaught and stranded turtles. Eight out of 27 nests were located in minor or occasional nesting areas, which were genetically described for the first time. We assessed intraspecific genetic diversity and performed phylogenetic analysis to reconstruct evolutionary relationships amongst matrilines and the time frame of radiation of haplogroups and mtDNA haplotypes within haplogroups. Using additional, published loggerhead mtDNA genome sequences, we provided an overview of mitogenomic diversity for *C. caretta* in the Central Mediterranean. We argue that our findings will help supporting detailed inference in mixed-stock analysis and foster thorough investigation of the phylogeographic history of loggerhead turtles.

Materials and methods

Sampling, DNA isolation and sequencing by synthesis

A total of 61 tissue biopsies were obtained from loggerhead turtles sampled along the coastline of the Italian peninsula and the island of Linosa and in the sea waters between Sicily and Tunisia from 2008 to 2021 (Fig. 1). Samples were collected from stranded turtles in Veneto and Emilia Romagna ($N=4$), Tuscany ($N=1$) and Latium ($N=2$) and from bycaught turtles in the Strait of Sicily ($N=27$). Samples were also collected from dead embryos from four nests in Tuscany, two nests in Latium, 19 nests in Calabria and two nests on the island of Linosa (Table S1). Numbered tags were applied to bycaught turtles before release.

Each sample was preserved in a cryovial with 95% ethanol and subsequently stored at − 80 °C. Whole DNA was extracted following a standard phenol:chloroform:isoamyl alcohol procedure (Green and Sambrook 2012). Integrity

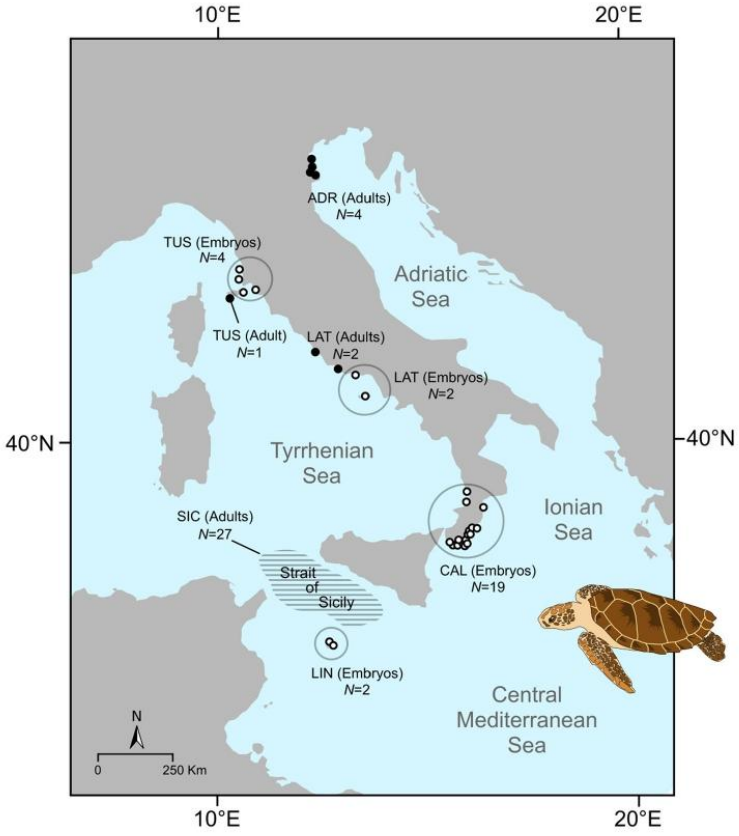


Fig. 1 Map of the study area. A total of 61 samples of *Caretta caretta* were collected between 2008 and 2021 along the coastline of the Italian peninsula and the Island of Linosa and in the sea waters of the Strait of Sicily. Nesting sites (circles) and stranded turtles (black dots) were sampled in Veneto and Emilia Romagna (ADR), Tuscany (TUS), Latium (LAT), Calabria (CAL) and the Island of Linosa (LIN). The horizontal shaded region shows location of the bycatch area between Sicily and Tunisia

of DNA was assessed by 1% agarose gel electrophoresis and DNA concentration was measured using a Qubit 4 fluorometer Broad Range Assay (Invitrogen). Mitochondrial DNA whole genome sequencing was performed by first generating dual-indexed (i5 and i7) genomic libraries using a PCR-free Tagmentation Kit and IDT for Illumina indexes set A (Illumina) according to the manufacturer's protocol. We envisaged a 1 × target nuclear genome coverage for all samples in order to obtain a high depth coverage (> 200 ×) for the mitochondrial DNA, which is represented by a much higher number of copies per cell than nuclear DNA (Michaels et al. 1982; Robin and Wong 1988; Wiesner et al. 1992). Considering a nuclear DNA genome size of 2.13 Gb (Chang et al. 2023) and a mtDNA genome size of 16,737 bp for *C. caretta* (Drosopoulou et al. 2012), we sequenced the libraries paired-end on an Illumina NovaSeq 6000 System using a 300-cycle NovaSeq 6000 SP Reagent Kit v 1.5 for an expected yield of approximately 250 Gb.

Mitogenomes reconstruction and alignment

Demultiplexing and conversion of sequencing data from bcl to fastq format were performed using bcl2fastq version 2.20 (Illumina). FastQC 11.9 (Andrews 2010) was used to check for reads quality. Adaptors and low-quality reads were removed using Trimmomatic 0.39 (Bolger et al. 2014). Reads were mapped to the *C. caretta* complete annotated mitochondrial genome (GenBank accession number: FR694649.1) published by Drosopoulou et al. (2012) using Bowtie2-2.5.0 (Langmead and Salzberg 2012). This mitogenome was chosen as reference for it was the only one published for *C. caretta* from the Mediterranean basin. Sequence Alignment Map (SAM) files were compressed in Binary Alignment Map (BAM) format, sorted and indexed using Samtools 1.16.1 (Danecek et al. 2021). Variant calling was performed with BCFtools 1.16 (Danecek et al. 2021) and Tablet alignment viewer (Milne et al. 2013) was used to visually check variants and relative coverage. Consensus

sequences were extracted from Variant Calling Format (VCF) files using BCFtools by filtering out variants in the first 10 bp and those with a QUAL value < 30.

Mitogenomes alignment was performed by including five additional loggerhead mtDNA genome sequences available in GenBank (see also Table S2): FR694649.1 (Greece; used as reference mitogenome), JX454983.1 (Florida), JX454988.1 (Hawaii), MF554690.1 and MF579505.1 (Colombia) (Drosopoulou et al. 2012; Duchene et al. 2012; Hernández-Fernández and Delgado Cano 2018). The MUSCLE algorithm was used to produce a multiple sequence alignment, which was trimmed at position 16,419, ahead of the (TATAT)n microsatellite repeat and annotated using Geneious Prime 2022.2. We then partitioned the alignment in (i) 12S and 16S rRNAs, (ii) tRNAs, (iii) the control region and (iv) coding genes. The final nucleotide dataset had no ambiguous positions and was numbered with reference to the FR694649.1 mitogenome (Drosopoulou et al. 2012). Genbank accession numbers are reported in Table S1 for each complete mtDNA haplotype sequence described in this study.

Mitogenomic haplotype nomenclature

Mitochondrial haplotypes were named following the criteria set by Shamblin et al. (2012a) and the guidelines of the Archie Carr Center for Sea Turtles Research (https://accstr.ufl.edu/resources/mtdna-sequences/). Complete mtDNA haplotypes were defined by three serial numbers. The first two numbers refer to the 815 bp long *C. caretta* control region haplotype code, whilst the third number refers to variants recorded in other genes of the mitogenome outside of the control region, so that complete mtDNA haplotype names are basically variants of the current control region haplotype nomenclature. For example, CC-A2.1.1, CC-A2.1.2 and CC-A2.1.3 all refer to complete mtDNA haplotype variants of the control region haplotype CC-A2.1.

Nomenclature for groups of complete mtDNA haplotypes also followed that of *C. caretta* control region variants, which identify three main haplotype clusters named haplogroups IA, IB, and II (Shamblin et al. 2014). Haplogroup IA includes haplotypes characterising Pacific Ocean nests, haplogroup IB includes haplotypes which are found in Atlantic rookeries and one single haplotype recorded in Indian Ocean nests (Shamblin et al. 2014). Finally, haplogroup II comprises haplotypes from both Atlantic and Mediterranean rookeries.

Mitogenomic diversity and phylogenetic analysis

Number of haplotypes, haplotypic diversity, number of variants, nucleotide diversity and GC content were computed using DnaSP (Librado and Rozas 2009) and the R package pegas v1.1 (Paradis 2010). Analysis was performed on whole mitogenome sequences and then single mitochondrial genes using the entire dataset, first and then on haplogroup II only. McDonald–Kreitman test was implemented using DnaSP to check for neutrality of substitutions in the coding regions (McDonald and Kreitman 1991). The nonsynonymous (dN) to synonymous (dS) substitution rate ratio amongst haplogroup II sequences was then compared to the dN/dS ratio between haplogroup II sequences and mitogenomes of haplogroup IB (obtained by this study and from GenBank, accession number: MF579505.1). In order to rule out parallel patterns of nucleotide variations between haplogroup II and IB, the dN/dS ratio amongst haplogroup II variants was measured against the dN/dS between haplogroup II and the mitogenome of *Lepidochelys olivacea* as an outgroup (GenBank accession number: JX454987.1). Finally, differences in dN/dS amongst mitogenomes of the entire *C. caretta* dataset of our study were compared to dN/dS between our dataset and the mtDNA genome of *L. olivacea.*

Two haplotype networks were constructed using statistical parsimony (Templeton et al. 1992). The method, implemented in the TCS programme (Clement et al. 2000), links haplotypes with the smaller number of differences as defined by a 95% confidence criterion. The first network was based on 815 bp control region haplotypes only, whilst the second one was a network of complete mitogenome haplotypes.

A Maximum Parsimony (MP) graph based on the entire dataset was also constructed using MEGA11 (Kumar et al. 2018) with the "use all sites" option in order to visualise similarities between sequences and polarise mutational changes. We then applied a Bayesian divergence time analysis to provide a time frame for the process of radiation of haplogroup II with respect to the other two major *C. caretta* mitochondrial haplogroups IA and IB. Phylogeny and divergence time analyses provided poor resolution when using *L. olivacea* as outgroup. We therefore conducted MP and Bayesian inference of phylogeny using haplogroup IB as reference in order to better resolve intraspecific topology of haplogroup II, which included the vast majority of haplotypes defined in our study. We used BEAST suite v.1.10.4 (Suchard et al. 2018) on four mitogenome partitions (12S and 16S rRNAs, tRNAs, the control region and coding genes). We performed 20 independent test-runs in order to optimise priors for chain convergence and defined, for each partition, the "unlink" option flagged for substitution model and for molecular clock and the "link" option flagged for generation of the tree. The HKY (Hasegawa et al. 1985) substitution model and a strict molecular clock were selected for each partition. The "unlink parameters on codon positions" option was flagged for the fourth partition. The HKY model was selected because the tests conducted using more complex substitution models gave a poor Effective Sample Size (ESS), possibly due to over-parametrization. Likewise,

we chose a strict molecular clock because multiple tests conducted using more complex uncorrelated relaxed clocks resulted in poor ESS and zero values for the parameters uced. mean and meanRate, which indicate a *clock-like* behaviour of the data. No prior was set on the time to the most recent common ancestor (TMRCA) and no divergence time estimate was assessed for the root of the tree. We converted the substitutions value from (site × million year)$^{-1}$ to (site × generation)$^{-1}$, where the generation time is calculated as age of maturity plus half of reproductive longevity (Pianka 1973). We used 25 years as age of maturity and 22 years as reproductive longevity which resulted in a 36-year generation time (Margaritoulis et al. 2020; Baldi et al. 2023). We set a normal distribution (mean = 1.2E-7, S.D. = 4.0E-8) as prior on the substitution rates for each mitogenomic partition. The distribution was truncated with a lower bound equal to 0.0 and an upper bound of 2.41E-7 (corresponding to 6.7E-3 million year^{-1}), which is the highest value for mitochondrial substitution rates recorded in marine turtles (Dutton et al. 1999). Chain length was set to 10E8 steps with a 10% initial burn-in. We monitored the ESS for each output parameter after each run using Tracer v.1.7.1 (Rambaut et al. 2018). The overall BEAST output was summarised using TreeAnnotator (Drummond and Rambaut 2007) and the statistics related to the consensus tree nodes including node height, 95% Height Posterior Density (HPD) and posterior, were assessed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Mitogenomic diversity and complete mtDNA haplotype assignment

Whole mtDNA sequences were obtained for all 61 samples with an average genome coverage (×) of 272.4 ± 37.1SE (see Table S3). Mitogenome sequences were trimmed at position 16,425 so that the $(TATAT)_n$ microsatellite repeat positioned at the 3' end of the control region was discarded from the analysis.

We recorded two mtDNA sequences in sample SIC04. We reported this pattern as a possible occurrence of heteroplasmy and assigned to SIC04 the mitogenomic haplotype name of the mtDNA variant (G > A at position 4671 of the genome) that was confirmed in 70% of the reads (Table S1).

All samples belonged to haplogroup II apart from the stranded adult TUS05, of unknown origin, which was assigned to haplogroup IB, control region haplotype CC-A1.1. We named this mitogenomic haplotype CC-A1.1.1. Downloaded mitogenome sequences FR694649.1 (our reference genome), JX454983.1 and MF554690.1 were assigned to haplogroup II, while MF579505.1 and JX454988.1 were assigned to haplogroups IB and IA, respectively (Table S2).

Haplogroup II sequence variants and their distribution across sampling sites are reported in Tables 1 and 2, respectively. The loggerhead mitogenomes characterised in this study showed a total of 27 polymorphic sites with respect to the *C. caretta* reference mtDNA FR694649.1. Of these, two segregating sites were recorded across all 61 samples of our dataset and in particular an indel in position 1631 in the 16S rRNA gene and a T > C substitution in position 9974 in the ND4L gene (Table1). Mitogenomic haplotype CC-A2.1.1 was the most abundant both in embryos and adult turtles (found in 26 out of 61 samples). The second most abundant haplotype was CC-A2.1.8, found in six samples from the Calabrian nests, one sample from the Island of Linosa and three samples collected from bycaught adults in the Strait of Sicily. Seven embryos from different nests in Calabria carried mitogenomic haplotype CC-A20.1.1. This mitogenome was not found elsewhere in the study area. Haplotype CC-A2.1.3 was found in three specimens in Tuscany and southern Italy. All the other mitogenomic haplotypes were represented by either one or two individuals. The nests from Tuscany, Latium and Linosa Island were characterised by haplotypes CC-A2.1.1 (TUS01, TUS02, LAT01, LAT02), CC-A2.1.3 (TUS03), CC-A2.1.9 (TUS04), CC-A2.1.8 (LIN01) and CC-A2.9.2 (LIN02) (Table S1).

Overall, we characterised 286 variable sites with respect to 55 polymorphisms recorded by comparing control region sequences alone and assigned samples to 23 different mitogenomic haplotypes. Ten haplotypes only were defined by control region sequence analysis (Table 3). Within haplogroup II, mitogenome sequences identified 27 variable sites and 20 haplotypes compared to six variable sites and seven haplotypes recorded by control region sequence comparison (Table 3; Fig. 2).

Genetic diversity analysis of whole mtDNA revealed a much higher number of haplotypes and higher haplotypic and nucleotide diversity than values recorded by control region sequences comparison (Table 3). Amongst the 16 gene partitions of the *C. caretta* mitogenome, the ND5 gene showed the highest number of segregating sites and haplotypic diversity (Tables S4 and S5). The same pattern was recovered by the analysis of haplogroup II mitogenomes for the high abundance (28%) of mitogenomic haplotypes CC-A2.1.8 and CC-A2.1.20 carrying the variant T at site 13,373 of the ND5 gene. Overall, nucleotide diversity (π) showed a threefold variation along the mtDNA molecule, with peak values at positions 3000–5000, 9000 to 10,000 and 13,000 to 14,000 (Fig. S1). When haplogroup II sequences only were considered, π values were the highest at positions 9000–10,000 and 13,000–14,000. The NADH dehydrogenase and ATP synthase subunit genes displayed the highest π values. Conversely, analysis restricted to haplogroup II

Table 1 Mitochondrial genome polymorphic sites for *Caretta caretta* haplogroup II mitogenomic haplotypes

Mitogenomic haplotypes	MtDNA gene													
	16S	ND1		ND2		Cys	COI	COII	ATP8	ATP6		COIII		Gly
	1631	2936	3045	4277	4671	5303	5847	7433	7900	8398	8661	9102	9440	9482
FR694649.1	T	C	C	G	G	A	C	C	C	C	T	G	T	A
MF554690.1														
JX454983.1	–						A							
CC-A2.1.1	–													
CC-A2.1.2	–				A									
CC-A2.1.3	–								T					
CC-A2.1.4	–										C			
CC-A2.1.5	–											A		
CC-A2.1.6	–													
CC-A2.1.7	–													
CC-A2.1.8	–													
CC-A2.1.9	–													
CC-A2.1.10	–	T							T					
CC-A2.1.11	–		T											G
CC-A2.9.1	–									T			A	
CC-A2.9.2	–					G				T			A	
CC-A3.1.1	–													
CC-A10.4.1	–													
CC-A20.1.1	–													
CC-A43.1.1	–			A				T						

Mitogenomic haplotypes	MtDNA gene												
	ND3	ND4L	ND4	Leu	ND5		CYTB	Control region					
	9610	9974	10,591	11,800	13,287	13,373	14,373	15,562	15,645	15,731	15,800	15,970	15,975
FR694649.1	T	T	T	G	G	C	G	T	C	A	G	A	A
MF554690.1				A									
JX454983.1		C											–
CC-A2.1.1		C											
CC-A2.1.2		C											
CC-A2.1.3		C											
CC-A2.1.4		C											
CC-A2.1.5		C											
CC-A2.1.6		C	C										
CC-A2.1.7		C			A								
CC-A2.1.8		C				T							
CC-A2.1.9		C					A						
CC-A2.1.10		C											
CC-A2.1.11		C											
CC-A2.9.1		C						C					
CC-A2.9.2		C						C					
CC-A3.1.1		C									A		
CC-A10.4.1		C										G	
CC-A20.1.1		C				T			T				
CC-A43.1.1	C	C								G			

Numbers refer to nucleotide positions of the *C. caretta* reference mtDNA FR694649

Table 2 Number of whole mtDNA and control region haplotypes recorded for 61 loggerhead samples collected along the coastline of the Italian peninsula and the island of Linosa and in the sea waters of the Strait of Sicily

MtDNA	Control region	ADR	TUS	LAT	CAL	LIN	SIC	Total
CC-A1.1.1	CC-A1.1	0	1	0	0	0	0	1
CC-A2.1.1	CC-A2.1	2	2	3	4	0	15	26
CC-A2.1.2		0	0	0	0	0	1	1
CC-A2.1.3		0	1	0	1	0	1	3
CC-A2.1.4		0	0	0	0	0	1	1
CC-A2.1.5		0	0	0	0	0	1	1
CC-A2.1.6		0	0	1	0	0	0	1
CC-A2.1.7		0	0	0	1	0	0	1
CC-A2.1.8		0	0	0	6	1	3	10
CC-A2.1.9		0	1	0	0	0	1	2
CC-A2.1.10		0	0	0	0	0	1	1
CC-A2.1.11		0	0	0	0	0	1	1
CC-A2.9.1	CC-A2.9	1	0	0	0	0	0	1
CC-A2.9.2		0	0	0	0	1	0	1
CC-A3.1.1	CC-A3.1	0	0	0	0	0	1	1
CC-A10.4.1	CC-A10.4	1	0	0	0	0	0	1
CC-A20.1.1	CC-A20.1	0	0	0	7	0	0	7
CC-A43.1.1	CC-A43.1	0	0	0	0	0	1	1
Total		4	4	2	19	2	27	61

ADR Veneto and Emilia Romagna, *TUS* Tuscany, *LAT* Latium, *CAL* Calabria, *LIN* Island of Linosa, *SIC* Strait of Sicily

Table 3 Genetic diversity parameters for whole mitogenomes and control region sequences only

		N	Sequence length (bp)	*k*	*P*	*h* ± SE	π ± SE	GC
All samples	MtDNA	66	16,425	23	286	0.816 ± 0.005	9.44E-4 ± 5.82E-5	0.392
	Control region	66	919	10	55	0.419 ± 0.009	3.26E-3 ± 2.46E-4	0.340
Haplogroup II	MtDNA	63	16,425	20	27	0.798 ± 0.006	9.58E-5 ± 8.09E-10	0.392
	Control region	63	919	7	6	0.362 ± 0.009	3.92E-4 ± 2.29E-8	0.340

Values are reported for 61 samples plus five mitogenomes from published literature and a subset of haplogroup II mtDNA only (60 samples from this study and three sequences from published literature)

N number of samples, *k* number of haplotypes, *P* number of polymorphic sites, *h* haplotypic diversity ± Standard Error, π nucleotide diversity ± Standard Error, *GC* content of GC bases

mitogenomic haplotypes showed the highest π values for the ATP8 gene only (Tables S4 and S5).

The TCS haplotype networks clustered both the partial mtDNA control region sequences and the mitogenomes in three unlinked haplogroups with 95% probability (Fig. 2). Links between haplogroups were obtained with a lower probability threshold by fixing connexion limits at 100 and 200 steps for control region and mitogenome haplotypes, respectively. Two internal ambiguities in the mitogenome cladogram were resolved by applying criteria derived from empirical validation of the coalescent theory (Crandall and Templeton 1993; Templeton and Sing 1993; Fetzner and Crandall 2003). Accordingly, a haplotype connexion was maintained to high-frequency haplotypes with an interior position in the network (Castelloe and Templeton 1994) rather than to low-frequency ones located in tip clades. Internal ambiguities could not be resolved for the mtDNA control region haplotypes network. We therefore maintained the three haplogroups unlinked. Haplogroup IA was represented by a single sequence in both reconstructions, while haplogroup IB included two haplotypes separated by 12 and 30 possible unobserved intermediate haplotypes in the partial mtDNA control region and mitogenomic network, respectively. The haplogroup II network based on control region sequences had a star-like configuration with a core CC-A2.1 haplotype and six equidistant derived haplotypes. A similar but more complex pattern was described by the mitogenomic analysis. Haplogroup II showed 10 equidistant haplotypes deriving from CC-A2.1.1, two haplotypes (CC-A2.1.10 and CC-A20.1.1) deriving from CC-A2.1.3 and CC-A2.1.8, respectively and five additional

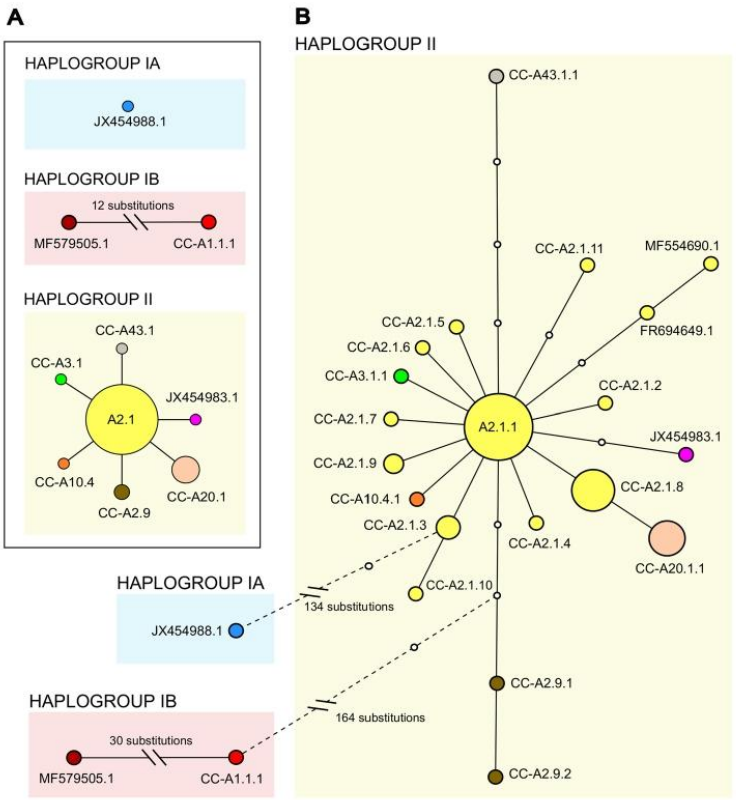


Fig. 2 TCS haplotype networks based on partial mtDNA control region (**A**) and complete mitogenome (**B**) sequences. Dashed lines represent links with probability lower than 95%

branches with haplotypes linked to CC-A2.1.1 by unobserved intermediate haplotype states.

Variation in the protein-coding regions

The McDonald-Kreitman neutrality test returned no significant values when comparing dN/dS of haplogroup II with dN/dS between haplogroup II and either haplogroup IB or the mitogenome of *L. olivacea* (Table S4). We found an excess of synonymous substitutions in all genes when comparing haplogroup II with haplogroup IB sequences and a similar result was obtained when comparing haplogroup II variants to the mitogenome of *L. olivacea*. Ten variants in the coding regions of haplogroup II sequences were synonymous and six variants were non-synonymous. In particular, ND2, ND3 and CYTB genes displayed no synonymous substitutions and a non-synonymous one. Comparison of dN/dS of our complete *C. caretta* dataset with dN/dS between *C. caretta* mitogenomes and *L. olivacea* mtDNA revealed an excess of non-synonymous substitutions for ND1, ND2, ATP6, ND3, ND5 and CYTB genes. Nevertheless, a significant value was shown for the ND3 gene only.

Phylogenetic analyses

The phylogenetic tree inferred by Maximum Parsimony resolved equally parsimonious topologies whereby haplogroup II (Atlantic and Mediterranean) and haplogroup IA (Pacific) shared the same cluster, while haplogroup IB (Atlantic) included the most phylogenetically divergent sequences (Fig. 3). All haplogroup II sequences coalesced to a long branch defined by 55 substitutions. Of these, 18 substitutions occurred in the control region, three were found in rRNA genes, one in a tRNA and 33 in coding genes with 28 synonymous and five non-synonymous substitutions (Table S6).

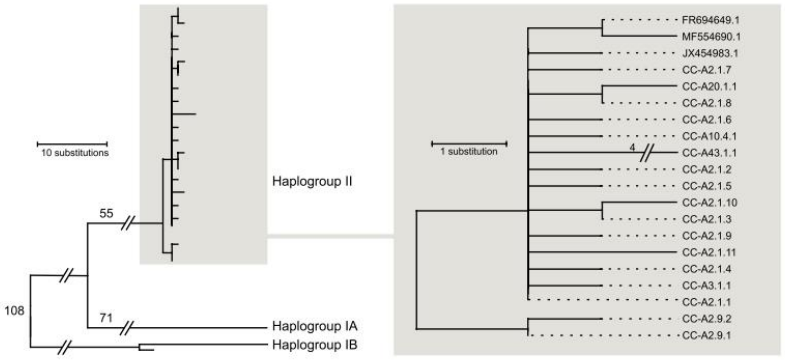


Fig. 3 Unrooted maximum parsimony (MP) tree of *Caretta caretta* obtained using 61 mitogenomes from this study and five sequences from published literature (left). Haplogroup IA (Pacific Ocean) corresponds to Genbank accession number JX454988.1. Haplogroup IB (Atlantic Ocean) includes Genbank accession number MF579505.1 and mitogenomic haplotype TUS05 from this study (see Tables S1 and S2). Details of the haplogroup II subtree (right) show phylogenetic relationships amongst mtDNA sequences of 60 samples from this study (Mediterranean Sea) and three sequences from published literature (Mediterranean Sea and Atlantic Ocean). Numbers on tree branches represent number of nucleotide substitutions

Two main clades were defined within haplogroup II (Fig. 3). The clade characterised by the rare Mediterranean control region haplotype CC-A2.9 differed from the other haplogroup II sequences by the control region variant C at position 15,562, the derived state C > T at position 8398 in the ATP6 gene and the ancestral state A at position 9440 in the COIII gene. Mitogenome analysis further distinguished mitogenomic haplotypes CC-A2.9.1 and CC-A2.9.2 as a result of a A/G variant at position 5303 in the Cys tRNA gene (Fig. 3, Table S7). The second and most conspicuous clade was defined by a basal A > T transversion at position 9440 and included all CC-A2.1 mitogenomic variants as well as haplotypes corresponding to control region haplotypes CC-A3.1, CC-A10.4, CC-A20.1 and CC-A43.1.

Of the CC-A2.1 mitogenomic variants, six of them (CC-A2.1.2 to CC-A2.1.7 and CC-A2.1.9) derived each by one private variant from the basal CC-A2.1.1 mitogenomic haplotype. Similarly, haplotypes CC-A3.1.1 and CC-A10.4.1 each differed from CC-A2.1.1 by a single control region variant. Mitogenomic haplotype CC-A2.1.10 stemmed from CC-A2.1.3 by a synonymous substitution in the ND1 gene, while haplotype CC-A2.1.11 differed from the basal CC-A2.1.1 by two variants. Mitogenomic haplotype CC-A43.1.1 was the most divergent sequence characterised by two non-synonymous, two synonymous substitutions and the control region variant. Haplotype CC-A20.1.1 derived from CC-A2.1.8, which in turn stemmed from the basal clade by a C > T variant in position 13,373 in the ND5 gene (Table S7). All CC-A20.1 sequences were identical, denoting a recent common origin. Finally, the Atlantic mitogenomes JX454983.1 and MF554690.1 derived from CC-A2.1.1 by a single synonymous substitution and from the Mediterranean reference FR694649.1, respectively (Fig. 3, Tables S6 and S7).

The Bayesian tree confirmed the branching order of the haplogroups recovered by the MP tree (Fig. 4 and Table 4). The root of the tree (node I) separated haplogroup IB from haplogroups IA and II. The second node, positioned at 67% of the height of the total tree, divided haplogroup IA from haplogroup II mitogenomic haplotypes. According to Bayesian phylogeny, radiation of haplogroup IB (node III) occurred well before the branching of haplogroup II, which occupies only 4% of the total tree height. The Bayesian analysis also confirmed the basal position of the CC-A2.9 clade within haplogroup II. The only discrepancy between MP phylogeny and Bayesian reconstruction was the position of mitogenomic haplotype CC-A43.1.1, which resulted basal, instead of derived, with respect to CC-A2.1 variants. Other nodes that reached strong support (> 95% posterior probability) were those bifurcating CC-A20.1.1 and CC-A2.1.8, CC-A2.9.1 and CC-A2.9.2, CC-A2.1.3 and CC-A2.1.10, MF554690.1 and FR694649.1. All these nodes confirmed the phylogenetic clustering of the MP tree. The radiation of these mitogenomic haplotypes resulted to be a very recent event for they occupied just 1% of the total tree height. In particular, the most recently diverged mitogenomic sequences were FR694649.1 and MF554690.1, sampled in Greece and in the Atlantic Ocean, respectively (Table 4). Substitution rate estimates obtained for each mitogenome partition are reported in Table S8. The highest substitution rate (number of substitutions × (site × million year)$^{-1}$) was recorded for the control region. The coding regions had a substitution rate three times lower than the control region while the rRNAs' partition showed the lowest values.

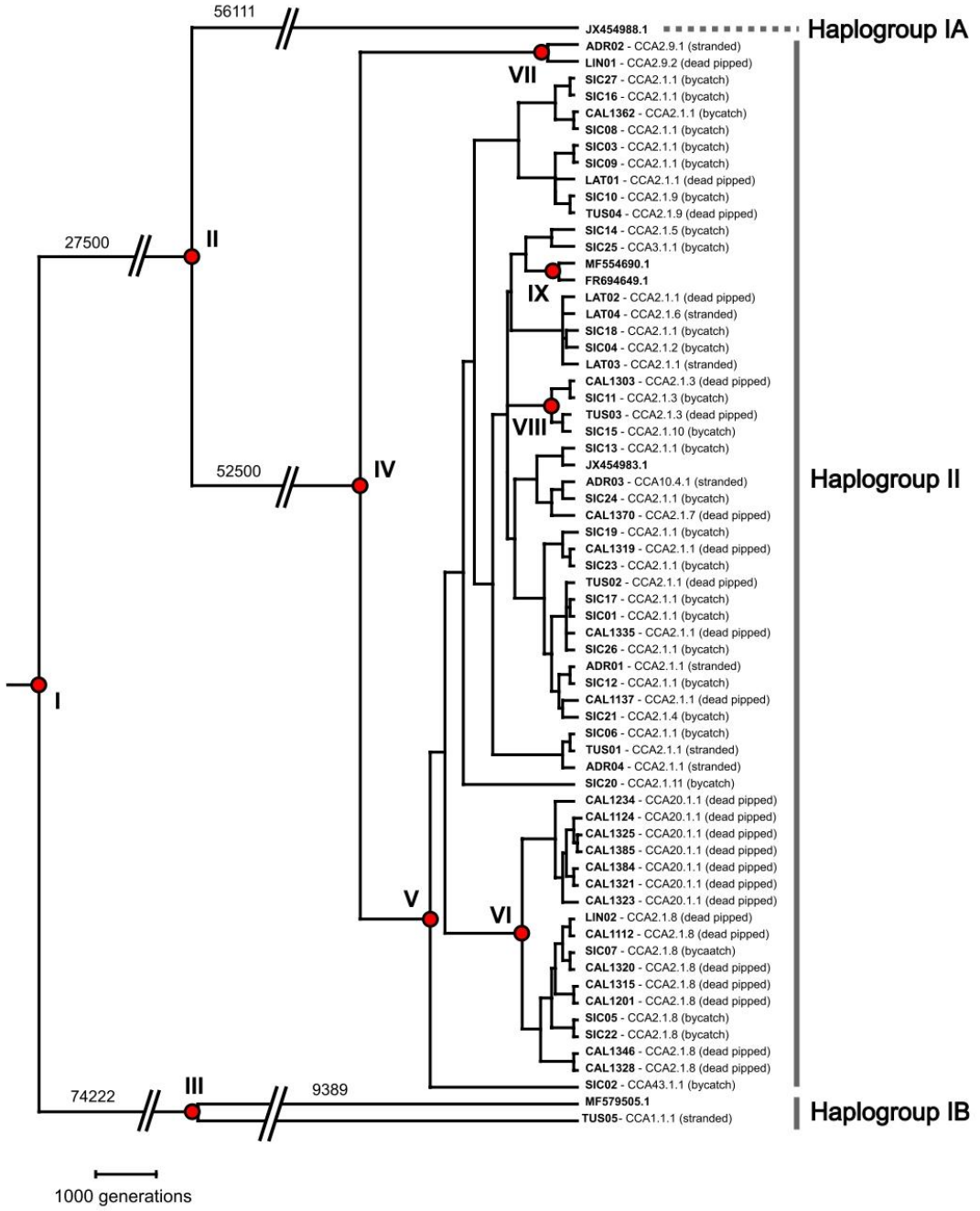


Fig. 4 Bayesian tree of *Caretta caretta* based on 61 mitogenomes from this study and five sequences from published literature. Nodes supporting > 95% posterior probability are shown in red. BEAST coerces multifurcations of identical sequences, so that nodes separating sequences assigned to the same mitogenomic haplotype should not be considered. Numbers on tree branches represent number of generations. Sample names and haplotypes assignment as for Fig. 1 and Table S1, respectively

Table 4 Bayesian tree's height estimates and 95% Height Posterior Density (HPD) for tree nodes with > 90% support. Estimates are reported in number of *Caretta caretta* generations (1 generation = 36 years) and in thousands of years before present (KYBP)

Node	Node description	Median (generations)	Median (KYBP)	95% HPD (generations)	95% HPD (KYBP)
I	Haplogroup IB – Haplogroups IA and II	83,611	3.01E3	49,444–132,500	1.78E3–4.77E3
II	Haplogroup IA – Haplogroup II	56,111	2.02E3	32,222–902,777	1.16E3–3.25E3
III	CCA1.1.1—MF579505.1	9389	338	4500–16,555	162–596
IV	Haplotypes CCA2.9 – other mitotypes in Haplogroup II	3611	130	1555–6388	56–230
V	CCA43.1.1—other mitotypes in Haplogroup II	2444	88	1166–4305	42–155
VI	CCA20.1.1—CCA2.1.8	917	33	278–1889	10–68
VII	CCA2.9.1—CCA2.9.2	500	18	8–1694	0.28–61
VIII	CCA2.1.3—CCA2.1.10	389	14	3–1778	0.10–64
IX	MF554690.1 – FR694649.1	278	10	3–833	0.10–30

Discussion

Improved resolution in loggerhead mitochondrial genetic diversity

We characterised the mitochondrial genomes of 61 loggerhead turtles sampled between 2008 and 2021 along the coastline of Italy, on the Island of Linosa and in sea waters of the Strait of Sicily and provided a first dataset on whole mtDNA sequences for Mediterranean *C. caretta* populations. Mitogenome analysis allowed an improved resolution of loggerhead matrilines with respect to previous studies based on partial mtDNA control region sequences comparison. The TCS haplotype networks based on statistical parsimony delineated similar star-like topologies in haplogroup II for both partial mtDNA control region and complete mtDNA analysis. However, statistical parsimony showed the increased resolution of the mitogenome haplotype network with respect to the 815 bp control region reconstruction, particularly for haplotype CC-A2.1.

The control region haplotype CC-A2.1 is the most abundant amongst the 13 main rookeries so far described in the Mediterranean Sea and it is also widespread in Atlantic nesting sites (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). The ubiquity of haplotype CC-A2.1 has often hampered the characterisation of rookeries for turtles of unknown origin. In particular, mixed-stock analyses appeared to suffer from large confidence intervals associated to estimates of contribution by rookeries to specific mixed stocks. Confidence intervals with a lower value equal to zero most probably underestimated contribution of minor but perhaps genetically important rookeries (Garofalo et al. 2013; Clusa et al. 2014; Splendiani et al. 2017; Tolve et al. 2018; Loisier et al. 2021). In our study, 52 individual turtles all shared the same control region haplotype CC-A2.1 whilst whole mitogenome analysis recovered 11 distinct mitogenomic haplotypes (CC-A2.1.1-CC-A2.1.11), which significantly improved genetic characterisation of geographically distinct nesting sites. While the CC-A2.1.1 variant resulted the most abundant in our dataset with 28 records out of a total of 61 samples, characterisation of the other 10 mitogenomic haplotypes enabled distinction of nesting sites from approximately 30 km down to only 5 km apart along the coastline of Tuscany, Latium and Calabria. These nests could not be distinguished by mtDNA control region sequences analysis alone.

Similarly, haplotype CC-A1.1, which was carried by TUS05 and never found in Mediterranean rookeries, is very abundant and widespread in the Atlantic and therefore poorly informative of turtles' provenance (Shamblin et al. 2012b). It is well known that CC-A1.1 turtles enter the western Mediterranean to forage and then migrate back to the Atlantic for reproduction (Bolten et al. 1998; Laurent et al. 1998; Carreras et al. 2006, 2011). It may therefore be difficult to envisage a Mediterranean origin for TUS05. Nevertheless, future mtDNA characterisation for Atlantic and Mediterranean loggerheads may unveil a number of mitogenomic variants of CC-A1.1 and allow description of differences in mtDNA genome frequencies across rookeries to improve resolution of turtles' origin.

Our mitogenomic approach also improved the resolution of matrilines based on mtDNA control region haplotype CC-A2.9. This is a rare Mediterranean haplotype so far recorded only in Libya and in a small Israeli rookery. The two mitogenomic haplotypes CC-A2.9.1 and CC-A2.9.2 were found in a stranded adult in the northern Adriatic Sea and most important in one nest on the Island of Linosa, respectively. This is also the first evidence of a female carrying the rare haplotype nesting in a rookery other than Libya or Israel.

Mitogenomes comparison showed that in Calabrian nests there are at least five different mitogenomic haplotypes instead of only two variants (CC-A2.1 and CC-A20.1)

detected by the analysis of mtDNA control region sequences. Of these, haplotype CC-A20.1.1 remains exclusive to the Calabrian rookery as for previous studies reporting on mtDNA control region haplotype CC-A20.1. However, we did not record mitogenomes characterised by control region haplotype CC-A31.1, also private to Calabrian nests (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Clusa et al. 2014).

Genetic characterisation of occasional and minor nesting sites

Major Mediterranean loggerhead rookeries are found in the eastern and southern sections of the basin from Greece through Turkey, Cyprus, Lebanon, Israel to Libya (Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014). The western and central parts of the Mediterranean have been so far home of occasional nests and it was only in the last decade that nesting events were recorded on a more frequent and regular basis (Casale et al. 2018). According to Hochscheid et al. (2022), the number of turtle nests west of the main rookeries has increased from three nests per year in 2013 to 84 nests recorded in 2020, suggesting a 1300 km north-westward range expansion of loggerhead nesting distribution (Mancino et al. 2022). Samples collected from nests in Tuscany and Latium are evidence of such latitudinal dispersal, which seems now possible also as a result of a steady increase in water temperature (Casale et al. 2018; Hochscheid et al. 2022). All four nests recovered in Tuscany and the two nests from Latium had the same control region haplotype CC-A2.1, a highly shared sequence across both the Mediterranean basin and the Atlantic. On the other hand, analysis of mitogenome sequences recovered the presence of at least three distinct nesting females in this region of the northern Tyrrhenian Sea. Of these, a female that laid eggs in southern Tuscany (TUS03, see Table S1) carried mitogenomic haplotype CC-A2.1.3, which was also found in embryos from a nest on the Ionian coast of the Calabrian rookery (CAL1303). This pattern may suggest a southern Italian origin of the occasional nests recorded in Tuscany possibly as a result of the ongoing northward expansion of the distribution range of loggerheads in the Mediterranean (Mancino et al. 2022). As much as our evidence is based on a single observation, it nevertheless provides valuable information to further investigate dispersal of pioneer female loggerheads and the propagation of specific matrilines for the establishment of new rookeries.

In the Pelagian archipelago, Mingozzi et al. (2006) recorded nesting events on the Islands of Lampedusa and Linosa between 1985 and 1999 and regular, but not annual, nesting activities from 2000 to 2004. Nevertheless, genetic data was not produced during these surveys, so that ours represents the first study characterising the mitochondrial signature of the region. The two nests sampled in 2006 and 2008 were assigned to different matrilines, suggesting a relatively high degree of genetic diversity, given the limited size (540 ha) of Linosa Island. This is nevertheless coherent with the mitogenomic analysis of samples collected in the nearby Strait of Sicily where 11 mitogenomic haplotypes were recorded in 27 bycaught adult loggerheads. Additional surveys should assess the relevance of Linosa island as an additional nesting area for female loggerheads carrying a rare CC-A2.9.2 haplotype, considering that the CC-A2.9 matriline had so far been recorded in Libyan and Israeli nests only and in several individuals foraging in Mediterranean waters (e.g. Garofalo et al. 2013; Splendiani et al. 2017; Tolve et al. 2018). Moreover, the higher resolution offered by a mitogenomic approach could allow analysis of the genetic diversity and genomic differentiation between Linosa and the Island of Lampedusa, for which genetic characterisation of nesting sites is yet to be provided.

Heteroplasmy occurrence and patterns of substitutions in the coding regions

The deep sequencing coverage we obtained for *C. caretta* mitochondrial genomes revealed the presence of two mtDNA variants in one individual bycaught in the Strait of Sicily (SIC04). Heteroplasmy consisted of the coexistence of the CC-A2.1.1 mitogenomic haplotype with a mitogenome sequence carrying a truncating variant at mtDNA position 4671 on the ND2 gene, turning a tryptophan codon into a stop codon. The tryptophan codon located in the ND2 gene is a well-conserved sequence across vertebrates (Avise 1986) and the presence of a termination codon instead of a coding sequence during translation would suggest a possible detrimental effect on gene functioning (Yan et al. 2019). Although a mitogenome carrying such variant may not persist in homoplasy, we assigned a distinct mitogenomic haplotype name (CC-A2.1.2) to specimen SIC04. Tikochinski et al (2020) argued how intra-individual heteroplasmy may not be such a rare occurrence in sea turtles and could instead play a role in the evolution of population genetic diversity across generations. Unfortunately, the limitations of Sanger sequencing in the characterisation of low-frequency variants have impeded the development of studies on heteroplasmy in sea turtles. Nevertheless, regular use of high-throughput sequencing techniques should now lead to an increase in studies investigating this phenomenon and its evolutionary implications.

We found the highest nucleotide diversity in NADH dehydrogenase subunits genes. A similar pattern was previously recorded for genes encoding the ND1 and ND3 subunits (Novelletto et al. 2016). In particular, the ND5 subunit gene showed a relatively high proportion of polymorphism and the highest haplotype diversity. We also found a high

nucleotide diversity in the genes encoding the subunits of ATP synthases across mitogenomes of haplogroup II, suggesting a pattern of sequence radiation within this clade. However, we did not record a significant excess of amino acid substitutions with respect to synonymous substitutions in haplogroup II mitogenomes. This would argue against adaptive selection acting on haplogroup II.

Mediterranean loggerhead phylogeography

The topologies of both Maximum Parsimony and Bayesian trees showed a marked star-like phylogeny of the main clade within haplogroup II, suggesting a fast and recent radiation of mitogenomic haplotypes from CC-A2.1.1 to CC-A2.1.10, CC-A3.1.1, CC-A10.4.1, CC-A20.1.1 and CC-A43.1.1. On the other hand, the sister branch defined by haplotypes CC-A2.9.1 and CC-A2.9.2 was basal in our phylogenetic reconstruction. A similar pattern was found for the CC-A2.9 mtDNA control region haplotype (Novelletto et al. 2016; Splendiani et al. 2017; Loisier et al. 2021), which was until now considered to be exclusive to Libyan and Israeli rookeries. This pattern supports the hypothesis of Clusa et al. (2013) by which the first loggerhead colonies of the Mediterranean settled in Libya during the Pleistocene, approximately 65 KYBP. The southern Mediterranean was probably a glacial refugium from where loggerhead turtles would have dispersed after the Last Glacial Maximum (LGM). Although our dataset only partially represents the Mediterranean mtDNA diversity, the time estimate we obtained for the divergence of the CC-A2.9 haplotype clade was consistent with a Pleistocenic colonisation of Libya. Our phylogeny did not show additional matrilines deriving from mtDNA carrying the control region haplotype CC-A2.9, which was previously recorded to have a close phylogenetic relationship with haplotypes CC-A26.1 and CC-A66.1 (Novelletto et al. 2016). A more comprehensive sampling including Ionian and eastern Mediterranean rookeries may reveal additional mitogenomic variants of the CC-A2.9 clade, which has so far been reported to have a limited radiation with respect to the sister branch CC-A2.1 (Garofalo et al. 2009; Yilmaz et al. 2011; Saied et al. 2012; Carreras et al. 2014; Clusa et al. 2014).

A more recent, second colonisation event of the Mediterranean from the Atlantic during the Holocene was suggested by Clusa et al. (2013) to explain the origin and expansion of the Calabrian rookery in southern Italy. This hypothesis was based on the presence of mtDNA control region haplotype CC-A20.1 in both Calabria and three rookeries in Florida (Shamblin et al. 2012b). The mitogenomic haplotype CC-A20.1.1 described in our study shared a recent ancestor with CC-A2.1.8 and both were recorded only in the southern part of our study area including Calabria, Linosa Island and the Strait of Sicily. This may suggest a local origin of these matrilines after the end of the LGM. Comparison of whole mitogenomic sequences from both Mediterranean and Atlantic stocks will be crucial to help understand origin and dispersal patterns of the CC-A20.1 clade.

On a wider, trans-oceanic scale, a number of biogeographical scenarios have been suggested advocating a first colonisation of the Mediterranean Sea by haplogroup II from Atlantic rookeries (Shamblin et al. 2014) or through a South African route (Bowen et al. 1994; Baltazar-Soares et al. 2020). Our topology recovered by both maximum parsimony and Bayesian analysis provided preliminary information on mitogenome shared ancestry. We envisage that further analysis based on e.g. an Approximate Bayesian Computation approach (see Baltazar-Soares et al. 2020) that makes use of whole mtDNA sequences from a wider sample set including individuals from the Atlantic, could help better elaborate current hypotheses and build a robust phylogeographic scenario of colonisation and dispersal of loggerheads in the Mediterranean Sea.

Conclusions

Analysis of complete mitogenome sequences can improve population genetic and phylogeographic studies, particularly for marine species that exhibit female natal philopatry. In this study, we investigated patterns of mtDNA diversity in loggerhead sea turtles and presented a high-coverage, complete mitogenome dataset for *C. caretta* in the central Mediterranean Sea along with a first genetic description of the Italian nesting sites of Tuscany, Latium and the Island of Linosa. We also proposed the first nomenclature for *C. caretta* mitogenomic haplotypes and described a larger number of variants than those so far recovered by mtDNA control region sequence comparisons. Mitogenome analysis provided a more comprehensive definition of matriline phylogeny and improved genetic characterisation of rookeries. Our study represents a first example of the potential of whole mitogenome analysis for the study of the evolutionary history of Mediterranean loggerheads and fine-scale characterisation of nesting sites to help mixed-stock analysis and the conservation of extant rookeries.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00227-023-04325-x.

Acknowledgements We are grateful to the following collaborators involved in sample collection: Nicola Novarini of the Natural History Museum of Venice, Lisa Poppi, Caterina Raso, Luca Marini and collaborators of the NGO TartaLazio and all voluntary workers and former students involved in the TartaCare Project of Calabria University including Sandra Berlingeri, Giulia Cambiè, Francesca Crispino, Concetta Denaro, Simona Fabiano, Elena Giardinazzo, Teresa Malito, Carmela Mancuso, Giampiero Masciari, Alessandro Massolo, Nunzia Micò, Gianni Parise, Brunella Pisani, Manuela Policastrese, Patrizia

Rima, Manuela Russo, Salvatore Salice and Salvatore Urso. We also thank Luana Papetti and collaborators of the NGO tartAmare and Isa Tonso and collaborators of Legambiente Arcipelago Toscano for support during collection of pipped eggs from nesting sites in Tuscany. A special thanks to Enrico Negrisolo for valuable suggestions during the preparation of the manuscript.

Author contributions LT, AI, ANo, LG and CCi designed and conceived the study. LG, ACD, ADL, CCo, CM, EM, GT, IC, LM, MALZ, SN and TM were involved in sample collection. CCi, AF, LM and CM secured funding for laboratory work and consumables. LT, AI and CN performed genetic analyses. LT, AI, ANo, ANi, FI, LG and MF contributed to data analysis. LT, AI, ANo, LG and CCi wrote the paper. All authors revised and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. This study was funded by the Tuscany Regional Authority NATura Network Toscana agreement in compliance to the Habitats Directive 92/43/EEC. Additional support was provided by the Italian Ministry of University and Research through the National Biodiversity Future Center, part of the National Recovery and Resilience Plan, Mission 4, Component 2, Investment 1.4, Project CN00000033. LT was granted a graduate research scholarship supported by the European Structural and Investment Funds through the National Operational Programme (PON) of the Italian Ministry of University and Research.

Data availability Mitochondrial DNA haplotype sequences presented in this study were deposited in the GenBank database (accession numbers: OR166582 - OR166647).

Declarations

Conflict of interest Authors declare no conflict of interest.

Ethical standards Sampling was conducted in compliance to the exemption provided by Article VII, paragraph 6 of CITES to the University of Pisa and the University of Siena (register of scientific institutions codes IT026 and IT007, respectively) and in accordance to permit 27676 of 16 March 2021 granted to MALZ by the Italian Ministry for the Ecological Transition. The study complies with institutional, national, and international ethics guidelines concerning the use of animals in research and/or the sampling of endangered species. None of the procedures used in the study met the criteria to define them "experiments" as defined in article 2 of the Council of European Communities Directive 86/609/EEC regarding the protection of animals used for experimental and other scientific purposes.

References

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

Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Philos Trans R Soc Lond B 312:325–342. https://doi.org/10.1098/rstb.1986.0011

Baker CS, Steel D, Calambokidis J, Falcone E, González-Peral U, Barlow J, Burdin AM, Clapham PJ, Ford JK, Gabriele CM (2013) Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. Mar Ecol Prog Ser 494:291–306. https://doi.org/10.3354/meps10508

Baldi G, Furii G, Del Vecchio M, Salvemini P, Vallini C, Angelini V, Pari S, Lombardi Moraes K, Margaritoulis D, Çurri A, Lazar B, Casale P (2023) Growth rates and age at maturity of Mediterranean loggerhead sea turtles estimated from a single-population foraging ground. Mar Biol 170:36. https://doi.org/10.1007/s00227-023-04189-1

Baltazar-Soares M, Klein JD, Correia SM, Reischig T, Taxonera A, Roque SM, Dos Passos L, Durão J, Lomba JP, Dinis H, Cameron SJK, Stiebens VA, Eizaguirre C (2020) Distribution of genetic diversity reveals colonization patterns and philopatry of the loggerhead sea turtles across geographic scales. Sci Rep 10:18001. https://doi.org/10.1038/s41598-020-74141-6

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolten AB, Bjorndal KA, Martins HR, Dellinger T, Biscoito MJ, Encalada SE, Bowen BW (1998) Transatlantic developmental migrations of loggerhead sea turtles demonstrated by mtDNA sequence analysis. Ecol Appl 8:1–7. https://doi.org/10.1890/1051-0761(1998)008[0001:TDMOLS]2.0.CO;2

Bowen BW, Kamezaki N, Limpus CJ, Hughes GR, Meylan AB, Avise JC (1994) Global phylogeography of the loggerhead turtle (*Caretta caretta*) as indicated by mitochondrial DNA haplotypes. Evolution 48:1820–1828. https://doi.org/10.1111/j.1558-5646.1994.tb02217.x

Carreras C, Pont S, Maffucci F, Pascual M, Barceló A, Bentivegna F, Cardona L, Alegre F, SanFélix M, Fernández G, Aguilar A (2006) Genetic structuring of immature loggerhead sea turtles (*Caretta caretta*) in the Mediterranean Sea reflects water circulation patterns. Mar Biol 149:1269–1279. https://doi.org/10.1007/s00227-006-0282-8

Carreras C, Pascual M, Cardona L, Marco A, Bellido JJ, Castillo JJ, Tomás J, Raga JA, Sanfélix M, Fernández G, Aguilar A (2011) Living together but remaining apart: Atlantic and Mediterranean loggerhead sea turtles (*Caretta caretta*) in shared feeding grounds. J Hered 102:666–677. https://doi.org/10.1093/jhered/esr089

Carreras C, Pascual M, Tomas J (2018) Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). Sci Rep 8:1435. https://doi.org/10.1038/s41598-018-19887-w

Carreras C, Rees AF, Broderick AC, Godley BJ, 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. Sci Mar 78: 115–124. https://doi.org/10.3989/scimar.03865.27B

Casale P, Broderick AC, Camiñas JA, Cardona L, Carreras C, Demetropoulos A, Fuller WJ, Godley BJ, Hochscheid S, Kaska Y (2018) Mediterranean sea turtles: current knowledge and priorities for conservation and research. Endanger Species Res 36:229–267. https://doi.org/10.3354/esr00901

Casale P, Hochscheid S, Kaska Y, Panagopoulou A (2020) Sea Turtles in the Mediterranean Region: MTSG Annual Regional Report 2020. Report of the IUCN-SSC Marine Turtle Specialist Group

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol Phylogenet Evol 3:102–113. https://doi.org/10.1006/mpev.1994.1013

Chang G, Jones S, Leelakumari S, Ashkani J, Culibrk L, O'Neill K, Tse K, Cheng D, Chuah E, McDonald H, Kirk H, Pandoh P, Pari S, Angelini V, Kyle C, Bertorelle G, Zhao Y, Mungall A, Moore R, Vilaça ST, Jones S (2023) The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758 [version 2; peer review: 2 approved]. F1000Research 12. https://doi.org/10.12688/f1000research.131283.2

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Clusa M, Carreras C, Pascual M, Demetropoulos A, Margaritoulis D, Rees AF, Hamza AA, Khalil M, Aureggi M, Levy Y (2013) Mitochondrial DNA reveals Pleistocenic colonisation of the Mediterranean by loggerhead turtles (*Caretta caretta*). J Exp Mar Biol Ecol 439:15–24. https://doi.org/10.1016/j.jembe.2012.10.011

Clusa M, Carreras C, Pascual M, Gaughran SJ, Piovano S, Giacoma C, Fernández G, Levy Y, Tomás J, Raga JA, Maffucci F, Hochscheid S, Aguilar A, Cardona L (2014) Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. Mar Biol 161:509–519. https://doi.org/10.1007/s00227-013-2353-y

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134:959–969. https://doi.org/10.1093/genetics/134.3.959

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008. https://doi.org/10.1093/gigascience/giab008

Drosopoulou E, Tsiamis G, Mavropoulou M, Vittas S, Katselidis KA, Schofield G, Palaiologou D, Sartsidis T, Bourtzis K, Pantis J (2012) The complete mitochondrial genome of the loggerhead turtle *Caretta caretta* (Testudines: Cheloniidae): Genome description and phylogenetic considerations. Mitochondrial DNA 23:1–12. https://doi.org/10.3109/19401736.2011.637109

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. https://doi.org/10.1186/1471-2148-7-214

Duchene S, Frey A, Alfaro-Núñez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. Mol Phylogenet Evol 65:241–250. https://doi.org/10.1016/j.ympev.2012.06.010

Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK (1999) Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). J Zool 248:397–409. https://doi.org/10.1111/j.1469-7998.1999.tb01038.x

Feldheim KA, Gruber SH, DiBattista JD, Babcock EA, Kessel ST, Hendry AP, Pikitch EK, Ashley MV, Chapman DD (2014) Two decades of genetic profiling yields first evidence of natal philopatry and long-term fidelity to parturition sites in sharks. Mol Ecol 23:110–117. https://doi.org/10.1111/mec.12583

Fetzner JW Jr, Crandall KA (2003) Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). Evolution 57:2101–2118. https://doi.org/10.1111/j.0014-3820.2003.tb00388.x

Feutry P, Kyne PM, Pillans RD, Chen X, Naylor GJ, Grewe PM (2014) Mitogenomics of the Speartooth Shark challenges ten years of control region sequencing. BMC Evol Biol 14:1–9. https://doi.org/10.1186/s12862-014-0232-x

Feutry P, Kyne PM, Pillans RD, Chen X, Marthick JR, Morgan DL, Grewe PM (2015) Whole mitogenome sequencing refines population structure of the Critically Endangered sawfish *Pristis pristis*. Mar Ecol Prog Ser 533:237–244. https://doi.org/10.3109/19401736.2015.1018208

Frandsen HR, Figueroa DF, George JA (2020) Mitochondrial genomes and genetic structure of the Kemp's ridley sea turtle (*Lepidochelys kempii*). Ecol Evol 10:249–262. https://doi.org/10.1002/ece3.5891

Garofalo L, Mingozzi T, Mico A, Novelletto A (2009) Loggerhead turtle (*Caretta caretta*) matrilines in the Mediterranean: further evidence of genetic diversity and connectivity. Mar Biol 156:2085–2095. https://doi.org/10.1007/s00227-009-1239-5

Garofalo L, Mastrogiacomo A, Casale P, Carlini R, Eleni C, Freggi D, Gelli D, Knittweis L, Mifsud CR, Mingozzi T, Novarini N, Scaravelli D, Scillitani G, Oliverio M, Novelletto A (2013) Genetic characterization of central Mediterranean stocks of the loggerhead turtle (*Caretta caretta*) using mitochondrial and nuclear markers, and conservation implications. Aquat Conserv 23:868–884. https://doi.org/10.1002/aqc.2338

Green M, Sambrook J (2012) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY

Hasegawa M, Kishino H, Yano T-a (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174. https://doi.org/10.1007/BF02101694

Hernández-Fernández J, Delgado Cano DA (2018) Complete mitochondrial genome of the three Colombian Caribbean loggerhead sea turtles, using next-generation sequencing. Mitochondrial DNA B 3:626–627. https://doi.org/10.1080/23802359.2018.1450664

Hochscheid S, Maffucci F, Abella E, Bradai MN, Camedda A, Carreras C, Claro F, de Lucia GA, Jribi I, Mancusi C, Marco A, Marrone N, Papetti L, Revuelta O, Urso S, Tomás J (2022) Nesting range expansion of loggerhead turtles in the Mediterranean: Phenology, spatial distribution, and conservation implications. Glob Ecol Conserv 38:e02194. https://doi.org/10.1016/j.gecco.2022.e02194

Jensen MP, FitzSimmons NN, Dutton PH, Michael P (2013) Molecular genetics of sea turtles. In: Wyneken J, Lohmann K, Musick JA (eds) The biology of sea turtles, vol III. CRC Press. Boca Raton, FL, pp 135–161

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923

Laurent L, Casale P, Bradai MN, Godley BJ, Broderick G, Gerosa AC, Schroth W, Schierwater B, Levy AM, Freggi D, Abd EM, Hadoud DA, Gomati HE, Domingo M, Hadjichristophprou M, Kornaraky L, Demirayak F, Gautier CH (1998) Molecular resolution of marine turtle stock composition in fishery bycatch: a case study in the Mediterranean. Mol Ecol 7:1529–1542. https://doi.org/10.1046/j.1365-294x.1998.00471.x

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Loisier A, Savelli M-P, Arnal V, Claro F, Gambaiani D, Sénégas JB, Cesarini C, Sacchi J, Miaud C, Montgelard C (2021) Genetic composition, origin and conservation of loggerhead sea turtles (*Caretta caretta*) frequenting the French Mediterranean coasts. Mar Biol 168:52. https://doi.org/10.1007/s00227-021-03855-6

Maffucci F, Corrado R, Palatella L, Borra M, Marullo S, Hochscheid S, Lacorata G, Iudicone D (2016) Seasonal heterogeneity of ocean warming: a mortality sink for ectotherm colonizers. Sci Rep 6:23983. https://doi.org/10.1038/srep23983

Mancino C, Canestrelli D, Maiorano L (2022) Going west: range expansion for loggerhead sea turtles in the Mediterranean Sea

under climate change. Glob Ecol Conserv 38:e02264. https://doi.org/10.1016/j.gecco.2022.e02264
Margaritoulis D, Dean CJ, Lourenço G, Rees AF, Riggall TE (2020) Reproductive longevity of loggerhead sea turtles nesting in Greece. Chelonian Conserv Biol 19:133–136. https://doi.org/10.2744/CCB-1437.1
McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654. https://doi.org/10.1038/351652a0
Michaels G, Hauswirth W, Laipis P (1982) Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 94:246–251. https://doi.org/10.1016/0012-1606(82)90088-4
Miller J (1997) Reproduction in sea turtles. In: Musick JA (ed) Lutz P. The biology of sea turtles. CRC Press, Boca Raton, FL, pp 51–81
Milne I, Stephen G, Bayer M, Cock P, Pritchard L, Cardle L, Shaw P, Marshall D (2013) Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193–202. https://doi.org/10.1093/bib/bbs012
Mingozzi T, Masciari G, Paolillo G, Pisani B, Russo M, Massolo A (2006) Discovery of a regular nesting area of loggerhead turtle *Caretta caretta* in southern Italy: a new perspective for national conservation. In: Hawksworth D, Bull A (eds) Biodiversity conservation in Europe Topics in biodiversity and conservation, vol 7. Springer, Dordrecht, pp 277–299
Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Abel Nielsen SC, Rasmussen M, Willerslev E, Gilbert MT, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res 20:908–916. https://doi.org/10.1101/gr.102954.109
Moritz C (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. Mol Ecol 3:401–411. https://doi.org/10.1111/j.1365-294X.1994.tb00080.x
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. Physiol Biochem Zool 89:473–486. https://doi.org/10.1086/688679
Otálora K, Hernández-Fernández J, Carlos Salcedo-Reyes J (2018) Complete Colombian Caribbean loggerhead turtle mitochondrial genome: tRNA structure analysis and revisited marine turtle phylogeny. Univ Sci 23:355–381
Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420. https://doi.org/10.1093/bioinformatics/btp696
Pianka ER (1973) Evolutionary ecology. Harper & Row, New York, NY
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032
Robin E, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507–513. https://doi.org/10.1002/jcp.1041360316
Saied A, Maffucci F, Hochscheid S, Dryag S, Swayeb B, Borra M, Ouerghi A, Procaccini G, Bentivegna F (2012) Loggerhead turtles nesting in Libya: an important management unit for the Mediterranean stock. Mar Ecol Prog Ser 450:207–218. https://doi.org/10.3354/meps09548
Shamblin BM, Bjorndal KA, Bolten AB, Hillis-Starr ZM, Lundgren I, Naro-Maciel E, Nairn CJ (2012a) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. Mol Ecol 21:2330–2340. https://doi.org/10.1111/j.1365-294X.2012.05530.x
Shamblin BM, Bolten AB, Bjorndal KA, Dutton PH, Neilen JT, Abreu-Grobois FA, Reich KJ, Witherington BE, Bagley DA, Ehrhart LM, Tucker AD, Addison DS, Arenas A, Johnson C, Carthy RR, Lamont MM, Dodd MG, Gaines MS, LaCasella E, Nairn CJ (2012b) Expanded mitochondrial control region sequences increase resolution of stock structure among North Atlantic loggerhead turtle rookeries. Mar Ecol Prog Ser 469:145–160. https://doi.org/10.3354/meps09980
Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, Clusa M, Monzón-Argüello C, Nairn CJ, Nielsen JT (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:e85956. https://doi.org/10.1371/journal.pone.0085956
Sheridan CM, Spotila JR, Bien WF, Avery HW (2010) Sex-biased dispersal and natal philopatry in the diamondback terrapin, Malaclemys terrapin. Mol Ecol 19:5497–5510. https://doi.org/10.1111/j.1365-294X.2010.04876.x
Shields WM (1982) Philopatry, inbreeding, and the evolution of sex. State University of New York Press, New York, NY
Splendiani A, Fioravanti T, Giovannotti M (2017) Mitochondrial DNA reveals the natal origin of *Caretta caretta* (Testudines: Cheloniidae) stranded or bycaught along the Southwestern Adriatic coasts. Eur Zool J 84:566–574. https://doi.org/10.1080/24750263.2017.1400597
Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4. https://doi.org/10.1093/ve/vey016
Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619–633. https://doi.org/10.1093/genetics/132.2.619
Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics 134:659–669. https://doi.org/10.1093/genetics/134.2.659
Tikochinski Y, Carreras C, Tikochinski G, Vilaça ST (2020) Population-specific signatures of intra-individual mitochondrial DNA heteroplasmy and their potential evolutionary advantages. Sci Rep 10:211. https://doi.org/10.1038/s41598-019-56918-6
Tolve L, Casale P, Formia A, Garofalo L, Lazar B, Natali C, Novelletto A, Vallini C, Bužan E, Chelazzi G, Gaspari S, Fortuna C, Kocijan I, Marchiori E, Novarini N, Poppi L, Salvemini P, Ciofi C (2018) A comprehensive mitochondrial DNA mixed-stock analysis clarifies the composition of loggerhead turtle aggregates in the Adriatic Sea. Mar Biol 165:68. https://doi.org/10.1007/s00227-018-3325-z
Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional Management Units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5:e15465. https://doi.org/10.1371/journal.pone.0015465
Wiesner R, Rüegg J, Morano I (1992) Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. Biochem Biophys Res Commun 183:553–559. https://doi.org/10.1016/0006-291x(92)90517-o
Yan K, An T, Zhai M, Huang Y, Wang Q, Wang Y, Zhang R, Wang T, Liu J, Zhang Y, Zhang J, Wang K (2019) Mitochondrial miR-762 regulates apoptosis and myocardial infarction by

impairing ND2. Cell Death Dis 10:500. https://doi.org/10.1038/s41419-019-1734-7

Yilmaz C, Turkozan O, Bardakci F (2011) Genetic structure of loggerhead turtle (*Caretta caretta*) populations in Turkey. Biochem Syst Ecol 39:266–276. https://doi.org/10.1016/j.bse.2011.08.018

Authors and Affiliations

Livia Tolve[1] · Alessio Iannucci[1,2] · Luisa Garofalo[3] · Andrea Ninni[4,5] · Andrea Capobianco Dondona[6] · Ilaria Ceciarini[7] · Cristiano Cocumelli[3] · Alessandra De Lucia[8] · Mattia Falconi[4] · Angela Formia[1,9] · Federico Iacovelli[4] · Cecilia Mancusi[7,10] · Erica Marchiori[11] · Letizia Marsili[7] · Toni Mingozzi[12] · Stefano Nannarelli[8] · Chiara Natali[1] · Giuliana Terracciano[13] · Marco A. L. Zuffi[14] · Andrea Novelletto[5] · Claudio Ciofi[1]

✉ Livia Tolve
livia.tolve@unifi.it

✉ Alessio Iannucci
alessio.iannucci@unifi.it

[1] Department of Biology, University of Florence, 50019 Sesto Fiorentino, FI, Italy

[2] National Biodiversity Future Center, 90133 Palermo, Italy

[3] Istituto Zooprofilattico Sperimentale del Lazio e Della Toscana "M. Aleandri", 00178 Rome, Italy

[4] Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy

[5] PhD Program in Evolutionary Biology and Ecology, Tor Vergata University of Rome, 00133 Rome, Italy

[6] Farm4Trade, 66041 Atessa, CH, Italy

[7] Department of Physical Sciences, Earth and Environment, University of Siena, 53100 Siena, Italy

[8] Hydrosphera Onlus, Rome, Italy

[9] African Aquatic Conservation Fund, Chilmark, MA 02535, USA

[10] Agenzia Regionale Per La Protezione Ambientale Toscana (ARPAT), 57125 Leghorn, Italy

[11] Department of Comparative Biomedicine and Food Science, University of Padua, Agripolis, 35020 Legnaro, PD, Italy

[12] Department of Biology, Ecology and Earth Sciences, University of Calabria, 87030 Rende, CS, Italy

[13] Istituto Zooprofilattico Sperimentale del Lazio E Della Toscana "M. Aleandri", 56123 Pisa, Italy

[14] Natural History Museum, University of Pisa, 56011 Calci, PI, Italy

Marine Biology (2025) 172:59
https://doi.org/10.1007/s00227-025-04614-7

CORRECTION

Correction to: whole mitochondrial genome sequencing provides new insights into the phylogeography of loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea

Livia Tolve[1] · Alessio Iannucci[1] · Luisa Garofalo[2] · Andrea Ninni[3,4] · Andrea Capobianco Dondona[5] · Ilaria Ceciarini[6] · Cristiano Cocumelli[2] · Alessandra De Lucia[7] · Mattia Falconi[3] · Angela Formia[1,8] · Federico Iacovelli[3] · Cecilia Mancusi[6,9] · Erica Marchiori[10] · Letizia Marsili[6] · Toni Mingozzi[11] · Stefano Nannarelli[7] · Chiara Natali[1] · Giuliana Terracciano[12] · Marco A. L. Zuffi[13] · Andrea Novelletto[4] · Claudio Ciofi[1]

Published online: 25 March 2025

Correction to: Mar Biolhttps://doi.org/10.1007/s00227-023-04325-x

In the original manuscript, sample TUS05 (Genbank accession number OR166594) was assigned to haplogroup IB and identified as a mitogenomic variant of the control region haplotype CC-A1.1. Consequently, the corresponding mitogenomic haplotype was named CC-A1.1.1. The 800 bp control region haplotype of TUS05, however, does not correspond to CC-A1.1, nor the 380 bp control region haplotype of TUS05 corresponds to CC-A1. According to the Archie Carr Center for Sea Turtles Research database (https://accstr.ufl.edu/resources/mtdna-sequences/), the control region haplotype described for TUS05 was yet to be described. We therefore renamed the mitogenomic haplotype of TUS05 as A79.1.1. Tables 2 and 4 and Table S1 should read CC-A79.1.1 and CC-A79.1 instead of CC-A1.1.1 and CC-A1.1, respectively. The corrected and uncorrected Fig. 2 is produced here.

The original article can be found online at https://doi.org/10.1007/s00227-023-04325-x.

✉ Livia Tolve
livia.tolve@unifi.it

✉ Alessio Iannucci
alessio.iannucci@unifi.it

[1] Department of Biology, University of Florence, 50019 Sesto Fiorentino (FI), Florence, Italy

[2] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana M. Aleandri, Rome 00178, Italy

[3] Department of Biology, Tor Vergata University of Rome, Rome 00133, Italy

[4] PhD Program in Evolutionary Biology and Ecology, Tor Vergata University of Rome, Rome 00133, Italy

[5] Farm4Trade, Farm4Trade, Atessa (CH), Italy66041

[6] Department of Physical Sciences, Earth and Environment, University of Siena, Siena 53100, Italy

[7] Hydrosphera onlus, Hydrosphera onlus, Rome, Italy53100

[8] African Aquatic Conservation Fund, MA 02535 Chilmark, USA

[9] Agenzia Regionale per la Protezione Ambientale Toscana (ARPAT), Livorno 57125, Italy

[10] Department of Comparative Biomedicine and Food Science, University of Padua, Agripolis, Legnaro (PD) 35020, Italy

[11] Department of Biology, Ecology and Earth Sciences, University of Calabria, Rende (CS) 87030, Italy

[12] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana, M. Aleandri, Pisa 56123, Italy

[13] Natural History Museum, University of Pisa, Calci (PI) 56011, Italy

Corrected Figure:

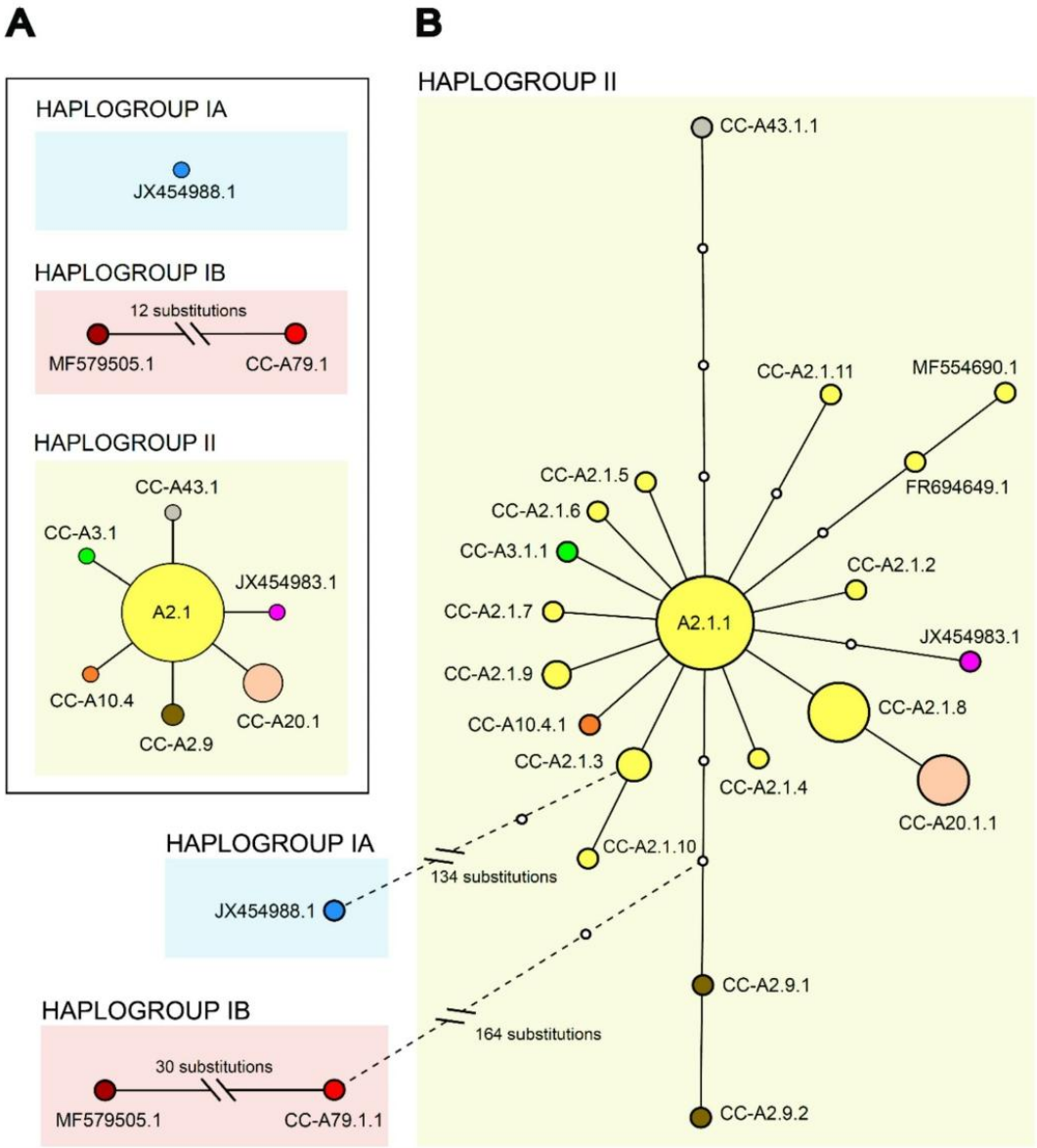


Figure 2 TCS haplotype networks based on partial mtDNA control region (**A**) and complete mitogenome (**B**) sequences. Dashed lines represent links with probability lower than 95%.

Uncorrected figure:

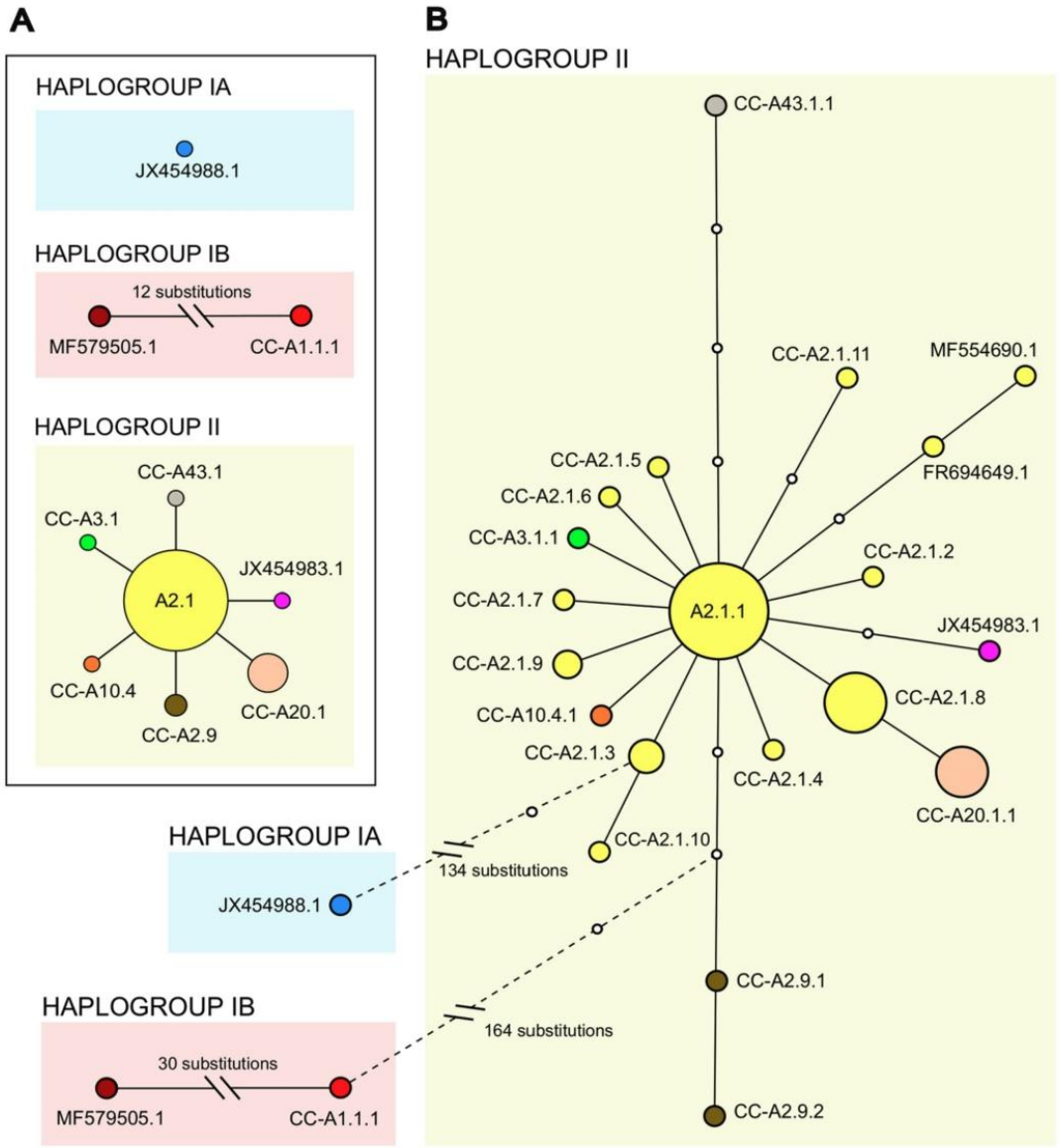


Fig. 2 TCS haplotype networks based on partial mtDNA control region (**A**) and complete mitogenome (**B**) sequences. Dashed lines represent links with probability lower than 95%.

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020, risorse FSE REACT-EU
Azione IV.4 “Dottorati e contratti di ricerca su tematiche dell’innovazione”
e Azione IV.5 “Dottorati su tematiche Green”