

# Complete Colombian Caribbean loggerhead turtle mitochondrial genome: tRNA structure analysis and revisited marine turtle phylogeny

## Supplementary information 1. Loggerhead mitogenome annotation results

### Protein-coding genes

All protein encoding genes, found in other loggerhead turtles were described in the mitogenome of the Cc-AO-C turtle (Florentz *et al.* 2003, Jung *et al.* 2006, Suzuki *et al.* 2011). All genes were located on the H-strand, except the ND6 gen, that is located on the L-strand. This is in accordance with previous loggerhead turtle mitogenome annotations (Kumazawa & Nishida, 1999, Duchene *et al.* 2012, Drosopoulou *et al.* 2012). All of the protein coding genes use ATG as a start codon except for the COI gene that uses GTG (Supplementary Table 1). This is a feature shared by all sea turtles (Kumazawa & Nishida 1999, Duchene *et al.* 2012, Drosopoulou *et al.* 2012). GTG is also a start codon in the prokaryote genomes but is absent in some mitogenomes of mammals and invertebrates (Bagatharia *et al.* 2013). Two different types of stop codons were identified in the Cc-AO-C mitogenome: TAA present in four genes (COII, ATP8, ND4L, and ND5); and AGG present in two genes (COI and ND6) (Supplementary Table 1). The other seven genes (ND1, ND2, ATP6, COII, ND3, ND4 and Cytb) do not have stop codons, they use T or TA as stops. This is common and typical for mtDNA in vertebrates, where the stop codon is probably complemented by post-transcriptional polyadenylation (Drosopoulou *et al.* 2012, Bagatharia *et al.* 2013) (Supplementary Table 1). The differences in stop codons seem to be a tendency of the mitogenomes of turtles (Seutin *et al.* 1994, Jung *et al.* 2006), also shared by the mitochondrial genome of mammals and fish (Kim *et al.* 2005).

The sequence the ND3 gene presented an insertion A base at position 175. This programmed frameshift mutation has only been reported in the Greek Cc-MS-G mitogenome (Drosopoulou *et al.* 2012), and it is a likely neutral variant since it does not lead to protein sequence changes. Moreover, the same mutation has been described in other turtle species, in reptiles and birds, and it is considered as relatively ancestral (Russel & Beckenbanch 2008). Additionally of the 13 protein encoding genes of the Cc-AO-C turtle, three regions ATP8-ATP6, ND4L-ND4 and ND5-ND6 partially overlap in the H-strand in 10, 7 and 6 nucleotides were identified. This is a tendency of vertebrates to use the mitogenomes in a more efficient way (Attardi 1985, Broughton *et al.* 2001, Kim *et al.* 2005) and is present in the mitogenomes of sea turtles (Kumazawa & Nishida 1999, Duchene *et al.* 2012, Drosopoulou *et al.* 2012).

### Non-protein coding genes

In the Cc-AO-C mitogenome, ribosomal subunit genes rRNA12S and rRNA16S were of size 970 and 1 611 bp, respectively (Supplementary Table 1). These genes are located between the tRNA<sup>Phe</sup> and tRNA<sup>Leu</sup> (UUR) and are separated by the tRNA<sup>Val</sup>, as in other freshwater and sea turtles, rays, and mammals (Florentz *et al.* 2003, Kim *et al.* 2005, Jung *et al.* 2006, Kumazawa & Nishida 1999, Suzuki *et al.* 2011, Drosopoulou *et al.* 2012).

The replication origin of the L-strand of the Cc-AO-C turtle (OL) is 28 bp long (TTTTCCTCCGCTCTATAAAAAGCGGGAAAA) and is situated between the tRNA<sup>Asn</sup> and tRNA<sup>Cys</sup> (Supplementary Table 1 and Supplementary Fig. 1). This region is potentially capable of forming a stable stem loop, common and crucial for replicating the L-strand of mitochondrial DNA in vertebrates (Shadel & Clayton 1997).

### tRNAs

tRNAs have unique structural characteristics. Their 2D structure has a cloverleaf-shape with five main regions (D-loop, TΨC, anti-codon, 5' and 3' extremes) with well-defined features, including stem size, loops, and highly conserved structural regions in strategic positions (Florentz *et al.* 2003). The heavy tRNAs were: tRNA<sup>Arg</sup>, tRNA<sup>Asp</sup>, tRNA<sup>Gly</sup>, tRNA<sup>His</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Leu</sup> (CUN), tRNA<sup>Leu</sup> (UUR), tRNA<sup>Lys</sup>, tRNA<sup>Met</sup>, tRNA<sup>Phe</sup>, tRNA<sup>Ser</sup> (AGY), tRNA<sup>Thr</sup>, tRNA<sup>Trp</sup>. The other tRNAs were in the L-strand: tRNA<sup>Val</sup>, tRNA<sup>Ala</sup>, tRNA<sup>Asn</sup>, tRNA<sup>Cys</sup>, tRNA<sup>Gln</sup>, tRNA<sup>Glu</sup>, tRNA<sup>Pro</sup>, tRNA<sup>Ser</sup> (UCN), tRNA<sup>Tyr</sup> (Table 4). All tRNAs of the Cc-AO-C turtle were of similar size (66-77 bp), in recognition of the codon (Table 4) and in the nucleotide composition to other sea turtles (*Chelonia mydas*

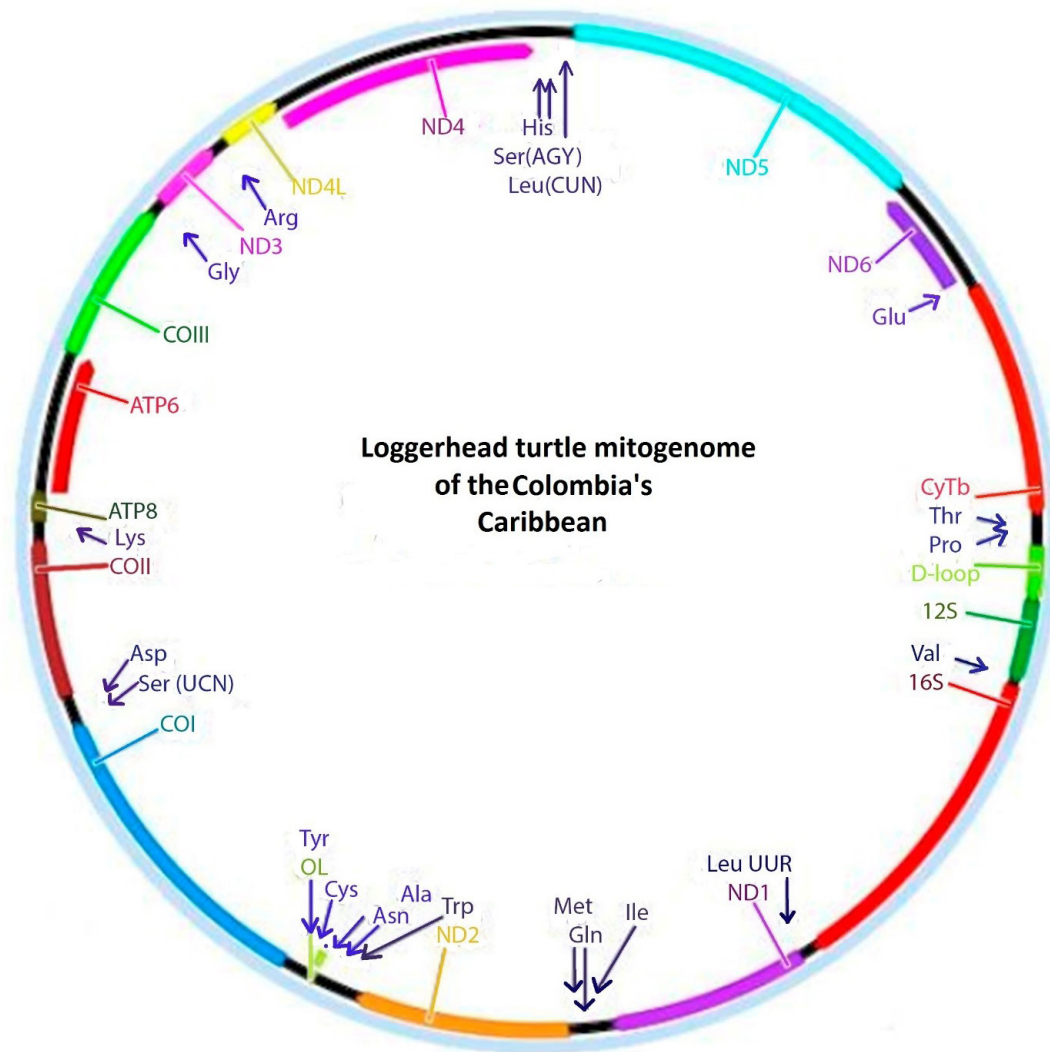
and Cc-MS-G) (Kumazawa & Nishida 1999, Drosopoulou *et al.* 2012). The tRNA<sup>Val</sup> encodes in the L-strand differing from what was described by Drosopoulou *et al.* (2012) for loggerhead turtle. The tRNA<sup>Ser</sup> with AGY anti-codon presented a typology III according to Suzuki *et al.* (2011), this typology is characterized by the absence of the D-loop region, which agrees with diverse analysis of 2D structure of the tRNA<sup>Ser</sup> (AGY) for vertebrates (Florentz *et al.* 2003, Jung *et al.* 2006).

**Table S1.** Organization of the *C. caretta* Colombian Caribbean mitochondrial genome.

Gene/Element	Abbreviation	Strand	Size (bp)	Position	Start codon	Stop codon	Codon recognized	Aminoacids
tRNA <sup>Phe</sup>	Ph	H	69	1-69			TTC	
12S ARN ribosomal	12S	H	973	70-1042				
tRNA <sup>Val</sup>	V	L	71	1112-1042			GTA	
16S ARN ribosomal	16S	H	1.610	1111-2720				
tRNA <sup>Leu</sup> (UUR)	L1	H	77	2721-2797				
NADH dehydrogenase subunit 1	ND1	H	978	2797-3774	ATG	TA		322
tRNA <sup>Ile</sup>	I	H	71	3773-3843			ATC	
tRNA <sup>Gln</sup>	G	L	71	3912-3842			CAA	
tRNA <sup>Met</sup>	M	H	69	3912-3980			ATG	
NADH dehydrogenase subunit 2	ND2	H	1042	3981-5022	ATG	T		346
tRNA <sup>Trp</sup>	T	H	77	5023-5099			TGA	
tRNA <sup>Ala</sup>	A	L	69	5169-5101			GCA	
tRNA <sup>Asn</sup>	As	L	73	5243-5171			AAC	
Replication origin of the light strand	O <sub>L</sub>	-	28	5244-5271				
tRNA <sup>Cys</sup>	C	L	66	5335-5270				
tRNA <sup>Tyr</sup>	Tr	L	71	5406-5336			TGC	515
cytochrome c oxidase subunit 1	COI	H	1549	5408-6955	GTG	AGG	TAC	
tRNA <sup>Ser</sup> (UCN)	S1	L	71	7017-6947			TCA	
tRNA <sup>Asp</sup>	Ap	H	70	7020-7089			GAC	
cytochrome c oxidase subunit 2	COII	H	681	7090-7770	ATG	TAA		256

tRNA <sup>Lys</sup>	Ly	H	74	7772-7845			AAA	
ATP synthase F0 subunit 8	ATP8	H	165	7846-8010	ATG	TAA		54
ATP synthase F0 subunidad 6	ATP6	H	684	8001-8684	ATG	TA		227
cytochrome c oxidase, subunit 3	COIII	H	784	8684-9467	ATG	T		261
tRNA <sup>Gly</sup>	Gl	H	68	9468-9535			GGA	
NADH dehydrogenase subunit 3	ND3	H	352	9536-9887	ATG	T		116
tRNA <sup>Arg</sup>	Ar	H	68	9887-9954				
NADH dehydrogenase subunit 4L	ND4L	H	297	9956-10252	ATG	TAA		99
NADH dehydrogenase subunit 4	ND4	H	1381	10246-11626	ATG	T		435
tRNA <sup>Leu</sup> (CUN)	L2	H	71	11629-11699			CTA	
tRNA <sup>Ser</sup> (AGY)	S2	H	66	11699-11764			AGC	
tRNA <sup>His</sup>	H	H	69	11764-11832			CAC	
NADH dehydrogenase subunit 5	ND5	H	1797	11833-13629	ATG	TAA		596
NADH dehydrogenase subunit 6	ND6	L	525	14150-13626	ATG	AGG		175
tRNA <sup>Glu</sup>	Gu	L	68	14218-14151			GAA	
cytochrome b	CytB	H	1145	14222-15366	ATG	T		382
tRNA <sup>Thr</sup>	Th	H	71	15367-15437			ACA	
tRNA <sup>Pro</sup>	Pr	L	71	15508-15438			CCA	
D-loop (control region)	D-loop	-	854	15509-16362				

**Figure S1.** *C. caretta* mitochondrial genome. Protein coding genes and tRNAs are shown with standard abbreviations. Genes for tRNAs are labeled by a three first letters for corresponding amino acid.



## Supplementary references

List of references cited only in this supplementary material, for the remainder see reference list in the main text.

Bagatharia SB, Joshi MN, Pandya RV, Pandit AS, Patel RP, Desai SM, Saxena AK. Complete metagenome of Asiatic lion resolves phylogenetic status within Panthera, *BMC genomics*, 14(1): 572. 2013.

doi: [10.1186/1471-2164-14-572](https://doi.org/10.1186/1471-2164-14-572)

Broughton RE, Milam JE, Roe BA. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome evolutionary patterns in vertebrate mitochondrial DNA, *Genome Research*, 11(11): 1958-1967, 2001.

doi: [10.1101/gr.156801](https://doi.org/10.1101/gr.156801)

Committee on the Status of Endangered Wildlife in Canada, COSEWIC assessment and status report on the Loggerhead Sea Turtle *Caretta caretta* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa. 2010

Retrieved: [https://sararegistry.gc.ca/document/doc2025/ind\\_e.cfm](https://sararegistry.gc.ca/document/doc2025/ind_e.cfm)

Shadel GS, Clayton DA. Mitochondrial DNA maintenance in vertebrates, *Annual Review of Biochemistry*, 66: 409-435, 1997.

doi: [10.1146/annurev.biochem.66.1.409](https://doi.org/10.1146/annurev.biochem.66.1.409)

Seutin G, Lang BF, Mindell DP, Morais R. Evolution of the WANCY region in amniote mitochondrial DNA, *Molecular Biology and Evolution*, 11(3): 329-340, 1994.

doi: [10.1093/oxfordjournals.molbev.a040116](https://doi.org/10.1093/oxfordjournals.molbev.a040116)