



GENOME NOTE

REVISED The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758 [version 2; peer review: 2 approved]

Glenn Chang ^{1,2}, Samantha Jones ², Sreeja Leelakumari², Jahanshah Ashkani², Luka Culibrk², Kieran O'Neill², Kane Tse², Dean Cheng², Eric Chuah², Helen McDonald², Heather Kirk², Pawan Pandoh², Sauro Pari³, Valeria Angelini³, Christopher Kyle^{4,5}, Giorgio Bertorelle⁶, Yongjun Zhao², Andrew Mungall², Richard Moore², Sibelle Vilaça ⁵, Steven Jones^{2,7}

¹Genome Science and Technology Graduate Program, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

²Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, V5Z 4S6, Canada

³Fondazione Cetacea Onlus, Riccione, RN, 47838, Italy

⁴Forensic Science Department, Trent University, Peterborough, Ontario, K9L 0G2, Canada

⁵Environmental and Life Sciences Graduate Program, Trent University, Peterborough, Ontario, K9L 0G2, Canada

⁶Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, FE, 44121, Italy

⁷Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

v2 First published: 27 Mar 2023, 12:336
<https://doi.org/10.12688/f1000research.131283.1>
 Latest published: 27 Jun 2023, 12:336
<https://doi.org/10.12688/f1000research.131283.2>

Abstract

We present a genome assembly of *Caretta caretta* (the Loggerhead sea turtle; Chordata, Testudines, Cheloniidae), generated from genomic data from two unrelated females. The genome sequence is 2.13 gigabases in size. The assembly has a busco completion score of 96.1% and N50 of 130.95 Mb. The majority of the assembly is scaffolded into 28 chromosomal representations with a remaining 2% of the assembly being excluded from these.

Keywords

Caretta caretta, Loggerhead sea turtle, genome sequence, chromosomal, reptile



This article is included in the **Genomics and Genetics** gateway.

Open Peer Review

Approval Status

	1	2
version 2 (revision) 27 Jun 2023		 view
version 1 27 Mar 2023	 view	 view

1. **Richard Challis** , Wellcome Sanger Institute, Hinxton, UK

2. **Cinta Peguerols** , Universitat de Barcelona, Barcelona, Spain

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Steven Jones (sjones@bcgsc.ca)

Author roles: **Chang G:** Data Curation, Formal Analysis, Investigation, Software, Visualization, Writing – Original Draft Preparation; **Jones S:** Data Curation, Visualization, Writing – Original Draft Preparation; **Leelakumari S:** Data Curation, Investigation; **Ashkani J:** Formal Analysis, Investigation, Software; **Culibrk L:** Formal Analysis, Investigation, Software, Writing – Review & Editing; **O'Neill K:** Formal Analysis, Investigation, Software, Writing – Review & Editing; **Tse K:** Investigation; **Cheng D:** Investigation; **Chuah E:** Investigation; **McDonald H:** Investigation; **Kirk H:** Investigation; **Pandoh P:** Investigation; **Pari S:** Data Curation, Investigation, Project Administration; **Angelini V:** Data Curation, Investigation, Project Administration; **Kyle C:** Project Administration; **Bertorelle G:** Funding Acquisition; **Zhao Y:** Investigation; **Mungall A:** Investigation; **Moore R:** Investigation; **Vilaça S:** Data Curation, Project Administration, Supervision, Writing – Review & Editing; **Jones S:** Conceptualization, Funding Acquisition, Supervision

Competing interests: No competing interests were disclosed.

Grant information: Sequencing of the loggerhead sea turtle genome was supported through the Canadian BioGenome Project (Grant ID 18107, Genome Canada) and CanSeq150 program of Canada's Genomics Enterprise (www.cgen.ca), as well as the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement 844756 (TurtleHyb). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2023 Chang G *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Chang G, Jones S, Leelakumari S *et al.* **The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758 [version 2; peer review: 2 approved]** F1000Research 2023, 12:336 <https://doi.org/10.12688/f1000research.131283.2>

First published: 27 Mar 2023, 12:336 <https://doi.org/10.12688/f1000research.131283.1>

REVISED Amendments from Version 1

Based on suggestions made by reviewers, we have made several revision and clarifications to improve the clarity and precision of our findings.

We utilized RepeatMasker to analyze repetitive elements and have now included the findings in the result section. Additionally, we have specified the parameters used for each software in Table 3. We have rephrased the gene annotation section to clarify results for both the RefSeq and Ensemble annotation pipelines. We clarified that JupyterPlot is used for scaffold-level alignment and synteny plots in the syntenic analysis. Latly, QC metrics are specified in the abstract.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archelosauria; Testudinata; Testudines; Cryptodira; Durocryptodira; Americhelydia; Chelonioidea; Cheloniidae; Caretta; *Caretta caretta* Linnaeus 1758 (NCBI txid 8467).

Introduction

The loggerhead sea turtle, *Caretta caretta*, is one of only seven extant marine turtle species and is globally distributed throughout the subtropical and temperate regions of the Mediterranean Sea and Pacific, Indian and Atlantic Oceans (Wallace *et al.*, 2010, Casale and Tucker, 2015). The species is divided in various Regional Management Units (RMUs) and management units (MUs) that vary greatly by population size, geographic range, and population trends (Wallace *et al.*, 2010, Casale and Tucker, 2015, Shamblin *et al.*, 2014). Events such as fisheries bycatch (Caracappa *et al.*, 2018, Pulcinella *et al.*, 2019), human intrusion and disturbance (Mazaris *et al.*, 2009), oceanic pollution (Savoca *et al.*, 2018), and climate change and severe weather (Alduina *et al.*, 2020) have caused the global population to continuously decline

Table 1. Genome data for *Caretta caretta*, rCarCar2.

Project accession data	
Assembly identifier	rCarCar2
Species	<i>Caretta caretta</i>
Specimen	SJ_126, SJ_184
NCBI Taxonomy ID	8467
BioProject	PRJNA826225
BioSample ID	SAMN28968396, SAMN27958248
Isolate Information	SJ_184/204:Loco2, SJ_126:Eziel1
Raw data accessions	
Oxford Nanopore PromethION	SRX15677840, SRX15677841
Hi-C Illumina	SRX15677843
Illumina short-read	SRX15677842
Genome assembly	
Assembly accession	GCA_023653815.1
Assembly name	GSC_CCare_1.0
Span (Mb)	2,134
Number of contigs	2,753
Contig N50 length (Mb)	18,214
Number of scaffolds	2,008
Scaffold N50 length (Mb)	130,956
Longest scaffold (Mb)	345.7
BUSCO* genome score	C:96.1%[S:95.2%,D:0.9%],F:0.4%,M:3.5%,n:7480

*BUSCO scores based on the sauropsida_odb10 BUSCO set using v5.0.0. C=complete [S=single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison.

(Casale and Tucker, 2015). Consequently, the highly migratory *C. caretta* requires the collaborative efforts of numerous international conservation and protection organizations (Species at Risk Act, 2002), and is currently listed as Vulnerable by the International Union for the Conservation of Nature (IUCN) (Casale and Tucker, 2015). The genome of *C. caretta* was sequenced as part of the Canadian BioGenome Project (CBP) and CanSeq150 initiatives. The *C. caretta* genome will provide insights into genomic diversity and architecture, and inform conservation genomics applications.

Methods

Sample collection

Blood samples from an adult female and a juvenile of unknown sex were collected from the Fondazione Cetacea (43.9940 N, 12.6745 E) by Nicola Ridolfi (veterinarian; Fondazione Cetacea). Animal husbandry and welfare were

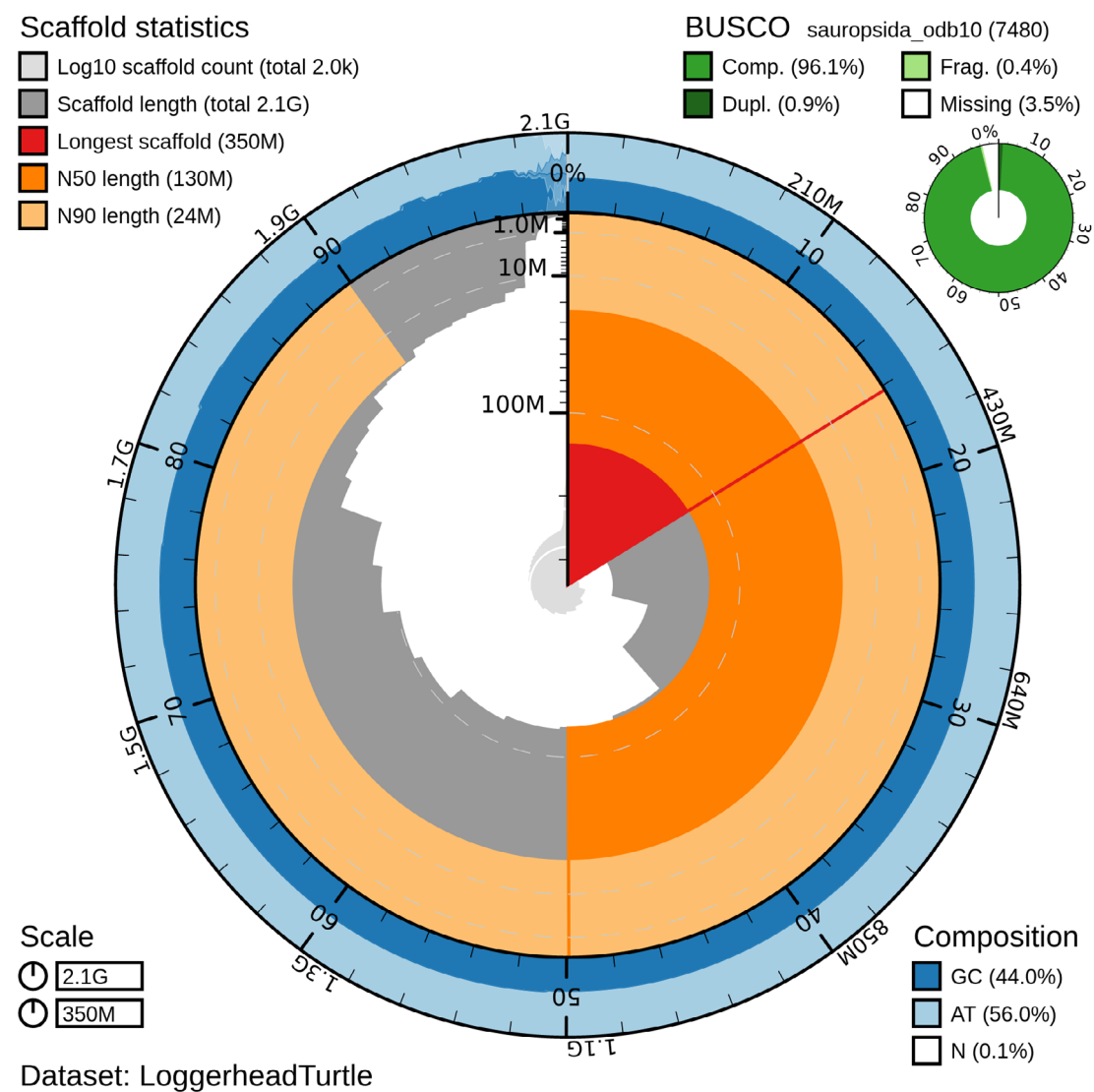


Figure 1. Genome assembly of *Caretta caretta*, rCarCar2: metrics. Snail plot showing N50 metrics, base pair composition and BUSCO gene completeness for *C. caretta* (rCarCar2) generated from Blobtoolkit v.2.6.4 (Challis *et al.*, 2020). The plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 2,134,012,717 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (345,741,823 bp) shown in red. Orange and pale-orange arcs show the N50 and N90 chromosome lengths (130,956,235 and 23,648,662 bp, respectively). The pale grey spiral shows the cumulative chromosome count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot displays the distribution of GC (blue), AT (pale blue) and N (white) percentages using the same bins as the inner plot. A summary of complete (96.1%), fragmented (0.4%), duplicated (0.9%), and missing (3.5%) BUSCO genes in the sauropsida_odb10 set is shown in the top right.

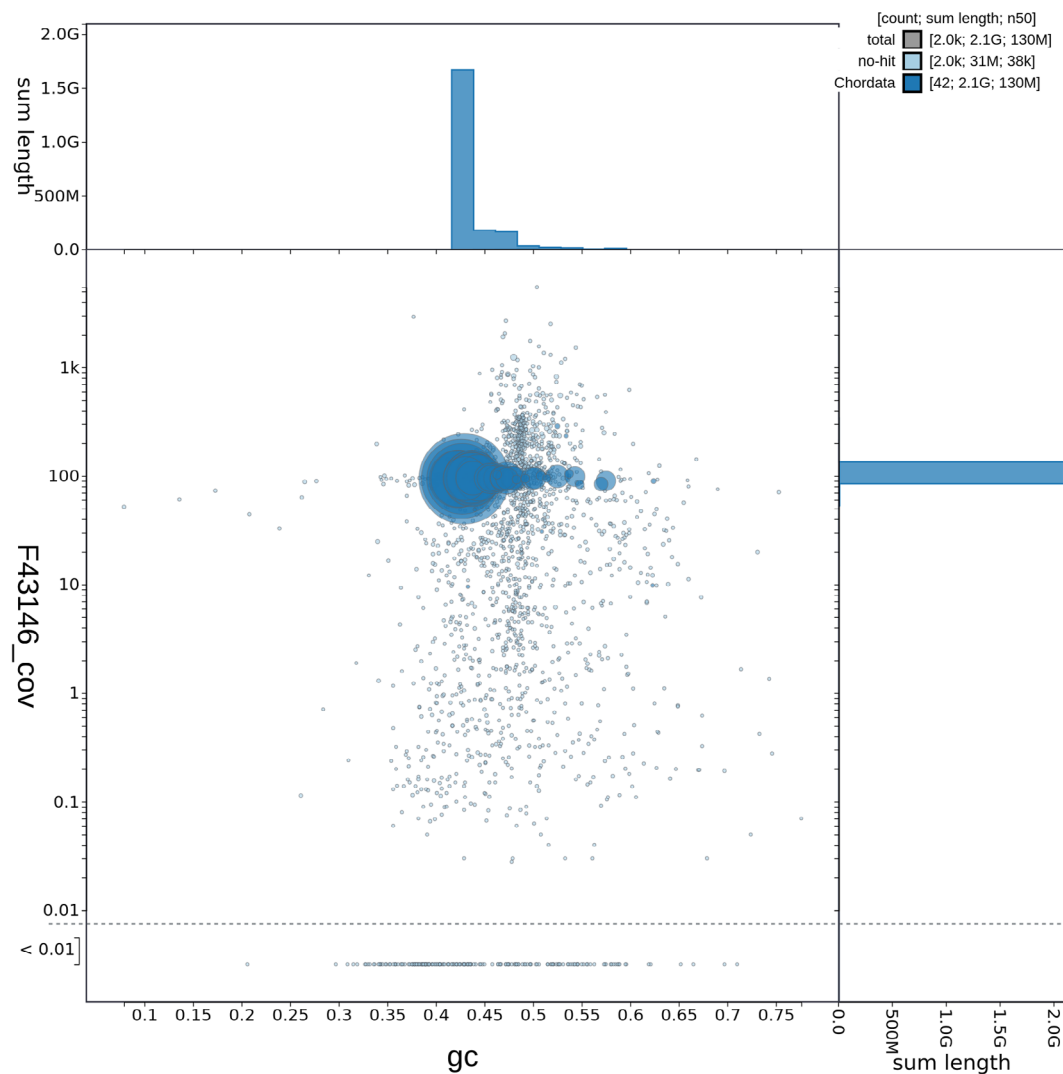


Figure 2. Genome assembly of *Caretta caretta*, rCarCar2: GC-content. GC-coverage plot of *C. caretta* (rCarCar2) generated from Blobtoolkit v.2.6.4 (Challis *et al.*, 2020). Scaffolds are coloured by phylum with Chordata represented by blue and no-hit represented by pale blue. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis.

overseen by Fondazione Cetacea. The specimens were transferred to Canada with two CITES permits between institutions (IT002 and CA027).

Sample extraction, library construction and sequencing

High-molecular weight (HMW) DNA was extracted from nucleated blood using the MagAttract HMW DNA kit (QIAGEN, Germantown, MD, USA). Nanopore genome libraries were constructed according to manufacturer instructions and sequenced using the PromethION instrument (Oxford Nanopore Technologies). A PCR-free genome library was sequenced in a multiplexed pool of an Illumina NovaSeq 6000 instrument S4 flowcell with paired-end 150 bp (PE150) reads. A Hi-C library was constructed using the Arima-HiC kit 2.0 (Arima Genomics, San Diego, CA) and the Swift Biosciences Accel-NGS 2S Plus DNA Library Kit (Integrated DNA Technologies, Mississauga, ON, Canada) and subjected to PE150 sequencing on an Illumina NovaSeq 6000 instrument. All lab work were performed at Canada's Michael Smith Genome Sciences Centre at BC Cancer.

Genome assembly

Assembly was carried out using Redbean (Ruan and Li, 2019), followed by four rounds of racon (Vaser *et al.*, 2017) polishing and medaka (medaka, n.d.) polishing. Scaffolding with Hi-C data was carried out using nf-core/hic workflow

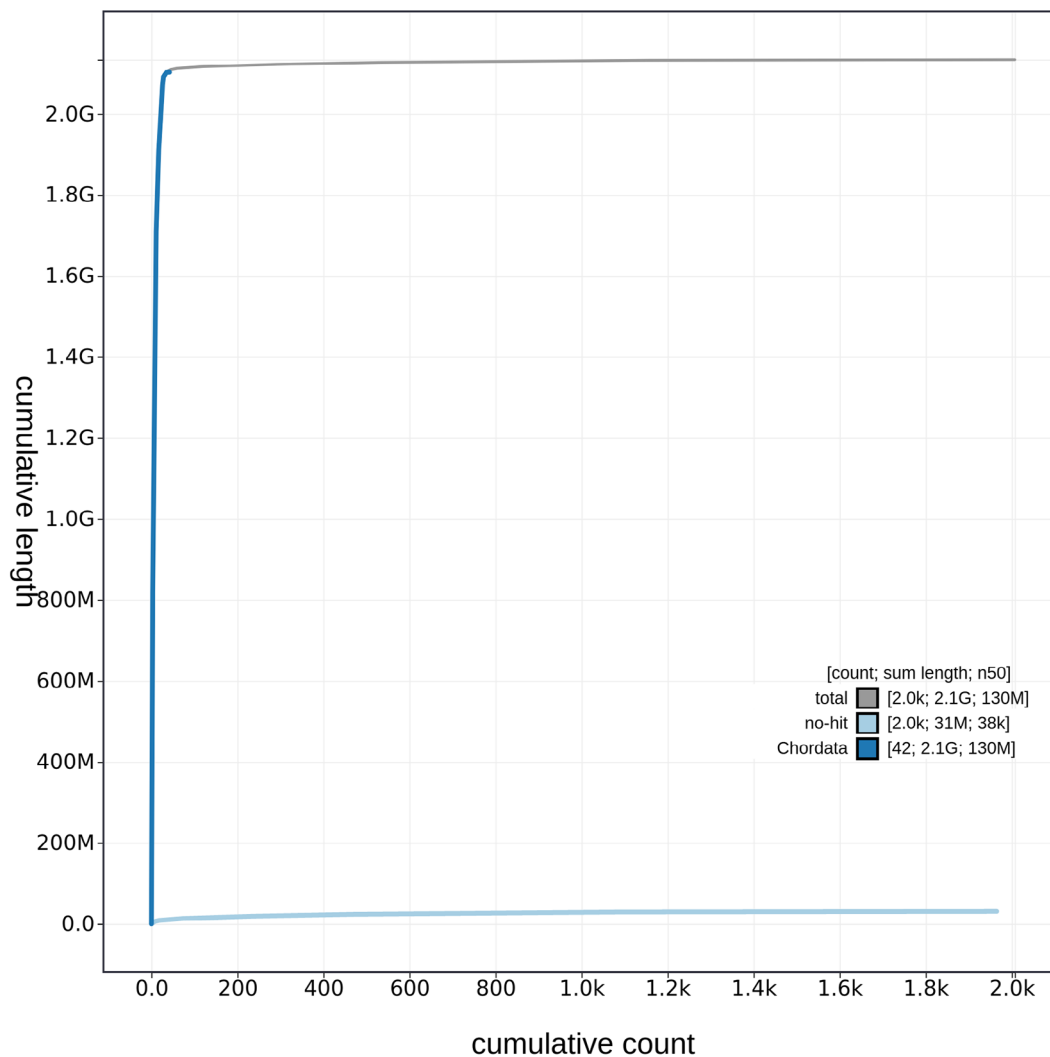


Figure 3. Genome assembly of *Caretta caretta*, rCarCar2: cumulative sequence length. Cumulative sequence length of *C. caretta* (rCarCar2) generated from Blobtoolkit v.2.6.4 (Challis *et al.*, 2020). The grey line shows the cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the BUSCO genes tax rule, with Chordata represented by blue and no-hit represented by pale blue.

(Servant and Peltzer, 2019), Salsa (Ghurye *et al.*, 2019) and LongStitch (Coombe *et al.*, 2021). The Hi-C scaffolded assembly was polished using Illumina short-reads using Pilon (Walker *et al.*, 2014). Four rounds of manual assembly curation and re-scaffolding with nf-core/hic workflow (Servant and Peltzer, 2019) and Salsa (Ghurye *et al.*, 2019) corrected 54 missing/misjoins. The changes were visualized with a Hi-C contact map using Juicer (Durand *et al.*, 2016b). JupiterPlots (Chu, 2018) was used to perform scaffold-level alignment with Green turtle reference genome and generate synteny plot for synteny analysis. The final sequence was analyzed using BlobToolKit (Challis *et al.*, 2020) for quality assessment and RepeatMasker (Tarailo-Graovac & Chen, 2009) for annotation of repetitive regions. The parameter and version number of software tools are listed in Table 3.

Results

Genome sequence report

The genomes of two unrelated loggerhead sea turtles were sequenced from the same population collected from the Fondazione Cetacea hospital, Riccione, Italy. A total of 39-fold coverage in Nanopore PromethION long reads were generated from a single adult female. Approximately 50-fold coverage in Illumina NovaSeq6000 150 bp paired-end (PE150) reads and 18-fold coverage in Illumina NovaSeq6000 Hi-C sequencing were generated from a second individual. Primary assembly contigs from Nanopore data were further polished with Illumina PE150 shotgun sequencing data and scaffolded with Hi-C data. The final assembly has a total length of 2.13 Gb in 2007 sequence scaffolds with a scaffold N50 of 130.95 Mb (Table 1). The majority (98.0%) of the assembly sequence was assigned to 28 chromosomal-

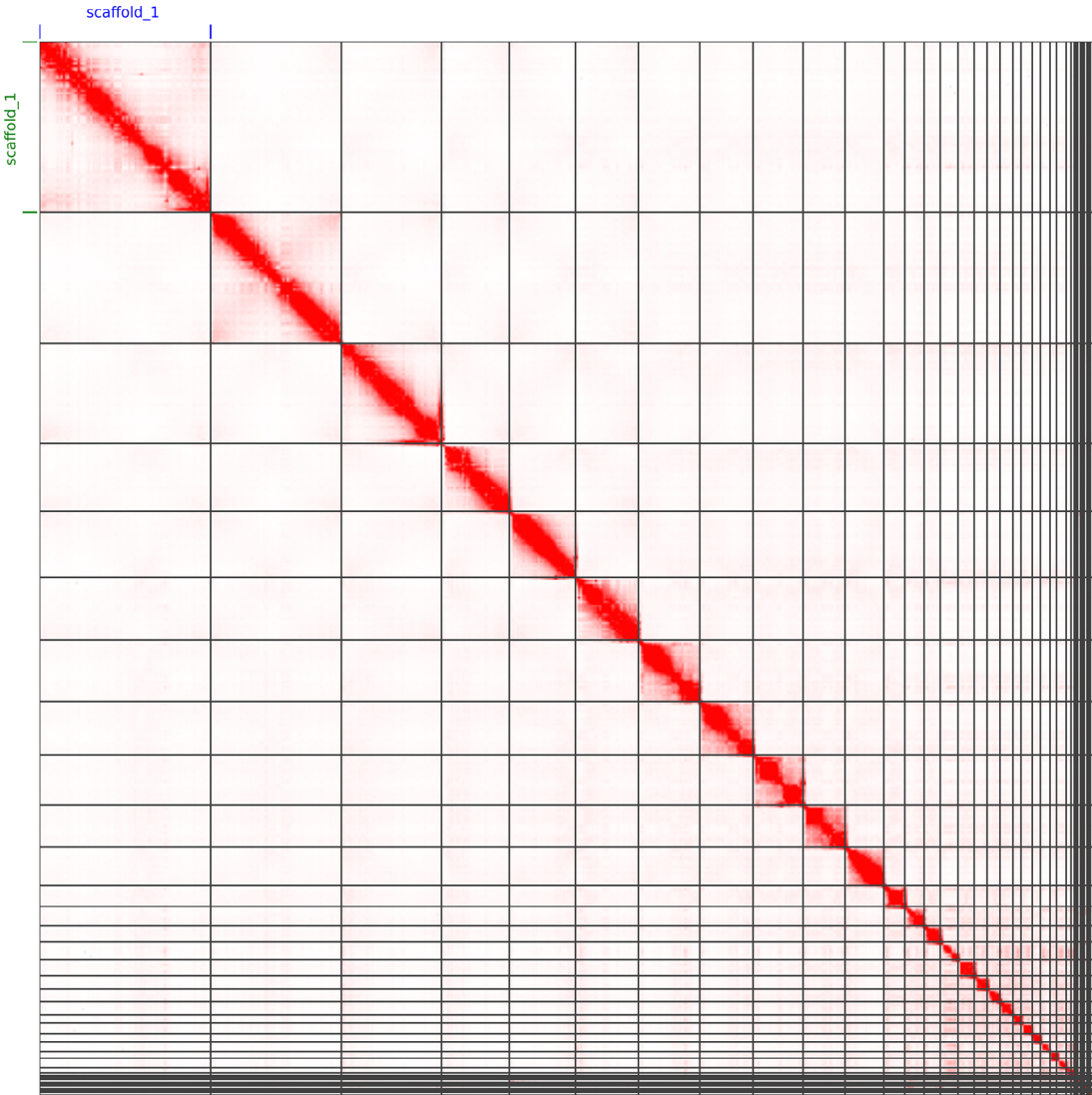


Figure 4. Genome assembly of *Caretta caretta*, rCarCar2: Hi-C contact map. HiC contact map of rCarCar2 assembly visualized using JuiceBox v2.13.07 (Durand *et al.*, 2016a). Chromosomes are shown in order of size from left to right and top to bottom. As an additional confirmation for the quality of the assembly, the microchromosomes are visible as a cluster of spatially-associated contigs in the lower right, as reported in by Waters *et al.*, 2021.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Caretta caretta*, rCarCar2.

RefSeq sequence	Chromosome	Size (Mb)	GC%
NC_064473.1	1	345.74	42.86
NC_064474.1	2	265.32	42.62
NC_064475.1	3	208.08	42.71
NC_064476.1	4	135.63	42.34
NC_064477.1	5	130.96	42.42
NC_064478.1	6	128.66	43.74
NC_064479.1	7	123.31	43.74

Table 2. *Continued*

RefSeq sequence	Chromosome	Size (Mb)	GC%
NC_064480.1	8	108.54	43.66
NC_064481.1	9	101.34	43.68
NC_064482.1	10	85.28	44.40
NC_064483.1	11	76.53	43.00
NC_064484.1	12	43.19	43.81
NC_064485.1	13	38.20	47.24
NC_064486.1	14	35.79	45.97
NC_064487.1	15	33.48	45.53
NC_064488.1	16	25.69	46.28
NC_064489.1	17	24.70	45.64
NC_064490.1	18	23.65	46.93
NC_064491.1	19	20.21	48.10
NC_064492.1	20	19.04	47.85
NC_064493.1	21	18.99	46.81
NC_064494.1	22	17.93	52.48
NC_064495.1	23	16.78	47.24
NC_064496.1	24	16.65	49.92
NC_064497.1	25	16.37	50.20
NC_064498.1	26	13.31	54.27
NC_064499.1	27	12.55	57.47
NC_064500.1	28	5.34	57.00

Table 3. Software tools used.

Software	Version	Parameters	Source
Racon	1.4.13	Default parameters	Vaser <i>et al.</i> , 2017
Medaka	1.2.0	Default parameters	https://github.com/nanoporetech/medaka
Pilon	1.23	Default parameters	Walker <i>et al.</i> , 2014
Salsa	2.3	-m CLEAN -e GATC,GATC,CTNAG,TTAA	Ghurye <i>et al.</i> , 2019
BlobToolKit	2.6.4 (BTK pipeline) 3.1.0 (Blobtoolkit)	Default parameters	Challis <i>et al.</i> , 2020
nf-core/hic	1.1.0	-restriction_site '^GATC,G^ANTC,C^TNAG,T^TAA' -ligation_site 'GATCGATC,GANTGATC,GANTANTC,GATCANTC' -skip_tads	Servant and Peltzer, 2019
Juicer Tools	2.13.06	Default parameters	Durand <i>et al.</i> , 2016b
Juice Box	2.13.06	Default parameters	Durand <i>et al.</i> , 2016a
Redbean	2.5	Default parameters	Ruan and Li, 2019
LongStitch	1.0.1	tigmint-ntLink-arks G=2e9 z=100	Coombe <i>et al.</i> , 2021
Jupiter Plot	1.0	ng=98	Chu, 2018
Busco	5.2.2	-l sauropsida_odb10	Manni <i>et al.</i> , 2021
Quast	5.0.2	Default parameters	Gurevich <i>et al.</i> , 2013
RepeatMasker	4.1.5	-species "Caretta caretta"	Tarailo-Graovac & Chen, 2009

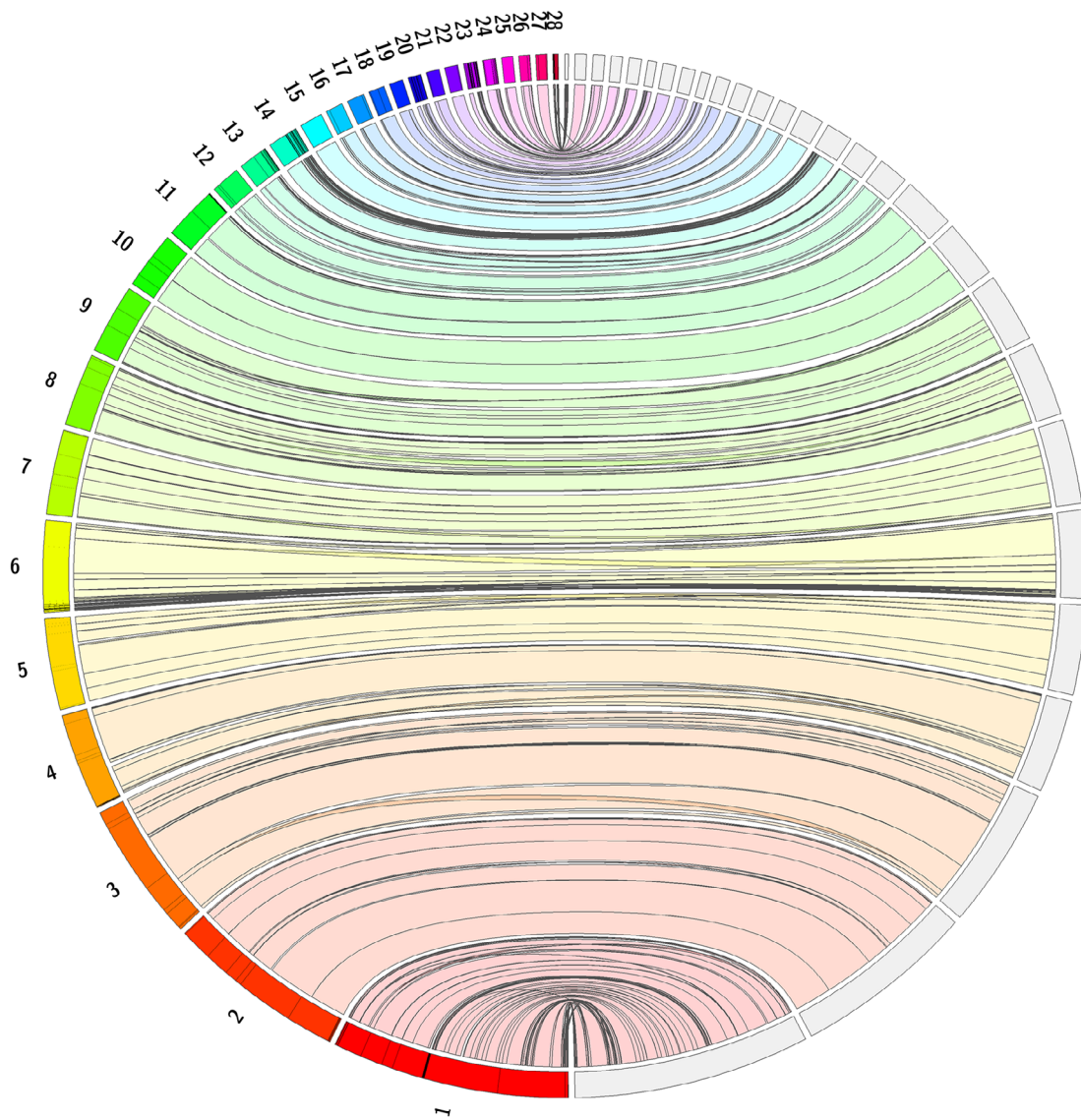


Figure 5. Jupiter plot alignment of *Caretta caretta* with *Chelonia mydas* (green sea turtle). Full genome alignment of *Caretta caretta* genome, rCarCar2 (right), and *Chelonia mydas* (green sea turtle) genome (primary haplotype v2), rCheMyd1 (left), generated using Jupiter Plot (Chu, 2018). The left of the circle shows 28 green sea turtle chromosomes and the right of the circle shows 28 loggerhead sea turtle chromosomes. Coloured bands represent synteny between the genomes, and lines crossing the circle indicate genomic rearrangements, or break points in the scaffolds.

level scaffolds representing the species' known 28 autosomes (Kamezaki, 1989, Machado *et al.*, 2020) (numbered by sequence length; Figure 1–Figure 4; Table 2). Aligned reads from the second turtle to the final assembly had an estimated heterozygosity of 0.11% (2,449,606 heterozygous hits). Determining gene coverage using BUSCO, we estimated 96.1% gene completeness using the sauropsida_odb10 reference set (Manni *et al.*, 2021). The assembly was compared to a previous chromosome-scale assembly of the closely-related green sea turtle, *Chelonia mydas* (Wang *et al.*, 2013), which has been reported to hybridize with the loggerhead sea turtle (James *et al.*, 2004, Vilaça *et al.*, 2012). The loggerhead sea turtle assembly showed strong synteny to the green sea turtle assembly, as shown in Figure 5. The primary haplotype (rCheMyd1.pri.v2) of the green sea turtle was downloaded from NCBI on July 16, 2022. The proportions of SINEs, LINEs, LTR elements, and DNA transposons within the genomic sequences were determined to be 1.55%, 8.75%, 0.13%, and 1.10%, respectively.

Genome annotation

The loggerhead sea turtle genome assembly was annotated by both RefSeq annotation pipeline (Li *et al.*, 2020) and Ensembl gene annotation system (Aken *et al.*, 2016). The RefSeq annotation pipeline includes 24,923 genes and pseudogenes, and 54,583 mRNA transcripts (NCBI *Caretta caretta* Annotation Release). The Ensembl annotation includes 19,633 coding genes, 4,161 non-coding genes and 42,302 mRNA transcripts (*Caretta caretta* - Ensembl Rapid Release).

Data availability

Underlying data

National Centre for Biotechnology Information BioProject: Loggerhead Sea turtle (*Caretta caretta*) genome sequencing and assembly, rCarCar2. Accession number: [PRJNA826225](#).

The genome sequence is released openly for reuse. The *C. caretta* genome sequencing initiative is part of the Canadian BioGenome Project and CanSeq150 Projects initiatives. All raw sequence data and the assembly have been deposited in INSDC databases. The genome is annotated through the Reference Sequence (RefSeq) database in BioProject accession number [PRJNA853764](#). Raw data and assembly accession identifiers are reported in [Table 1](#).

References

- Aken BL, *et al.*: **The Ensembl gene annotation system**. *Database*. 2016; **2016**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Alduina R, Gambino D, Presentato A, *et al.*: **Is *Caretta caretta* a carrier of antibiotic resistance in the Mediterranean Sea?** *Antibiotics*. 2020; **9**(3): 116.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Caracappa S, Persichetti M, Piazza A, *et al.*: **Incidental catch of loggerhead sea turtles (*Caretta caretta*) along the Sicilian coasts by longline fishery**. *PeerJ*. 2018; **6**: e5392.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Casale P, Tucker A: ***Caretta caretta* (amended version of 2015 assessment)**. IUCN red list of threatened species. 2015.
[Publisher Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – Interactive quality assessment of genome assemblies**. *G3: Genes, Genomes, Genetics*. 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chu J: **Jupiter Plot: A Circos-based tool to visualize genome assembly consistency (1.0)**. *Zenodo*. 2018.
[Publisher Full Text](#)
- Coombe L, Li J, Lo T, *et al.*: **LongStitch: High-quality genome assembly correction and scaffolding using long reads**. *BMC Bioinformatics*. 2021; **22**(1): 534.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Durand N, Robinson J, Shamim M, *et al.*: **Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom**. *Cell Systems*. 2016a; **3**(1): 99–101.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Durand N, Shamim M, Machol I, *et al.*: **Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments**. *Cell Systems*. 2016b; **3**(1): 95–98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ghurye J, Rhie A, Walenz B, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly**. *PLoS Comput. Biol.* 2019; **15**(8): e1007273.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gurevich A, Saveliev V, Vyahhi N, *et al.*: **QUAST: Quality assessment tool for genome assemblies**. *Bioinformatics*. 2013; **29**(8): 1072–1075.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- James M, Martin K, Dutton P: **Hybridization between a green turtle, *Chelonia mydas*, and Loggerhead Turtle, *Caretta caretta*, and the first record of a Green Turtle in Atlantic Canada**. *The Canadian Field-Naturalist*. 2004; **118**(4): 579.
[Publisher Full Text](#)
- Kamezaki N: **Karyotype of the loggerhead turtle, *Caretta caretta*, from Japan**. *Zool. Sci.* 1989; **6**: 421–422. Retrieved 4 August 2022.
[Reference Source](#)
- Li W, O'Neill KR, Haft DH, *et al.*: **RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with protein family model curation**. *Nucleic Acids Res.* 2020; **49**(D1): D1020–D1028.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Machado CR, Glugoski L, Domit C, *et al.*: **Comparative cytogenetics of four sea turtle species (Cheloniidae): G-banding pattern and in situ localization of repetitive DNA units**. *Cytogenet. Genome Res.* 2020; **160**(9): 531–538.
[PubMed Abstract](#) | [Publisher Full Text](#)
- medaka: **Sequence correction provided by ONT Research**. Accessed 4 August 2022.
[Reference Source](#)
- Manni M, Berkeley M, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol. Biol. Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mazaris A, Matsinos G, Pantis J: **Evaluating the impacts of coastal squeeze on sea turtle nesting**. *Ocean Coast. Manag.* 2009; **52**(2): 139–145.
[Publisher Full Text](#)
- Pulcinella J, Bonanomi S, Colombelli A, *et al.*: **Bycatch of loggerhead turtle (*Caretta caretta*) in the Italian Adriatic midwater pair trawl fishery**. *Front. Mar. Sci.* 2019; **6**: 365.
[Publisher Full Text](#)
- Ruan J, Li H: **Fast and accurate long-read assembly with wtdbg2**. *Nat. Methods*. 2019; **17**(2): 155–158.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Savoca D, Arculeo M, Barreca S, *et al.*: **Chasing phthalates in tissues of marine turtles from the Mediterranean Sea**. *Mar. Pollut. Bull.* 2018; **127**: 165–169.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Servant N, Peltzer A: **nf-core/hic: Initial release of nf-core/hic (v1.0)**. *Zenodo*. 2019.
[Publisher Full Text](#)
- Shamblin BM, Bolten AB, Abreu-Grobois FA, *et al.*: **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*. 2014; **9**(1): e85956.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Species at Risk Act: **SC 2002, c 29**.
- Tarailo-Graovac M, Chen N: **Using RepeatMasker to identify repetitive elements in genomic sequences**. *Curr. Protoc. Bioinformatics*. 2009; **25**(1): 4.10.1.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vaser R, Sović I, Nagarajan N, *et al.*: **Fast and accurate de novo genome assembly from long uncorrected reads**. *Genome Res.* 2017; **27**(5): 737–746.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vilaça ST, Vargas SM, Lara-ruiz P, *et al.*: **Nuclear markers reveal a complex introgression pattern among marine turtle species on the Brazilian coast**. *Mol. Ecol.* 2012; **21**(17): 4300–4312.
[PubMed Abstract](#) | [Publisher Full Text](#)

Walker B, Abeel T, Shea T, *et al.*: **Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement.** *PLoS One.* 2014; **9**(11): e112963.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wallace B, DiMatteo A, Hurley B, *et al.*: **Regional management units for marine turtles: A novel framework for prioritizing conservation and research across multiple scales.** *PLoS One.* 2010; **5**(12): e15465.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wang Z, Pascual-Anaya J, Zadissa A, *et al.*: **The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan.** *Nat. Genet.* 2013; **45**(6): 701–706.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Waters P, Patel H, Ruiz-Herrera A, *et al.*: **Microchromosomes are building blocks of bird, reptile, and mammal chromosomes.** *Proc. Natl. Acad. Sci.* 2021; **118**(45): e2112494118.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 12 July 2023

<https://doi.org/10.5256/f1000research.151837.r181940>

© 2023 Pegueroles C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Cinta Pegueroles 

Department of Genetics, Microbiology and Statistics, and Institute for Research on Biodiversity (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain

I thank the authors for their thorough revisions, which carefully addressed the comments raised. Congratulations for generating this high quality genome of *Caretta caretta*. It is an importance resource for the scientific community and the management of this vulnerable species.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 27 April 2023

<https://doi.org/10.5256/f1000research.144107.r168077>

© 2023 Pegueroles C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Cinta Pegueroles 

Department of Genetics, Microbiology and Statistics, and Institute for Research on Biodiversity (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain

This manuscript describes the sequencing and annotation of the *Caretta caretta* genome, which is already available in public data bases. It is a high quality genome that for sure is positively impacting the sea turtles community.

The analyses are appropriate and results are sound (despite I miss more details, see below). A high percentage of contigs were assembled into chromosomes, and assembled chromosomes overall showed conserved synteny with the green turtle.

I found surprising that there is no information about repetitive elements. Where they annotated? I strongly recommend to report the levels and type of repetitive elements found within the genome. They can be easily annotated using the repeatMasker software.

In the abstract I recommend to briefly report the quality of the genome assembly, for instance by adding the percentage of complete BUSCO.

Despite genome notes are short by definition, in general I miss more details of how analyses were performed. For instance, there is no information of the parameters used when running the programs and it is not explained how the syntenic analyses were performed, neither the annotation of the genome.

Regarding the annotation of the genome, I do not understand this sentence, "The loggerhead sea turtle assembly was also annotated for 54,583 protein sequences using RefSeq (GCF_023653815.1, PRJNA853764)" since there are 19,633 protein coding genes annotated in Ensembl.

The mitochondrial genome is not reported in this genome note but it is provided in NCBI. I think it should be mentioned here including the tools that were used for its assembly and annotation.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

No

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Author Response 20 Jun 2023

Glenn Chang

Dear Dr. Cinta Pegueroles,

Thank you for your thorough review of our paper. We have carefully considered your comments and have made the following changes to the revised version of the genome note:

1. **Repetitive elements:** We have now used RepeatMaster to annotate the repetitive elements within the genome. The revised paper reports that we found 1.55% SINES, 8.75% LINES, 0.13% LTR elements, and 1.10% DNA transposons within the genomic sequences, as mentioned in the Results section.
2. **Abstract QC metrics:** We have included the busco score and N50 in the abstract to provide quality control metrics right from the beginning.
3. **Software Parameters:** Table 3 now includes the parameters used for each software involved in the genome assembly. In addition to the software name, version, and source, we have added a new column specifically stating the parameters used.
4. **Syntenic analyses:** We have made it more explicit in the paper that JupyterPlot was specifically used to perform scaffold-level alignment and synteny plots for the syntenic analysis.
5. **Gene Annotation pipeline:** The revised paper now clearly states that this genome underwent two annotation pipelines, namely the RefSeq annotation pipeline and the Ensembl gene annotation system. We have provided clearer results for both annotations in the genome annotation section.
6. **Mitochondria:** We did not examine the mitochondrial genome in this study. However, it was automatically grouped with our genome by NCBI. The other mitochondrial genome study can be found here:
<https://pubmed.ncbi.nlm.nih.gov/22295859/>

Thank you once again for your time and valuable feedback during the review process. We believe that these changes have strengthened our paper and addressed your suggestions appropriately.

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 April 2023

<https://doi.org/10.5256/f1000research.144107.r168076>

© 2023 Challis R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Richard Challis 

Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

Chang et al. present a chromosomal genome assembly of the Loggerhead sea turtle, *Caretta caretta*, using a combination of Nanopore long reads, HiC and Illumina. The conservation importance of having a genome assembly for this globally distributed but vulnerable species is made very clear.

As the second chromosomal assembly of a marine turtle it is informative to see a synteny plot comparing this to the green sea turtle, *Chelonia mydas*. This highlights the strongly conserved synteny, similarity in overall assembly span and relative chromosome sizes between these species while maintaining the concise focussed approach typical of a Genome Note.

Overall the article was very clearly presented, however the presentation of summary information about the 2 sets of gene annotation was slightly inconsistent and I found myself referring to the RefSeq annotation page to compare the numbers of coding vs no-coding genes with the values presented for the Ensembl annotation.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 20 Jun 2023

Glenn Chang

Dear Dr. Richard Challis,

Thank you for reviewing our genome note and providing valuable comments. We have carefully considered your comments and made the necessary revisions to address your concerns.

In particular, we have taken steps to clarify the genome annotation sections. We have made the results of the RefSeq and Ensembl annotation pipelines more distinct in the paper. Additionally, we have provided hyperlinks to both sets of results, allowing readers to access them directly.

Once again, we sincerely appreciate your time and effort in reviewing our genome note. We believe that the changes we have made effectively address your concerns and improve the clarity of our paper.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research