



Divergence and conservation of satellite DNA in the loggerhead sea turtle (*Caretta caretta*): a comparative cytogenomic analysis

Katerina V. Tishakova¹ · Anastasia A. Proskuryakova¹ · Dmitry Yu. Prokopov¹ · Sergey A. Marchenko¹ · Alessio Iannucci² · Claudio Ciofi² · Livia Tolve² · Paolo Casale² · Pasquale Salvemini³ · Natalia A. Serdyukova¹ · Guzel I. Davletshina¹ · Vladimir A. Trifonov¹ · Svetlana A. Romanenko¹

Received: 26 November 2025 / Revised: 17 February 2026 / Accepted: 17 February 2026
© The Author(s), under exclusive licence to Springer Nature B.V. 2026

Abstract Repetitive DNA sequences, particularly satellite DNA (satDNA), are key components of eukaryotic genomes with important structural and potential functional roles. In this study, we used long-read assemblies and short-read sequencing to perform de novo identification of the most abundant satDNAs in the sea turtle *Caretta caretta*. We combined a bioinformatic approach and chromosome mapping to describe the distribution of the major tandem repetitive DNA sequences of this species. Additionally, we mapped major ribosomal DNA clusters, telomeric sequences and performed Chromomycin A3-DAPI-after G-banding (CDAG) staining to visualise AT-/

GC-rich chromatin composition. We found that the *C. caretta* satellitome contains at least 18 satDNAs. Most of the satDNAs are localised in pericentromeric regions of chromosomes. We observed a differential, non-uniform distribution of repetitive DNA sequences between chromosomes: while most chromosomes carry blocks consisting of a single chromosome-specific satellite, some chromosomes contain complex heterochromatic blocks composed of multiple satellites. Comparative analysis showed inconsistencies between in situ and in silico satDNA mapping. These new data add to the knowledge about the diversity, distribution, and evolution of satDNA in the *C. caretta* genome.

Responsible Editor: Gustavo Kuhn

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10577-026-09794-6>.

K. V. Tishakova (✉) · A. A. Proskuryakova · N. A. Serdyukova · G. I. Davletshina · V. A. Trifonov · S. A. Romanenko
Laboratory of Diversity and Evolution of Genomes, Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia
e-mail: tishakova@mcb.nsc.ru

D. Y. Prokopov
Independent Researcher, London, UK

S. A. Marchenko
Department of Natural Science, Novosibirsk State University, Novosibirsk, Russia

Keywords Assembly · FISH · Repeats · Reptilia · Testudines

A. Iannucci · C. Ciofi · L. Tolve
Department of Biology, University of Florence, Sesto Fiorentino, Italy

P. Casale
Department of Biology, University of Pisa, Pisa, Italy

P. Salvemini
Associazione Panda Molfetta, Molfetta, Bari, Italy

Introduction

Eukaryotic genomes have a complex and dynamic structure, where a significant portion is composed of repetitive DNA sequences (Richard et al. 2008; Biscotti et al. 2015). Among these, satellite DNA (satDNA) is a fundamental component, characterized by long, head-to-tail tandem repeats that can span millions of base pairs (bp) (Plohl et al. 2012). The classification of satDNA is primarily based on repeat unit length, and these sequences are divided into microsatellites (repeats 1–9 bp long), minisatellites (repeats 10–100 bp long), and satellites (repeats hundreds of bp long) with an ongoing discussion about the length limits for each group (Ruiz-Ruano et al. 2016). Arrays of satDNA may be organized into high-order repeats (HORs), where larger repeat units are composed of multiple basic repeat units (Sujiwattananarat et al. 2015). The entire collection of satDNAs families in a genome is referred to as “satellitome” (Ruiz-Ruano et al. 2016).

Historically, satDNA has been categorized as “junk DNA” and considered a passive, selfish element with no discernible function, simply cluttering genomes (Ohno 1972). This perception was fueled by the enigmatic characteristics of satDNA: its predominant localization in genetically silent heterochromatic regions — specifically centromeres, pericentromeric regions, and telomeres (Charlesworth et al. 1994). Advances in genomic and bioinformatic technologies have revealed the structural and functional role of satDNA in the genome. As a structural component of genomes, satellite repeats are crucial for maintaining genomic architecture: they form ribosomal clusters (Davidian et al. 2022), constitute centromeric region (e.g., human alpha-satellite as a part of HORs) (Sundararajan and Straight 2022), and make up subtelomeric regions (Trofimova and Krasikova 2016). As a functional component of the genome, satDNA is required for correct chromosome segregation during cell divisions and may influence gene expression regulation (Klein and O’Neill 2018; Liao et al. 2023). Transcription of satDNA may produce non-coding RNAs that play key roles in heterochromatin formation (Grewal and Elgin 2007) and in the maintenance of chromosomal integrity (Jacquier 2009); however, overexpression of satDNA may lead to genomic instability (Ting et al. 2011).

SatDNA exhibits high variability in both copy number and sequence among species, even closely related, reflecting rapid evolution (Modi et al. 2003; Bolsheva et al. 2019; Ivanova et al. 2022; De Lima and Ruiz-Ruano 2022). While, in some cases, nucleotide sequences are preserved for millions of years due to concerted evolution, homogenization, and selective pressure (Plohl et al. 2010; Palacios-Gimenez et al. 2020; Dos Santos et al. 2021), the composition of satDNA can diverge considerably during speciation. Several models of satDNA evolution have been proposed to explain this variability. For example, the library hypothesis suggests that related species inherit a common “library” or repertoire of many different satDNA families from a common ancestor (Fry and Salser 1977; Meštrović et al. 1998; Ugarković and Plohl 2002). Following speciation, different subsets of these families are selectively amplified in different evolutionary lineages, leading to the dramatic divergence in satDNA composition observed between closely related species (Fry and Salser 1977; Ugarković and Plohl 2002). This process of differential amplification operates in tandem with other mechanisms. The molecular drive concept explains how non-random mechanisms like unequal crossing-over and gene conversion lead to the concerted evolution and rapid homogenization of the amplified repeats within a species (Ugarković and Plohl 2002; Dover 2002). Furthermore, the centromere drive hypothesis posits that an evolutionary arms race between centromeric satellites (which favor transmission in female meiosis) and centromeric proteins (which suppress this drive) acts as a selective pressure on which sequences from the library are amplified and maintained (Henikoff et al. 2001). New satDNA can originate *de novo* via tandem amplification of specific genomic segments through unequal crossing-over, slipped-strand mispairing, rolling-circle replication, and repair processes including non-homologous end joining and non-allelic homologous recombination (Ugarković and Plohl 2002; Mazurczyk and Rybaczek 2015; Zattera and Bruschi 2022). In addition, existing satDNA sequences can diminish or be lost through mutational decay, deletions, and rapid turnover (Camacho et al. 2022).

Turtles (Testudines), an order of Reptiles, are one of the ancient amniotic groups (Crawford et al. 2015), having diverged from other vertebrates approximately 320 million years ago (MYA) (Shaffer et al. 2017).

Their genomes have a bimodal structure and demonstrate conservative genetic characteristics, such as relatively low rates of changes in chromosome number (ranging from $2n=26$ to $2n=68$), chromosomal morphology, and G-banding patterns (Badenhorst et al. 2015; Montiel et al. 2016).

The current knowledge of satDNA in turtle genomes is exceptionally limited. Excluding the standard cytogenetic mapping of rRNA genes and telomeric repeats (Henikoff et al. 2001; Martinez et al. 2009; Kawagoshi et al. 2012; Cavalcante et al. 2020; Machado et al. 2020b; Clemente et al. 2021), satDNA has been characterized in only a few instances. These include the PSI-Bgl family in the Chinese soft-shelled turtle (*Pelodiscus sinensis*) (Yamada et al. 2005), the SCR-AluI-09 and SCR-AluI-01 families in the black marsh turtle (*Siebenrockiella crassicolis*), with the latter also found in two other species (Kawagoshi et al. 2012). Six major satellite sequences were revealed in the genome of *Trachemys scripta elegans* in our recent study (Romanenko et al. 2025). Non-telomeric microsatellites have been identified in the heterochromatin of four sea turtle species (Machado et al. 2020a).

The loggerhead sea turtle (*Caretta caretta*) is one of seven marine turtle species (superfamily Chelonioidae), and is found in the Indian, Pacific, and Atlantic (including Mediterranean Sea) oceans (Wallace et al. 2010). Loggerhead is a vulnerable species whose populations have declined sufficiently to reduce its range, primarily due to bycatch, coastal development, and poaching (Wallace et al. 2011). Temperature-dependent sex determination, where higher incubation temperature promotes female development, makes it a climate-vulnerable species (Yntema and Mrosovsky 1982). The karyotype of *C. caretta* has been described using conventional cytogenetic methods and revealed 12 macrochromosome pairs and 16 microchromosome pairs ($2n=56$) (López et al. 2008; Machado et al. 2020b). Previous C-banding of *C. caretta* metaphase chromosomes (Machado et al. 2020a) revealed small heterochromatic blocks in pericentromeric regions of several macrochromosomes and microchromosome pairs, and in the long arms of three microchromosome pairs. The genome of *C. caretta* has been investigated repeatedly: at least three long-read-based genome assemblies were published recently (Chang et al. 2023; Arantes et al. 2025; Yen et al. 2025). However,

satellite DNAs in the genome of *C. caretta* have been poorly studied, only distribution of several microsatellites has been described in its karyotype using fluorescent in situ hybridization (FISH) (Machado et al. 2020a).

In this study, we aimed to fill the existing gaps in our understanding of the diversity, distribution, and evolution of the major fraction of repetitive DNA in the *C. caretta* genome using a combination of cytogenetic and bioinformatic approaches.

Methods

Sample collection, cell culturing and metaphase chromosome preparation

Cell cultures of two female specimens of *C. caretta* (CCAR1f and CCAR2f) were obtained from biopsies. Samples were obtained under permits Prot. 0004851 PNM (07/03/2016) issued by the Italian ministry MATTM (Ministero dell'ambiente e della tutela del territorio e del mare).

Cell lines of both individuals were cultivated in α MEM (Gibco) supplemented with 15% FBS (HyClone), 100 mg/L ampicillin, 100 mg/L streptomycin, and 2.5 mg/L amphotericin B at 28–30 °C in an atmosphere of 5% CO₂ and deposited in the “Cryobank of cell cultures” IMCB SB RAS. Chromosome suspensions were prepared from the cell cultures in the Laboratory of diversity and evolution of genomes, IMCB SB RAS, Novosibirsk, Russia. Three and half hours before cell harvesting colcemid (KaryoMAX, Gibco) to a final concentration of 0.1 μ g/ml and ethidium bromide to a final concentration of 3.5 μ g/ml were added. Cells were dissociated by 0.25% trypsin with 0.2% EDTA and centrifuged for 5 min at $0.6\times g$. The pellet was gently resuspended in hypotonic solution (62 mM KCl with 2% of FBS) and incubated for 1 h at 29 °C. For prefixation 1/20 volume of fresh ice-cold fixative (methanol:acetic acid, 3:1) was added, mixed carefully and incubated for 12 min at 4 °C. Then cells were centrifuged for 5 min at $0.6\times g$ and supernatant was discarded. For fixation, the pellet was covered by ice-cold fixative (–20 °C) and kept for 30 min at –20 °C without mixing. Cells were then centrifuged for 5 min at $0.6\times g$ and resuspended in ice-cold fixative. The chromosome suspensions were stored long-term at –20 °C.

DNA Extraction, Library Preparation, and Whole-Genome DNA Sequencing

Genomic DNA was extracted from CCAR2f cell cultures using the GeneJET Genomic DNA Purification (Thermo Fisher Scientific Baltics, Vilnius, Lithuania) according to the manufacturer's protocol. The extracted genomic DNA was then fragmented and a library was prepared for CCAR2f as described previously (Romanenko et al. 2022). Library sequencing was performed on a MGISEQ-2000 platform using the DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100) at the SB RAS Genomics Core Facility (ICBFM SB RAS, Novosibirsk, Russia). Data were deposited in the NCBI SRA database under the accession number SRR26671777.

Repetitive DNA identification and annotation

In silico identification of putative satDNAs from short reads was performed using RepeatExplorer2 and TAREAN (Novák et al. 2013, 2017) implemented on the Galaxy platform (Afgan et al. 2016). We analyzed publicly available whole-genome sequencing data for *C. caretta* (SRR15328383, female) alongside the sequencing data generated in this study. Reads were filtered for length and quality using the “Pre-processing of FASTQ paired-end reads” tool. Filtering by length, adapter presence, and quality scores was performed as previously described (Novák et al. 2020). For “RepeatExplorer2 clustering” tool the “Database of taxon and protein domain” parameter was set to “Metazoa version 3.0”. The reads were randomly selected and analysed using RepeatExplorer2 and TAREAN. From the 19,954,894 input reads in SRR15328383, 4,229,461 high-quality reads (150 bp length) were randomly selected ($\sim 0.3 \times$ coverage of the reference genome). From the 11,749,647 input reads in SRR26671777 (this study), 4,960,258 high-quality reads (80 bp length) were randomly selected ($\sim 0.2 \times$ coverage). These read numbers and coverages are within the limits recommended by the software developers ($0.01\text{--}0.50 \times$ genome coverage) (Novák et al. 2013, 2017). The selected reads were analyzed using RepeatExplorer2 and TAREAN.

For in silico identification of putative satDNA in long-read genome assemblies we used the publicly available genome data for *C. caretta* (GCA_023653815.1 and GCA_965140235.1). The

analysis was performed using TideCluster (version 1.6) implemented on the Galaxy platform (Novák et al. 2017; Gao et al. 2019, <https://doi.org/10.5281/zenodo.7885625>) with default parameters.

All satellite sequences were annotated using RepBase (Kohany et al. 2006), Tandem Repeat Finder (TRF) (Benson 1999), BLAST (Johnson et al. 2008), and the Dfam databases (Hubley et al. 2016). Manual annotation was performed using default parameters, except that RepBase searches were performed with the “Report simple repeats” option and Dfam database searches were performed with “Other” and “*Chrysemys picta bellii*” as query organisms. The satellite sequences were queried against the ‘core_nt’, ‘nr/nt’, and ‘refseq_reference_genomes’ BLAST NCBI database. For ‘refseq_reference_genomes’ database, the organism filter was set to ‘*Caretta caretta* (taxid:8467)’ and ‘sea turtles (taxid:8456)’. Each sequence in the search field was iterated 5 times to prevent edge effects. Sequence cut off was set manually (E-value < 0.01; Identity $\geq 70\%$; Query Coverage $\geq 70\%$). All satellites identified for *C. caretta* were compared with previously published satDNA from *Trachemys scripta elegans* (Romanenko et al. 2025) using BLAST.

RepeatMasker version 4.1.7-p1 (Smit et al. 2015) was used to estimate the abundance for each satDNA variant. The analysis was run with the options ‘-engine rmbblast’, ‘-a’, ‘-s’, ‘-nolow’, ‘-no_is’, and ‘-xsmall’. The input library contained duplicated consensus sequences of each satDNA. Based on the RepeatMasker *.out file, the coverage of each masked region (cov_reg) was calculated as $(q_{\text{end}} - q_{\text{begin}} + 1) / (\text{length of the consensus sequence})$. Data were filtered using $\text{cov_reg} > 70\%$ and $\text{perc_div} < 30\%$, meaning only regions with sequence identity to the consensus exceeding 70% and coverage exceeding 70% were considered. For each satellite, its genomic representation was calculated as: $(\text{total cov_reg across all filtered regions}) * 100 / (\text{assembly length})$. Genomic data was visualized using the Integrative Genomics Viewer (IGV) version 2.18.4 (Robinson et al. 2011; Thorvaldsdóttir et al. 2013).

Diversity and spatial distribution of the detected satDNA in the assembled *C. caretta* reference genome (GCA_965140235.1) was determined using the SatXplor pipeline with default parameters (Volarić et al. 2025). The assessment of monomer variability and satDNA organization was conducted based on the

graphical images generated by the pipeline, in accordance with the authors' recommendations.

Probes generation

Primers for PCR amplification and labelling of the satellite probes were designed using the Primer3 web tool and Gene Runner software (version 6.5.52) (see Supplementary Table 1). PCR amplification was performed as described previously (Biltueva et al. 2017). The CcarSat-06–50 probe was synthesized and directly labeled with a Cy3 fluorophore by the Biolabmix company (Novosibirsk, Russia).

The telomeric probe was generated by PCR using the oligonucleotides (TTAGGG)₅ and (CCCTAA)₅ (Ijdo et al. 1991). Clones of human ribosomal DNA (rDNA) containing a partial 18S rRNA gene, the full 5.8S rRNA gene, a part of the 28S rRNA gene, and two internal transcribed spacers were obtained as described elsewhere (Maden et al. 1987).

CDAG-staining, GTG-banding, fluorescence in situ hybridization (FISH), and image capture

Chromomycin A3-DAPI-after G-banding (CDAG-banding) was performed as previously described (Lemskaya et al. 2018).

GTG-banding was performed prior to FISH according to the method of Seabright (Seabright 1971) with modifications from Graphodatsky and Radjabli (Graphodatsky and Radjabli 1988). After microscopic examination, image capture, and recording the coordinates of each metaphase on the microscope stage, the slides were washed to remove the immersion oil and Giemsa stain as described previously (Yang and Graphodatsky 2017).

FISH was performed on metaphase plates of CCAR1f and CCAR2f as described previously (Yang and Graphodatsky 2017) with some modifications. Slides were incubated in 2×SSC and PBS (RT) for 4 min and then fixed in 0.5% formaldehyde in PBS and 50 mM MgCl₂ (RT) for 10 min. Slides were rinsed in PBS and washed in 2×SSC for 4 min. A 0.01% RNase A solution in 2×SSC was applied to each slide for 40 min at 40 °C. Then slides were washed in 2×SSC (RT) for 4 min, rinsed in distilled water, and dried with a rubber bulb. Slides were denatured in 70% formamide with 2×SSC at 70 °C for 1 min 10 s, immediately quenched in ice-cold 70% ethanol, and

dehydrated in ice-cold 80%, 96% ethanol for 2 min each. The slides were rinsed with 96% (RT) ethanol and dried using a rubber bulb.

The hybridization probe mixture containing 50% formamide, 2×SSC and 20 ng of each probe were denatured at 96 °C for 5 min, immediately quenched on a PCR-Cooler (Eppendorf), and applied to the slides. The slides were covered with coverslips and the edges of the coverslips were sealed with rubber glue. Slides were incubated overnight at 40 °C in a wet chamber.

For post-hybridization washing, the slides were incubated in 2×SSC (RT) for 4 min, immersed in 0.4×SSC at 72 °C for 2 min, and rinsed in 4×SSCT (0.05% Triton X-100 in 4×SSC) at RT. Pre-detection blocking was performed for 15 min in 0.5% BSA in 4×SSCT, detection of biotin- and digoxigenin-labelled probes was performed as previously described (Yang and Graphodatsky 2017). Counterstaining was performed with DAPI in VECTASHIELD Antifade Mounting Medium (Vector Laboratories).

Images were captured using JenOptic charge-coupled device (CCD; Jenoptic, Germany) camera mounted on an Olympus BX53 microscope (Olympus, Japan) and processed VideoTest Karyo 3.1 (version 3.1.0.153; VideoTest, Russia) and VideoTest FISH 3.1 (version 3.1.0.153; VideoTest, Russia) software.

Results

Karyotype description

The karyotypes of both *C. caretta* specimens consisted of 56 chromosomes (12 macrochromosome pairs and 16 microchromosome pairs), corresponding to the previously described karyotype (López et al. 2008). CDAG-staining (Chromomycin A3-DAPI-after G-banding) (Lemskaya et al. 2018) revealed that macrochromosomes were predominantly AT-rich (DAPI-positive) (Supplementary Fig. 1). Small DAPI-positive blocks were also detected in the *C. caretta* microchromosomes. Intense CMA₃-positive (GC-rich) signals were observed in the pericentromeric and telomeric regions of most macrochromosomes, and across entire microchromosomes (Supplementary Fig. 1). As expected, we observed bright

Table 1 Putative satellites revealed in *Caretta caretta* genome

SatDNA family name	Source	SatDNA monomer length	Method	GC-content	Proportion in the reference genome, %
CcarSat-01–2848	SRR15328383	2848	TAREAN	48.6	1.28
CcarSat-02–2869	GCA_965140235.1	2869	TideCluster	48.59	0.472
	GCA_023653815.1	2863	TideCluster		
	SRR26671777	2862	TAREAN		
	SRR15328383	2869	TAREAN		
CcarSat-03–1149	GCA_965140235.1	1149	TideCluster	53.61	0.065
CcarSat-04–1334	GCA_965140235.1	1334	TideCluster	53.4	0.049
	GCA_023653815.1	1334	TideCluster		
CcarSat-05–665	GCA_965140235.1	665	TideCluster	61.35	0.048
CcarSat-06–50	GCA_965140235.1	50	TideCluster	68	0.015
CcarSat-07–51	GCA_965140235.1	51	TideCluster	64.71	0.003
CcarSat-08–67	GCA_965140235.1	67	TideCluster	67.16	0.004
CcarSat-09–316	SRR26671777	316	TAREAN, RepeatExplorer2	64.9	0.004
	SRR15328383	315	TAREAN, RepeatExplorer2		
CcarSat-10–2402	SRR15328383	2402	TAREAN, RepeatExplorer2	46.8	0.035
CcarSat-11–2207	GCA_965140235.1	2207	TideCluster	46.08	0.035
	SRR26671777	2168	TAREAN, RepeatExplorer2		
	SRR15328383	2170	TAREAN, RepeatExplorer2		
CcarSat-12–2736	GCA_965140235.1	2736	TideCluster	49.2	0.026
CcarSat-13–2115	GCA_965140235.1	2115	TideCluster	50.26	0.008
CcarSat-14–1917	GCA_965140235.1	1917	TideCluster	50.67	0.007
CcarSat-15–2877	GCA_023653815.1	2877	TideCluster	45.6	0.001
CcarSat-16–1319	GCA_965140235.1	1319	TideCluster	70.66	0.001
CcarSat-17–823	SRR26671777	823	TAREAN, RepeatExplorer2	40.5	0.001
CcarSat-18–27	GCA_965140235.1	27	TideCluster	59.26	0.0003
CcarSat-19–1598*	SRR15328383	1598	TAREAN, RepeatExplorer2	53.6	-

* Subsequent bioinformatic and cytogenetic analysis excluded CcarSat-19–1598 from *C. caretta* satellitome

CMA₃-positive signals in the nucleolus organized region (NOR) (Supplementary Fig. 1). The localization of ribosomal and telomeric probes was also performed using FISH (Supplementary Fig. 2a). The ribosomal probe specific to the major ribosomal RNA gene cluster was hybridized to the NOR-bearing chromosome pair 13. The telomeric probe marked the distal parts of all chromosomal arms and did not yield any clear or reproducible interstitial signals.

Bioinformatics search and analysis of repetitive DNA

Bioinformatic satDNA identification of various genomic datasets using multiple tools revealed 26 putative satellites. Analysis of long-read assemblies using TideCluster revealed three putative satDNA

sequences with consensus lengths of 1334, 2863, and 2877 bp in the GCA_023653815.1 assembly and thirteen putative satDNAs (27, 50, 51, 67, 665, 1149, 1319, 1334, 1917, 2115, 2207, 2736, and 2869 bp) in the GCA_965140235.1 assembly (Table 1).

Analysis of *C. caretta* short reads (SRR15328383) using TAREAN within the RepeatExplorer2 pipeline identified four putative satellites with consensus lengths of 315, 1598, 2170, and 2402 bp. A separate analysis of the same dataset using the standalone TAREAN tool identified satellites with consensus lengths of 2170, 2402, 2869, and 2848 bp.

Analysis of CCAR2f dataset (SRR26671777) using TAREAN within RepeatExplorer2 revealed three potential satellites (316, 823, and 2168 bp); additionally, one repeat with a consensus length of

2402 bp was classified as an LTR element. A separate analysis of this dataset using standalone TAREAN identified only the satellites with consensus lengths of 316 and 2862 bp.

SatDNAs sequences discovered in different genomic data that exhibited >90% query coverage and percentage identity were grouped into distinct satDNA families. Following the nomenclature suggested by Ruiz-Ruano et al. (2016), we numbered the satDNA in order of decreasing abundance in the reference genome GCA_965140235.1, as determined by RepeatMasker (Table 1). A summary file of identified satellite DNAs in the genomic data of *C. caretta* is available in the Zenodo repository (Tishakova et al. 2026).

The satDNA genomic proportion varied between 0.0003% (CcarSat-18–27) and 1.28% (CcarSat-01–2848), and the consensus unit length varied from 27 to 2877 bp. The GC-content ranged from 40% (CcarSat-17–823) to 70.7% (CcarSat-16–1319) (Table 1).

Pairwise BLAST alignment of the identified satDNAs indicated that several repeats from different families share sequence homology. For example, segments of CcarSat-02–2869 show high similarity to the consensus sequences with CcarSat-17–823 (100% coverage, 99% identity) and CcarSat-19–1598 (96% coverage, 98.7% identity) (Supplementary Fig. 3). CcarSat-02–2869 and CcarSat-12–2736, both identified in GCA_965140235.1 using Tide-Cluster, also exhibit a high homology (96% coverage and 97.6% identity). The sequence of CcarSat-17–823 showed 90% coverage and 70% identity with CcarSat-11–2207, while CcarSat-05–665 shares homology (76% coverage, 84% identity) with CcarSat-03–1149. A 2402 bp long consensus sequence was identified as an LTR in SRR26671777, but as a satellite in SRR15328383; we refer to this sequence as CcarSat-10–2402.

Self-alignments of satDNA sequences revealed internal repeats within several monomers (Supplementary Fig. 4). CcarSat-09–316, CcarSat-14–Sat1917, CcarSat-15–2877, CcarSat-16–1319, and CcarSat-17–823 (Supplementary Fig. 4b, f–i) contained a relatively small number of internal repeats, whereas CcarSat-10–2402, CcarSat-11–2207, and CcarSat-13–2115 (Supplementary Fig. 4c–e) showed multiple repetitive segments throughout their lengths. CcarSat-02–2869 contained internal

repeats spanning approximately half of its sequence (Supplementary Fig. 4a).

BLAST alignment of satDNA consensus sequences against the *C. caretta* reference genome (GCA_965140235.1) mapped most of them to the assembly, with the exception of CcarSat-18–27, CcarSat-06–50, and CcarSat-08–67. BLAST analysis also mapped most of the consensus satDNA monomers to other available sea turtle genomes.

Annotation of the retrieved satellite sequences revealed no significant similarity to known repetitive elements in Repbase (<https://www.girinst.org/repbase/>) or DFAM (<https://www.dfam.org/>) databases, except for CcarSat-05–665 and CcarSat-03–1149. These satellites were identified as fragments of the Tx1 non-LTR retrotransposon, previously annotated in the *Chrysemys picta bellii* genome. Furthermore, TRF analysis identified simple repeated units, including (TAA)₁₀ in CcarSat-02–2869, an insertion of (ATGTTGTGGGTCGTGTTGTGTTGAGTAGTGTGAGTTGTGACG)₂ units in CcarSat-19–1598, a simple (T)₂₇ tract in CcarSat-03–1149, and an insertion of (GCGT)₉ in CcarSat-16–1319. Comparison of *C. caretta* satDNA with previously published major satellite sequences of *Trachemys scripta elegans* revealed homology between TSE-1959 (Romanenko et al. 2025) and CcarSat-14–1917 (coverage 95%, identity 74.6%).

We used RepeatMasker to determine the coordinates of CcarSat-02–2869, CcarSat-08–67, CcarSat-12–2736, CcarSat-17–823, and CcarSat-19–1598 arrays in available long-read genome assemblies GCA_023653815.1, GCA_965117115.1, and GCA_965140235.1. Only regions with coverage and identity to the consensus exceeding 70% were considered; these regions were visualized using IGV (Supplementary Fig. 5). Under these criteria, CcarSat-19–1598 was not detected in any dataset, and all detected CcarSat-17–823 arrays were shorter than 2.6 kb. No overlaps were observed among the arrays of different satDNAs. An extended array of CcarSat-08–67 was found only on chromosome 5 in GCA_965140235.1. A single CcarSat-12–2736 array was identified on chromosome 26. For CcarSat-02–2869, extended arrays were located on chromosomes 12, 22, and 25, with shorter arrays on chromosomes 15 and 26. Additionally, scaffolds NW_027439743.1 and NW_027439744.1 appeared to

be composed almost entirely of CcarSat-02–2869 (Supplementary Fig. 5).

Most satellite sequences were identified in the *C. caretta* reference genome assembly (GCA_965140235.1) using the SatXplor pipeline, with the exception of CcarSat-08–67. The pipeline identified between 6 (for CcarSat-15–2877) and 3731 (for CcarSat-17–823) monomers in the reference genome assembly. For sequences grouped into individual satDNA families, SatXplor identified a similar number of monomers with identical chromosomes/scaffolds localization (data not shown). Table 2 summarizes the number of mapped monomers, array localization, and satellite analysis results.

Similarity plots comparing consensus monomers to their individual mapped monomers, revealed satDNA with low, moderate, high, and extra-high monomer variability (Table 2., Fig. 1a). UMAP-based clustering of satDNA monomers showed distinct patterns of monomer clustering. The satellite CcarSat-02–2869 revealed a high degree of intrachromosomal similarity with limited interchromosomal mixing (Fig. 1b, Example 1). CcarSat-05–665 and CcarSat-03–1149 displayed several clusters characterized by high interchromosomal mixing of monomers, alongside several chromosome-specific subgroups of monomers (Fig. 1b, Example 4). Repeats CcarSat-18–27 and CcarSat-06–50 contained both chromosome specific monomer subgroups and subgroups with interchromosomal distribution (Fig. 1b, Example 2). CcarSat-17–823 represents an example of an extremely high degree of inter- and intrachromosomal exchange, which prevented efficient clustering (Supplementary Fig. 6). Other satDNA sequences exhibited distinct segregation of monomer subgroups with high or moderate internal similarity (Fig. 1b, Example 3). Only three satDNAs (CcarSat-18–27, CcarSat-06–50, and CcarSat-07–51) showed homogenous organization of their arrays. In contrast, most satDNAs exhibited a complex structure of array and may be organized as a part of HORs (Table 2., Fig. 1c).

Localization of major repeated sequences on turtles' chromosomes

Probes of most satDNA sequences were localized on the metaphase chromosomes of both individuals of *C. caretta* using FISH (Fig. 2, Supplementary Fig. 2,

7). Due to the high sequence similarity between CcarSat-02–2869 and CcarSat-12–2736, we could not design specific oligonucleotides for each satellite separately; therefore, only a probe for CcarSat-02–2869 was generated and localized (Supplementary Table 1). We observed no significant differences in hybridization patterns between the two individuals, except for CcarSat-18–27, which produced a brighter signal on the chromosomes of CCAR2f. Differences in signal size on homologous chromosomes were observed for both CcarSat-04–1334 and CcarSat-08–67 in both individuals (Fig. 2b, d).

Following the FISH results, the probe CcarSat-09–316 labeled an extended interstitial region of chromosome 13 and partially overlapping with the signal of the ribosomal probe (Fig. 2a). CcarSat-15–2877 produced weak signals in the pericentromeric region of chromosome 1, whereas the probe CcarSat-04–1334 labeled pericentromeric region of chromosomes 4, 11, and 13 (Fig. 2b). The CcarSat-02–2869 probe colocalized with the CcarSat-17–823 probe, labeling approximately seven chromosome pairs (Supplementary Fig. 2d). CcarSat-02–2869 clearly marked the p-arm of chromosome 4 and an interstitial region of chromosome 13 (Supplementary Fig. 2d), whereas CcarSat-17–823 showed very weak fluorescent signals on these chromosomes. Both probes labeled the p-arm of chromosome 16 and pericentromeric regions of chromosomes 15, 17, 23, and 26 with the same intensity (Supplementary Fig. 2d). Additionally, CcarSat-02–2869 colocalized with a CcarSat-09–316 on chromosome 13 (Supplementary Fig. 2c). The localization of CcarSat-11–2207 and CcarSat-10–2402 was restricted to the pericentromeric regions of chromosomes 2 and 3, respectively (Fig. 2c). The probe CcarSat-08–67 localized to the pericentromeric region of chromosome 4 (Fig. 2d). The probes CcarSat-06–50, CcarSat-07–51, CcarSat-05–665, CcarSat-03–1149, CcarSat-16–1319, and CcarSat-13–2115 produced dispersed signals (Supplementary Fig. 2e and Supplementary Fig. 7a, b, d). However, FISH revealed that CcarSat-05–665 and CcarSat-06–50 also formed distinct clusters on chromosomes 27 and 12, respectively (Supplementary Fig. 7b, d). The CcarSat-14–1917 probe showed signals in the pericentromeric region of chromosome 19 and the subtelomeric regions of chromosomes 23 (Fig. 2d). The CcarSat-18–27

Table 2 The SatXplor pipeline analysis

SatDNA family name	Number of monomers in GCA_965140235.1	Major arrays localization on chromosome/Scaffold in GCA_965140235.1	Monomer variability, SatDNA organization
CcarSat-01-2848	9769	8 19	Moderate monomer variability -
CcarSat-02-2869	3490	12 21 22 25 NW027439743.1 NW027439744.1	Low monomer variability Complex array organization
CcarSat-03-1149	1164	4 17	Extremely high monomer variability Complex array organization
CcarSat-04-1334	312	11	High monomer variability Complex array organization
CcarSat-05-665	2472	1 3 4 5 7 8 10 14 17 23	Moderate monomer variability Complex array organization
CcarSat-06-50	2481	2 5 9 14	Moderate monomer variability Homogeneous array
CcarSat-07-51	1091	5 12	Moderate monomer variability Homogeneous array
CcarSat-08-67	-	-	-
CcarSat-09-316	114	28	Low monomer variability Complex array organization
CcarSat-10-2402	223	3	Low monomer variability Complex array organization
CcarSat-11-2207	342	2	Moderate monomer variability Complex array organization
CcarSat-12-2736	3005	12 21 22 25 26 NW027439743.1 NW027439744.1	Low monomer variability Complex array organization
CcarSat-13-2115	52	14	High monomer variability Complex array organization
CcarSat-14-1917	54	21 22	Extremely high monomer variability Complex array organization
CcarSat-15-2877	6	1	Extremely high monomer variability -
CcarSat-16-1319	91	2	Low monomer variability Complex array organization

Table 2 (continued)

SatDNA family name	Number of monomers in GCA_965140235.1	Major arrays localization on chromosome/Scaffold in GCA_965140235.1	Monomer variability, SatDNA organization
CcarSat-17-823	3731	12 21 22 25 26 NW027439743.1 NW027439744.1	Moderate monomer variability Complex array organization
CcarSat-18-27	2133	21	Moderate monomer variability Homogeneous array
CcarSat-19-1598*	3708	12 21 22 25 26 NW027439743.1 NW027439744.1	Low monomer variability Complex array organization

* Subsequent bioinformatic and cytogenetic analysis excluded CcarSat-19-1598 from the *C. caretta* satellitome

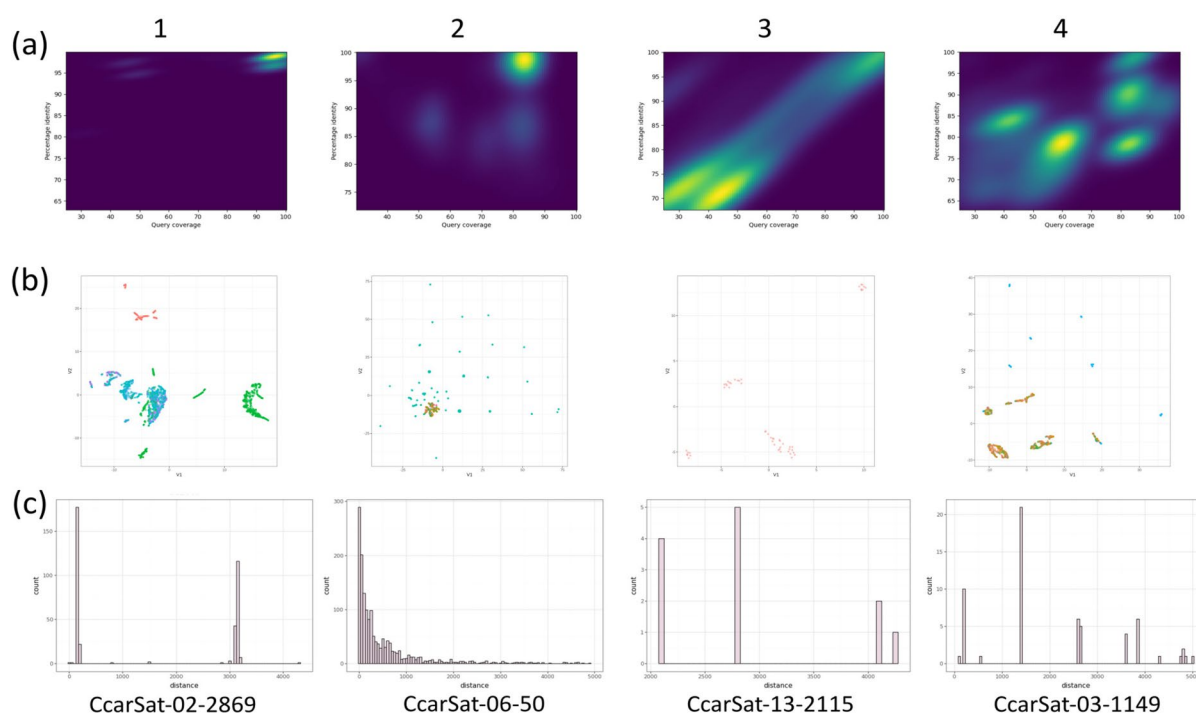


Fig. 1 Examples of SatXplor reports. **a)** Monomer variability plot. Depending on the characteristics of the satDNAs, we observed repeats with low (1), moderate (2), high (3), and extra-high (4) monomer variability. **b)** SatDNA monomer clustering using UMAP. Each color represents a certain chromosome. The figure shows examples of: (1) monomer subgroups with a high degree of intrachromosomal similarity and limited interchromosomal mixing; (2) monomer subgroups with both chromosome-specific and interchromosome distribution;

(3) distinct segregation of monomers subgroups, and (4) high degree of monomer interchromosome mixing alongside several chromosome-specific monomer subgroups. **c)** Complex satDNA organization. CcarSat-06-50 has homogeneous array (2). In contrast, repeats CcarSat-02-2869, CcarSat-13-2115, and CcarSat-03-1149 show complex array organisation where satDNA arrays interspersed with other DNA sequences (1, 3–4)

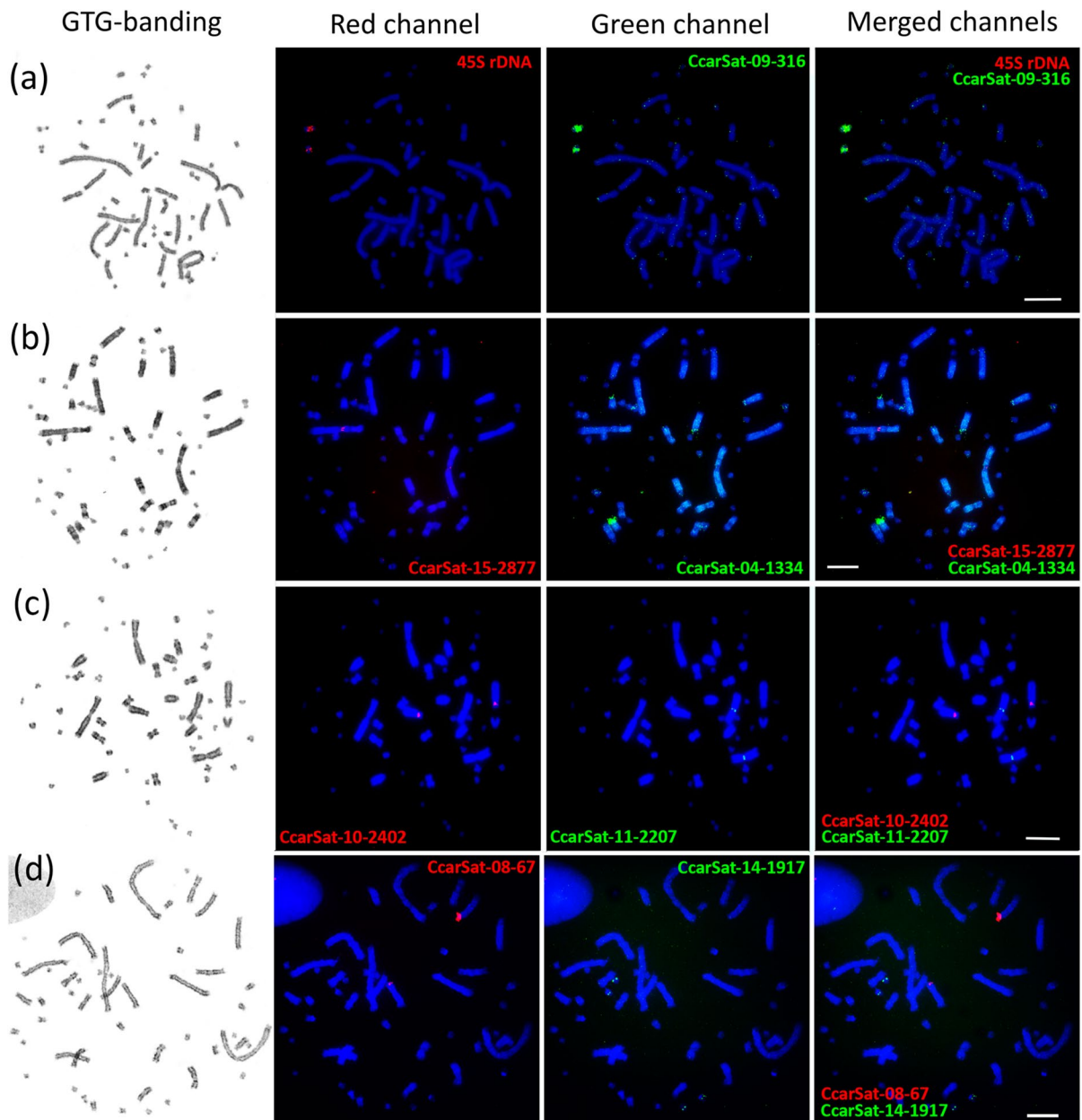


Fig. 2 FISH localization of repetitive DNA probes on *C. caretta* If chromosomes: **a**) ribosomal DNA probe (red) and CcarSat-09-316 (green), **b**) CcarSat-15-2877 (red)

and CcarSat-04-1334 (green), **c**) CcarSat-10-2402 (red) and CcarSat-11-2207 (green), **d**) CcarSat-08-67 (red) and CcarSat-14-1917 (green). Scale bar 10 μ m

probe localized closer to the centromere than CcarSat-14-1917 on chromosome 21 (Supplementary Fig. 2e). The most abundant sequence, CcarSat-01-2848, produced signals in the subtelomeric and pericentromeric regions of most chromosomes (Supplementary Fig. 7c).

Discussion

Repetitive elements continue to be the “dark matter of the genome”, despite accumulating evidence on their structural and functional roles. In turtles, however, knowledge of satellitomes remains still limited

to only a few species (Yamada et al. 2005; Kawagoshi et al. 2012; Machado et al. 2020a; Romanenko et al. 2025). In this study, we provided the first description of the loggerhead turtle (*Caretta caretta*) satellitome, based on an integrated analysis of short reads and long-read assemblies generated here and on data retrieved from publicly available sources.

The *C. caretta* satellitome comprises at least eighteen satDNA families. Fifteen consensus sequences of satDNAs (e.g., CcarSat-18–27, CcarSat-06–50, and others) were identified exclusively in a single dataset. Most of these “unique” satDNA were discovered in the reference genome assembly GCA_965140235.1 using TideCluster tool (Table 1). At least two scenarios could explain this finding. First, the low proportion of these “unique” satDNAs may hinder their detection in low-coverage short-read data when using tools like TAREAN. Second, the detection of “unique” satDNA may reflect interindividual variation, as we utilized genomic data from different *C. caretta* individuals. The highest number of satellites was also identified in the long-read-based reference genome assembly. However, the most abundant repeat, CcarSat-01–2848, was initially found only in short reads (SRR15328383). These results underscore the importance of using multiple independent genomic datasets and combining different bioinformatic approaches to obtain the most complete satDNA library for a species.

GC-content varied significantly relative to monomer length. Satellites with monomer lengths between 27 and 1334 bp (excluding CcarSat-17–823) were GC-rich (>50% GC), while the remaining repeats had either balanced (~50% GC) or AT-rich (<50% GC) composition. (Table 1). Historically, satDNA was considered predominantly AT-rich, especially in centromeres (Fitzgerald et al. 1994). However, this rule is not universal and GC-rich satDNA have been described in a wide range of species (Comings and Mattoccia 1970; Petrović et al. 2009; Giovannotti et al. 2009; Ruiz-Ruano et al. 2016; Peona et al. 2023).

BLAST analysis indicated that CcarSat-17–823 is a component of other satellite units. For instance, a combination of sequences CcarSat-17–823 and CcarSat-19–1598 formed CcarSat-02–2869 (Supplementary Fig. 3). However, bioinformatic analysis suggests that CcarSat-19–1598 should not be considered a separate satellite family. We propose that satellite

CcarSat-17–823 may exist both as part of a longer satellite, CcarSat-02–2869, and as an independent repeat. Arrays composed solely of CcarSat-17–823, however, are likely too small to be visualized by FISH. Satellite CcarSat-17–823 also showed homology with CcarSat-11–2207 (90% coverage, 70% identity), yet FISH localization of these probes revealed no overlap. A similar pattern was observed for satellites CcarSat-05–665 and CcarSat-03–1149 (76% coverage, 84.81% identity). These results suggest that the evolution of tandemly arranged repetitive elements can involve the merging of distinct sequences or their components, followed by amplification.

We verified the chromosomal localization of the analyzed satDNA sequences both cytogenetically, using FISH on metaphase chromosomes of *C. caretta* (Fig. 2, 3, Supplementary Fig. 2, 7), and bioinformatically, on the reference genome assembly (GCA_965140235.1) (Table 2).

FISH mapping of probes, revealed three patterns of satDNA distribution:

1. Dispersed signals across chromosomes (e.g., CcarSat-03–1149, CcarSat-06–50, CcarSat-07–51, CcarSat-13–2115, and CcarSat-16–1319) (Supplementary Fig. 2b, 7a, e);
2. Accumulated signals forming distinct clusters (represented in Fig. 2, 3, Supplementary Fig. 2);
3. Mixed pattern, showing both accumulated and dispersed signals (e.g. CcarSat-01–2848, CcarSat-05–665) (Supplementary Fig. 7b).

SatDNA with an accumulated organization were typically localized in the centromeric and pericentromeric regions of both macro- and microchromosomes, consistent with their classic heterochromatic localization (Charlesworth et al. 1994). For satDNAs forming distinct accumulated FISH signals, their GC/AT content partially correlated with CDAG staining results (Table 1, Supplementary Fig. 1). Specifically, we detected GC-rich bands on chromosomes 4, 11–13, 16, 17, 21, 23, and 27, while AT-rich bands were identified on chromosomes 1, 3, 16, and 21.

Notably, although no prominent heterochromatic blocks were detected on chromosomes 2 and 26 by CDAG staining, FISH revealed accumulated patterns of CcarSat-11–2207 and CcarSat-02–2869 on these chromosomes, respectively. This discrepancy may be due to the small size of these satellite

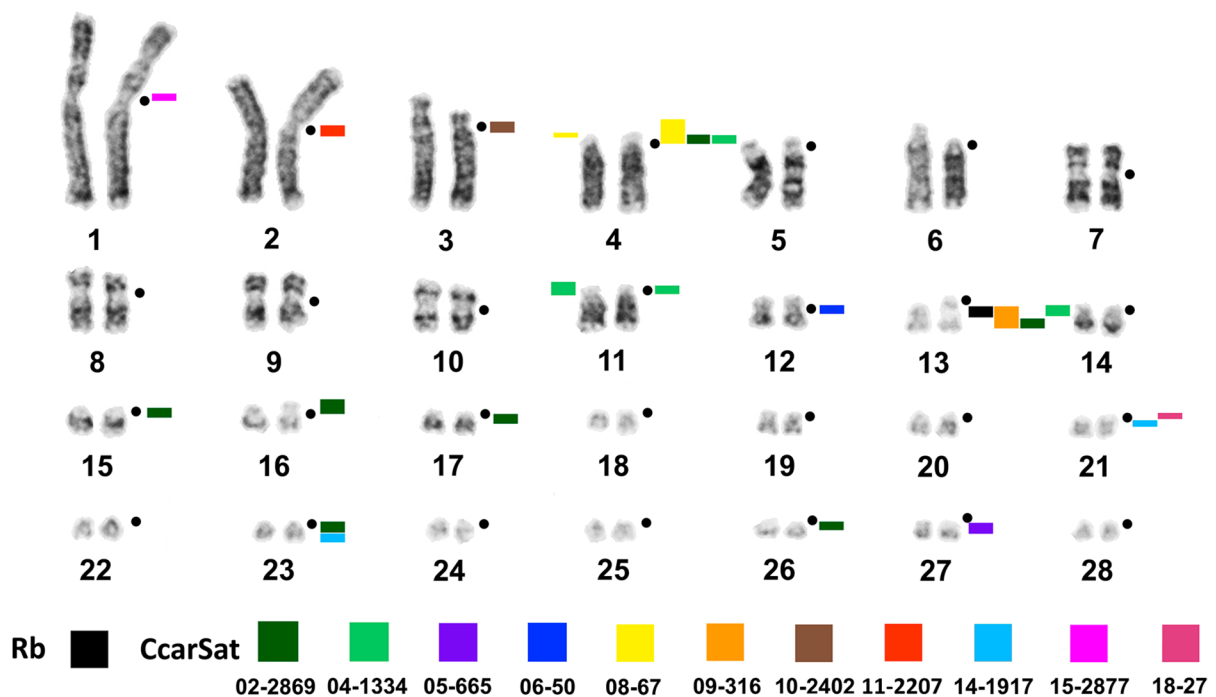


Fig. 3 Major repetitive DNA probe localization on the GTG-banded *Caretta caretta* karyotype. Each cluster of repeated sequences is assigned a specific color code (shown below chromosomes). Block sizes roughly represent signal intensity. When different signal intensities were detected on homologous chromosomes, the corresponding cluster is depicted separately

to the right and left of the chromosome pair. Black dots indicate centromere positions. The satellite CcarSat-01-2848, which marked all chromosomes predominantly in subtelomeric and pericentromeric regions, is not shown on the figure. Repeats demonstrating dispersed distribution on chromosomes are also omitted

clusters, which could fall below the detection threshold of molecular cytogenetic techniques. In this case, the FISH results were consistent with the bioinformatic analysis, which confirmed the presence of the CcarSat-11-2207 array on chromosome 2 in GCA_965140235.1, and CcarSat-02-2869 arrays on several microchromosomes in the reference assembly (Table 2.).

According to SatXplor analysis, most satellites displayed complex array structures and may be organized as part of HORs. FISH analysis confirmed this complexity, revealing the colocalization of distinct satellite arrays on chromosomes 4, 13, 21, and 23. In contrast, most other chromosomes (1–3, 15–17, 26, and 27) harbored single satellite clusters, each composed of sequence from only one satellite family. The distribution of CcarSat-02-2869 arrays on acrocentric microchromosomes 15–17, 23, and 26, tended to be equilocal—that is, occupying analogous position on nonhomologous chromosomes. It has been proposed

that such equilocality may be facilitated by telomeres-mediated interactions and, in the case of acrocentric chromosomes, centromeric reunion in bouquet at first meiotic prophase. These interactions could promote unequal crossing-over and sequence exchange between non-homologous chromosomes (Ruiz-Ruano et al. 2016; Guarracino et al. 2023).

UMAP-based clustering analysis of satDNA monomers revealed different scenarios of monomer evolution. Satellites with high intrachromosomal similarity and limited interchromosomal mixing (Fig. 1b, Example 1) or those with high degree of interchromosome mixing alongside chromosome-specific subgroups (Fig. 1b, Example 4) may have undergone several monomer expansion events. Repeats exhibiting both chromosome-specific and interchromosome distributed monomer subgroups (Fig. 1b, Example 2), likely experienced active and dynamic sequence exchange within and between chromosomes. SatDNAs showing distinct segregation into subgroups

with high or moderate degree of monomer similarity inside the subgroups (Fig. 1b, Example 3) reflect sequence variability maintained by concurrent molecular divergence and homogenization mechanisms (Volarić et al. 2025).

Comparative analysis of satDNA localization via FISH, SatXplor pipeline, and IGV revealed several discrepancies between in situ mapping and in silico genome assembly position (Table 2., Fig. 3). While assemblies often produce scaffolds whose sizes do not correspond to physical chromosome lengths due to a focus on euchromatic ordering, some observed differences cannot be attributed solely to nomenclature issues. The most striking discrepancy involves CcarSat-09–316, which was localized by FISH to chromosome 13, but mapped by SatXplor to chromosome 28. This may be explained by the presence of the nucleolus organizer region (NOR) on chromosome 13, or by HORs, both of which remain challenging to assemble accurately even with long-read sequences (Peona et al. 2023). Notably, probes CcarSat-04–1334 and CcarSat-02–2869 were also localized by FISH on chromosome 13, yet were not detected on the corresponding assembly scaffold (chromosome 28), as would be expected if the cytogenetic and assembly-based chromosomes were congruent.

Further inconsistencies were also observed for CcarSat-08–67, CcarSat-04–1334, and CcarSat-02–2869, which were physically localized on chromosome 4 (Fig. 3). In the GCA_965140235.1 assembly, however, these satellites were either absent from macrochromosomes in SatXplor and Repeat-Masker analysis or assigned to a different chromosome (CcarSat-08–67 was detected on chromosome 5 using IGV). Finally, the previously reported FISH localization of rDNA probe to chromosome 14 (Machado et al. 2020b) may also reflect a discrepancy in chromosome nomenclature.

All satellites with dispersed FISH signals were predicted by SatXplor to have an accumulated organization (Table 2.). The absence of pronounced cluster signals for these repeats on metaphase chromosomes reflect individual variability in satellite accumulation or the limited resolution of FISH. The standard FISH protocol can reliably detect regions as small as 30 kb (Raap 1997), while detecting smaller regions required specialized modification, such as the tyramide signal amplification system, oligopaint probes, and so on

(Raap 1997; Bienko et al. 2013; Maslova and Krasikova 2021).

Satellite CcarSat-14–1917 showed homology with the previously described satellite TSE-1959 of *Trachemys scripta elegans* (95% coverage, 74.6% identity) (Romanenko et al. 2025). The distribution of this satellite is similar on two homologous microchromosomes of these species (our unpublished data), however satellite TSE-1959 was also localized on five chromosome pairs in the *T. scripta elegans* genome. The conservation of CcarSat-14–1917 over approximately 100 MYA (timetree.org) and the identification of most *C. caretta* satDNA sequences in other available sea turtle genomes suggest that the sea turtle clade is a promising model for testing the library hypothesis of satDNA evolution.

In conclusion, this study expands our knowledge of the diversity and distribution of major satDNA clusters in the *C. caretta* genome. A comprehensive satDNA library is best constructed through the combined long and short reads analysis from the same species, supplemented by the use of multiple bioinformatic tools for repeat detection. Since satDNA still remains part of the “dark matter of genome”, studying it solely through bioinformatic approaches is difficult. The integration of cytogenetic and bioinformatic methods contributes to linking the physical chromosomes of *C. caretta* to its genome assemblies. The high-quality assembly of the *C. caretta* genome allowed us to bioinformatically map most satellites to chromosomes, with only a minor fraction assigned to unanchored scaffolds. However, the discrepancies observed between assembly data and FISH results highlight the persistent difficulty in assembling highly repetitive genomic regions. A comprehensive analysis of the sea turtle satellitome is now needed to provide new deeper insights into the evolutionary dynamic and history of satDNA in this clade.

Acknowledgements The research was completed using equipment and materials of the large-scale research facilities “Cryobank of cell cultures” Institute of Molecular and Cellular Biology SB RAS (Novosibirsk, Russia). NGS was performed in the Genomics Core Facility (ICBFM SB RAS, Novosibirsk). Computational resources were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure, and the computational resources provided by the Shared Services Center “Data Center of FEB RAS” (Khabarovsk) (Sorokin et al. 2017). We thank two anonymous reviewers for their insightful comments and suggestions.

Author contributions Material collection: AI, CC, LT, PC, PS. Cell culture establishment and cultivation: SAR, AI. Conceptualization: KVT, SAR. Experimental procedure: KVT, SAR, DYuP, SAM, AAP, NAS, GID. Supervision: VAT. Review and the draft of the manuscript: KVT, SAR, AAP. All authors contributed to the study conception and design, commented on previous versions, and approved the final manuscript.

Funding This research was funded by Russian Science Foundation grant No 25–24–00171 (KVT).

Data availability The sequenced genomic reads were deposited in the NCBI SRA database under the accession number SRR26671777. The obtained repetitive element consensus sequences were deposited in Zenodo (Tishakova et al. 2026).

Declarations

Samples were obtained under permits Prot. 0004851 PNM (07/03/2016) issued by the Italian ministry MATTM (Ministero dell'ambiente e della tutela del territorio e del mare).

Competing of interest The authors declare no competing interests.

References

- Afgan E, Baker D, van den Beek M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 44:W3–W10. <https://doi.org/10.1093/nar/gkw343>
- Arantes LS, Brown T, De Panis D et al (2025) Haplotype-resolved reference genomes of the sea turtle clade unveil ultra-syntenic genomes with hotspots of divergence. *Gigascience* 14:giaf105. <https://doi.org/10.1093/gigascience/giaf105>
- Badenhorst D, Hillier LDW, Literman R et al (2015) Physical mapping and refinement of the painted turtle genome (*Chrysemys picta*) inform amniote genome evolution and challenge turtle-bird chromosomal conservation. *Genome Biol Evol* 7:2038–2050. <https://doi.org/10.1093/gbe/evv119>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>
- Bienko M, Crosetto N, Teytelman L et al (2013) A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat Methods* 10:122–124. <https://doi.org/10.1038/nmeth.2306>
- Biltueva LS, Prokopov DY, Makunin AI et al (2017) Genomic organization and physical mapping of tandemly arranged repetitive DNAs in Sterlet (*Acipenser ruthenus*). *Cytogenet Genome Res* 152:148–157. <https://doi.org/10.1159/000479472>
- Biscotti MA, Olmo E, Heslop-Harrison JS (2015) Repetitive DNA in eukaryotic genomes. *Chromosom Res* 23:415–420. <https://doi.org/10.1007/s10577-015-9499-z>
- Bolsheva NL, Melnikova NV, Kirov IV et al (2019) Characterization of repeated DNA sequences in genomes of blue-flowered flax. *BMC Evol Biol* 19:49. <https://doi.org/10.1186/s12862-019-1375-6>
- Camacho JPM, Cabrero J, López-León MD et al (2022) Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. *BMC Biol* 20:36. <https://doi.org/10.1186/s12915-021-01216-9>
- Cavalcante MG, Nagamachi CY, Pieczarka JC, Noronha RCR (2020) Evolutionary insights in Amazonian turtles (Testudines, Podocnemididae): co-location of 5S rDNA and U2 snRNA and wide distribution of Tc1/Mariner. *Biol Open* 9:bio049817. <https://doi.org/10.1242/BIO.049817>
- Chang G, Jones S, Leelakumari S et al (2023) The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758. *F1000Res* 12:336. <https://doi.org/10.12688/f1000research.131283.2>
- Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215–220. <https://doi.org/10.1038/371215A0>
- Clemente L, Mazzoleni S, Pensabene E et al (2021) Cytogenetic analysis of the asian box turtles of the genus *Cuora* (Testudines, Geoemydidae). *Genes* 12:156. <https://doi.org/10.3390/genes12020156>
- Comings DE, Mattoccia E (1970) Studies of microchromosomes and a G-C rich DNA satellite in the quail. *Chromosoma* 30:202–214. <https://doi.org/10.1007/BF00282001>
- Crawford NG, Parham JF, Sellas AB et al (2015) A phylogenomic analysis of turtles. *Mol Phylogenet Evol* 83:250–257. <https://doi.org/10.1016/j.ympev.2014.10.021>
- Davidian AG, Dyomin AG, Galkina SA et al (2022) 45S rDNA repeats of turtles and crocodiles harbor a functional 5S rRNA gene specifically expressed in oocytes. *Mol Biol Evol* 39:msab324. <https://doi.org/10.1093/molbev/msab324>
- De Lima LG, Ruiz-Ruano FJ (2022) In-depth satellitome analyses of 37 *Drosophila* species illuminate repetitive DNA evolution in the *Drosophila* genus. *Genome Biol Evol* 14:evac064. <https://doi.org/10.1093/gbe/evac064>
- Dos Santos RZ, Calegari RM, Silva DMZDA et al (2021) A long-term conserved satellite DNA that remains unexpanded in several genomes of Characiformes fish is actively transcribed. *Genome Biol Evol* 13:evab002. <https://doi.org/10.1093/gbe/evab002>
- Dover G (2002) Molecular drive. *Trends Genet* 18:587–589. [https://doi.org/10.1016/S0168-9525\(02\)02789-0](https://doi.org/10.1016/S0168-9525(02)02789-0)
- Fitzgerald DJ, Dryden GL, Bronson EC et al (1994) Conserved patterns of bending in satellite and nucleosome positioning DNA. *J Biol Chem* 269:21303–21314. [https://doi.org/10.1016/s0021-9258\(17\)31963-4](https://doi.org/10.1016/s0021-9258(17)31963-4)
- Fry K, Salser W (1977) Nucleotide sequences of HS- α satellite DNA from kangaroo rat *dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* 12:1069–1084. [https://doi.org/10.1016/0092-8674\(77\)90170-2](https://doi.org/10.1016/0092-8674(77)90170-2)
- Gao Y, Liu B, Wang Y, Xing Y (2019) TideHunter: efficient and sensitive tandem repeat detection from noisy long-reads using seed-and-chain. *Bioinformatics* 35:i200–i207. <https://doi.org/10.1093/bioinformatics/btz376>
- Giovannotti M, Nisi Cerioni P, Caputo V, Olmo E (2009) Characterisation of a GC-rich telomeric satellite DNA

- in *eumeces schneideri* Daudin (Reptilia, Scincidae). Cytogenet Genome Res 125:272–278. <https://doi.org/10.1159/000235933>
- Graphodatsky AS, Radjabli SI (1988) Chromosomes of farm and laboratory mammals. ~ Nauka (ed. Baranov OK), Novosibirsk, pp 14–15
- Grewal SIS, Elgin SCR (2007) Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406. <https://doi.org/10.1038/NATURE05914>
- Guarracino A, Buonaiuto S, de Lima LG et al (2023) Recombination between heterologous human acrocentric chromosomes. Nature 617:335–343. <https://doi.org/10.1038/s41586-023-05976-y>
- Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102. <https://doi.org/10.1126/science.1062939>
- Hubley R, Finn RD, Clements J et al (2016) The Dfam database of repetitive DNA families. Nucleic Acids Res 44:D81–D89. <https://doi.org/10.1093/nar/gkv1272>
- Ijdo JW, Wells RA, Baldini A, Reeders ST (1991) Improved telomere detection using a telomere repeat probe (TTA GGG)_n generated by PCR. Nucleic Acids Res 19:4780. <https://doi.org/10.1093/nar/19.17.4780>
- Ivanova NG, Kartavtseva IV, Stefanova VN et al (2022) Tandem repeat diversity in two closely related hamster species—the Chinese hamster (*Cricetulus griseus*) and striped hamster (*Cricetulus barabensis*). Biomedicines 10:925. <https://doi.org/10.3390/biomedicines10040925>
- Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet 10:833–844. <https://doi.org/10.1038/NRG2683>
- Johnson M, Zaretskaya I, Raytselis Y et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–9. <https://doi.org/10.1093/nar/gkn201>
- Kawagoshi T, Nishida C, Matsuda Y (2012) The origin and differentiation process of X and Y chromosomes of the black marsh turtle (*Siebenrockiella crassicolis*, Geoemydidae, Testudines). Chromosom Res 20:95–110. <https://doi.org/10.1007/s10577-011-9267-7>
- Klein SJ, O'Neill RJ (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosom Res 26:5–23. <https://doi.org/10.1007/s10577-017-9569-5>
- Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7:474. <https://doi.org/10.1186/1471-2105-7-474>
- Lemskaya NA, Kulemzina AI, Beklemisheva VR et al (2018) A combined banding method that allows the reliable identification of chromosomes as well as differentiation of AT- and GC-rich heterochromatin. Chromosom Res 26:307–315. <https://doi.org/10.1007/s10577-018-9589-9>
- Liao X, Zhu W, Zhou J et al (2023) Repetitive DNA sequence detection and its role in the human genome. Commun Biol 6:954. <https://doi.org/10.1038/S42003-023-05322-Y>
- López EA, Hernández-Fernández J, Bernal-Villegas J (2008) Lymphocyte culture and partial karyotype of the marine turtle *Caretta caretta* (Testudines: Cheloniidae) in Santa Marta, Colombian Caribbean. Rev Biol Trop 56:1459–1469
- Machado CRD, Domit C, Pucci MB et al (2020) Heterochromatin and microsatellites detection in karyotypes of four sea turtle species: interspecific chromosomal differences. Genet Mol Biol 43:1–9. <https://doi.org/10.1590/1678-4685-GMB-2020-0213>
- Machado CRD, Glugoski L, Domit C et al (2020) Comparative cytogenetics of four sea turtle species (Cheloniidae): G-banding pattern and in situ localization of repetitive DNA units. Cytogenet Genome Res 160:531–538. <https://doi.org/10.1159/000511118>
- Maden BEH, Dent CL, Farrell TE et al (1987) Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones. Biochem J 246:519–527. <https://doi.org/10.1042/bj2460519>
- Martínez PA, Boeris JM, Sánchez J et al (2009) Karyotypic characterization of *Trachemys dorsignii* (Testudines: Emydidae) and *Chelonoidis (Geochelone) donosobarrosi* (Testudines: Testudinidae), two species of Cryptodiran turtles from Argentina. Genetica 137:277–283. <https://doi.org/10.1007/s10709-009-9377-8>
- Maslova A, Krasikova A (2021) Fish going meso-scale: a microscopic search for chromatin domains. Front Cell Dev Biol 9:753097. <https://doi.org/10.3389/fcell.2021.753097>
- Mazurczyk M, Rybaczek D (2015) Replication and re-replication: different implications of the same mechanism. Biochimie 108:25–32. <https://doi.org/10.1016/j.biochi.2014.10.026>
- Meštrović N, Pohl M, Mravinac B, Ugarković D (1998) Evolution of satellite DNAs from the genus *Palorus* - experimental evidence for the “library” hypothesis. Mol Biol Evol 15:1062–1068. <https://doi.org/10.1093/oxfordjournals.molbev.a026005>
- Modi WS, Serdyukova NA, Vorobieva NV, Graphodatsky AS (2003) Chromosomal localization of six repeated DNA sequences among species of *Microtus* (Rodentia). Chromosom Res 11:705–713. <https://doi.org/10.1023/A:1025922813756>
- Montiel EE, Badenhorst D, Lee LS et al (2016) Cytogenetic insights into the evolution of chromosomes and sex determination reveal striking homology of turtle sex chromosomes to amphibian autosomes. Cytogenet Genome Res 148:292–304. <https://doi.org/10.1159/000447478>
- Novák P, Neumann P, Pech J et al (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. <https://doi.org/10.1093/bioinformatics/btt054>
- Novák P, Robledillo LÁ, Kobližková A et al (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res 45:e111. <https://doi.org/10.1093/nar/gkx257>
- Novák P, Neumann P, Macas J (2020) Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat Protoc 15:3745–3776. <https://doi.org/10.1038/s41596-020-0400-y>
- Ohno S (1972) So much “junk” DNA in our genome. Brookhaven Symp Biol 23:366–370
- Palacios-Gimenez OM, Milani D, Song H et al (2020) Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library

- hypothesis. *Genome Biol Evol* 12:88–102. <https://doi.org/10.1093/gbe/evaa018>
- Peona V, Kutschera VE, Blom MPK et al (2023) Satellite DNA evolution in Corvoidea inferred from short and long reads. *Mol Ecol* 32:1288–1305. <https://doi.org/10.1111/mec.16484>
- Petrović V, Pérez-García C, Pasantes JJ et al (2009) A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* 124:63–71. <https://doi.org/10.1159/000200089>
- Plohl M, Petrović V, Luchetti A et al (2010) Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Heredity* 104:543–551. <https://doi.org/10.1038/hdy.2009.141>
- Plohl M, Meštrović N, Mravinac B (2012) Satellite DNA evolution. *Genome Dyn* 7:126–152. <https://doi.org/10.1159/000337122>
- Raap AK (1997) Overview of fluorescence in situ hybridization techniques for molecular cytogenetics. *Curr Protoc Cytom* 1:8.1.1–8.1.6. <https://doi.org/10.1002/0471142956.cy0801s00>
- Richard G-F, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72:686–727. <https://doi.org/10.1128/mmr.00011-08>
- Robinson JT, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>
- Romanenko SA, Prokopov DY, Proskuryakova AA et al (2022) The cytogenetic map of the Nile crocodile (*Crocodylus niloticus*, Crocodylidae, Reptilia) with fluorescence *in situ* localization of major repetitive DNAs. *Int J Mol Sci* 23:13063. <https://doi.org/10.3390/ijms232113063>
- Romanenko SA, Prokopov DY, Marchenko SA et al (2025) In situ and in silico localization of major satellite DNAs in the genome of the red-eared slider (*Trachemys scripta elegans*, Emydidae, Testudines). *Cytogenet Genome Res* 165:162–174. <https://doi.org/10.1159/000544908>
- Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci Rep* 6:28333. <https://doi.org/10.1038/srep28333>
- Seabright M (1971) A rapid banding technique for human chromosomes. *Lancet* 298:971–972. [https://doi.org/10.1016/S0140-6736\(71\)90287-X](https://doi.org/10.1016/S0140-6736(71)90287-X)
- Shaffer HB, McCartney-Melstad E, Near TJ et al (2017) Phylogenomic analyses of 539 highly informative loci dates a fully resolved time tree for the major clades of living turtles (Testudines). *Mol Phylogenet Evol* 115:7–15. <https://doi.org/10.1016/j.ympev.2017.07.006>
- Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0 2013–2015. <https://www.repeatmasker.org/>. Accessed Jan 2026
- Sorokin AA, Makogonov SV, Korolev SP (2017) The information infrastructure for collective scientific work in the Far East of Russia. *Sci Tech Inf Process* 44:302–304. <https://doi.org/10.3103/S0147688217040153>
- Sujiwattanarat P, Thapana W, Srikulnath K et al (2015) Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci Rep* 5:10315. <https://doi.org/10.1038/srep10315>
- Sundararajan K, Straight AF (2022) Centromere identity and the regulation of chromosome segregation. *Front Cell Dev Biol* 10:914249. <https://doi.org/10.3389/fcell.2022.914249>
- Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
- Ting D, Lipson D, Paul S et al (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331:593–596. <https://doi.org/10.1126/SCIENCE.1200801>
- Tishakova K, Prokopov D, Marchenko S, Romanenko S (2026) *In silico* identified satellite DNAs in genomic data of *Caretta caretta*. Dataset on Zenodo. <https://zenodo.org/records/17707690>. Accessed 22 Jan 2026
- Trofimova I, Krasikova A (2016) Transcription of highly repetitive tandemly organized DNA in amphibians and birds: a historical overview and modern concepts. *RNA Biol* 13:1246–1257. <https://doi.org/10.1080/15476286.2016.1240142>
- Ugarković D, Plohl M (2002) Variation in satellite DNA profiles – causes and effects. *EMBO J* 21:5955–5959. <https://doi.org/10.1093/emboj/cdf612>
- Volarić M, Meštrović N, Despot-Slade E (2025) SatXplor—a comprehensive pipeline for satellite DNA analyses in complex genome assemblies. *Brief Bioinform* 26:bbae660. <https://doi.org/10.1093/bib/bbae660>
- Wallace BP, DiMatteo AD, Bolten AB et al (2011) Global conservation priorities for marine turtles. *PLoS ONE* 6:e24510. <https://doi.org/10.1371/journal.pone.0024510>
- Wallace BP, DiMatteo AD, Hurley BJ et al (2010) Regional Management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. *PLoS ONE* 5:1–11. <https://doi.org/10.1371/journal.pone.0015465>
- Yamada K, Nishida-Umehara C, Matsuda Y (2005) Molecular and cytogenetic characterization of site-specific repetitive DNA sequences in the Chinese soft-shelled turtle (*Pelodiscus sinensis*, Trionychidae). *Chromosome Res* 13:33–46. <https://doi.org/10.1007/s10577-005-2351-0>
- Yang F, Graphodatsky AS (2017) Animal Probes and ZOO-FISH. In: Springer Protocols, pp 395–415
- Yen EC, Gilbert JD, Balard A et al (2025) Chromosome-level genome assembly and methylome profile yield insights for the conservation of endangered loggerhead sea turtles. *Gigascience* 14:giaf054. <https://doi.org/10.1093/gigascience/giaf054>
- Yntema CL, Mrosovsky N (1982) Critical periods and pivotal temperatures for sexual differentiation in loggerhead sea turtles (*Caretta caretta*). *Can J Zool* 60:1012–1016. <https://doi.org/10.1139/z82-141>
- Zattera ML, Bruschi DP (2022) Transposable elements as a source of novel repetitive DNA in the eukaryote genome. *Cells* 11:3373. <https://doi.org/10.3390/cells11213373>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.