Genomics 116 (2024) 110771

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

A comprehensive phylogenomic study unveils evolutionary patterns and challenges in the mitochondrial genomes of Carcharhiniformes: A focus on Triakidae

Jessica C. Winn [a], Simo N. Maduna [b], Aletta E. Bester-van der Merwe [a,*]

[a] Molecular Breeding and Biodiversity Group, Department of Genetics, Stellenbosch University, Stellenbosch, Western Cape 7602, South Africa
[b] Department of Ecosystems in the Barents Region, Svanhovd Research Station, Norwegian Institute of Bioeconomy Research, 9925 Svanvik, Norway

ARTICLE INFO

Keywords:
Comparative mitogenomics
Convergent evolution
Reproductive mode
Taxonomy
Gene duplication
Multi-species coalescence
Houndsharks
Mustelus

ABSTRACT

The complex evolutionary patterns in the mitochondrial genome (mitogenome) of the most species-rich shark order, the Carcharhiniformes (ground sharks) has led to challenges in the phylogenomic reconstruction of the families and genera belonging to the order, particularly the family Triakidae (houndsharks). The current state of Triakidae phylogeny remains controversial, with arguments for both monophyly and paraphyly within the family. We hypothesize that this variability is triggered by the selection of different *a priori* partitioning schemes to account for site and gene heterogeneity within the mitogenome. Here we used an extensive statistical framework to select the *a priori* partitioning scheme for inference of the mitochondrial phylogenomic relationships within Carcharhiniformes, tested site heterogeneous CAT + GTR + G4 models and incorporated the multi-species coalescent model (MSCM) into our analyses to account for the influence of gene tree discordance on species tree inference. We included five newly assembled houndshark mitogenomes to increase resolution of Triakidae. During the assembly procedure, we uncovered a 714 bp-duplication in the mitogenome of *Galeorhinus galeus*. Phylogenetic reconstruction confirmed monophyly within Triakidae and the existence of two distinct clades of the expanded *Mustelus* genus. The latter alludes to potential evolutionary reversal of reproductive mode from placental to aplacental, suggesting that reproductive mode has played a role in the trajectory of adaptive divergence. These new sequences have the potential to contribute to population genomic investigations, species phylogeography delineation, environmental DNA metabarcoding databases and, ultimately, improved conservation strategies for these ecologically and economically important species.

1. Introduction

Mitochondrial phylogenomics (mitophylogenomics), the use of entire or partial mitochondrial genomes (mitogenomes) within a phylogenetic framework, has proven to be more powerful in taxonomic, phylogenetic, population, and evolutionary studies than individual gene sequence comparison [1,2,3,4]. The typical fish mitogenome is a double-stranded, closed-circular molecule ranging in size from 15 kb to 20 kb that encodes 37 genes [13 protein-coding genes (PCGs), two ribosomal RNA genes (16S and 12S), and 22 transfer RNA genes] and two non-coding regions known as the light-strand origin of replication (O_L) and the control region (CR), respectively [5,6]. However, nonstandard duplications of various regions within the fish mitogenome have previously been reported [5,7,8]. The use of the mitogenome as a marker of choice for phylogenetic and evolutionary analysis stems not only from its fast evolutionary rates, but also from its small genome size, maternal inheritance, low sequence recombination, haploidy, and high copy number [9,10,11]. Because mitochondrial regions are linked, it stands to reasoning that they evolve in tandem with, and contain evidence of, species' population histories, and thus share a phylogeny. However, it is evident from previous literature that using different genomic regions (including the whole molecule and single genes) often produces incongruent results [12,13]. This has led to doubts about the classification, phylogeny, and divergence times of many taxonomic groups, including elasmobranchs [14,15,8]. Additionally, despite the great potential of mitogenomic data, there are currently no standards for their rigorous and objective use in phylogenetic analyses.

Mitogenomes exhibit heterogeneity in base composition and

* Corresponding author.
E-mail address: aeb@sun.ac.za (A.E. Bester-van der Merwe).

https://doi.org/10.1016/j.ygeno.2023.110771
Received 15 June 2023; Received in revised form 14 December 2023; Accepted 22 December 2023
Available online 24 December 2023

evolutionary rates at different scales across the molecule [16,17], which indicates that data partitioning would be advantageous for phylogenetic investigations of mitogenomes compared to data exclusion [18,19,20,21]. An increasing number of studies support the notion that data partitioning and model selection can significantly affect phylogenetic inference [22,23,24,19,25]. Beyond the effects it has on the estimation of genetic distances and branch lengths, are the effects on estimates of bootstrap support and posterior probabilities [26]. Model-fit is improved by partitioning alignments into relatively homogeneous sets of sites before selecting and optimizing a substitution model for each set independently [27,28,29]. However, this process may be unsatisfactory due to the subjective nature of the divisions that are typically predefined based on the properties of the sequences in the alignment, such as particular gene fragments, codon locations, rRNA structural components (*e.g.*, stems and/or loops), or a combination of these features [27,24,30]. As a result, there is a progressive shift from data partitioning based solely on biological or mechanistic processes towards statistically testing a pre-defined subset against every other pre-defined subset. The latter is done under a maximum likelihood framework and then proceeds to selecting the grouping that most improves an information theoretic criterion score (*e.g.*, Akaike Information Criterion, AIC, or Bayesian Information Criterion, BIC), after which the process is repeated until no more are plausible [19,28,29].

Although data partitioning and model selection is conducted under a statistical framework, it still requires an *a priori* partitioning scheme to be fed into the software [31,32,28,29], which can significantly influence the phylogenetic reconstruction results [14,33,21]. To that end, it is essential that several datasets with varying *a priori* partitioning strategies be applied to a single dataset and compared in order to establish the appropriate partitioning technique for mitophylogenomics investigations to avoid dividing the data into too many blocks *i.e.*, 'over-partitioning'. Overpartitioning leads to overparametrization of model parameters and yields well supported, but erroneous nodes in the tree, and is a real risk when working with mitogenome datasets [34,33,35].

A common alternative approach to applying different best-fitting 'site-homogeneous' substitution models (*e.g.*, GTR + G4) to different partitions of a supermatrix is to apply the Bayesian site-heterogeneous mixture model, known as the CAT model [36]. Site-heterogeneity refers to heterogeneous equilibrium frequencies across site compared to the homogeneous equilibrium frequency estimation in classical substitution models *i.e.*, site-homogeneous models. The CAT model utilizes a Dirichlet process prior to accommodate multiple categories of substitution profiles with diverse nucleotide or amino acid frequencies, along with a single set of exchange rates for the entire supermatrix. These exchange rates can either be fixed to flat values (referred to as CAT-F81 or CAT-Poisson; commonly abbreviated as "CAT") or independently estimated from data using general time reversible Markov processes (known as CAT-GTR), which provide more expressive capabilities [36,37]. While the inclusion of gamma-rate heterogeneity is typically not explicitly mentioned, all CAT models incorporate a component that accounts for variation in rates across sites, which is set to 4 by default (*i.e.*, G4). Undoubtedly, the CAT models have been extensively employed for phylogenomic reconstruction and have successfully addressed certain longstanding phylogenetic inquiries (*e.g.*, [38,39]). Nevertheless, their utility is impeded by the substantial computational load and convergence challenges when dealing with large datasets [40,38,41,37]. Furthermore, both partitioning with site-homogeneous models and CAT models presuppose that the substitution process is uniform across lineages and that all taxa and genes possess comparable compositions. Therefore, it is necessary to conduct a systematic comparison of the performance of partitioning with site-homogeneous models and CAT models in phylogenetic estimation to avoid misusing the models.

There has been a significant increase in studies examining the evolutionary history of chondrichthyans (Class Chondrichthyes, cartilaginous fish; sharks, rays, and chimeras) using mitophylogenomics, and these studies are providing unique insights from a biogeographic and phylogenetic perspective [42,12,13,15,8]. However, these investigations also reveal several inconsistencies at several levels, from superorders and families down to the genera within families, as well as their interrelationships. There is phylogenetic discordance in particular regarding the status of the topmost species-rich order of sharks comprised of eight families, the Carcharhiniformes (ground sharks; [43]). Mitophylogenomic inferences of Carcharhiniformes have confirmed family-level paraphyly (Carcharhinidae, Scyliorhinidae, and Triakidae) and genera-level polytomy (*Carcharhinus*) [12,13]. *Galeocerdo cuvier* consistently clusters as a sister to a clade including Sphyrnidae and Carcharhinidae and has subsequently been reassigned to its own family, Galeocerdonidae to make Carcharhinidae monophyletic (see [44]). Markedly, Kousteni et al. [15] and Wang et al. [8] recovered conflicting topologies for Carcharhiniformes with the same mitogenomic dataset (Fig. 1). Kousteni et al. [15] recovered the family Triakidae (houndsharks; [45]) as a monophyletic group while Wang et al. [8] recovered Triakidae as a paraphyletic group, both with high branch support values. Wang et al. [8] found that the triakid *Hemitriakis japonica* clustered to a main clade including Hemigaleidae + Sphyrnidae + Carcharhinidae instead of the sister-group relationship with the genus *Mustelus* in Triakidae as recovered by Kousteni et al. [15]. Upon review, it was evident that Kousteni et al. [15] and Wang et al. [8] applied different partitioning strategies involving data partitioning by gene boundaries and codon positions and/or data exclusion (Table 1). This necessitates a comprehensive analysis of the dataset to discover the causes of these inconsistencies and the effects of partitioning techniques on tree topology. (See Fig. 2.)

Our objectives in this study were fivefold, including methodological and research goals. First, we set out to develop a time- and cost-effective whole mitogenome sequencing strategy on the Ion GeneStudioTM S5 Prime System. Second, we assembled and characterized complete mitogenomes for five species in the Triakidae family using a combined assembly approach. Third, using previously published Carcharhiniform mitogenomes, we compared genome structure, base composition, substitution, and evolutionary rates among representative families. Fourth, we tested the effect of different partitioning strategies on the inference of phylogenetic relationships within Carcharhiniformes and determined the best data partitioning strategy for analyzing mitogenome data of Carcharhiniformes in a maximum likelihood framework. Within the scope of this objective, we aimed to test the following hypotheses (*sensu* [46]):

Hypothesis H1. Different partitioning strategies result in different topologies due to a lack of or varying levels of phylogenetic signal in the respective partition datasets.

Hypothesis H2. The lack of phylogenetic concordance is caused by gene tree/species tree issues, which means that the evolutionary history of genes differs from that of the species.

Hypothesis H3. Discordance in phylogenetic reconstruction arises from violations of the assumptions underlying the applied models.

Here, we tested Hypothesis H1 by closely examining branch support in phylogenetic inferences based on different partitioning schemes of the same nucleotide dataset and compared the topology to the amino acid dataset and other published studies for a set of *a priori* defined clades in the focal order Carcharhiniformes. If it was found that the same clades are well supported and concordant in topologies derived from other datasets (Table 1), we could reject this hypothesis and conclude that the discordant results are not due to a lack of phylogenetic signal. Whereas to test Hypothesis H2, we compared the topology of individual phylogenies obtained with different datasets. If it was discovered that the differences in topology were primarily due to analyses of different datasets, this likely implied that we were dealing with gene tree *versus* species tree issues and that analyses using the multispecies coalescent model (MSCM) were required. Hypothesis H3 was tested by surveying the datasets for evolutionary signatures which could be violating the

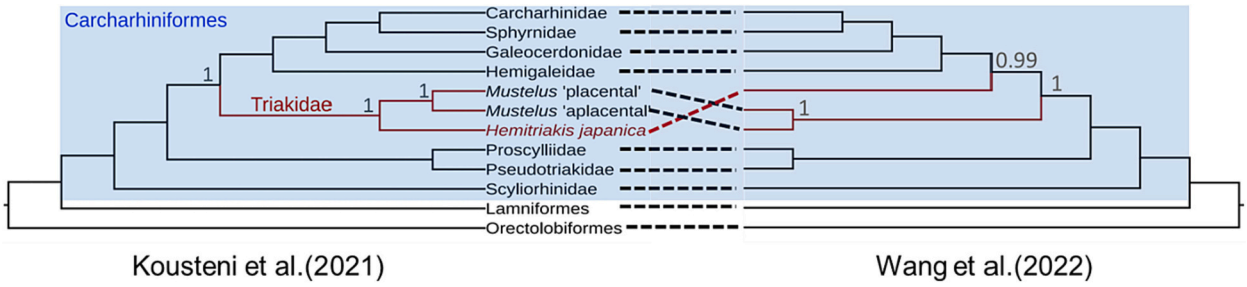

Fig. 1. Two alternative phylogenetic hypotheses for the grouping of members of the family Triakidae based on mitogenome datasets. Branch support values are displayed next to the nodes relating to Triakidae.

Table 1
Mitophylogenomic investigations with representatives from the focal order Carcharhiniformes.

Study	Partition strategy*	Model selection	Inference method	Main findings
Díaz-Jaimes et al. [13]	traditional - gene	jModelTest version? AIC or BIC?	RAxML-HPC v.8, MrBayes v.3.0b4	The phylogeny supported the association ((Lamniformes, Carcharhiniformes) Orectolobiformes) with Heterodontiformes in a basal position. There was paraphyly in the family Carcharhinidae. *Scoliodon macrorhynchos* is placed as a sister group to Triakidae, Sphrynidae and the other Carcharhinidae.
da Cunha et al. [12]	none	jModelTest v.2, AIC	PhyML v.3.0 MrBayes v.3.0b4	The phylogeny supported the association ((Lamniformes, Carcharhiniformes) Orectolobiformes) with Heterodontiformes in a basal position. There was paraphyly in the order Carcharhinidae.
Amaral et al. [42]	traditional - gene-codon: (i) all codon positions, (ii) first and second codon positions, and (iii) the third codon position	jModelTest v.0.1.1, AIC	PhyML v.3 and Mega v.6, MrBayes2	Highly convergent topologies independent of the use of different codon positions, except for the position of the species *Scoliodon macrorhynchos* within the Carcharhiniformes. Heterodontiformes was found as a sequential sister taxon between the Orectolobiformes and the Carcharhiniformes in the analysis using codon positions 1 and 2. In all other topologies it is basal to (Lamniformes, Carcharhiniformes) Orectolobiformes). There is polytomy in the genus *Carcharhinus*.
Kousteni et al. [15]	traditional - gene	JModelTest v.2.1.7, BIC	PhyML v.3, MrBayes v.3.2.6	Galeomorphii (all modern sharks except the dogfish and its relatives, *i.e.*, superorder containing the Heterodontiformes, Orectolobiformes, Lamniformes, and Carcharhiniformes) was represented by four monophyletic orders in the BI tree - the basal Heterodontiformes, the Orectolobiformes, and the Carcharhiniformes–Lamniformes group. However, the ML tree topology placed Orectolobiformes at the most basal position of the group. According to BI analysis Carcharhiniformes was represented by seven families, with Scyliorhinidae at the basal position of the order. Carcharhinidae was paraphyletic. The *Carcharhinus* genus showed polytomy in ML analysis.
Wang et al. [8]	*a priori* - first codon + second codon (excl. ND6) + two rRNA genes	ModelFinder, BIC	IQ-Tree v.1.6.2, MrBayes v.3.2.6	Scyliorhinidae, Triakidae and Carcharhinidae were recovered as non-monophyletic families. *Hemitriakis japonica* clustered to a main clade including Hemigaleidae; Sphyrnidae and Carcharhinidae instead of the genus *Mustelus* in Triakidae. Three species in Scyliorhinidae clustered with other families. Carcharhinidae was placed as a sister clade to Sphyrnidae, apart from *Galeocerdo cuvier* which clustered as a main clade, including Sphyrnidae and other Carcharhinidae species. The results show many inconsistencies in the phylogenetic position of some species with the traditional views.

* We consider a data partitioning scheme solely based on biological prior knowledge as a traditional approach compared to statistically determining an optimal partitioning scheme from *a priori* partitioning scheme.

assumptions of the applied evolutionary models utilized for phylogenetic reconstruction. If certain taxa in the tree contain a disproportionate amount of the nucleotide/amino acid variation, the assumption of stationarity is violated leading to the inference of false, yet highly supported, topologies [47,48]. Similarly, if there is heterogeneity in nucleotide or amino acid frequencies across sites, the assumptions of homogeneity and reversibility are violated and phylogenetic reconstructions are biased [49]. If patterns of compositional heterogeneity were detected, site heterogenous models needed to be employed for phylogenetic analyses.

After confirming the informativeness of the dataset for phylogenetic reconstruction, it was important to identify patterns in the empirical datasets that might violate the assumptions of phylogenetic reconstruction techniques. We could not reliably state that gene tree issues were the cause of incongruencies if our methodological approach was flawed because the MSCM assumes that incomplete lineage sorting is the only source of discordance between gene trees and the species tree. Comparably, compositional heterogeneity in the dataset must be detected before using site heterogenous models to make phylogenetic inferences. Finally, we performed a phylogenomic analysis to determine the higher-level phylogenetic relationships within Carcharhiniformes, with a focus on Triakidae, and compared the recovered inferences to previous hypotheses. We analyse the phylogenomic outcomes for features of convergent evolution of adaptations related to reproductive mode in Triakidae. We hypothesize that our phylogenetic analysis will recover the study triakids embedded within the Triakidae clade, within which respective mustelids will assign to two alternate reproductive mode subclades.

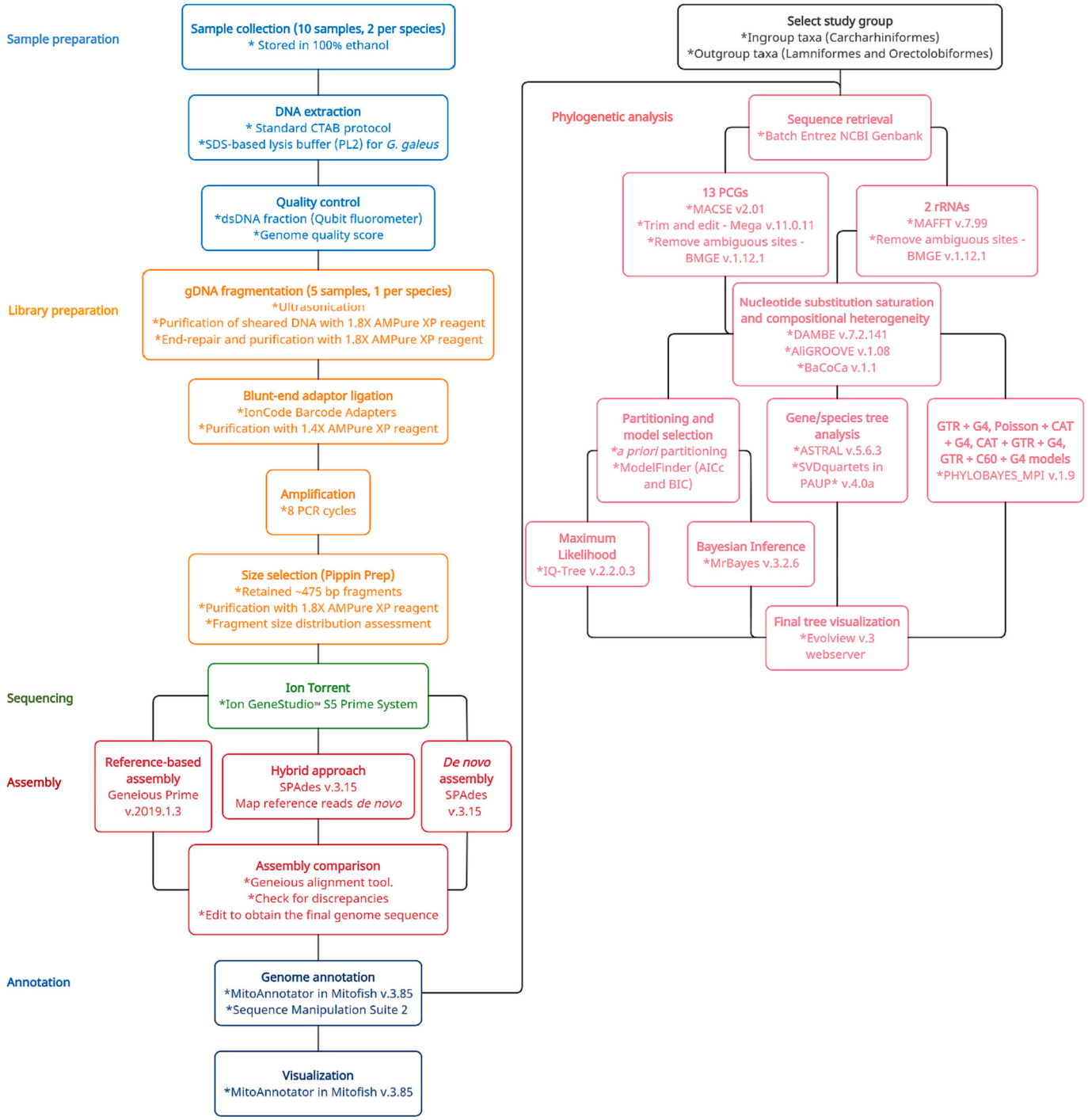

Fig. 2. Schematic summary and representation of the assembly of five houndshark mitogenomes and Carcharhiniform phylogenetic analysis.

2. Methods

2.1. Sample collection, DNA extraction and quality control

Fin clip tissue samples of *Galeorhinus galeus*, *Mustelus palumbes*, and *Triakis megalopterus* were collected along the coast of South Africa and stored in 100% ethanol. *Mustelus asterias* was sampled in Wales and *Mustelus mosis* was sampled off the coast of the Sultanate of Oman. DNA extraction was conducted with a standard cetyltrimethylammonium bromide (CTAB) protocol [50]. For *G. galeus*, DNA was extracted using an SDS-based lysis buffer (PL2) from the NucleoSpin Plant II mini kit for DNA from plants (MACHEREY-NAGEL, Dueren, Germany) as per the manufacturer's protocol (Instruction-NucleoSpin-Plant-II.pdf (mn-net.com). Quality control included the assessment of double stranded DNA (dsDNA) fraction using the Qubit 4.0 fluorometer with the QubitTM 1× dsDNA HS assay kit (ThermoFisher Scientific) according to the manufacturer's protocol (MAN0017455) at the Central Analytical Facilities (CAF) at Stellenbosch University. Quantification of intact genomic DNA was indicated by the genome quality score (GQS) determined with electrophoresis on the LabChip® GXII Touch (PerkinElmer, Waltham, MA, USA), using the DNA Extended Range Chip and Genomic DNA Reagent Kit (PerkinElmer) according to the manufacturer's protocol, CLS140166.

2.2. Sequencing, quality assessment and assembly

One sample per species that passed quality control analysis was submitted to CAF for library preparation using the Ion Plus Fragment Library Kit (ThermoFisher Scientific) according to the manufacturer's protocol, Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide (MAN0009847 K.0). Briefly, DNA was fragmented using the Covaris S2 focused ultrasonicator (Covaris, Inc.; Woburn, MA, USA) with a 10% duty cycle, 5% intensity and 200 cycles/burst in two cycles; each consisting of 60 s. Sheared DNA was purified using 1.8× volume AMPure XP reagent (Beckman Coulter, Brea, CA, USA). The fragmented DNA was end-repaired at room temperature for 20 mins and purified using 1.8× volume AMPure XP reagent (Beckman Coulter). End-repaired DNA was blunt-end ligated to IonCode Barcode Adapters at 22 °C for 15 mins and purified, using 1.4× volume AMPure XP reagent (Beckman Coulter). The adapter-ligated library was amplified across eight polymerase chain reaction (PCR) cycles. The amplified library was size-selected on the Pippin Prep (Sage Science, Beverly, MA, USA) using 2% dye-free gel cassettes with marker L, to retain DNA fragments of ~475 bp. The size selected library was purified, using 1.8× volume AMPure XP reagent (Beckman Coulter) and assessed for fragment size distribution on the LabChip GX Touch 24 Nucleic Acid Analyzer (PerkinElmer, Waltham, MA, USA), using the X-Mark DNA LabChip and HT DNA NGS 3 K Reagent Kit according to the manufacturer's protocol (NGS 3 K Assay Quick Guide). The libraries were quantified using the Ion Library TaqMan™ Quantitation Kit according to the manufacturer's protocol (Ion Library TaqMan Quantitation Kit User Guide). Quantitative PCR amplification was performed using the StepOnePlus™ Real-Time PCR System (ThermoFisher Scientific).

Barcoded libraries were diluted to a target concentration of 40 pM and then combined in equimolar amounts of 25 μl for template preparation using the Ion 510™, Ion 520™, and Ion 530™ Chef Kit (ThermoFisher Scientific). Enriched, template positive ion sphere particles were loaded onto an Ion 530™ Chip (ThermoFisher Scientific). Two sample libraries were loaded per chip. Massively parallel sequencing was performed on the Ion GeneStudio™ S5 Prime System using the Ion S5™ Sequencing Solutions and Sequencing Reagent Kits according to the protocol.

After sequencing, flow space calibration and BaseCaller analyses were performed using default analysis parameters in the Torrent Suite v.5.16. Reads from each species' library were pulled from the pool based on their unique molecular barcode. Quality was checked in FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptors and poor-quality bases (Phred score below 20) were trimmed and reads shorter than 25 base pairs (bp) removed.

Finally, we employed three different assembly techniques, namely a reference-based approach, a hybrid technique whereby reads aligned to the reference were extracted and then mapped *de novo* and lastly, a straight *de novo* assembly. We chose to perform a more rigorous assembly procedure to account for reference allele bias that can result from a reference-based approach and to increase the chances of detecting duplications [51]. For the first approach, we mapped quality-filtered reads to the published *Mustelus mustelus* mitochondrial genome (NC_039629.1; [52]) to obtain a consensus sequence using the Geneious read mapper with medium sensitivity settings and five iterations in Geneious Prime (v.2019.1.3) [53]. *M. mustelus* was selected since the species falls unequivocally in the Triakidae clade and we could verify that it is a high-quality assembly. Reads that mapped to the reference genome were extracted and fed into a *de novo* pipeline in SPAdes v.3.15 [54]. *De novo* assembly was performed in SPAdes with the input set for unpaired Ion Torrent reads with 8 threads, kmers 21,33,55,77,99,127, the careful option to reduce the number of mismatches and short indels and all other parameters left as default. The same parameters were set for the straight *de novo* approach. All three assemblies were aligned to each other using the Geneious alignment tool with default parameters. The alignment was checked for discrepancies and edited manually to obtain the final genome sequence.

2.3. Genome annotation

Protein coding genes (PCGs), ribosomal (r)RNA and transfer (t)RNA genes were annotated using MitoAnnotator in MitoFish v.3.85 [55,56]. The reading frame of each coding region was checked by translating the sequences into amino acids in the Sequence Manipulation Suite 2 [57]. The annotated sequences were checked in Geneious to ensure completeness and to manually count overlapping regions and intergenic spaces between PCGs, rRNAs, tRNAs, and non-coding regions. MitoAnnotator was used to construct circular annotated mitogenome visuals. The annotated mitochondrial genomes were deposited on the GenBank [58] repository under the accession numbers ON075075, ON075076, ON075077, ON652873, and ON652874.

DAMBE v.7.2.141 [59] was used to calculate the A + T content and relative synonymous codon usage (RSCU) of the PCGs in each mitogenome. To measure the base composition skewness of nucleotide sequences, the following formula was used: AT-skew = [A-T]/[A + T] and GC-skew = [G-C]/[G + C] [60]. ARWEN v.1.2.3 (Björn Canbäck Bioinformatics) [61] and the tRNAscan-SE webserver v.2.0 (http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE2.cgi) [62] were used to predict the secondary structure of tRNAs using the generalized vertebrate mitochondrial tRNA settings. The control regions were inspected using the "Tandem Repeat Finder" webserver (https://tandem.bu.edu/trf/trf.html) [63] maintaining default settings.

2.4. Sequence alignment and concatenation

For representation of the Triakidae family, we included the four houndshark mitogenomes available on GenBank together with the five newly assembled mitogenomes. We also selected representative mitogenomes from each family in the order Carcharhiniformes and four representative species each from the orders Lamniformes and Orectolobiformes to complete the Galeomorphi cluster (Table 2). Selections were made in view of our intent to resolve the phylogenetic placement of the Triakidae family, and address whether the mitochondrial genome can offer new insight into intra-familial relationships. We produced codon-aware multiple sequence alignments for each of the 13 PCGs using MACSE v.2.01 [64]. We inspected and manually trimmed each set of alignments using MEGA v.11.0.11 [65], and any remaining ambiguously aligned sites were then further trimmed using BMGE v.1.12.1, with a sliding window size of 3, and maximum entropy of 0.5 [66].

In order to incorporate the structural information of ribosomal rRNAs during the alignment, we aligned the 16S rRNA and 12S rRNA using MAFFT v.7.299 [67,68] with the Q-INS-i algorithm of Katoh and Toh [69], and removed ambiguously aligned sites using BMGE. Before phylogenetic analysis, we produced three concatenated mitogenomic datasets from (i) the aligned individual PCGs datasets (Dataset 1: 13PCGs_NT dataset), (ii) the 13 PCGs plus the two rRNA genes (Dataset 2: 13PCGs_rRNAs_NT dataset) with the R package concatipede v1.0.1 [70], and (iii) derived the third mitogenomic dataset by translating the 13PCGs_NT dataset in MEGA (Dataset 3: 13PCGs_AA dataset).

2.5. Substitution saturation and data partitioning schemes

To examine the degree of nucleotide substitution saturation, we performed two-tailed tests of substitution saturation [71] for each gene as well as each codon position of the 13 PCGs considering the proportion of invariant sites as recommended by Xia and Lemey [72] in DAMBE v.7.2.141 [59]. We visually inspected substitution saturation by plotting the number of transitions (s) and transversions (v) *versus* divergence based on genetic distances derived from the Kimura two-parameter (K2P or K80) substitution model [73] (Supplementary material S1: part 1). The K80 substitution model accommodates transition/transversion rate bias and the 'K80 distance' is thus expected to increase linearly with

Table 2
General information and nucleotide composition for the 64 mitochondrial genomes of the Carcharhiniform order (including the five newly assembled Triakidae mitogenomes) and eight outgroups used in this study.

	Family	Species	Size (bp)	Accession Number	Whole genome composition						
					A	C	G	T	AT.%	AT skew	GC skew
Order	Carcharhinidae	*Carcharhinus acronotus*	16,719	NC 024055	5258	4210	2202	5049	61.6	0.020	−0.313
	Carcharhinidae	*Carcharhinus albimarginatus*	16,706	NC 047239	5271	4281	2182	4972	61.3	0.029	−0.325
	Carcharhinidae	*Carcharhinus amblyrhynchoides*	16,705	NC 023948	5249	4185	2198	5073	61.8	0.017	−0.311
	Carcharhinidae	*Carcharhinus amblyrhynchos*	16,705	NC 047238	5260	4218	2198	5029	61.6	0.022	−0.315
	Carcharhinidae	*Carcharhinus amboinensis*	16,704	NC 026696	5274	4167	2181	5082	62.0	0.019	−0.313
	Carcharhinidae	*Carcharhinus brachyurus*	16,704	NC 057525	5232	4190	2208	5073	61.7	0.015	−0.310
	Carcharhinidae	*Carcharhinus brevipinna*	16,706	NC 027081	5237	4224	2212	5033	61.5	0.020	−0.313
	Carcharhinidae	*Carcharhinus falciformis*	16,677	NC 042256	5244	4222	2201	5010	61.5	0.023	−0.315
	Carcharhinidae	*Carcharhinus leucas*	16,704	NC 023522	5256	4063	2190	5195	62.6	0.006	−0.300
	Carcharhinidae	*Carcharhinus limbatus*	16,705	NC 057057	5243	4198	2206	5058	61.7	0.018	−0.311
	Carcharhinidae	*Carcharhinus longimanus*	16,706	NC 025520	5260	4240	2192	5014	61.5	0.024	−0.318
	Carcharhinidae	*Carcharhinus macloti*	16,701	NC 024862	5279	4372	2175	4875	60.8	0.040	−0.336
	Carcharhinidae	*Carcharhinus melanopterus*	16,706	NC 024284	5227	4232	2226	5023	61.3	0.020	−0.311
	Carcharhinidae	*Carcharhinus obscurus*	16,706	NC 020611	5238	4232	2210	5026	61.4	0.021	−0.314
	Carcharhinidae	*Carcharhinus perezii*	16,709	MW528216	5274	4265	2175	4995	61.5	0.027	−0.325
	Carcharhinidae	*Carcharhinus plumbeus*	16,706	NC 024596	5221	4266	2226	4993	61.1	0.022	−0.314
	Carcharhinidae	*Carcharhinus sorrah*	16,707	NC 023521	5254	4306	2201	4946	61.1	0.030	−0.323
	Carcharhinidae	*Carcharhinus tjutjot*	16,705	NC 026871	5272	4406	2178	4849	60.6	0.042	−0.338
	Carcharhinidae	*Glyphis fowlerae*	16,704	NC 028342	5265	4412	2174	4853	60.6	0.041	−0.340
	Carcharhinidae	*Glyphis garricki*	16,702	NC 023361	5265	4386	2160	4891	60.8	0.037	−0.340
	Carcharhinidae	*Glyphis glyphis*	16,701	NC 021768	5267	4338	2168	4928	61.0	0.033	−0.334
	Carcharhinidae	*Lamiopsis temminckii*	16,708	NC 028341	5247	4316	2179	4966	61.1	0.028	−0.329
	Carcharhinidae	*Lamiopsis tephrodes*	16,705	NC 028340	5254	4297	2178	4976	61.2	0.027	−0.327
	Carcharhinidae	*Loxodon macrorhinus*	16,702	NC 029843	5296	4309	2194	4903	61.1	0.039	−0.325
	Carcharhinidae	*Prionace glauca*	16,705	NC 022819	5276	4076	2184	5169	62.5	0.010	−0.302
	Carcharhinidae	*Rhizoprionodon acutus*	16,693	NC 046016	5280	3991	2184	5238	63.0	0.004	−0.293
	Carcharhinidae	*Scoliodon laticaudus*	16,695	NC 042504	5266	3959	2199	5271	63.1	0.000	−0.286
	Carcharhinidae	*Scoliodon macrorhynchos*	16,693	NC 018052	5288	3994	2176	5235	63.0	0.005	−0.295
	Carcharhinidae	*Triaenodon obesus*	16,700	NC 026287	5240	4305	2203	4952	61.0	0.028	−0.323
	Galeocerdonidae	*Galeocerdo cuvier*	16,703	NC 022193	5310	3984	2174	5235	63.1	0.007	−0.294
	Hemigaleidae	*Hemigaleus microstoma*	16,701	NC 029400	5179	4411	2257	4854	60.1	0.032	−0.323
	Hemigaleidae	*Hemipristis elongata*	16,691	NC 032065	5296	4024	2154	5217	63.0	0.008	−0.303
	Pentanchidae	*Galeus melastomus*	16,706	NC 049881	5170	3783	2364	5388	63.2	−0.021	−0.231
	Pentanchidae	*Halaelurus buergeri*	19,100	NC 031811	5868	4784	2642	5806	61.1	0.005	−0.288
	Pentanchidae	*Parmaturus melanobranchus*	16,687	NC 056784	5067	3915	2348	5357	62.5	−0.028	−0.250
	Proscyllidae	*Proscyllium habereri*	16,708	NC 030216	5159	3968	2370	5211	62.1	−0.005	−0.252
	Pseudotriakidae	*Pseudotriakis microdon*	16,700	NC 022735	5227	3799	2276	5398	63.6	−0.016	−0.251
	Scyliorhinidae	*Cephaloscyllium fasciatum*	16,703	MZ424309	5111	3938	2419	5235	61.9	−0.012	−0.239
	Scyliorhinidae	*Cephaloscyllium umbratile*	16,698	NC 029399	5174	4000	2331	5193	62.1	−0.002	−0.264
	Scyliorhinidae	*Poroderma pantherinum*	16,686	NC 043830	5176	4175	2312	5023	61.1	0.015	−0.287
	Scyliorhinidae	*Scyliorhinus canicula*	16,697	NC 001950	5143	3986	2358	5210	62.0	−0.006	−0.257
	Scyliorhinidae	*Scyliorhinus torazame*	17,861	AP019520	5511	4205	2629	5516	61.7	0.000	−0.231
	Sphyrnidae	*Eusphyra blochii*	16,727	NC 031812	5279	4296	2180	4972	61.3	0.030	−0.327
	Sphyrnidae	*Sphyrna lewini*	16,726	NC 022679	5254	4406	2204	4862	60.5	0.039	−0.333
	Sphyrnidae	*Sphyrna mokarran*	16,719	NC 035491	5280	4273	2179	4987	61.4	0.029	−0.325
	Sphyrnidae	*Sphyrna tiburo*	16,723	NC 028508	5227	4338	2244	4914	60.6	0.031	−0.318
	Sphyrnidae	*Sphyrna zygaena*	16,731	NC 025778	5277	4187	2209	5058	61.8	0.021	−0.309
	Triakidae	***Galeorhinus galeus***	**17,487**	**ON652874**	**5444**	**4278**	**2366**	**5399**	**62.0**	**0.004**	**−0.288**
	Triakidae	*Hemitriakis japanica*	17,301	KJ617039	5379	4612	2316	4994	60.0	0.037	−0.331
	Triakidae	***Mustelus asterias***	**16,708**	**ON652873**	**5138**	**4107**	**2321**	**5142**	**61.5**	**0.000**	**−0.278**
	Triakidae	*Mustelus griseus*	16,754	NC 023527	5149	4178	2352	5075	61.0	0.007	−0.280
	Triakidae	*Mustelus manazo*	16,707	NC 000890	5139	4077	2319	5172	61.7	−0.003	−0.275
	Triakidae	***Mustelus mosis***	**16,755**	**ON075077**	**5125**	**4208**	**2370**	**5052**	**60.7**	**0.007**	**−0.279**
	Triakidae	*Mustelus mustelus*	16,755	NC 039629	5123	4193	2376	5063	60.8	0.006	−0.277
	Triakidae	***Mustelus palumbes***	**16,708**	**ON075076**	**5141**	**4114**	**2319**	**5135**	**61.5**	**0.001**	**−0.279**
	Triakidae	***Triakis megalopterus***	**16,746**	**ON075075**	**5182**	**4167**	**2322**	**5075**	**61.3**	**0.010**	**−0.284**
Lamniformes	Alopiidae	*Alopias pelagicus*	16,692	NC 022822	5354	4230	2218	4890	61.4	0.045	−0.312
	Lamnidae	*Lamna ditropis*	16,699	NC 024269	4944	4518	2466	4771	58.2	0.018	−0.294
	Lamnidae	*Carcharodon carcharias*	16,744	NC 022415	5119	4498	2326	4801	59.2	0.032	−0.318
	Odontaspididae	*Carcharias taurus*	16,773	NC 023520	5326	4415	2212	4820	60.5	0.050	−0.332
	Hemiscylliidae	*Chiloscyllium griseum*	16,755	NC 017882	5612	3962	2091	5090	63.9	0.049	−0.309
Orectolobiformes	Orectolobidae	*Orectolobus japonicus*	16,706	NC 022148	5440	4017	2216	5033	62.7	0.039	−0.289
	Rhincodontidae	*Rhincodon typus*	16,928	KC633221	5665	4111	2166	4986	62.9	0.064	−0.310
	Stegostomatidae	*Stegostoma fasciatum*	16,658	NC 029480	5677	4311	2089	4581	61.6	0.107	−0.347

divergence time. We also conducted the substitution saturation analysis for the 13PCGs_NT and 13PCGs_rRNAs_NT datasets. According to the observed index of substitution saturation (I_{SS}), all genes and codon positions, except for ND3_pos3, showed little saturation [$p < 0.05$: $I_{SS} < I_{SS.cSym}$ (assuming a symmetrical topology) and $I_{SS} < I_{SS.cAsym}$ (assuming an asymmetrical topology); see Supplementary material S2: Table S1].

Given the discordance in family-level phylogenetic inferences within the Carcharhiniformes based on mitochondrial phylogenomics [15,8], we systematically surveyed *a priori* partitioning schemes with varying degrees of complexity to determine the effects on phylogeny estimation. Based on the substitution saturation analyses and frequently used *a priori* partitioning schemes in mitochondrial phylogenomics, we designated

eight different partitioning schemes (Table 3). For each partition scheme, we used ModelFinder [32] to select the best-fitting partitioning scheme and models of evolution using the corrected Akaike Information Criterion (AICc) and the edge-linked proportional partition model [74] as implemented in IQ-Tree v.2.2.0.3 [75]. This was also conducted using the Bayesian Information Criterion (BIC) to see how the choice of optimality criterion affected the results. We applied the new model selection procedure (−m MF + MERGE), which additionally implements the FreeRate heterogeneity model inferring the site rates directly from the data instead of being drawn from a gamma distribution (−cmax 20; [76]). To reduce the computational burden, the top 30% partition merging schemes were inspected using the relaxed clustering algorithm (−rcluster 30), as described in Lanfear et al. [25]. We also applied a secondary model selection for the best-fitting partitioning identified by ModelFinder under the FreeRate heterogeneity model to select the next best model for Bayesian inference by rerunning ModelFinder with options: -m TESTONLY -mset mrbayes.

2.6. Phylogenetic reconstruction

We reconstructed phylogenies based on the Maximum Likelihood (ML) criterion in IQ-Tree and Bayesian inference (BI) in MrBayes v.3.2.6 [77,78]. For ML analysis, we used the substitution models indicated by ModelFinder (Table 4). To determine if there was systematic bias in phylogenetic inference, we conducted eight independent runs with the best-fit partitioning scheme identified in ModelFinder per *a priori* partitioning scheme. For all runs, we used the Nearest Neighbor Interchange (NNI) approach to search for tree topology and for computing branch supports with 1000 replicates of the Shimodaira-Hasegawa approximate likelihood-ratio test (SH-aLRT; [79]) and 1000 bootstrapped replicates of the ultrafast bootstrapping (UFBoot2) approach [80]. We considered clades with UFBoot2 $\geq$ 95 and SH-aLRT $\geq$80 as being well supported [75]. We then compared the trees from the eight runs to determine significant differences with the approximately unbiased (AU) tree topology test [81] also implemented in IQ-Tree. For BI, we ran a pair of independent searches for 5 million generations, with trees saved every 1000 generations and the first 2500 sampled trees of each search discarded as burn-in. Convergence diagnostics involved the screen of the model parameter summary statistics Estimated Sample Size (ESS) and Potential Scale Reduction Factor (PSRF), where convergence occurred at ESS >200, and PSRF ~1.0. Finally, a consensus tree showing all compatible groupings was constructed. We performed the BI analysis on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway portal v.3.3 (www.phylo.org) at the San Diego Supercomputer Center [82].

In addition to UFBoot2 and SH-aLRT (IQ-Tree) and posterior probabilities (MrBayes), we investigated topological conflict around each branch of the species tree using gene (gCF) and site concordance factor (sCF), also implemented in IQ-Tree. The gCF and sCF for each branch of the species tree represent the percentage of decisive gene trees and alignment sites that contain that branch [83]. Assuming incomplete lineage sorting (ILS) is the only source of discordance between gene and species trees, the multi-species coalescent model predicts that the likelihood of a gene tree quartet matching the species tree topology is greater than the likelihood of matching the two alternatives. Furthermore, the two alternatives will have comparable frequencies [84,85]. We calculated the relative frequency of branch quartets surrounding focal clades based on this expectation. sCF values have a minimum value of ~30% because they are calculated by comparing the three possible resolutions of a quartet around a node, so when the data are completely equivocal about these resolutions, we expect an sCF value of 33%. However, gCF values can reach 0% if no gene tree contains a branch present in the species tree because they are calculated from full gene trees, such that there are many more than three possible resolutions around a node [83]. This can happen when a combination of biology and stochastic error leads to gene-tree discordance.

We then used a χ2-test to see if the frequency of gene trees (gCF) and sites (sCF) supporting the two alternative topologies differed significantly as implemented in Lanfear's R script ([83]; http://www.robertlanfear.com/blog/files/concordance_factors.html) in R v.4.1.2 [86]. Non-significant *p*-values ($p > 0.05$) indicate that the hypothesis of equal frequencies was not rejected, implying that discordance between gene trees and/or sites is most likely due to ILS. We also constructed plots comparing gCF and sCF values with UFBoot2 values using Lanfear's R script.

Trees were visualized in Figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and, finally, a consensus tree was constructed in Evolview v.3 webserver [87,88].

2.7. Gene/species tree analyses

Phylogenetic inference based on the traditional concord approach, which groups genes into a supermatrix, produces high support for incorrect branches when incomplete lineage sorting is present because it assumes that all genes/sites share the same evolutionary history [89,90]. Therefore, we employed the summary method, Accurate Species TRee ALgorithm (ASTRAL v.5.6.3) [91], and the site-based method, SVDQuartets [89,92], to estimate the effects of gene-tree conflict on species-tree inference under the multispecies coalescent model. These quartet-based methods seek to find the tree that maximizes the number of induced quartet trees in gene trees that are shared by the species tree [91]. We began by estimating individual gene trees for the 13 PCGs and 2 rRNAs based on the ML criterion in IQ-Tree with a greedy model selection strategy (−m MFP). We used the NNI approach to search for tree topology and for computing branch support with 1000 bootstrapped replicates of the UFBoot2 approach [80]. Gene trees for the amino acid alignment were created in the same manner. We then created a combined file with all the gene trees and ran this through ASTRAL using the default options. We also tested for polytomies with the method of Sayyari and Mirarab [85], which is based on a Chi-Square test among quartet frequencies for nodes, implemented with the -t 10 command.

The SVDQuartets analysis was conducted on Dataset 2 (13PCGs_rRNAs_NT) in PAUP* v.4.0a 169 (http://phylosolutions.com/paup-test/) [93], implementing the multispecies coalescent tree model with random quartet sampling of 100,000 replicates and 1000 bootstrap replicates. We conducted this step using the merged gene partitions identified by ModelFinder above (Table 4, PS5). We compared the coalescence-based trees with those from concatenated phylogenies, for both our datasets and the two most recent Carcharhiniform phylogenomic studies [15,8].

2.8. Testing site heterogenous models

To test for compositional heterogeneity among lineages in the nucleotide and amino acid datasets, we used AliGROOVE v.1.08 [94] with default settings. For nucleotide alignments, we ran the software with the -N option to treat indels as ambiguous characters and tested it without invoking -N so that indels were treated as a 5th character trait.

Table 3
A list of *a priori* partitioning schemes tested in this study. PCG, Protein-coding gene; pos, position.

PS01 - None (1 partition)
PS01AA - None (1 partition)
PS02 - Codon: pos1 + pos2 + pos3 + 2 rRNAs (5 partitions)
PS03 - Codon: pos1_pos2 + pos3 + 2 rRNAs (4 partitions)
PS04 - Codon: pos1 + pos3 + 2 rRNAs (4 partitions)
PS05 - Gene: 13 PCGs +2 rRNAs (15 partitions)
PS05AA - Gene: 13 PCGs (13 partitions)
PS06 - GeneXCodon: 13 PCGs pos1 + 13 pos2 + 13 pos2 + 2 rRNAs (41 partitions)
PS07 - GeneXCodon: 13 PCGs pos1_pos2 + 13 PCGs pos3 + 2 rRNAs (28 partitions)
PS08 - GeneXCodon: 13 PCGs pos1 + 13 PCGs pos3 + 2 rRNAs (28 partitions)

Table 4
Best-fit partition schemes and substitution models determined using ModelFinder for Maximum likelihood (ML) and Bayesian Inference (BI) phylogenies.

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
PS1	None (1 partition)	All	GTR + F + I + I + R6	GTR + F + I + G4	All	GTR + F + I + I + R5	GTR + F + I + G4
PS2	Codon: pos1 + pos2 + pos3 + 2 rRNAs (5 partitions)	13PCGs_pos1	GTR + F + R4	GTR + F + I + G4	13PCGs_pos1	GTR + F + I + I + R5	GTR + F + I + G4
		13PCGs_pos2	GTR + F + R3	GTR + F + I + G4	13PCGs_pos2	TVM + F + I + I + R3	GTR + F + I + G4
		13PCGs_pos3	GTR + F + R6	GTR + F + I + G4	13PCGs_pos3	GTR + F + I + I + R5	GTR + F + I + G4
		12S	GTR + F + R5	GTR + F + I + G4	12S_16S	GTR + F + R5	GTR + F + I + G4
		16S	GTR + F + R4	GTR + F + I + G4			
PS3	Codon: pos1_pos2 + pos3 + 2 rRNAs (4 partitions)	13PCGs_pos1_pos2	TIM2 + F + R5	GTR + F + I + G4	13PCGs_pos1_pos2	TIM2 + F + I + I + R4	GTR + F + I + G4
		13PCGs_pos3	GTR + F + R6	GTR + F + I + G4	13PCGs_pos3	GTR + F + I + I + R4	GTR + F + I + G4
		12S	GTR + F + R5	GTR + F + I + G4	12S_16S	GTR + F + R5	GTR + F + I + G4
		16S	GTR + F + R4	GTR + F + I + G4			
PS4	Codon: pos1 + pos3 + 2 rRNAs (4 partitions)	13PCGs_pos1	GTR + F + R4	GTR + F + I + G4	13PCGs_pos1	GTR + F + I + I + R5	GTR + F + I + G4
		13PCGs_pos3	GTR + F + R6	GTR + F + I + G4	13PCGs_pos3	GTR + F + I + I + R5	GTR + F + I + G4

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
		12S	GTR + F + R5	GTR + F + I + G4	12S	GTR + F + I + R4	GTR + F + I + G4
		16S	GTR + F + R4	GTR + F + I + G4	16S	TIM2 + F + I + R3	GTR + F + I + G4
		ATP6	TIM2 + F + R4	GTR + F + I + G4	ATP6_ATP8_COX2_COX3_ND3	TIM2 + F + R5	GTR + F + I + G4
		ATP8	TPM2 + F + I + G4	HKY + F + I + G4	COX1	TIM2 + F + I + R5	GTR + F + I + G4
		COX1	GTR + F + R4	GTR + F + I + G4	CYTB_ND1_ND2_ND4_ND4L_ND5	TIM2 + F + I + R5	GTR + F + I + G4
		COX2	TIM2 + F + R4	GTR + F + I + G4	ND6	TIM + F + I + G4	GTR + F + I + G4
PS5	Gene: 13 PCGs +2 rRNAs (15 partitions)	COX3_ND3	TIM2 + F + R5	GTR + F + I + G4	12S_16S	GTR + F + R5	GTR + F + I + G4
		CYTB_ND1_ND4_ND4L_ND5	TIM2 + F + R6	GTR + F + I + G4			
		ND2	TIM2 + F + R5	GTR + F + I + G4			
		ND6	GTR + F + I + G4	GTR + F + I + G4			
		12S	GTR + F + R5	GTR + F + I + G4			
		16S	GTR + F + R4	GTR + F + I + G4			
PS6	GeneXCodon: 13PCGs pos1 + 13 pos2 + 13 pos2 + 2	ATP6_pos1_ND1_pos1_ND3_pos1_ND4L_pos1	TIM2 + F + R4	GTR + F + I + G4	ATP6_pos1_CYTB_pos1_ND1_pos1_ND2_pos1_ND3_pos1_ND4_pos1_ND4L_pos1_ND5_pos1	GTR + F + I + R4	GTR + F + I + G4

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
	rRNAs (41 partitions)			GTR + F + I + G4			GTR + F + I + G4
		ATP6_pos2_ND1_pos2_ND4L_pos2	TVM + F + R3		ATP6_pos2_CYTB_pos2_ND1_pos2_ND2_pos2_ND3_pos2_ND4_pos2_ND4L_pos2_ND5_pos2	TVM + F + I + R3	
		ATP6_pos3_ATP8_pos3	TIM2 + F + I + G4	GTR + F + I + G4	ATP6_pos3_ATP8_pos3_COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	GTR + F + I + G4
		ATP8_pos1_ATP8_pos2	TIM2e + I + G4	GTR + F + I + G4	ATP8_pos1_ATP8_pos2_12S_16S	GTR + F + R5	GTR + F + I + G4
		COX1_pos1	TVM + F + I + G4	SYM + I + G4	COX1_pos1_COX2_pos1_COX3_pos1	SYM + I + I + R3	SYM + I + G4
		COX1_pos2_COX2_pos2	GTR + F + I + G4	GTR + F + I + G4	COX1_pos2_COX2_pos2_COX3_pos2	K3Pu + F + I + G4	GTR + F + I + G4
		COX1_pos3	GTR + F + R4	GTR + F + I + G4	COX1_pos3	TIM2 + F + I + R5	GTR + F + I + G4
		COX2_pos1_COX3_pos1	SYM + R3	SYM + I + G4	CYTB_pos3_ND1_pos3_ND2_pos3_ND4_pos3_ND4L_pos3_ND5_pos3	TIM2 + F + I + R4:	GTR + F + I + G4
		COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	HKY + F + I + G4	ND6_pos1_ND6_pos2	K3Pu + F + I + G4	HKY + F + I + G4
		COX3_pos2	K3Pu + F + I + G4	HKY + F + I + G4	ND6_pos3	TIM3 + F + I + G4	GTR + F + I + G4
		CYTB_pos1_ND4_pos1_ND5_pos1	GTR + F + R4	GTR + F + I + G4			
		CYTB_pos2_ND4_pos2	GTR + F + R3	GTR + F + I + G4			
		CYTB_pos3	GTR + F + R4	GTR + F + I + G4			
		ND1_pos3_ND2_pos3	TIM2 + F + R4	GTR + F + I + G4			

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
		ND2_pos1	GTR + F + I + G4	GTR + F + I + G4			
		ND2_pos2_ND3_pos2	GTR + F + R3	GTR + F + I + G4			
		ND4_pos3_ND4L_pos3	TIM2 + F + R5	GTR + F + I + G4			
		ND5_pos2	GTR + F + I + G4	GTR + F + I + G4			
		ND5_pos3	TIM2 + F + R5	GTR + F + I + G4			
		ND6_pos1	TIM2 + F + I + G4	GTR + F + I + G4			
		ND6_pos2	TVM + F + I + G4	HKY + F + I + G4			
		ND6_pos3	TIM3 + F + I + G4	GTR + F + I + G4			
		12S	GTR + F + R5	GTR + F + I + G4			
		16S	GTR + F + R4	GTR + F + I + G4			
PS7	GeneXCodon: 13 PCGs pos1_pos2 + 13 pos3 + 2 rRNAs (28 partitions)	ATP6_pos1_pos2_ND1_pos1_pos2_ND3_pos1_pos2	GTR + F + I + G4	GTR + F + I + G4	ATP6_pos1_pos2_CYTB_pos1_pos2_ND1_pos1_pos2_ND2_pos1_pos2_ND3_pos1_pos2_ND4_pos1_pos2_ND4L_pos1_pos2_ND5_pos1_pos2	TIM2 + F + R4	GTR + F + I + G4
		ATP6_pos3_ATP8_pos3	TIM2 + F + I + G4	GTR + F + I + G4	ATP6_pos3_ATP8_pos3_COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	GTR + F + I + G4
		ATP8_pos1_pos2	TIM2 + F + I + G4	GTR + F + I + G4	ATP8_pos1_pos2_12S_16S	GTR + F + R5	GTR + F + I + G4

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
		COX1_pos1_pos2	TIM2 + F + I + G4	GTR + F + I + G4	COX1_pos1_pos2_COX2_pos1_pos2_COX3_pos1_pos2	TIM2 + F + I + G4	GTR + F + I + G4
		COX1_pos3	GTR + F + R4	GTR + F + I + G4	COX1_pos3	TIM2 + F + I + I + R5	GTR + F + I + G4
		COX2_pos1_pos2_COX3_pos1_pos2	GTR + F + R3	GTR + F + I + G4	CYTB_pos3_ND1_pos3_ND2_pos3_ND4_pos3_ND4L_pos3_ND5_pos3,	TIM2 + F + I + I + R4	GTR + F + I + G4
		COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	HKY + F + I + G4	ND6_pos1_pos2	K3Pu + F + I + G4	GTR + F + I + G4
		CYTB_pos1_pos2_ND4L_pos1_pos2_ND5_pos1_pos2	TIM2 + F + R4	GTR + F + I + G4	ND6_pos3	TIM3 + F + I + G4	GTR + F + I + G4
		CYTB_pos3	GTR + F + R4	GTR + F + I + G4			
		ND1_pos3_ND2_pos3	TIM2 + F + R4	GTR + F + I + G4			
		ND2_pos1_pos2_ND4_pos1_pos2	TIM2 + F + R4	GTR + F + I + G4			
		ND4_pos3_ND4L_pos3	TIM2 + F + R5	GTR + F + I + G4			
		ND5_pos3	TIM2 + F + I + G4	GTR + F + I + G4			
		ND6_pos1_pos2	GTR + F + R4	GTR + F + I + G4			
		ND6_pos3	TIM3 + F + I + G4	GTR + F + I + G4			
		12S	GTR + F + R5	GTR + F + I + G4			

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
		16S	GTR + F + R4	GTR + F + I + G4			
		ATP6_pos1_ND1_pos1	TIM2 + F + I + G4	GTR + F + I + G4	ATP6_pos1_CYTB_pos1_ND1_pos1_ND2_pos1_ND3_pos1_ND4_pos1_ND4L_pos1_ND5_pos1	GTR + F + I + I + R4	GTR + F + I + G4
		ATP6_pos3_ATP8_pos3	TIM2 + F + I + G4	GTR + F + I + G4	ATP6_pos3_ATP8_pos3_COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	GTR + F + I + G4
		ATP8_pos1	TN + F + G4	GTR + F + G4	ATP8_pos1_12S_16S	GTR + F + R5	GTR + F + I + G4
		COX1_pos1	TIM2e + I + G4	SYM + I + G4	COX1_pos1_COX2_pos1_COX3_pos1	SYM + I + I + R3	SYM + I + G4
		COX1_pos3	GTR + F + R4	GTR + F + I + G4	COX1_pos3	TIM2 + F + I + I + R5	GTR + F + I + G4
PS8	GeneXCodon: 13 PCGs pos1 + 13 pos3 + 2 rRNAs (28 partitions)	COX2_pos1_COX3_pos1_ND3_pos1	SYM + I + G4	SYM + I + G4	CYTB_pos3_ND1_pos3_ND2_pos3_ND4_pos3_ND4L_pos3_ND5_pos3	TIM2 + F + I + I + R4	GTR + F + I + G4
		COX2_pos3_COX3_pos3_ND3_pos3	TIM2 + F + R5	HKY + F + I + G4	ND6_pos1	TPM2u + F + I + G4	GTR + F + I + G4
		CYTB_pos1_ND4_pos1_ND4L_pos1_ND5_pos1	GTR + F + R4	GTR + F + I + G4	ND6_pos3	TIM3 + F + I + G4	GTR + F + I + G4
		CYTB_pos3_ND1_pos3_ND2_pos3	TIM2 + F + R4	GTR + F + I + G4			
		ND2_pos1	GTR + F + I + G4	GTR + F + I + G4			
		ND4_pos3_ND4L_pos3_ND5_pos3	GTR + F + R6	GTR + F + I + G4			
		ND6_pos1	TIM2 + F + I + G4	GTR + F + I + G4			

(*continued on next page*)

Table 4 (*continued*)

Partition Scheme	Prior	Partition (AICc)	Best-fit substitution model		Partition (BIC)	Best fit substitution model	
			ML	BI		ML	BI
				GTR			
			TIM3	+ F			
			+ F + I	+ I			
		ND6_pos3	+ G4	+ G4			
				GTR			
				+ F			
			GTR +	+ I			
		12S	F + R5	+ G4			
				GTR			
				+ F			
			GTR +	+ I			
		16S	F + R4	+ G4			

This method establishes non-random similarity between any two sequences at each site in a matrix of pairwise comparisons relative to the variation in the entire set of sequences. BaCoCa v.1.1 [95] was used to test for biases in individual PCGs and rRNAs as well as the alignment as a whole which could be violating the assumptions of stationarity, reversibility, homogeneity and therefore misleading phylogenetic reconstructions. We implemented the *r* option of the program to generate heat maps in combination with hierarchical clustering. Base heterogeneity across taxa is assessed using the Relative Composition Frequency Variability (RCFV) value [96] based on nucleotide or amino acid frequencies. Besides individual character states, this is also done for classes of character states such as purines or polar amino acids as well as for ambiguity characters.

The PHYLOBAYES_MPI v.1.9 package [36] was used to conduct Bayesian Inference analyses using the *pb_mpi* program for the three concatenated mitogenomic datasets defined in our study: (i) the aligned individual PCGs datasets (Dataset 1: 13PCGs_NT dataset), (ii) the 13 PCGs plus the two rRNA genes (Dataset 2: 13PCGs_rRNAs_NT dataset), and (iii) the derived third mitogenomic dataset by translating the 13PCGs_NT dataset (Dataset 3: 13PCGs_AA dataset). For each dataset (supermatrix), we tested the GTR + G4 site homogenous model, CAT + Poisson + G4 model (exchange rates are fixed) and the CAT + GTR + G4 model (exchange rates are independently estimated from the data under general time reversible Markov processes). The C60 + GTR + G4 empirical profile mixture model, which provides 60 rate categories to describe rate variation among sites, was also tested for the amino acid dataset. Two independent Markov chain Monte Carlo (MCMC) chains were run per supermatrix until convergence was reached, whenever possible, due to computational time constraints. The accuracy of findings from Bayesian phylogenetic analyses is heavily dependent on two factors: ensuring that the chains have reached a state of stability, known as stationarity, and obtaining a substantial number of independent samples from the posterior distribution. We assessed convergence of independent runs using maximum difference (*maxdiff*) in all bipartitions and effective sample size (*effsize*) following the guidelines specified in the PHYLOBAYES_MPI manual. The *bpcomp* program was used to generate outputs of the largest *maxdiff* and mean difference (*meandiff*) discrepancies observed across all bipartitions, while the *tracecomp* program was used to estimate the *effsize* and discrepancy for the summary variables of the models. For both programs we used a conservative burn-in of 20% of the length of MCMC chain. Chain convergence was also checked using Tracer v.1.7.1 [97]. Subsequently, post-analyses were conducted with the *readpb_mpi* program to obtain site-specific log likelihood statistics using a burn-in of 20% of the length of MCMC chain, and thinning accordingly to obtain a total of at least 100 MCMC points (see results later) given the computational intensiveness of estimating site log likelihoods under the CAT models. Finally, for model comparison we collected the site-specific scores, widely applicable information criterion (*wAIC*) and the asymptotically equivalent leave one-out cross-validation (*LOO-CV*) [98], using the script *read_loocv_waic.py*, which is provided as part of the PHYLOBAYES_MPI package.

3. Results

3.1. Genome structure and composition

Ion Torrent sequencing yielded five complete mitochondrial genome sequences. Genome assembly using *de novo,* reference-based and a hybrid approach resulted in contiguous consensus sequences for all genomes except for *G. galeus*. The reference-based approach failed to detect a 714 bp-duplication located between $tRNA^{Pro}$ and the control region. Annotation of the *de novo* assembly using MitoAnnotator revealed that this duplication is made up of two parts, a 644 bp non-coding (NC_{644}) fragment at the 5′-end and a 70 bp $tRNA^{Thr}$ ($tRNA^{Thr'}$) at the 3′-end of the H-strand. We aligned the duplicated fragments with the *G. galeus* reference genome assembly that did not contain the

duplication in Geneious and found that NC_{644} aligned with the *cytochrome b* (*Cytb*) gene and $tRNA^{Thr'}$ aligned with $tRNA^{Thr}$ with some gaps. The duplication was confirmed using Sanger sequencing (Supplementary material S3).

The five newly assembled mitogenomes are between 16,708 and 17,488 bp long with a gene content and composition consistent with other elasmobranch mitogenomes (Fig. 3, Supplementary material S2: Table S2). They all encode 13 protein-coding genes (PCGs), two rRNA genes, and 22 tRNA genes ($tRNA^{Leu}$ and $tRNA^{Ser}$ are duplicated), with the exception of *G. galeus* which contains 23 tRNA genes ($tRNA^{Thr}$ is also duplicated). Nine genes (1 PCG and 8 tRNAs) are found on the L-strand, while the other 28 are encoded on the H-strand.

All sequenced Triakidae species have two non-coding regions, a 1067–1142 bp control region (*D-loop*) between $tRNA^{Pro}$ and $tRNA^{Phe}$ and 30–36 bp L-strand replication origin (O_L) between $tRNA^{Asn}$ and $tRNA^{Cys}$. Additionally, there are overlapping nucleotides and gaps between some PCGs and between PCGs and tRNAs. The control region was enriched in tandem repeat sequences in *G. galeus, M. asterias, M. manazo, M. palumbes* and *T. megalopterus* (Table S3). *G. galeus* has an additional 714 bp-intergenic spacer between $tRNA^{Pro}$ and the control region which has not been reported before in other fish species (Fig. 3, Supplementary information S3).

The nucleotide composition of the houndshark mitogenomes is similar to that of all other Carcharhiniformes (Table 2). The genomes are AT-rich with an overall A + T content of 60–62% and a positive AT-skew (0.0006–0.0371), indicative of a slight excess of adenine bases compared to thymine, except *M. asterias* and *M. manazo* which have a negative AT-skew, concordant with *Galeus melastomus*. Overall, all the genomes have a negative GC-skew, suggestive of C-biased nucleotide composition. *ND6* has the largest negative AT-skew and largest positive GC-skew of all the Triakidae genomes (Supplementary material S1: part 2). Like other Carcharhiniformes, *ATP8* has the highest AT-content in all the genomes, except *H. japonica* where the AT-content of the *D-loop* is higher. *G. galeus* also has a significantly higher AT-content in the *D-loop* in comparison with other Triakidae species (Fig. 3, Supplementary material S1: part 2).

3.2. Protein coding genes

Among the 13 protein-coding genes, there are three cases of overlapping reading frames in all Triakidae mitogenomes: *ATP6* and *ATP8* share 10 nucleotides on the same strand, *ND4* and *ND4L* overlap by seven nucleotides on the same strand, and *ND5* and *ND6* share four nucleotides on the opposite strand. All PCGs have the start codon ATG, except for *COI* which begins with GTG instead. TAA is the stop codon in *COI, ATP8, ND4L* and *ND5*; TAG is found in *ND6* and *ND2, ATP6* and *COIII* possess the truncated "TA" stop codon while "T" is found in *COII, ND3, ND4,* and *Cytb*. *H. japonica, M. griseus* and *M. manazo* possess a complete stop codon for *ATP6* like *Carcharhinus amboinensis* (Table S2).

There is significant bias towards A/T in codon usage, concordant with nucleotide composition estimates, in all the genomes. The RSCU analysis showed that codons were biased towards using more A/T nucleotides at the third codon (Supplementary material S2: Table S4). Among PCGs, leucine 1 (11.53–12.08%) and cysteine (0.63%) are the most and least frequently used amino acids, respectively, concordant with other Carcharhiniform species.

3.3. Ribosomal RNA and transfer RNA genes

The *12S rRNA* and *16S rRNA* genes are both encoded by the H-strand, with A + T contents of 56–58.2% and 62.4–63.1%, respectively. The *12S rRNA* in the five mitogenomes ranges from 951 to 954 bp and is located between $tRNA^{Phe}$ and $tRNA^{Val}$ (Table S2). The *16S rRNA* ranges from 1667 to 1671 bp and is found between $tRNA^{Val}$ and $tRNA^{Leu1}$. 14 tRNAs are encoded by the H-strand and the remaining eight are encoded by the L-strand (Table S2). In *G. galeus*, there is a 2 bp overlap between $tRNA^{Ile}$ and $tRNA^{Gln}$. The anticodon sequences of each of the tRNA genes is identical among the Triakidae mitogenomes and outgroups. All the tRNAs were capable of folding into a secondary cloverleaf structure in Arwen and tRNAscan-SE apart from $tRNA^{Ser1}$ which lacked a dihydrouridine (DHU) arm. In *G. galeus,* $tRNA^{Thr'}$ was predicted to be a $tRNA^{Thr}$-like structure by MitoAnnotator and ARWEN. To clarify the functionality of the duplicated tRNA, we further investigated its secondary structure by employing three different prediction tools considering both algorithmic and energy models for more accurate prediction [99] (Fig. 4). We began by using RNAfold v.2.4.17 webserver (http://rna.tbi.univie.ac.at/) which uses a dynamic programming technique to predict secondary structure by calculating the minimum free energy (MFE) [100,101]. A short-coming of this approach is that it relies on experimental techniques to determine scoring parameters and often fails to detect unconventional tRNA structures which are products of tandem duplication repeats in the mitochondrial genome [102]. To address this, machine learning approaches were developed to train scoring parameters based on reference structures [99]. We chose tRNA-Scan v.2.0 [62] because it has been trained to using a broad set of vertebrate mitochondrial tRNAs. The challenge with algorithmic models is that rich parameterization can cause overfitting to the training data, preventing robust predictions [103]. We therefore elected to also test a hybrid method, MXFold2 (http://ws.sato-lab.org/mxfold2/), which integrates folding scores calculated by a deep neural network trained with a large quantity of data but avoids overfitting rich-parameterized weight parameters to the training data by resorting to the thermodynamic parameters for assessing previously unobserved substructures [99].

As evident from the MFE scores and base pair knots in Fig. 4, RNAfold and MXfold2 produce far less stable secondary structures for $tRNA^{Thr'}$ than for $tRNA^{Thr}$, however tRNAscan-SE produces a $tRNA^{Thr'}$ that can form a cloverleaf structure when an internal loop is incorporated into the acceptor stem. Furthermore, the duplicated $tRNA^{Thr'}$ has maintained a 79% sequence similarity with $tRNA^{Thr}$ and the anticodon is conserved.

3.4. Phylogenetic reconstruction

The final concatenated and cleaned alignment of Triakidae and other representative Carcharhiniformes with two outgroup families, was constructed using the 13 PCGs and 2 rRNA genes and comprised 13,945 sites. The likelihood statistics used to test our eight partitioning strategies are very similar for AICc and BIC and they rank the partitioning schemes in the same order. Partitioning schemes with lower AICc and BIC scores generally have fewer partitions and free rate parameters (Table S5). Partitioning strategy 4 (PS4; codon pos1 + pos3 + 2 rRNAs; 4 partitions) is the best supported scheme for both ML and BI model selection with an AICc of 374,298.30 and a BIC of 375,208.20. The log likelihood scores for the tree also support PS4 and show a positive correlation with the information criterion scores. The best-fit partitioning schemes, along with the optimal substitution models, under ML and BI as selected by ModelFinder differ substantially for AICc and BIC (Table 4). AICc yields an increased number of partitions, however the models of substitution are very similar across partitions. AICc also finds more parameter-rich models, and these differ to those selected by BIC for different codon positions (P2-P4 & P6-P8) and genes (P5-P8), particularly for the rate (R) model. However, ModelFinder was demonstrated to be accurate regardless of the optimality criterion and, in our dataset, tree topologies constructed with BIC are the same as those constructed with AICc, with minor variations in bootstrap and posterior probability values. For this reason, the BIC trees are not displayed in the results.

Phylogenetic reconstruction based on the different partitioning schemes yielded consistent clustering of larger, deeper clades but there are some contrasting relationships for more recently diverged lineages (Fig. 5, Supplementary material S1: part 4). The Lamniform order is confirmed as a sister clade to the Carcharhiniformes with a UFBoot2, SH-aLRT and posterior probability of 100% for all partitioning schemes.

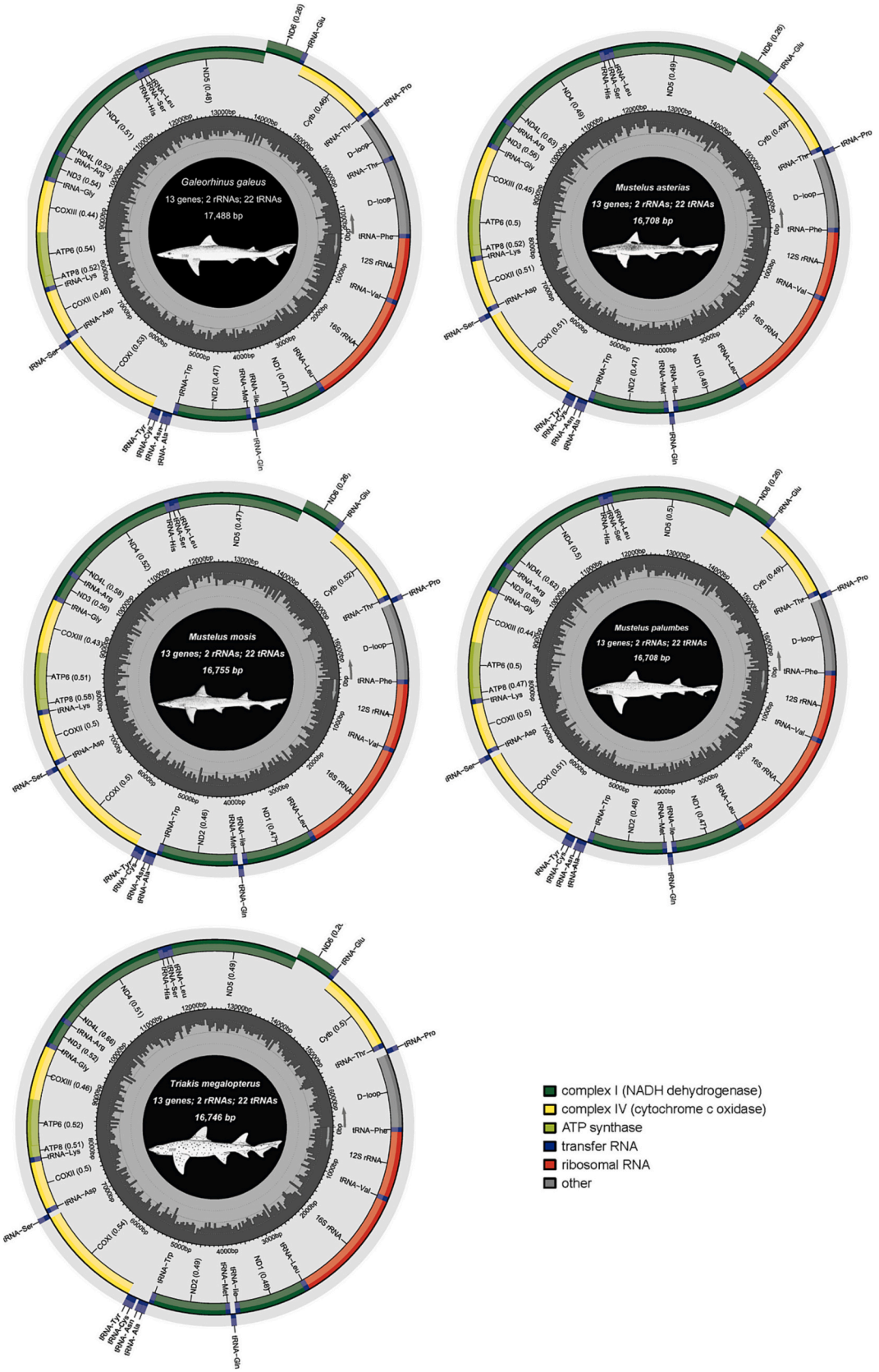

Fig. 3. Gene maps of the five new mitogenomes constructed in MitoAnnotator. In the outermost ring, the inner genes are transcribed clockwise, while the outer genes are transcribed counterclockwise. The middle grey circle shows the GC-content. Houndshark sketches were retrieved from the Food and Agriculture Organization (FAO). A. *Galeorhinus galeus*. *B. Mustelus asterias*. C. *Mustelus mosis*. D. *Mustelus palumbes*. E. *Triakis megalopterus*.

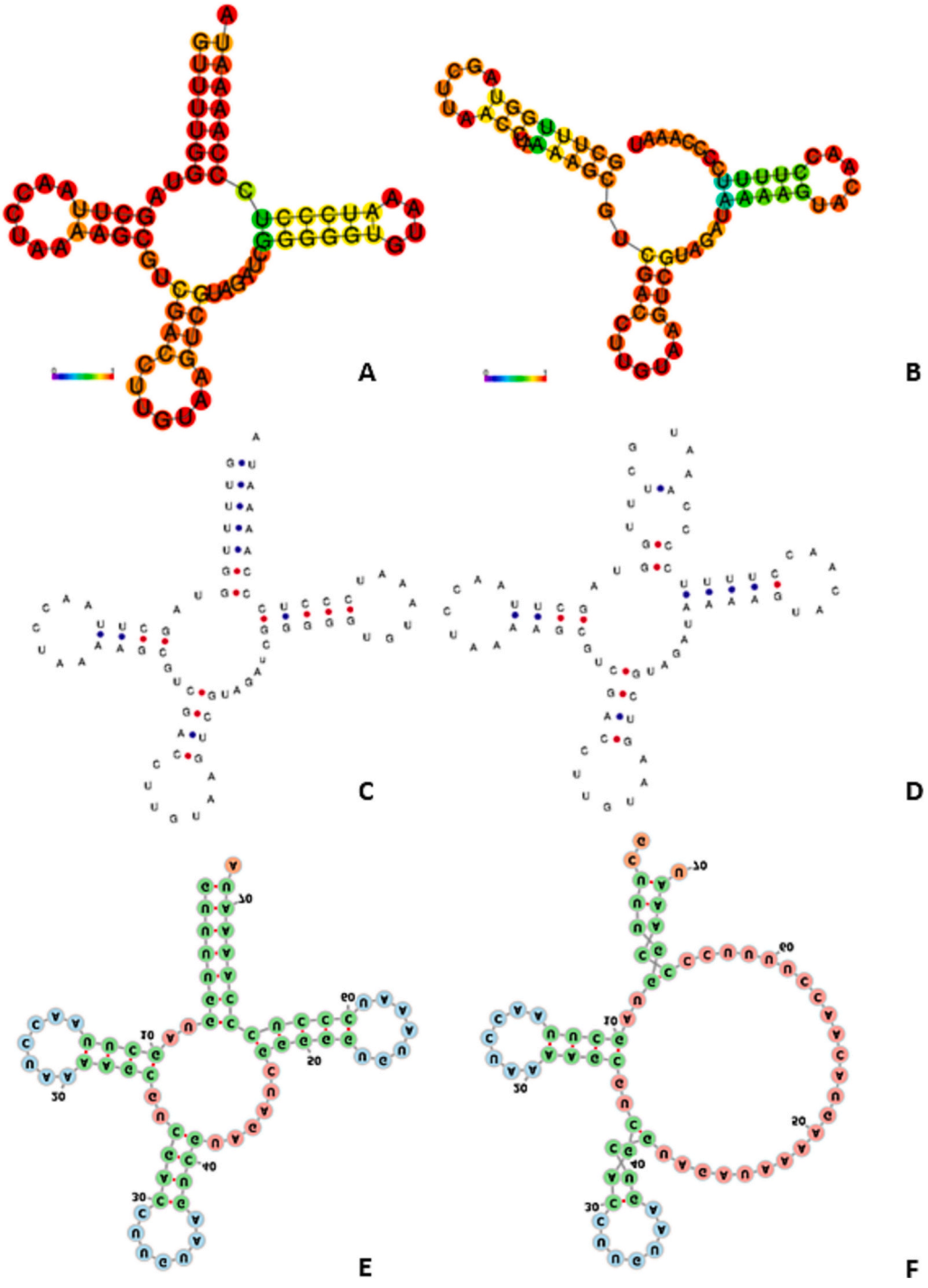

Fig. 4. RNA structure of the *tRNA*^{Thr} at 15,490–15,561 bases with a minimum free energy (MFE) of - 17.60 kcal/mol (A) and the duplicated pseudo-*tRNA*^{Thr} (*tRNA*$^{Thr'}$) in *Galeorhinus galeus* at 16,277–16,346 bases with a MFE of – 9.40 kcal/mol (B) discovered by RNAfold v.2.4.17. Base-pair probabilities are shown in the colour bar scale. More typical cloverleaf structures found by tRNAscan-SE v.2.0 for *tRNA*^{Thr} (C) and *tRNA*$^{Thr'}$ (D). Structure of *tRNA*^{Thr} with a MFE of −16.3 kcal/mol and *tRNA*$^{Thr'}$ with a MFE of −8 kcal/mol (F) created using MXFold2.

Galeocerdo cuvier was consistently recovered as a sister-group to Sphyrnidae and Carcharhinidae, which supports the decision to reclassify it from Carcharhinidae to its own family, Galeocerdonidae. The *Carcharhinus* genus shows evidence of polytomy in some of the BI variations (PS2, PS4 and PS8), but not in ML topologies. Scyliorhinidae is paraphyletic for all partitioning schemes, with *Parmaturus melanobranchus* clustering separately to the rest of the family.

PS1 (no partitioning) and PS5 (partitioning based on genes) result in the monophyletic grouping of Triakidae for the ML analysis, which is supported by the translated alignment phylogenies (AA1 and AA5). However, codon and geneXcodon partitioning schemes demonstrate paraphyly for Triakidae. PS5 supports Triakidae as a monophyletic clade with 100% UFBoot2, SH-aLRT and posterior probability values. *M. mosis* clusters with *M. mustelus* for the majority of ML topologies but clusters with *M. griseus* for PS2-PS4 and in all BI topologies. The paraphyletic PS4 topology has the highest AICc and BIC scores, however the UFBoot2 value is below 95% for the *M. griseus/M. mosis* split (77.7%), the *Mustelus* genus split (81.5%) and the isolated clustering of *H. japonica* (91.2%). BI

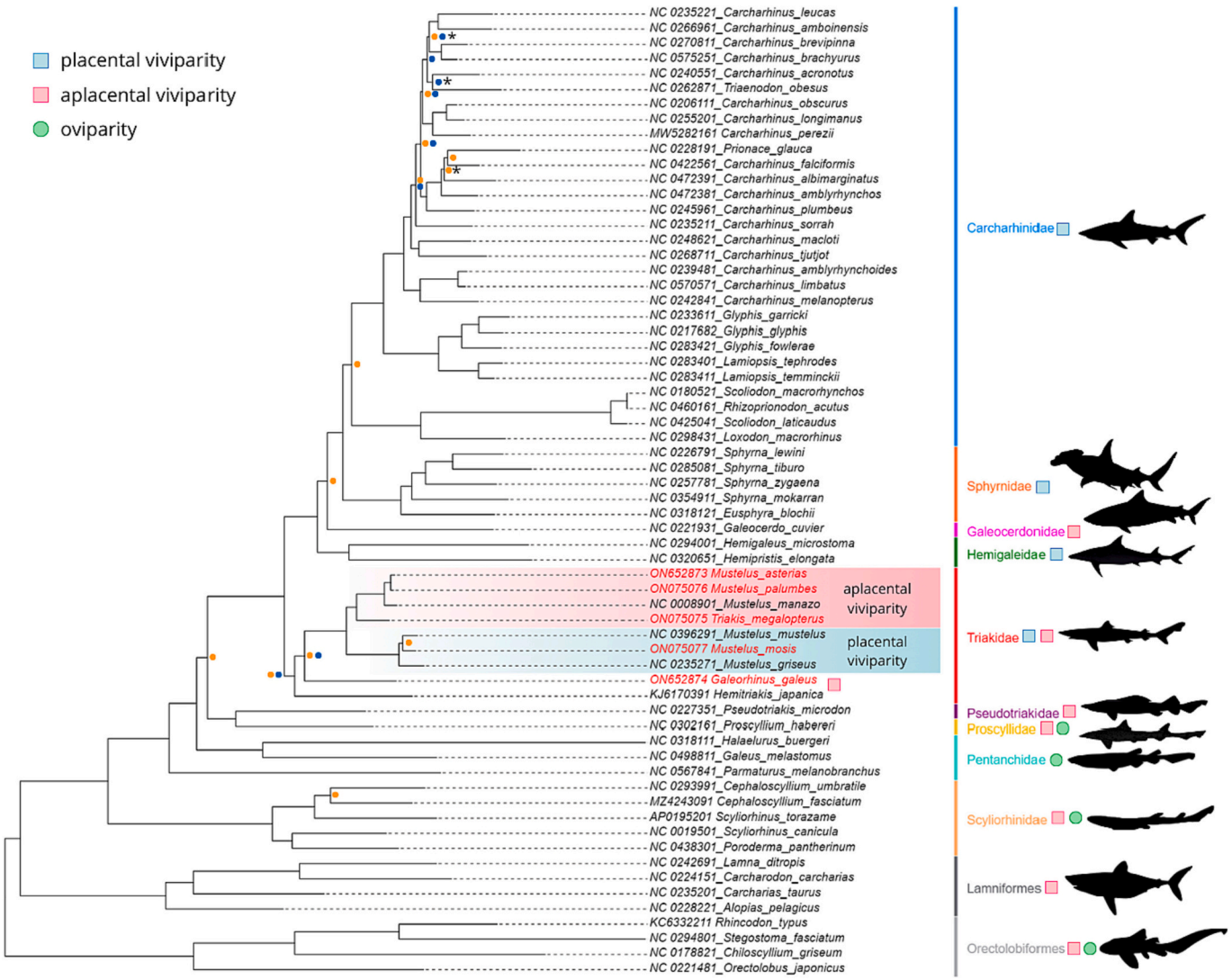

Fig. 5. Maximum Likelihood phylogeny of species in the order Carcharhiniformes (with representative Lamniform and Orectolobiform outgroups) constructed using 13 protein-coding genes and 2 rRNA genes with the PS5 gene partitioning scheme (13 PCGs +2 rRNAs; 15 partitions). Taxa in red text were sequenced in this study. Posterior probabilities and ultrafast bootstrapping UFBoot2 branch support values <95% are indicated at nodes with asterisks and orange dots respectively. SH-aLRT values below 80% are marked with blue dots. Reproductive modes are indicated in the key in the top left corner. Bayesian inference yielded the same topology, except *Triaenodon obesus* clusters with *C. macloti* and *C. tjutjot* and *M. mosis* clusters with *M. griseus* instead of *M. mustelus*. Shark graphics obtained from PhyloPic (Michael Keesey) (Creative Commons). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

trees result in a monophyletic Triakidae family with support values above 95% for all partition schemes, except PS4 and PS8 which both exclude codon position 2 (Supplementary material S1: part 4.2). The *Mustelus* genus clade also clusters into two groups, distinguished from each other by reproductive mode (placental and aplacental viviparity), with a UFBoot2, SH-aLRT, and posterior probability of 100% for all partitioning schemes. Monophyly of the *Mustelus* genus holds only if *Triakis megalopterus* is included, where it falls within the aplacental clade.

The tests used to assess confidence of tree selection for each partitioning scheme yielded non-significant *p*-values ($p > 0.05$) and log-likelihood values are comparable among partitioning schemes (Table 5). PS6 and PS8 had significant posterior weights for the bootstrap proportion using the RELL method (bp-RELL) and the Expected Likelihood Weight (c-ELW), likely attributable to overpartitioning.

Concordance factors, which estimate the topological variation in the branches of phylogenies, demonstrated interesting patterns when comparing the trees from different partitioning schemes (Supplementary material S1, part 4.3). sCF values are fixed for nodes found across the different trees. gCF values are 100 for all nodes of the unpartitioned (PS1 and AA1) topologies and the majority of gCFs are above 75 for codon partitioning schemes (PS2-PS4), except for some nodes in the families Triakidae and Carcharhinidae. gCF values are also generally higher than sCF values for all topologies, apart from some nodes in Carcharhinidae. Concordance values decreased significantly for the geneXcodon topologies (PS6-PS8), but on average gCF values are still larger than sCF values. None of the paraphyletic schemes (PS2 – PS4 and PS6 – PS8) are well supported for the *Mustelus* genus node or the *H. japonica* split, with gCF values of 25% or less and sCF values below 33.3%. Gene partition-based phylogenies (PS5 and AA5) displayed similar trends in *CF* values to codon partitions, however, sCF and gCF values supported monophyly for Triakidae. Notably, gCF is 100% for the node where Triakidae branches from Carcharhinidae + Sphrynidae + Hemigaleidae. Branching of *H. japonica* and *G. galeus* from the rest of the Triakidae taxa is less certain, with gCF values of 30% and 40% and sCF values of 36.9% and 43.4%, respectively. In Carcharhinidae, gCF values are 0% for very short branches across partitioning schemes.

The χ2-test used to determine whether gene and site tree frequencies

Table 5
Comparison of the log-likelihood values to assess the confidence of Maximum Likelihood tree selection using the eight partitioning schemes. Tree topology tests were run using 10,000 RELL replicates as implemented in IQ-Tree v.2.2.0.3. RELL, resampling of estimated log-likelihoods.

Tree	*logL*	deltaL	*bp-RELL*		*p-KH*		*p-SH*		*p-WKH*		*p-WSH*		*c-ELW*		*p-AU*	
PS1	−208,906	0	0.525	+	0.641	+	1	+	0.641	+	0.851	+	0.521	+	0.708	+
PS2	−208,914	7.8506	0.109	+	0.359	+	0.734	+	0.359	+	0.847	+	0.115	+	0.536	+
PS3	−208,914	7.8506	0.112	+	0.359	+	0.734	+	0.359	+	0.825	+	0.115	+	0.535	+
PS4	−208,933	27.449	0.051	+	0.172	+	0.324	+	0.163	+	0.395	+	0.0493	+	0.213	+
PS5	−208,919	13.672	0.121	+	0.275	+	0.605	+	0.275	+	0.624	+	0.117	+	0.378	+
PS6	−208,942	36.242	0.023	−	0.151	+	0.21	+	0.151	+	0.464	+	0.022	−	0.182	+
PS7	−208,939	33.133	0.0399	+	0.16	+	0.235	+	0.16	+	0.43	+	0.0393	+	0.208	+
PS8	−208,942	36.242	0.0194	−	0.151	+	0.21	+	0.151	+	0.476	+	0.022	−	0.182	+

deltaL: logL difference from the maximal logL in the set.
bp-RELL: bootstrap proportion using RELL method [104].
p-KH: p-value of one-sided Kishino-Hasegawa test [105].
p-SH: p-value of Shimodaira-Hasegawa test [81].
p-WKH: p-value of weighted KH test.
p-WSH: p-value of weighted SH test.
c-ELW: Expected Likelihood Weight [106].
p-AU: p-value of approximately unbiased (AU) test [81].
Plus, signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 10,000 resamplings using the RELL method.

supporting the two alternative tree topologies differed significantly, resulted in non-significant p-values ($p > 0.05$) for the most part (Table S6, Supplementary material S1, part 4.5). The majority of significant *p*-values across partitioning strategies are for branches in the genus *Carcharhinus*. However, overall, the hypothesis of equal frequencies cannot be rejected, implying that discordance among gene trees and/or sites is most likely due to incomplete lineage sorting rather than other phenomenon such as gene flow between lineages after separation or unequal rates of evolution of genes among different lineages.

3.5. Gene/species tree analyses

With ILS determined to be the culprit for discordance between genes and/or sites, we could move on to testing for discordance between gene/species trees using the multispecies coalescent model. Analysis of individual gene trees revealed substantial departure of many of the gene trees from the species tree (Supplementary material S1: part 4.6). The consensus NT and AA ASTRAL species trees, constructed from the gene trees, consistently support the deeper clades uncovered in the ML and BI phylogenies (Fig. 6). However, there are differences evident for some of the shallower clades in the family Carcharhinidae. The test for polytomy, gives insignificant p-values for all nodes indicating that, although some branches in *Carcharhinus* are very short, resolution is possible under the MSCM. Triakidae is found to be monophyletic and the clades within the family are concordant with those demonstrated in Fig. 5. A monophyletic Triakidae is also evident for the SVDquartets phylogeny (Supplementary material S1: part 4.7). This coalescent-based tree is consistent with the concatenated phylogenetic trees for deeper clades, but it is the only topology where the *Carcharhinus* genus is found to be monophyletic and *Galeocerdo cuvier* in the Carcharhinidae family, albeit not strongly supported.

3.6. Testing for compositional heterogeneity

The AliGROOVE tests showed very low heterogeneity among lineages in sequence composition for the 13PCG_NT and 13PCG_2rRNA_NT nucleotide datasets and no compositional heterogeneity for the 13PCG_AA amino acid matrix (Fig. 7). Similarly, the BaCoCa tests to detect compositional heterogeneity within taxa, showed very little compositional heterogeneity (Supplementary material S1: part 5.1). The heat maps for ambiguous characters and indels showed no heterogeneity at all and those for purines, pyrimidines, and AT-content showed very low compositional heterogeneity overall, with some of the outgroup taxa displaying moderate compositional bias.

The different models tested in PHYLOBAYES_MPI yielded contrasting convergence statistics and tree topologies (Supplementary material S3: Table S7). Overall, the complex CAT + GTR + G4 model was the best fitting model according to the weighted AIC scores. Furthermore, both the relative difference (*reldiff*) and *maxdiff* between chains well below 0.1 and the *effsizes* above 600 for Datasets 1 and 2. This model recovered triakids as paraphyletic with strong support for all datasets (Supplementary material S1: part 5.2). This was concerning for Dataset 3 because the runs never converged after over 19,500 MCMC and the Tracer plots show poor mixing for some parameters, particularly *statalpha* (Supplementary material S1: part 5.3). This hyperparameter of the CAT model determines the sparsity of the equilibrium frequency profiles across sites. The *reldiff* was also >0.3 and the *effsize* was <100 for some parameters. However, the less complex site heterogenous C60 model was recovered as the next best approximating model and converged, recovering triakids as monophyletic. The site homogenous GTR + G4 model recovered triakids as monophyletic for all datasets.

4. Discussion

4.1. Mitogenome assembly and annotation

Here we provided complete mitochondrial genome assemblies for five Triakidae species, enabling the expansion of the molecular phylogeny of the family. The utilization of an integrated assembly pipeline assisted with the detection of a 714 bp-duplication between $tRNA^{Pro}$ and the control region in the genome of *Galeorhinus galeus,* which failed to assemble with the reference-based approach (Fig. 3). This intergenic spacer appears to be a duplication of a portion of *Cytb* and $tRNA^{Thr}$. As evident from the Triakidae mitogenomes reported here, the gene content and organization of most fish mitogenomes is conserved, however, instances of nonstandard gene patterns have been uncovered [5]. A novel 594 bp non-coding region, located between *Cytb* and $tRNA^{Thr}$ as a duplication remnant of both the former and latter, was recently reported in the mitogenome of *Hemitriakis japonica* [8]. *Scyliorhinus torazame* was the first Carcharhiniform species found to contain a mitogenome duplication of the control region [7]. Additionally, an uncharacterized, non-functional 1060 bp control region remnant between $tRNA^{Thr}$ and $tRNA^{Pro}$ was reported in the Lamniform species, *Mitsukurina owstoni* (EU528659). Wang et al. [8] proposes that the *Cytb-Thr-Pro-D-loop* complex may be a gene rearrangement hotspot in shark mitogenomes. A likely explanation for these mitochondrial mutations is tandem

Fig. 6. Carcharhiniform phylogeny based on a multispecies coalescent model constructed using 13 protein-coding genes and 2 rRNA genes in ASTRAL. Representatives from the orders Lamniformes and Orectolobiformes form the outgroups. Taxa in red text were sequenced in this study. Posterior probabilities are indicated in red, but those above 0.99 are omitted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

duplication and random loss (TDRL), which assumes that some genes are duplicated to form a gene block and subsequent random loss of one of the redundant paralogs leads to the non-coding regions between complete genes [107,108]. Alternative mechanisms invoked to account for mitochondrial rearrangements include inversion [109] transposition [110], and intramolecular recombination [111], however, none of these explain the presence of duplicated pseudogenes, which act as intermediaries in changing gene orders under TDRL.

Based on the TDRL and the intramitochondrial recombination model, we speculate that the rearrangement process in the mitogenome of *G. galeus* is similar to that proposed for *H. japonica* [8]. Firstly, *Cytb* and *tRNA*^{Thr} underwent tandem duplication as a result of slipped-strand mispairing to form a *Cytb-tRNA*^{Thr}*-Cytb'- tRNA*$^{Thr'}$ dimeric block. Secondly, a random deletion occurred resulting in the loss of redundant fragments of *Cytb'* and *tRNA*$^{Thr'}$ and consequently a loss in function of these genes. Lastly, the remaining remnants of the duplication recombined into the *D-loop*.

Structural analysis of *tRNA*$^{Thr'}$ calculated using thermodynamic model parameters suggests that it likely does not fold into a functional molecule (Fig. 4), even though the algorithmic model (tRNAscan-SE) classifies the pseudogene as a functional tRNA. Previous studies investigating mitochondrial tRNAs have demonstrated that a low level of primary sequence conservation and exhibition of unusual secondary structures is not uncommon [102]. Furthermore, although partial gene region duplications are usually followed by rapid deletion of redundant material which may result in gene order rearrangement, some stabilized

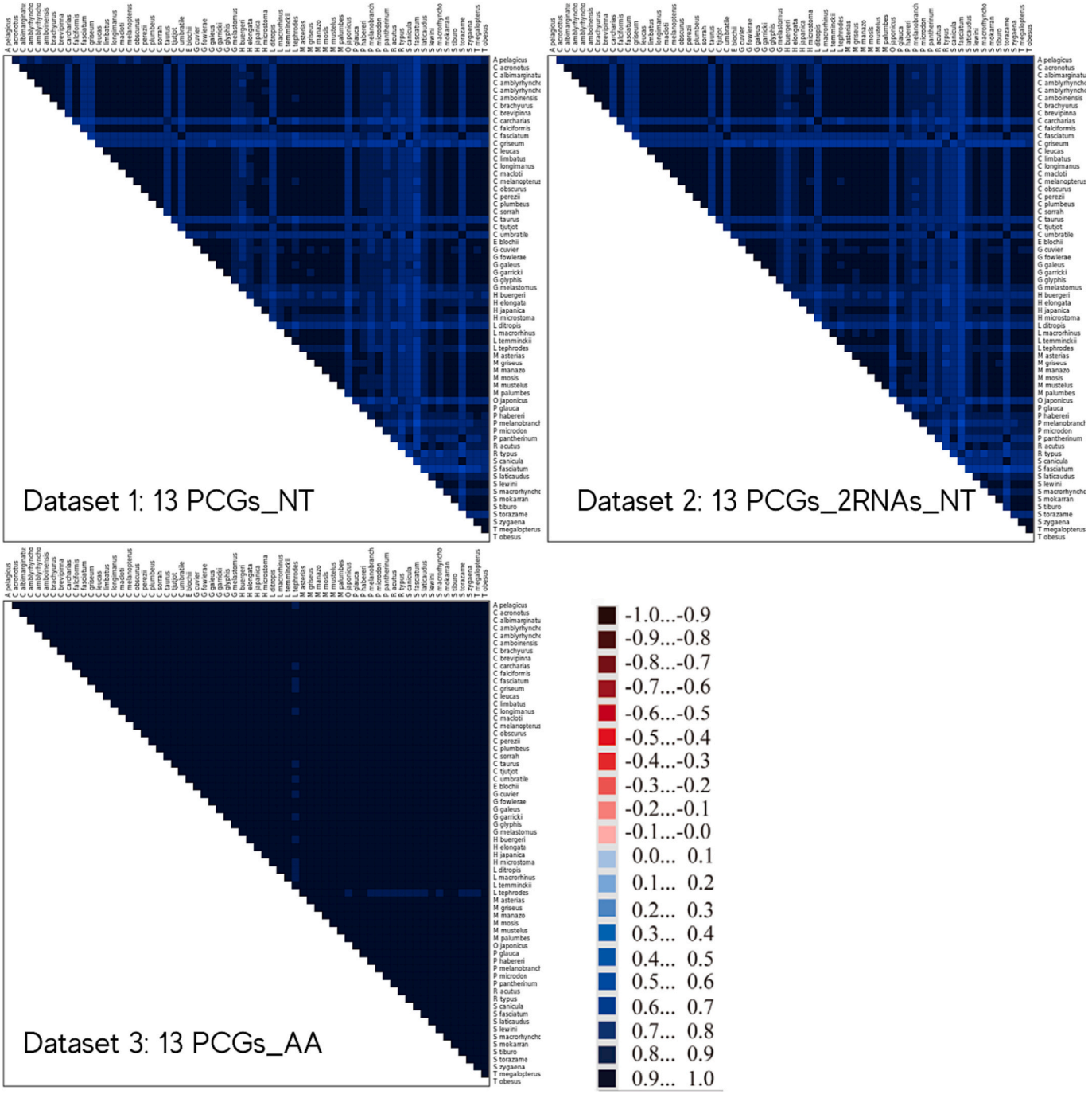

Fig. 7. Base compositional heterogeneity of mitogenome sequences for our three nucleotide and amino acid datasets. The scores range from −1 indicating full random similarity due to descent between pairwise sequence comparisons (dark blue), to +1 indicating non-random similarity (bright orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

duplications have been recorded. For example, Bakke et al. [112] reported a $tRNA^{Thr}$-$tRNA^{Pro}$ intergenic spacer shared by two groups of cod fishes, the common ancestor of which originated 40 million years ago (MYA). It is suggested that positive selection has maintained these non-coding regions due to retention of function as an alternative site of initiation of replication or transcription of adjacent genes [113] or heavy strand transcript termination [112]. The duplication remnant of $tRNA^{Thr}$ in *H. japonica* is located in front of the functional $tRNA^{Thr}$, however it is significantly more degraded and not detected as a functional tRNA by any of the prediction models. Mitogenome sequencing data from the closest relative of *G. galeus*, *Hypogaleus hyugaensis* may be useful in revealing the evolutionary pathway of this duplication. If it occurred in a common ancestor and has been retained in the two species, this may suggest that it has maintained its function. Nevertheless, further investigation of this control region duplication hotspot is required to understand the physiological implications of $tRNA^{Thr}$ in *G. galeus*, if any.

Although, intraspecific comparison of the Triakidae mitogenomes did not reveal additional distinct gene patterns, enrichment of the control region with tandem repeat sequences in *G. galeus, M. asterias, M. manazo, M. palumbes* and *T. megalopterus* (Table S3), confirms that this mitogenomic region exhibits extensive nucleotide and size polymorphism as demonstrated in several elasmobranchs [114,115,116]. Future studies should consider a more thorough characterization of the

control region, which has not yet been extensively analysed in Triakidae species.

4.2. Phylogenetic reconstruction

Equipped with five newly assembled houndshark mitogenomes, we conducted a robust phylogenomic analysis to determine the higher-level relationships within the Carcharhiniform order, with a focus on the problematic Triakidae family. We tested two hypotheses, described in Rota et al. [46], to uncover the root of mitochondrial phylogenomic discordance present in Triakidae for different partitioning strategies. We had to confirm that poor phylogenetic informativity in our dataset was not the instigator of phylogenomic discordance under Hypothesis 1 before we could test for biological signatures causing incongruencies at shallower nodes under Hypothesis 2.

Hypothesis H1. Different partitioning strategies result in different topologies due to a lack of or varying levels of phylogenetic signal in the respective partition datasets.

A review of the five mitophylogenomic studies on Carcharhiniformes (Table 1) revealed robust phylogenomic signals in higher-level relationships and our phylogenomic analyses uncovered the same major clades independent of the partitioning strategy used. On the level of orders, the ((Lamniformes, Carcharhiniformes) Orectolobiformes) association is consistent across studies. On a family level, there is strong support for: (1) Carcharhinidae, (2) Sphyrnidae, (3) Galeocerdonidae, (4) Hemigaleidae, (5) Triakidae, (6) Scyliorhinidae. These clades span across the Carcharhiniform order and are found across our nucleotide and amino acid tree topologies thus, we can conclude that there is sufficient phylogenetic signal in our dataset to resolve higher-level lineages. Consistency in branch length and high bootstrap support and posterior probability values for family level lineages across partitioning strategies and studies is indicative of high phylogenetic informativity [117,118]. Furthermore, all gene and codon positions showed low substitution saturation confirming that convergently evolved character states were not obscuring phylogenetic signal and underestimating branch lengths. Accordingly, we can reject Hypothesis 1 and conclude that a lack of phylogenetic signal is not the cause of incongruencies seen in the mitophylogenomic reconstructions of Carcharhiniformes.

Hypothesis H2. The lack of phylogenetic concordance is caused by gene tree/species tree issues, which means that the evolutionary history of genes differs from that of the species.

Although our dataset displays overall phylogenetic informativeness, it is still possible that subsets of the tree suffer from low phylogenetic signal. More recently diverged lineages within families display contrasting phylogenetic relationships among datasets and partitioning schemes and cannot be resolved with strong support for any of our partitioning strategies. The inferred internal branches within Carcharhinidae and Triakidae, shown to present contentious intrafamilial relationships (Table 1), are visibly shorter than those in other families, suggesting that these lineages may have experienced recent rapid radiations (Fig. 5, Supplementary material S1: part 4) [119,120]. Difficulties resolving these relationships is likely the consequence of incomplete lineage sorting, where not enough time has passed to confidently resolve some of these nodes because the standing genetic variation of the ancestral species inherited by the immediate descendant species has not been sorted out prior to speciation [121,122]. This is where over-partitioning can reduce the number of informative sites to group these taxa correctly [34].

Data partitioning has a profound impact on the phylogenetic patterns uncovered, particularly for recently diverged lineages [22,23,24,19,25]. This is demonstrated by the contrasting monophyletic *versus* paraphyletic topologies for Triakidae stemming from previous studies, which utilized very different partitioning approaches (Table 1), as well as our own study where there were noticeable discrepancies among the topologies constructed using different *a priori* partitioning schemes. Wang et al. [8] utilized a codon-informed *a priori* partitioning strategy which may have led to overparameterization and bootstrap inflation without accounting for the different evolutionary rates of the different genes in the mitogenome. Kousteni et al. [15] partitioned their dataset according to genes, resulting in lower support values, but accounting for gene evolution differences. In this study, the best supported partitioning schemes, PS4 (codon: pos1 + pos3 + 2rRNAs) and PS8 (geneXcodon: 13 PCGs pos1 + pos3 + 2rRNAs), excluded codon position 2 on the premise that it did not improve the informativity of the phylogenetic analysis as informed by the substitution saturation results (Table S1; Supplementary material S1: part 1). Examination of the AICc and BIC formulas, reveals that the fewer features (k) present, the lower the scores will be, suggesting that the outcome for PS4 and PS8 may not be a true reflection of phylogenetic robustness [123,124]. Moreover, although these two *a priori* strategies do account for evolutionary differences in genes, they divide the dataset up extensively which may reduce the signal available to resolve lineages experiencing ILS.

Although statistically guided selection of the best-fitting partitioning scheme and models of evolution in ModelFinder improves coherence between the data, models, and phylogenetic outcome [22,23,24,19,25], information criterion values are not always reliable indicators for selection of the best *a priori* partitioning scheme when ILS shapes the resolution of some subsets of the tree. Furthermore, it must be acknowledged that the process is a statistical estimation of the best-fitting models of evolution and thus may simplify the "true" evolutionary processes and/or fail to account for variations in historical relationships and contemporary forces between orders, families, genera, species, and even genes that influence molecular evolution on a finer scale [125,126,127]. This is particularly true for the diverse Carcharhiniformes which have had ~150 million years to accumulate complex evolutionary patterns and demonstrate a vast range of life history traits and dispersal habits [43,128].

Previous studies demonstrate that concordance factors are generally lower than bootstrap support values, particularly for large datasets where sampling variance is low [129,130,83,46,131], but these two types of values also responded differently to the choice of partitioning strategy for our dataset. In the geneXcodon partitioned phylogenies (PS6-PS8), bootstrap values were high for shallower lineages whereas gene concordance factors decreased significantly. Bootstrap values increase with the addition of data blocks and are thus sensitive to overparameterization which means that well-supported nodes in geneXcodon-based phylogenies were likely victims of inflated support values (Supplementary material S1, part 4) [34,33,35]. This means that, regardless of the suitability of a given partitioning strategy for the dataset, resampling may always return the same tree (*i.e.*, 100% bootstrap support), even though incomplete lineage sorting or other processes that lead to genealogical discordance are at work [83]. Concordance factors can provide a more comprehensive measure of support by considering the congruence of multiple independent partitions, more accurately capturing the heterogeneity of this mitogenomic dataset. Our *CF* values were higher for deeper lineages than those from previous studies [129,83,46,131], where short gene length, and therefore lack of sufficient phylogenetic signal, was described as the culprit for the lower values [83,46]. This further strengthens our decision to reject H1 and conclude that a lack of phylogenetic signal is not responsible for phylogenetic disagreement in this mitogenomic dataset.

To uncover possible causes for topological conflict around certain branches of the species tree, comparisons were made between gene and site concordance factors. Overall, gCF values were greater than sCF values, which is concordant with our finding that there is sufficient phylogenetic signal in our dataset to resolve deeper lineages. However, in Carcharhinidae and Triakidae, there are nodes where gCF is lower than sCF and for the shortest branches in *Carcharhinus*, gCF is 0 (Supplementary material S1, part 4.2). This indicates that no single gene tree contains the branch present in the species reference tree [80]. It is

possible that extensive ILS in Carcharhinidae lineages has led to unresolved nodes in the species tree [132]. Reconstructions of old radiations demonstrating explosive adaptive radiations in various tetrapod groups [133,134,135,136,137,138], and cichlid fish [119] contain one or more nodes that defy resolution, and the same may be at play for *Carcharhinus* lineages. This mirrors divergence estimates which suggest that there are increased speciation rates in coral reef-associated lineages within Carcharhinidae [139]. Seeing that phylogenetic signal does not seem to be a problem in our dataset, the low gCF values for these branches suggest that there is strong discordance between genes [83].

Furthermore, the gene and site tree frequencies supporting the two alternative topologies of gene tree quartets had comparable frequencies (Table S6), implying that discordance between gene trees and/or sites can be ascribed to ILS rather than phenomena such as gene flow between lineages after separation or unequal rates of evolution of genes among different lineages [84,85]. Accordingly, we moved onto assembling phylogenies based on the multispecies coalescent model. The input trees constructed for ASTRAL (Supplementary material S1, part 4.6) demonstrate the significant variability that exists between individual gene trees as a result of ILS. The species tree is identical to the concatenated ML tree constructed using PS5 and to the amino acid phylogenies (AA1 and AA5) for all families, however variations exist within the Carcharhinidae family. Although all nodes failed the polytomy test in ASTRAL, it appears that certain nodes in the Carcharhinidae family are defying phylogenomic resolution, even in the face of an increased volume of data from whole mitogenome sequencing [119]. As a result, our study cannot confidently resolve the relationships among Carcharhinidae considering that the topology varied between ML, BI, ASTRAL and SVDquartet reconstructions. However, we can conclude from the MSCM topologies that Triakidae is monophyletic with good support.

Hypothesis H3. Violations of the assumptions underlying the models applied during phylogenetic reconstruction are causing discordance.

To test whether discordance in recently diverged lineages of Carcharhiniforms for different datasets and studies was caused by compositional heterogeneity across lineages and/or among sites, we applied the site heterogenous CAT + GTR + G4, CAT + Poisson + G4 and C60 + GTR + G4 (amino acid dataset) models in PHYLOBAYES_MPI. Despite triakids being recovered as paraphyletic across all datasets with strong support in the site heterogenous analyses, we observed consistently high effective sample size values for the entropy of the relative exchange rates (*rrent*) and posterior means of the exchange abilities parameter (*rrmeans*) across chains and datasets (Table S7). Higher values for a specific pair of character states indicate a higher rate of substitution between those states implying that solely the GTR component of the CAT + GTR + G4 model was reliably converging across datasets [140], particularly in respect to Datasets 2 and 3. Furthermore, the failure of the amino acid dataset (Dataset 3) to converge after over 19,500 MCMC, suggests that the chosen model may not accurately represent the underlying evolutionary processes in the amino acid dataset. Many studies that have applied site-heterogeneous models have experienced difficulties in achieving acceptable metrics of convergence, even after running the chains for years [40,41,141,37].

The CAT + GTR + G4 model has been cited to supress long-branch attraction and address phylogenetic issues related to site heterogeneity better than site homogenous concatenation techniques [142,143]. However, when the CAT model is applied to a dataset without compositional heterogeneity, other information may be captured because CAT models are theoretically infinitely complex and may overestimate the number of substitutional categories in the dataset [140,37]. AliGROOVE detected almost negligible levels of heterogeneity across taxa which indicates that long branch attraction is not a problem in this dataset [118,141]. Furthermore, BaCoCa detected minimal taxon-specific base composition bias which means that taxa possess the same proportion of nucleotide bases or amino acids across the dataset, even though these may occur at different positions through time [95]. Consequently, our results indicate that the presence of compositional heterogeneity is unlikely to be in violation of the assumptions of the models applied to our datasets or those of previous studies and it is not the cause of the phylogenetic discordance observed in Triakidae.

4.2.1. *Taxonomic implications for the genus Mustelus*

The phylogenetic outcome of all partitioning schemes supports previous claims that the *Mustelus* genus clusters into two discrete clades representing different modes of reproduction (placental *versus* aplacental species) and should be expanded to incorporate *Triakis megalopterus* (Figs. 4 and 5) [144,145,146,147]. The patterns of variability in the modes of reproduction in Triakidae, suggest that the species belonging to the family are relics of an intermediate phase between the basal Carcharhiniformes (Pseudotriakidae, Proscyllidae, Pentanchidae, Scyliorhinidae; oviparity and aplacental viviparity) and "higher carcharhinids" (Carcharcharhinidae, Sphyrnidae, Hemigaleidae, and Triakidae; placental viviparity) [43]. Members of Triakidae display both placental viviparity and aplacental (limited-histotroph and yolk-sac) viviparity. It is suggested placental viviparity, the ancestral condition, was lost independently in the asterias clade of the mustelids and in the Galeorhinidae/Hemitriakidae lineages giving rise to aplacental gestation which mirrors the reproductive mode of basal Carcharhiniformes [144].

Our phylogenetic hypothesis has yielded new and valuable information regarding the coevolution of white spots and aplacental reproduction in the expanded *Mustelus* clade. Specifically, we have shown that *T. megalopterus*, despite belonging to the aplacental clade, has distinctive black spots on its dorsal surface. Therefore, given the basal placement of *T. megalopterus* in the aplacental clade, it is more likely to postulate that the emergence of white spots occurred at a later stage in the evolutionary trajectory of aplacental reproduction mode within the clade. However, our findings suggest that the previous hypotheses regarding the inference of the reproductive mode of *Mustelus* species, especially those without comprehensive reproductive investigations, based solely on the presence or absence of white spots [148,145], should be approached with caution. Additionally, the inferred phylogenetic positioning of *T. megalopterus* aligns with the previous hypothesis that the occurrence of black spots has independently emerged multiple times within the genus [148,145]. Nevertheless, it is necessary to perform additional phylogenomic research that encompasses more representative members from the genus *Triakis* and includes the non-spotted, but aplacental, *Schylliogaleus quecketti* with a sister-group relationship to *T. megaloterus* [145,149]. Furthermore, molecular dating and ancestral state reconstructions should be carried out in order to obtain more conclusive results.

Clustering of the placental mustelids showed some variation between topological reconstruction methods. For example, in PS2-PS4 topologies where *M. mosis* clusters with *M. griseus*, the sCF value is 28.9% compared to the 47.8% when it clusters with *M. mustelus* in the other trees. This could indicate that the branch is in a parameter space in which high levels of ILS may be misleading branch resolution for those partitioning schemes [150]. Alternatively, this branch may be unduly affected by a small number of influential sites that have extreme differences in likelihood between different resolutions of that split [151]. This effect may be amplified for partitioning schemes that do not account for gene boundaries [33,21]. Both ASTRAL and SVDquartets agree on a ((*M. mustelus, M. mosis*) *M. griseus*) grouping for the placental clade.

5. Conclusions

This study has provided five newly assembled houndshark mitochondrial genomes and demonstrated the importance of using a thorough assembly approach, incorporating *de novo* methods to ensure that contigs do not condense due to reference-bias. This has allowed us to detect a 714 bp-duplication in the control region of *Galeorhinus galeus*. We have also shed light on the effects of data partitioning in a Maximum

Likelihood and Bayesian Inference framework in the context of the mitochondrial phylogenomics of Carcharhiniformes and strongly recommend that future studies intending to clarify complexities in elasmobranch evolutionary relationships, carefully consider its impact on phylogenetic reconstruction. Furthermore, we have unpacked the root of the phylogenomic discordance present in Carcharhiniformes. The consistent phylogenetic patterns for higher-level lineages among partitioning schemes for our dataset and previous Carcharhiniform studies confirms that sufficient signal exists to resolve higher-level lineages. This is reiterated by the relatively high gene concordance values found across partitioning schemes. Furthermore, we demonstrated that compositional heterogeneity across lineages and/or sites is not the cause of phylogenomic discordance between our datasets and those of other studies and the use of the CAT + GTR + G4 model in this instance may overestimate the number of substitutional categories in a given dataset. The lack of phylogenetic concordance in shallower lineages can therefore be ascribed to gene tree conflicts affecting species tree inference. Accordingly, we reiterate the value of multispecies coalescent model analyses for resolution of phylogenies affected by incomplete lineage sorting. Furthermore, it is necessary to include gene boundaries in *a priori* partitioning strategies under an ML and BI framework to account for gene tree differences in this dataset. While site heterogenous models in PHYLOBAYES_MPI may be effective in handling site heterogeneity and mitigating long branch attraction, they do not consider gene boundaries, and it is therefore not regarded as an appropriate approach for addressing incomplete lineage sorting. MSCM topologies confirmed the monophyly of Triakidae as well as the separation of the *Mustelus* genus into a placental and aplacental clade containing *Triakis megalopterus*. However, they could not confidently resolve the topology of Carcharhinidae, indicating that the family contains clades that have likely undergone recent rapid radiations leading to high levels of incomplete lineage sorting. Expanding the study of the Carcharhiniform phylogeny to include nuclear genome data would be valuable, particularly to further understand the relationships within Triakidae and Carcharhinidae.

The phylogenomic field is expanding exponentially, with genomic data accessibility yielding significant improvements in signal and accuracy, however, analyses pipelines across studies still require some fine-tuning. This study presents a rigorous approach to tackling difficult-to-resolve phylogenies, accounting for factors that tend to produce erroneous and/or ill-informed inferences. We highlight the importance of thoroughly analyzing the evolutionary patterns present in the dataset to make an informed selection of appropriate phylogenomic methods. There is certainly room for expansion, but our hope is to see future studies in this space making use of carefully considered phylogenetic exploration. The addition of key species to our mitogenomic dataset will likely improve the resolution confidence for more recently diverged lineages.

Author contributions

J.C.W., S.N.M. and A.E.B. conceived of the study. J.C.W. and S.N.M wrote the manuscript with input from all other authors. J.C.W. conducted the wet laboratory work and conducted all bioinformatics and phylogenomic analyses with help from S.N.M. A.E.B. acquired funding for the study. Supervision and validation were undertaken by A.E.B. and S.N.M.

CRediT authorship contribution statement

Jessica C. Winn: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Simo N. Maduna:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Aletta E. Bester-van der Merwe:** Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The houndshark mitogenomes generated and analysed during the current study were deposited on the GenBank repository (GenBank accession numbers: ON075075, ON075076, ON075077, ON652873, and ON652874). Houndshark sketches in Fig. 3 were provided by the Food and Agriculture Organization. Galeomorphi graphics in Fig. 5 were downloaded from PhyloPic (Michael Keesey) (Creative Commons License). The Ion Torrent® next-generation sequencing data is presented in the research article "Ion Torrent data for the genome assembly and phylogenomic placement of mitochondrial genomes with a focus on houndsharks (Chondrichthyes: Triakidae)" in *Data in Brief*.

Acknowledgements

The authors wish to thank the following individuals, organisations, and institutions for providing biological samples: the South African Department of Forestry, Fisheries and Environment (DFFE), the Reel Science Coalition, Dr. Edward D. Farrell (University College Dublin; Killybegs Fishermen's Organization), Dr. Mikhail V. Chesalin (A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Russian Federation). We also extend our gratitude to Dr. Julianna Klein for partly assisting with DNA extractions of samples used in our study and to the Central Analytical Facility at Stellenbosch University for conducting the library preparation and Ion Torrent sequencing of all specimens. This study was funded by the National Research Foundation of South Africa.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110771.

References

[1] J.A. Baeza, F.J. García-De León, Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcharhinus falciformis*, BMC Genomics 23 (2022) 320, https://doi.org/10.1186/s12864-022-08482-z.

[2] J. Gretzinger, M. Molak, E. Reiter, S. Pfrengle, C. Urban, J. Neukamm, M. Blant, N.J. Conard, C. Cupillard, V. Dimitrijević, D.G. Drucker, E. Hofman-Kamińska, R. Kowalczyk, M.T. Krajcarz, M. Krajcarz, S.C. Münzel, M. Peresani, M. Romandini, I. Rufí, J. Soler, G. Terlato, J. Krause, H. Bocherens, V. J. Schuenemann, Large-scale mitogenomic analysis of the phylogeography of the late Pleistocene cave bear, Sci. Rep. 9 (2019) 10700, https://doi.org/10.1038/s41598-019-47073-z.

[3] L. Kistler, A. Ratan, L.R. Godfrey, B.E. Crowley, C.E. Hughes, R. Lei, Y. Cui, M. L. Wood, K.M. Muldoon, H. Andriamialison, J.J. McGraw, L.P. Tomsho, S. C. Schuster, W. Miller, E.E. Louis, A.D. Yoder, R.S. Malhi, G.H. Perry, Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs, J. Hum. Evol. 79 (2015) 45–54, https://doi.org/10.1016/j.jhevol.2014.06.016.

[4] Y. Wang, X. Liu, I.J. Garzón-Orduña, S.L. Winterton, Y. Yan, U. Aspöck, H. Aspöck, D. Yang, Mitochondrial phylogenomics illuminates the evolutionary history of Neuropterida, Cladistics 33 (2017) 617–636, https://doi.org/10.1111/cla.12186.

[5] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (1999) 1767–1780, https://doi.org/10.1093/nar/27.8.1767.

[6] D. Jameson, OGRe: a relational database for comparative analysis of mitochondrial genomes, Nucleic Acids Res. 31 (2003) 202–206, https://doi.org/10.1093/nar/gkg077.

[7] Y. Hara, K. Yamaguchi, K. Onimaru, M. Kadota, M. Koyanagi, S.D. Keeley, K. Tatsumi, K. Tanaka, F. Motone, Y. Kageyama, R. Nozu, N. Adachi, O. Nishimura, R. Nakagawa, C. Tanegashima, I. Kiyatake, R. Matsumoto, K. Murakumo, K. Nishida, A. Terakita, S. Kuratani, K. Sato, S. Hyodo, S. Kuraku, Shark genomes provide insights into elasmobranch evolution and the origin of

vertebrates, Nat. Ecol. Evol. 2 (2018) 1761–1771, https://doi.org/10.1038/s41559-018-0673-5.
[8] C. Wang, T. Lai, P. Ye, Y. Yan, P. Feutry, B. He, Z. Huang, T. Zhu, J. Wang, X. Chen, Novel duplication remnant in the first complete mitogenome of *Hemitriakis japanica* and the unique phylogenetic position of family Triakidae, Gene 820 (2022), 146232, https://doi.org/10.1016/j.gene.2022.146232.
[9] J.C. Avise, The history and purview of phylogeography: a personal reflection, Mol. Ecol. 7 (1998) 371–379, https://doi.org/10.1046/j.1365-294x.1998.00391.x.
[10] W.S. Moore, Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear-gene trees, Evolution 49 (1995) 718–726, https://doi.org/10.1111/j.1558-5646.1995.tb02308.x.
[11] D.R. Wolstenholme, Animal mitochondrial DNA: structure and evolution, in: International Review of Cytology, Elsevier, 1992, pp. 173–216, https://doi.org/10.1016/S0074-7696(08)62066-5.
[12] D.B. da Cunha, L.F. Rodrigues-Filho, J.B. de Sales, A review of the mitogenomic phylogeny of the Chondrichthyes, in: L.F. da Filho, J.B.L. de Sales (Eds.), Chondrichthyes - Multidisciplinary Approach, InTech, 2017, https://doi.org/10.5772/intechopen.70028.
[13] P. Díaz-Jaimes, N.J. Bayona-Vásquez, D.H. Adams, M. Uribe-Alcocer, Complete mitochondrial DNA genome of bonnethead shark, *Sphyrna tiburo*, and phylogenetic relationships among main superorders of modern elasmobranchs, Meta Gene 7 (2016) 48–55, https://doi.org/10.1016/j.mgene.2015.11.005.
[14] S. Duchêne, F.I. Archer, J. Vilstrup, S. Caballero, P.A. Morin, Mitogenome Phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation, PLoS One 6 (2011), e27138, https://doi.org/10.1371/journal.pone.0027138.
[15] V. Kousteni, S. Mazzoleni, K. Vasileiadou, M. Rovatsos, Complete mitochondrial DNA genome of nine species of sharks and rays and their phylogenetic placement among modern elasmobranchs, Genes 12 (2021) 324, https://doi.org/10.3390/genes12030324.
[16] M. Miya, M. Nishida, Use of Mitogenomic information in Teleostean molecular Phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion, Mol. Phylogenet. Evol. 17 (2000) 437–455, https://doi.org/10.1006/mpev.2000.0839.
[17] R. Zardoya, A. Meyer, Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene, Proc. Natl. Acad. Sci. 93 (1996) 5449–5454, https://doi.org/10.1073/pnas.93.11.5449.
[18] S.L. Cameron, C.L. Lambkin, S.C. Barker, M.F. Whiting, A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision, Syst. Entomol. 32 (2007) 40–59, https://doi.org/10.1111/j.1365-3113.2006.00355.x.
[19] D. Kainer, R. Lanfear, The effects of partitioning on phylogenetic inference, Mol. Biol. Evol. 32 (2015) 1611–1627, https://doi.org/10.1093/molbev/msv026.
[20] K.M. Kjer, R.L. Honeycutt, Site specific rates of mitochondrial genomes and the phylogeny of eutheria, BMC Evol. Biol. 7 (2007) 8, https://doi.org/10.1186/1471-2148-7-8.
[21] A.F.L.A. Powell, F.K. Barker, S.M. Lanyon, Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study, Mol. Phylogenet. Evol. 66 (2013) 69–79, https://doi.org/10.1016/j.ympev.2012.09.006.
[22] J.M. Brown, A.R. Lemmon, The importance of data partitioning and the utility of Bayes factors in Bayesian Phylogenetics, Syst. Biol. 56 (2007) 643–655, https://doi.org/10.1080/10635150701546249.
[23] T.R. Buckley, C. Simon, G.K. Chambers, Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support, Syst. Biol. 50 (2001) 67–86.
[24] P.B. Frandsen, B. Calcott, C. Mayer, R. Lanfear, Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates, BMC Evol. Biol. 15 (2015) 13, https://doi.org/10.1186/s12862-015-0283-7.
[25] R. Lanfear, B. Calcott, D. Kainer, C. Mayer, A. Stamatakis, Selecting optimal partitioning schemes for phylogenomic datasets, BMC Evol. Biol. 14 (2014) 82, https://doi.org/10.1186/1471-2148-14-82.
[26] A.R. Lemmon, E.C. Moriarty, The importance of proper model assumption in Bayesian Phylogenetics, Syst. Biol. 53 (2004) 265–277, https://doi.org/10.1080/10635150490423520.
[27] C. Blair, R.W. Murphy, Recent trends in molecular phylogenetic analysis: Where to next? J. Hered. 102 (2011) 130–138, https://doi.org/10.1093/jhered/esq092.
[28] R. Lanfear, B. Calcott, S.Y.W. Ho, S. Guindon, PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701, https://doi.org/10.1093/molbev/mss020.
[29] R. Lanfear, P.B. Frandsen, A.M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses, Mol. Biol. Evol. msw260 (2016), https://doi.org/10.1093/molbev/msw260.
[30] S.Y.W. Ho, R. Lanfear, Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks, Mitochondrial DNA 21 (2010) 138–146, https://doi.org/10.3109/19401736.2010.494727.
[31] R.R. Bouckaert, A.J. Drummond, bModelTest: Bayesian phylogenetic site model averaging and model comparison, BMC Evol. Biol. 17 (2017) 42, https://doi.org/10.1186/s12862-017-0890-6.
[32] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589, https://doi.org/10.1038/nmeth.4285.
[33] C. Li, G. Lu, G. Ortí, Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci, Syst. Biol. 57 (2008) 519–539, https://doi.org/10.1080/10635150802206883.
[34] J.R. Leavitt, K.D. Hiatt, M.F. Whiting, H. Song, Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: a phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study, Mol. Phylogenet. Evol. 67 (2013) 494–508, https://doi.org/10.1016/j.ympev.2013.02.019.
[35] J. Sullivan, P. Joyce, Model selection in phylogenetics, Annu. Rev. Ecol. Evol. Syst. 36 (2005) 445–466, https://doi.org/10.1146/annurev.ecolsys.36.102003.152633.
[36] N. Lartillot, H. Philippe, A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process, Mol. Biol. Evol. 21 (2004) 1095–1109, https://doi.org/10.1093/molbev/msh112.
[37] N.V. Whelan, K.M. Halanych, Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses, Syst. Biol. syw084 (2017), https://doi.org/10.1093/sysbio/syw084.
[38] C.E. Laumer, R. Fernández, S. Lemer, D. Combosch, K.M. Kocot, A. Riesgo, S.C. S. Andrade, W. Sterrer, M.V. Sørensen, G. Giribet, Revisiting metazoan phylogeny with genomic sampling of all phyla, Proc. R. Soc. B Biol. Sci. 286 (2019) 20190831, https://doi.org/10.1098/rspb.2019.0831.
[39] T.H. Struck, C. Paul, N. Hill, S. Hartmann, C. Hösel, M. Kube, B. Lieb, A. Meyer, R. Tiedemann, G. Purschke, C. Bleidorn, Phylogenomic analyses unravel annelid evolution, Nature 471 (2011) 95–98, https://doi.org/10.1038/nature09864.
[40] M.L. Borowiec, E.K. Lee, J.C. Chiu, D.C. Plachetzki, Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa, BMC Genomics 16 (2015) 987, https://doi.org/10.1186/s12864-015-2146-4.
[41] T. Nosenko, F. Schreiber, M. Adamska, M. Adamski, M. Eitel, J. Hammel, M. Maldonado, W.E.G. Müller, M. Nickel, B. Schierwater, J. Vacelet, M. Wiens, G. Wörheide, Deep metazoan phylogeny: when different genes tell different stories, Mol. Phylogenet. Evol. 67 (2013) 223–233, https://doi.org/10.1016/j.ympev.2013.01.010.
[42] C.R.L. Amaral, F. Pereira, D.A. Silva, A. Amorim, E.F. de Carvalho, The mitogenomic phylogeny of the Elasmobranchii (Chondrichthyes), Mitochondrial DNA Part A 29 (2018) 867–878, https://doi.org/10.1080/24701394.2017.1376052.
[43] G.J.P. Naylor, L.J.V. Compagno, Sharks of the Order Carcharhiniformes, Systematic Zoology 38 (1989) 292, https://doi.org/10.2307/2992292.
[44] D.A. Ebert, M. Dando, S. Fowler, Sharks of the World: A Complete Guide, Princeton University Press, 2021, https://doi.org/10.1515/9780691210872.
[45] Versuch einer Eintheilung der Fische nach den Ziihnen, Mag Neuste Phys Naturgesch 6 (1790) 28–38.
[46] J. Rota, V. Twort, A. Chiocchio, C. Peña, C.W. Wheat, L. Kaila, N. Wahlberg, The unresolved phylogenomic tree of butterflies and moths (Lepidoptera): assessing the potential causes and consequences, Syst. Entomol. 47 (2022) 531–550, https://doi.org/10.1111/syen.12545.
[47] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791, https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.
[48] L.S. Jermiin, V. Jayaswal, F.M. Ababneh, J. Robinson, Identifying optimal models of evolution, in: J.M. Keith (Ed.), Bioinformatics, Methods in Molecular Biology, Springer, New York, New York, NY, 2017, pp. 379–420, https://doi.org/10.1007/978-1-4939-6622-6_15.
[49] S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach, Mol. Biol. Evol. 18 (2001) 691–699, https://doi.org/10.1093/oxfordjournals.molbev.a003851.
[50] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 2001.
[51] A. Lloret-Villas, M. Bhati, N.K. Kadri, R. Fries, H. Pausch, Investigating the impact of reference assembly choice on genomic analyses in a cattle breed, BMC Genomics 22 (2021) 363, https://doi.org/10.1186/s12864-021-07554-w.
[52] K.L. Hull, S.N. Maduna, A.E. Bester-van der Merwe, Characterization of the complete mitochondrial genome of the common smoothhound shark, *Mustelus mustelus* (Carcharhiniformes: Triakidae), Mitochondr. DNA Part B 3 (2018) 962–963, https://doi.org/10.1080/23802359.2018.1507642.
[53] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647–1649, https://doi.org/10.1093/bioinformatics/bts199.
[54] A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes De Novo assembler, Curr. Protoc. Bioinformatics 70 (2020), https://doi.org/10.1002/cpbi.102.
[55] W. Iwasaki, T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, M. Nishida, MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline, Mol. Biol. Evol. 30 (2013) 2531–2540, https://doi.org/10.1093/molbev/mst141.
[56] Y. Sato, M. Miya, T. Fukunaga, T. Sado, W. Iwasaki, MitoFish and MiFish pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA Metabarcoding, Mol. Biol. Evol. 35 (2018) 1553–1555, https://doi.org/10.1093/molbev/msy074.
[57] P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences, Biotechniques 28 (2000) 1102–1104.

[58] E.W. Sayers, M. Cavanaugh, K. Clark, J. Ostell, K.D. Pruitt, I. Karsch-Mizrachi, GenBank, Nucleic Acids Res. gkz956 (2019), https://doi.org/10.1093/nar/gkz956.

[59] X. Xia, DAMBE7: new and improved tools for data analysis in molecular biology and evolution, Mol. Biol. Evol. 35 (2018) 1550–1552, https://doi.org/10.1093/molbev/msy073.

[60] NicoleT Perna, ThomasD Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J. Mol. Evol. 41 (1995), https://doi.org/10.1007/BF00186547.

[61] D. Laslett, B. Canback, ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences, Bioinformatics 24 (2008) 172–175, https://doi.org/10.1093/bioinformatics/btm573.

[62] T.M. Lowe, P.P. Chan, tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res. 44 (2016) W54–W57, https://doi.org/10.1093/nar/gkw413.

[63] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (1999) 573–580, https://doi.org/10.1093/nar/27.2.573.

[64] V. Ranwez, S. Harispe, F. Delsuc, E.J.P. Douzery, MACSE: multiple alignment of coding SEquences accounting for frameshifts and stop codons, PLoS One 6 (2011), e22594, https://doi.org/10.1371/journal.pone.0022594.

[65] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, Mol. Biol. Evol. 38 (2021) 3022–3027, https://doi.org/10.1093/molbev/msab120.

[66] A. Criscuolo, S. Gribaldo, BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments, BMC Evol. Biol. 10 (2010) 210, https://doi.org/10.1186/1471-2148-10-210.

[67] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Brief. Bioinform. 20 (2019) 1160–1166, https://doi.org/10.1093/bib/bbx108.

[68] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[69] K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program, Brief. Bioinform. 9 (2008) 286–298, https://doi.org/10.1093/bib/bbn013.

[70] M. Bruneaux, M. Vecchi, Tardipede/concatipede: First Release of Concatipede, 2021, https://doi.org/10.5281/ZENODO.5130603.

[71] X. Xia, Z. Xie, M. Salemi, L. Chen, Y. Wang, An index of substitution saturation and its application, Mol. Phylogenet. Evol. 26 (2003) 1–7, https://doi.org/10.1016/S1055-7903(02)00326-3.

[72] X. Xia, P. Lemey, Assessing substitution saturation with DAMBE, in: P. Lemey, M. Salemi, A.-M. Vandamme (Eds.), The Phylogenetic Handbook, Cambridge University Press, Cambridge, 2009, pp. 615–630, https://doi.org/10.1017/CBO9780511819049.022.

[73] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, J. Mol. Evol. 16 (1980) 111–120, https://doi.org/10.1007/BF01731581.

[74] O. Chernomor, A. von Haeseler, B.Q. Minh, Terrace aware data structure for Phylogenomic inference from Supermatrices, Syst. Biol. 65 (2016) 997–1008, https://doi.org/10.1093/sysbio/syw037.

[75] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (2020) 1530–1534, https://doi.org/10.1093/molbev/msaa015.

[76] J. Soubrier, M. Steel, M.S.Y. Lee, C. Der Sarkissian, S. Guindon, S.Y.W. Ho, A. Cooper, The influence of rate heterogeneity among sites on the time dependence of molecular rates, Mol. Biol. Evol. 29 (2012) 3345–3358, https://doi.org/10.1093/molbev/mss140.

[77] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics 17 (2001) 754–755, https://doi.org/10.1093/bioinformatics/17.8.754.

[78] F. Ronquist, Bayesian inference of character evolution, Trends Ecol. Evol. 19 (2004) 475–481, https://doi.org/10.1016/j.tree.2004.07.002.

[79] M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative, Syst. Biol. 55 (2006) 539–552, https://doi.org/10.1080/10635150600755453.

[80] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2018) 518–522, https://doi.org/10.1093/molbev/msx281.

[81] H. Shimodaira, An approximately unbiased test of phylogenetic tree selection, Syst. Biol. 51 (2002) 492–508, https://doi.org/10.1080/10635150290069913.

[82] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES science gateway for inference of large phylogenetic trees, in: 2010 Gateway Computing Environments Workshop (GCE). Presented at the 2010 Gateway Computing Environments Workshop (GCE), IEEE, New Orleans, LA, USA, 2010, pp. 1–8, https://doi.org/10.1109/GCE.2010.5676129.

[83] B.Q. Minh, M.W. Hahn, R. Lanfear, New methods to calculate concordance factors for Phylogenomic datasets, Mol. Biol. Evol. 37 (2020) 2727–2733, https://doi.org/10.1093/molbev/msaa106.

[84] E.S. Allman, S. Petrović, J.A. Rhodes, S. Sullivant, Identifiability of two-tree mixtures for group-based models, IEEE/ACM Trans. Comput. Biol. Bioinform. 8 (2011) 710–722, https://doi.org/10.1109/TCBB.2010.79.

[85] E. Sayyari, S. Mirarab, Testing for Polytomies in phylogenetic species trees using quartet frequencies, Genes 9 (2018) 132, https://doi.org/10.3390/genes9030132.

[86] R Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, 2021. https://www.R-project.org/.

[87] B. Subramanian, S. Gao, M.J. Lercher, S. Hu, W.-H. Chen, Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees, Nucleic Acids Res. 47 (2019) W270–W275, https://doi.org/10.1093/nar/gkz357.

[88] H. Zhang, S. Gao, M.J. Lercher, S. Hu, W.-H. Chen, EvolView, an online tool for visualizing, annotating and managing phylogenetic trees, Nucleic Acids Res. 40 (2012) W569–W572, https://doi.org/10.1093/nar/gks576.

[89] J. Chifman, L. Kubatko, Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites, J. Theor. Biol. 374 (2015) 35–47, https://doi.org/10.1016/j.jtbi.2015.03.006.

[90] S. Roch, M. Steel, Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent, Theor. Popul. Biol. 100 (2015) 56–62, https://doi.org/10.1016/j.tpb.2014.12.005.

[91] C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees, BMC Bioinform. 19 (2018) 153, https://doi.org/10.1186/s12859-018-2129-y.

[92] J. Chifman, L. Kubatko, Quartet inference from SNP data under the coalescent model, Bioinformatics 30 (2014) 3317–3324, https://doi.org/10.1093/bioinformatics/btu530.

[93] D.L. Swofford, Phylogenetic Analysis Using Parsimony (and Other Methods), PAUP*, 2002.

[94] P. Kück, S.A. Meid, C. Groß, J.W. Wägele, B. Misof, AliGROOVE – visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support, BMC Bioinform. 15 (2014) 294, https://doi.org/10.1186/1471-2105-15-294.

[95] P. Kück, T.H. Struck, BaCoCa – a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions, Mol. Phylogenet. Evol. 70 (2014) 94–98, https://doi.org/10.1016/j.ympev.2013.09.011.

[96] M. Zhong, B. Hansen, M. Nesnidal, A. Golombek, K.M. Halanych, T.H. Struck, Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids, BMC Evol. Biol. 11 (2011) 369, https://doi.org/10.1186/1471-2148-11-369.

[97] A. Rambaut, A.J. Drummond, D. Xie, G. Baele, M.A. Suchard, Posterior summarization in Bayesian Phylogenetics using tracer 1.7, Syst. Biol. 67 (2018) 901–904, https://doi.org/10.1093/sysbio/syy032.

[98] N. Lartillot, Identifying the best approximating model in Bayesian Phylogenetics: Bayes factors, cross-validation or wAIC? Syst. Biol. 72 (2023) 616–638, https://doi.org/10.1093/sysbio/syad004.

[99] K. Sato, M. Akiyama, Y. Sakakibara, RNA secondary structure prediction using deep learning with thermodynamic integration, Nat. Commun. 12 (2021) 941, https://doi.org/10.1038/s41467-021-21194-4.

[100] I.L. Hofacker, Vienna RNA secondary structure server, Nucleic Acids Res. 31 (2003) 3429–3431, https://doi.org/10.1093/nar/gkg599.

[101] R. Lorenz, S.H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I.L. Hofacker, ViennaRNA package 2.0. Algorithms, Mol. Biol. 6 (2011) 26, https://doi.org/10.1186/1748-7188-6-26.

[102] F. Jühling, J. Pütz, M. Bernt, A. Donath, M. Middendorf, C. Florentz, P.F. Stadler, Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements, Nucleic Acids Res. 40 (2012) 2833–2845, https://doi.org/10.1093/nar/gkr1131.

[103] E. Rivas, R. Lang, S.R. Eddy, A range of complex probabilistic models for RNA secondary structure prediction that includes the nearest-neighbor model and more, RNA 18 (2012) 193–212, https://doi.org/10.1261/rna.030049.111.

[104] H. Kishino, T. Miyata, M. Hasegawa, Maximum likelihood inference of protein phylogeny and the origin of chloroplasts, J Mol Evol 31 (1990) 151–160, https://doi.org/10.1007/BF02109483.

[105] H. Kishino, M. Hasegawa, Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea, J Mol Evol 29 (1989) 170–179, https://doi.org/10.1007/BF02100115.

[106] Strimmer Korbinian, Rambaut Andrew, Inferring confidence sets of possibly misspecified gene treesProc, R. Soc. Lond. (2002) B.269137–142, https://doi.org/10.1098/rspb.2001.1862.

[107] C. Moritz, W.M. Brown, Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA, Science 233 (1986) 1425–1427, https://doi.org/10.1126/science.3018925.

[108] D. San Mauro, D.J. Gower, R. Zardoya, M. Wilkinson, A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome, Mol. Biol. Evol. 23 (2006) 227–234, https://doi.org/10.1093/molbev/msj025.

[109] S.E. Smith, D.W. Au, C. Show, Intrinsic rebound potentials of 26 species of Pacific sharks, Mar. Freshw. Res. 49 (1998) 663, https://doi.org/10.1071/MF97135.

[110] J.R. Macey, A. Larson, N.B. Ananjeva, Z. Fang, T.J. Papenfuss, Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome, Mol. Biol. Evol. 14 (1997) 91–104, https://doi.org/10.1093/oxfordjournals.molbev.a025706.

[111] D.H. Lunt, B.C. Hyman, Animal mitochondrial DNA recombination, Nature 387 (1997) 247, https://doi.org/10.1038/387247a0.

[112] I. Bakke, G.F. Shields, S. Johansen, Sequence characterization of a unique intergenic spacer in Gadiformes mitochondrial DNA, Mar. Biotechnol. 1 (1999) 411–415, https://doi.org/10.1007/PL00011797.

[113] Y. Kumazawa, H. Ota, M. Nishida, T. Ozawa, Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster, Mol. Biol. Evol. 13 (1996) 1242–1254, https://doi.org/10.1093/oxfordjournals.molbev.a025690.
[114] A.L.F. Castro, B.S. Stewart, S.G. Wilson, R.E. Hueter, M.G. Meekan, P.J. Motta, B. W. Bowen, S.A. Karl, Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*), Mol. Ecol. 16 (2007) 5183–5192, https://doi.org/10.1111/j.1365-294X.2007.03597.x.
[115] T. Kitamura, A. Takemura, S. Watabe, T. Taniuchi, M. Shimizu, Mitochondrial DNA analysis for the cytochrome *b* gene and D-loop region from the bull shark *Carcharhinus leucas*, Fish. Sci. 62 (1996) 21–27, https://doi.org/10.2331/fishsci.62.21.
[116] D. Ramírez-Macías, R. Vázquez-Juárez, F. Galván-Magaña, A. Munguía-Vega, Variations of the mitochondrial control region sequence in whale sharks (*Rhincodon typus*) from the Gulf of California, Mexico, Fish. Res. 84 (2007) 87–95, https://doi.org/10.1016/j.fishres.2006.11.038.
[117] A.J. Aberer, D. Krompass, A. Stamatakis, Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and Webservice, Syst. Biol. 62 (2013) 162–166, https://doi.org/10.1093/sysbio/sys078.
[118] J.F. Fleming, A. Valero-Gracia, T.H. Struck, Identifying and addressing methodological incongruence in phylogenomics: a review, Evol. Appl. 16 (2023) 1087–1104, https://doi.org/10.1111/eva.13565.
[119] M.D. Scherz, P. Masonick, A. Meyer, C.D. Hulsey, Between a rock and a hard Polytomy: Phylogenomics of the rock-dwelling Mbuna cichlids of Lake Malaẅi, Syst. Biol. 71 (2022) 741–757, https://doi.org/10.1093/sysbio/syac006.
[120] J.B. Whitfield, P.J. Lockhart, Deciphering ancient rapid radiations, Trends Ecol. Evol. 22 (2007) 258–265, https://doi.org/10.1016/j.tree.2007.01.012.
[121] W.P. Maddison, L.L. Knowles, Inferring phylogeny despite incomplete lineage sorting, Syst. Biol. 55 (2006) 21–30, https://doi.org/10.1080/10635150500354928.
[122] G.P. Wallis, S.R. Cameron-Christie, H.L. Kennedy, G. Palmer, T.R. Sanders, D. J. Winter, Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes, Mol. Ecol. 26 (2017) 3116–3127, https://doi.org/10.1111/mec.14096.
[123] C.M. Hurvich, C.-L. Tsai, A corrected akaike information criterion for vector autoregressive model selection, J. Time Ser. Anal. 14 (1993) 271–279, https://doi.org/10.1111/j.1467-9892.1993.tb00144.x.
[124] G. Schwartz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[125] Y. Inagaki, A.J. Roger, Phylogenetic estimation under codon models can be biased by codon usage heterogeneity, Mol. Phylogenet. Evol. 40 (2006) 428–434, https://doi.org/10.1016/j.ympev.2006.03.020.
[126] A. Luo, H. Qiao, Y. Zhang, W. Shi, S.Y. Ho, W. Xu, A. Zhang, C. Zhu, Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets, BMC Evol. Biol. 10 (2010) 242, https://doi.org/10.1186/1471-2148-10-242.
[127] J.B. Miller, M.F. Whiting, J.S.K. Kauwe, P.G. Ridge, How codon usage bias affects our ability to recover the tree of life (preprint), Biology. (2019), https://doi.org/10.20944/preprints201910.0086.v1.
[128] D.A. Ebert, J.S. Bigman, J.M. Lawson, Biodiversity, life history, and conservation of northeastern Pacific Chondrichthyans, in: Advances in Marine Biology, Elsevier, 2017, pp. 9–78, https://doi.org/10.1016/bs.amb.2017.07.001.
[129] K.O. Chan, C.R. Hutter, P.L. Wood, L.L. Grismer, R.M. Brown, Target-capture phylogenomics provide insights on gene and species tree discordances in Old World treefrogs (Anura: Rhacophoridae), Proc. R. Soc. B Biol. Sci. 287 (2020) 20202102, https://doi.org/10.1098/rspb.2020.2102.
[130] R.J. Kallal, S.S. Kulkarni, D. Dimitrov, L.R. Benavides, M.A. Arnedo, G. Giribet, G. Hormiga, Converging on the orb: denser taxon sampling elucidates spider phylogeny and new analytical methods support repeated evolution of the orb web, Cladistics 37 (2021) 298–316, https://doi.org/10.1111/cla.12439.
[131] T. van Elst, T.H. Eriksson, J. Gadau, R.A. Johnson, C. Rabeling, J.E. Taylor, M. L. Borowiec, Comprehensive phylogeny of Myrmecocystus honey ants highlights cryptic diversity and infers evolution during aridification of the American southwest, Mol. Phylogenet. Evol. 155 (2021), 107036, https://doi.org/10.1016/j.ympev.2020.107036.
[132] S.V. Edwards, Is a new and general theory of molecular systematics emerging? Evolution 63 (2009) 1–19, https://doi.org/10.1111/j.1558-5646.2008.00549.x.
[133] E.L. Braun, J. Cracraft, P. Houde, Resolving the avian tree of life from top to bottom: The promise and potential boundaries of the Phylogenomic era, in: R.H. S. Kraus (Ed.), Avian Genomics in Ecology and Evolution, Springer International Publishing, Cham, 2019, pp. 151–210, https://doi.org/10.1007/978-3-030-16477-5_6.
[134] P.M. Hime, A.R. Lemmon, E.C.M. Lemmon, E. Prendini, J.M. Brown, R. C. Thomson, J.D. Kratovil, B.P. Noonan, R.A. Pyron, P.L.V. Peloso, M.L. Kortyna, J.S. Keogh, S.C. Donnellan, R.L. Mueller, C.J. Raxworthy, K. Kunte, S.R. Ron, S. Das, N. Gaitonde, D.M. Green, J. Labisko, J. Che, D.W. Weisrock, Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life, Syst. Biol. 70 (2021) 49–66, https://doi.org/10.1093/sysbio/syaa034.
[135] I. Irisarri, D. Baurain, H. Brinkmann, F. Delsuc, J.-Y. Sire, A. Kupfer, J. Petersen, M. Jarek, A. Meyer, M. Vences, H. Philippe, Phylotranscriptomic consolidation of the jawed vertebrate timetree, Nat. Ecol. Evol. 1 (2017) 1370–1378, https://doi.org/10.1038/s41559-017-0240-5.
[136] I. Irisarri, A. Meyer, The identification of the closest living relative(s) of Tetrapods: Phylogenomic lessons for resolving short ancient internodes, Syst. Biol. 65 (2016) 1057–1075, https://doi.org/10.1093/sysbio/syw057.
[137] S. Singhal, T.J. Colston, M.R. Grundler, S.A. Smith, G.C. Costa, G.R. Colli, C. Moritz, R.A. Pyron, D.L. Rabosky, Congruence and conflict in the higher-level Phylogenetics of Squamate reptiles: an expanded Phylogenomic perspective, Syst. Biol. 70 (2021) 542–557, https://doi.org/10.1093/sysbio/syaa054.
[138] A. Suh, The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves, Zool. Scr. 45 (2016) 50–62, https://doi.org/10.1111/zsc.12213.
[139] L. Sorenson, F. Santini, M.E. Alfaro, The effect of habitat on modern shark diversification, J. Evol. Biol. 27 (2014) 1536–1548, https://doi.org/10.1111/jeb.12405.
[140] B.E. Boudinot, M. Fikáček, Z.E. Lieberman, D. Kusy, L. Bocak, D.D. Mckenna, R. G. Beutel, Systematic bias and the phylogeny of Coleoptera—A response to Cai et al. (2022) following the responses to Cai et al. (2020), Syst. Entomol. 48 (2023) 223–232, https://doi.org/10.1111/syen.12570.
[141] T.H. Struck, A.R. Wey-Fabrizius, A. Golombek, L. Hering, A. Weigert, C. Bleidorn, S. Klebow, N. Iakovenko, B. Hausdorf, M. Petersen, P. Kück, H. Herlyn, T. Hankeln, Platyzoan Paraphyly based on Phylogenomic data supports a noncoelomate ancestry of Spiralia, Mol. Biol. Evol. 31 (2014) 1833–1849, https://doi.org/10.1093/molbev/msu143.
[142] N. Lartillot, H. Brinkmann, H. Philippe, Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model, BMC Evol. Biol. 7 (2007) S4, https://doi.org/10.1186/1471-2148-7-S1-S4.
[143] H. Philippe, H. Brinkmann, D.V. Lavrov, D.T.J. Littlewood, M. Manuel, G. Wörheide, D. Baurain, Resolving difficult phylogenetic questions: why more sequences are not enough, PLoS Biol. 9 (2011), e1000602, https://doi.org/10.1371/journal.pbio.1000602.
[144] J.A. López, J.A. Ryburn, O. Fedrigo, G.J.P. Naylor, Phylogeny of sharks of the family Triakidae (Carcharhiniformes) and its implications for the evolution of carcharhiniform placental viviparity, Mol. Phylogenet. Evol. 40 (2006) 50–60, https://doi.org/10.1016/j.ympev.2006.02.011.
[145] S.N. Maduna, K.L. Hull, E.D. Farrell, J.J. Boomer, A. Veríssimo, I.A.M. Marino, C. Mazzoldi, L. Zane, S.P. Wintner, M.V. Chesalin, C. da Silva, C. Gubili, S. Mariani, A.E. Bester-Van Der Merwe, Historical biogeography of smoothhound sharks (genus *Mustelus*) of Southern Africa reveals multiple dispersal events from the Northern hemisphere, Syst. Biodivers. 18 (2020) 633–645, https://doi.org/10.1080/14772000.2020.1787550.
[146] Gavin J.P. Naylor, J. Caira, K. Jensen, K. Rosana, N. Straube, C. Lakner, Elasmobranch phylogeny: A mitochondrial estimate based on 595 species, in: J. Carrier, J. Musick, M. Heithaus (Eds.), Biology of Sharks and their Relatives, Second Edition, Marine Biology, CRC Press, 2012, pp. 31–56, https://doi.org/10.1201/b11867-4.
[147] G.J.P. Naylor, J.N. Caira, K. Jensen, K.A.M. Rosana, W.T. White, P.R. Last, A DNA sequence–based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology, Bull. Am. Mus. Nat. Hist. 367 (2012) 1–262, https://doi.org/10.1206/754.1.
[148] J.J. Boomer, R.G. Harcourt, M.P. Francis, A.J. Stow, Genetic divergence, speciation and biogeography of Mustelus (sharks) in the central indo-Pacific and Australasia, Mol. Phylogenet. Evol. 64 (2012) 697–703, https://doi.org/10.1016/j.ympev.2012.05.024.
[149] S.N. Maduna, C. Rossouw, C. da Silva, M. Soekoe, A.E. Bester-van der Merwe, Species identification and comparative population genetics of four coastal houndsharks based on novel NGS-mined microsatellites, Ecol. Evol. 7 (2017) 1462–1486, https://doi.org/10.1002/ece3.2770.
[150] L.S. Kubatko, J.H. Degnan, Inconsistency of phylogenetic estimates from concatenated data under coalescence, Syst. Biol. 56 (2007) 17–24, https://doi.org/10.1080/10635150601146041.
[151] X.-X. Shen, C.T. Hittinger, A. Rokas, Contentious relationships in phylogenomic studies can be driven by a handful of genes, Nat. Ecol. Evol. 1 (2017) 0126, https://doi.org/10.1038/s41559-017-0126.