The Greenland shark (*Somniosus microcephalus*) genome provides insights into extreme longevity

Arne Sahm [1,2,3], Alexander Cherkasov [*,1], Hequn Liu [*,3], Danila Voronov [4], Kanstantsin Siniuk [1], Robert Schwarz [1], Oliver Ohlenschläger [5], Silke Förste [1], Martin Bens [6], Marco Groth [6], Ivonne Görlich [6], Sonia Paturej [7], Sven Klages [7], Bjoern Braendl [8], Jesper Olsen [9], Peter Bushnell [10], Amalie Bech Poulsen [11], Sara Ferrando [12], Fulvio Garibaldi [12], Davide Lorenzo Drago [13], Eva Terzibasi Tozzini [14], Franz-Josef Müller [15,8], Martin Fischer [1], Helene Kretzmer [15,16], Paolo Domenici [#,17], John Fleng Steffensen [#,11], Alessandro Cellerino [#,4,13], Steve Hoffmann [#,1]

1 Hoffmann Lab, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstrasse 11, 07745 Jena, Germany
2 Sahm Lab, Faculty of Biology and Biotechnology, Ruhr University Bochum, Universitätsstrasse 150, 44801 Bochum, Germany
3 Sahm Lab, IUF - Leibniz Research Institute for Environmental Medicine, Auf'm Hennekamp 50, 40225 Düsseldorf, Germany.
4 Cellerino Lab, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstrasse 11, 07745 Jena, Germany
5 Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstrasse 11, 07745 Jena, Germany
6 Core Facility Next Generation Sequencing, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstrasse 11, 07745 Jena, Germany
7 Sequencing Facility, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany
8 Department of Psychiatry and Psychotherapy, Christian-Albrechts-Universität zu Kiel, Niemannsweg 147, 24105 Kiel, Germany
9 AMS Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus, Denmark.
10 Department of Biological Sciences, Indiana University South Bend, 1700 Mishawaka Avenue, South Bend, IN, USA
11 Marine Biological Section, Department of Biology, University of Copenhagen, Strandpromenaden 5, 3000 Helsingør, Denmark
12 Department of Earth, Environmental and Life Sciences (DISTAV), University of Genoa, Corso Europa, 26, 16132 Genoa, Italy
13 Biology Laboratory (BIO@SNS), Scuola Superiore Normale, Piazza dei Cavalieri 7 – 56126, Pisa, Italy
14 Biology and Evolution of Marine Organisms Department (BEOM), Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
15 Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Ihnestraße 63, 14195 Berlin, Germany
16 Digital Health Cluster, Hasso Plattner Institute for Digital Engineering, Digital Engineering Faculty, University of Potsdam, Prof.-Dr.-Helmert-Straße 2-3, 14482 Potsdam, Germany
17 IBF-CNR, Consiglio Nazionale delle Ricerche, Area di Ricerca San Cataldo, Via G. Moruzzi 1, San Cataldo, 56124 Pisa, Italy
Corresponding authors
* Contributed equally

Abstract

The Greenland shark (*Somniosus microcephalus*) is the longest-lived vertebrate known, with an estimated lifespan of ~ 400 years. Here, we present a chromosome-level assembly of the 6.45 Gb Greenland shark, rendering it one of the largest non-tetrapod genomes sequenced so far. Expansion of the genome is mostly accounted for by a substantial expansion of transposable elements. Using public shark genomes as a comparison, we found that genes specifically duplicated in the Greenland shark form a functionally connected network enriched for DNA repair function. Furthermore, we identified a unique insertion in the conserved C-terminal region of the key tumor suppressor p53. We also provide a public browser to explore its genome.

Introduction

Life-history traits show significant variation in vertebrates, and the evolution of life-history traits is profoundly shaped by adaptations to habitat and ecological conditions. Lifespan shows robust covariation with adult size, metabolism, and age at sexual maturity across different taxa (Speakman 2005; Magalhães et al. 2007; Ricklefs 2010). Yet, long-lived outliers exist, e.g., Galapagos giant tortoise, bowhead whale, clownfish, rockfish, naked mole-rat, or Brandt's bat (Holtze et al. 2021). The term "negligible senescence" has been coined for such long-lived animals that appear to escape age-dependent deterioration (Finch 1998). However, only limited

support has been available for this hypothesis. It has been proposed that an overarching principle in the evolution of exceptional longevity is the occupancy of habitats that are not easily accessible. For instance, long-living species are found among troglomorph (Voituron et al. 2011) and hypogean (Edrey et al. 2011) taxa. In marine environments, depth may be associated with longevity, as suggested by the exceptional (> 100 y) lifespan of the rockfish (genus Sebastes) (Cailliet et al. 2001; Kolora et al. 2021). Consequently, molecular studies focusing on interspecies comparisons have been initiated to unravel the mechanisms and genetic architecture of life-history variation and identified signaling pathways that show signatures of selection in multiple clades that evolved exceptional longevity. Longevity-associated genes mainly relate to DNA repair, mitochondrial balance/mitochondrial biogenesis, immune system functions, oxidative stress resistance, and telomere maintenance (Holtze et al. 2021). For example, in Sebastes, copy number expansions in the immune modulatory butyrophilin gene family are associated with longevity (Kolora et al. 2021).

The Greenland shark (GLS, *Somniosus microcephalus*) is a giant (>5 m) iconic abyssal species that inhabits the North Atlantic depths and the Arctic Ocean, representing an extreme in the spectrum of life-history trait variation (Nielsen 2018). Based on radiocarbon dating, females show an extraordinary lifespan estimated at 392 ± 120 y (Nielsen et al. 2016; Nielsen 2018). Longevity in this species is coupled to a very slow growth rate of <1 cm/year (Hansen 1963) and sluggishness with a cruising speed of <1 m/s (Watanabe et al. 2012). Its size and the very low temperature of its environment result in a very low mass-specific metabolic rate (Ste-Marie et al. 2020; Smith et al. 2022).

Here, we present the first chromosome-scale assembly of the GLS genome, which will be instrumental to identify the genetic adaptations responsible for this remarkable animal's extreme longevity and peculiar traits. Our analysis suggests the presence of specific alterations in *TP53* and DNA repair.

Results

Chromosome-level genome assembly

To assemble the GLS genome, we used PacBio HiFi long-read sequencing. More than 50% of the initial contig assembly is covered by sequences longer than 1 Mb (**Table 1**). We performed high-throughput chromatin conformation capture (Hi-C) to assemble these sequences into scaffolds. We obtained 43 chromosome-scale scaffolds (>10Mb) accounting for a total of 6.06 Gb (94% of scaffolded nucleotides) ranging in size from 464.6 Mb for the largest- and 28.5 Mb for the smallest scaffold, respectively. A total of 12 scaffolds are longer than 200 Mb, while 17 are longer than 100 Mb (**Fig. 1a**). More than 50% of the nucleotides in our scaffold assembly are contained in sequences that are as long as the largest human chromosome (chromosome 1 with ~248 Mb vs. our N50 of 240 Mb). With its size of 6 GB, the GLS genome is the largest shark

genome assembled to date and shows the highest repeat content exceeding 70% (Table 1, compare, *e.g.*, to (Marra et al. 2019)). Using RNA-seq and known protein sequences from other species, we identified 22,634 protein-coding genes in the genome, with 20,274 functionally annotated. The genome, annotation, and gene expression data are publicly accessible through a dedicated genome browser (https://glshark.leibniz-fli.de/). The analysis of vertebrate single-copy orthologs via BUSCO (Simao et al. 2015) indicates a genome completeness of 91.8%. Thus, despite the large size and high repeat content, the assembly is as complete as the assemblies of other high-quality shark genomes that are publicly available (**Fig. 1b**).

Table 1. Greenland shark genome assembly statistics.

Total	Contigs				Scaffolds			
Len.	Num.	N50	N90	Max.	Num.	N50	N90	Max
6.45 Gb	13,991	1.1 Mb	234 Kb	9.2 Mb	2,111	240.4 Mb	47.6 Mb	464.6 Mb

With a repeat content of 58.5%, the Great white shark (GWS, Carcharodon carcharias) was previously reported to have the highest known repeat content among sharks (Marra et al. 2019). However, the GLS genome substantially exceeds this level with a repeat content of 70.6% (**Fig. 1c**). According to our data, the increase in total content is mainly due to the expansion of retrotransposable elements, namely long interspersed nuclear elements (LINEs) and, especially, long terminal repeats (LTRs).

Taken together, we provide a highly complete, chromosome-level genome with exceptionally high repeat content for a vertebrate.
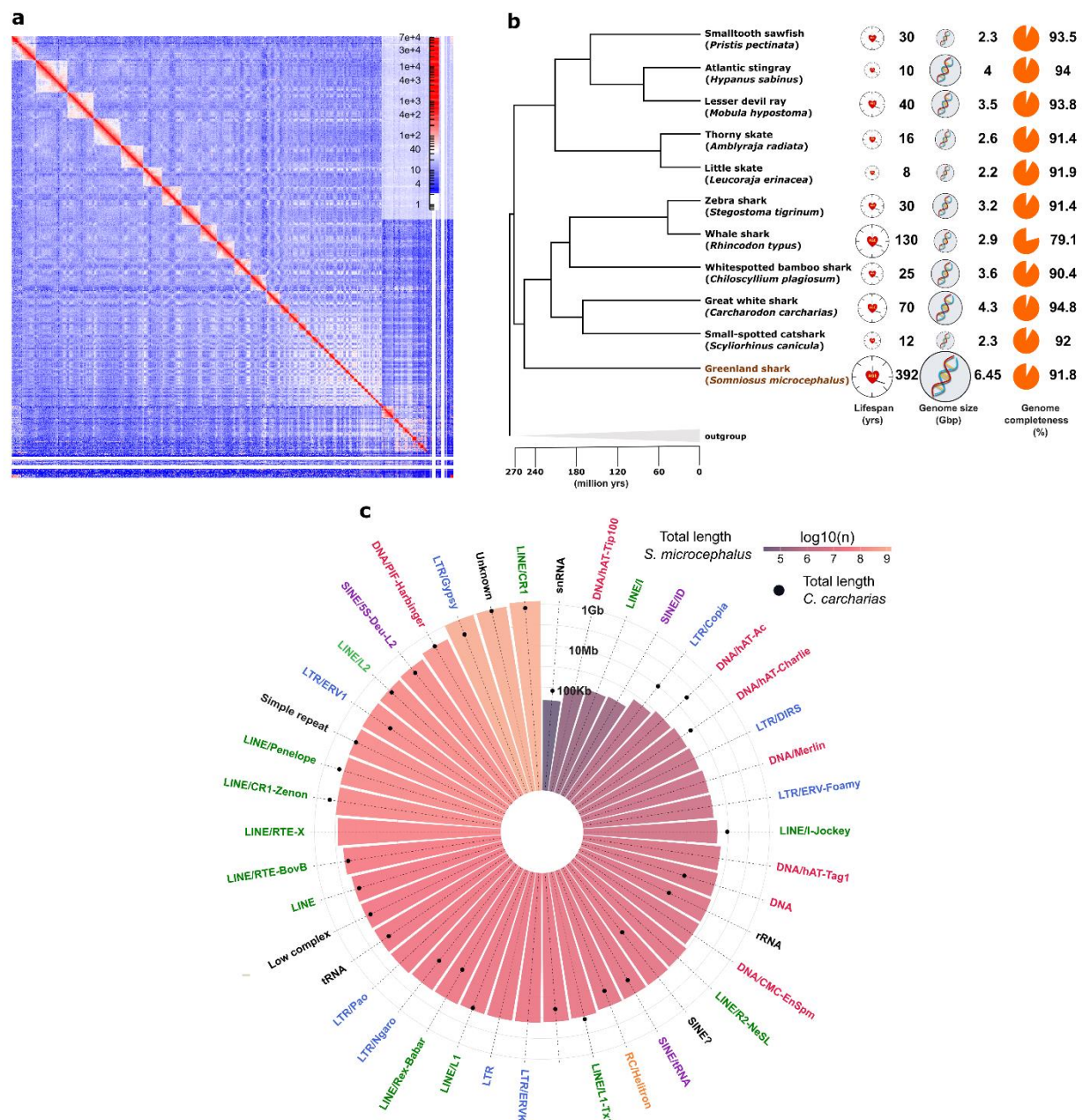
Fig 1. Characteristics of the Greenland shark genome assembly. a) Intensity signal heat map of HiC contacts. b) Phylogenetic tree of shark species with lifespan, genome size and genome completeness included for each taxon. Genome completeness was me

asured for all species by BUSCO against single-copy vertebrate genes. Divergence times were taken mainly from the timetree project (Kumar et al. 2017). Single remaining missing divergence times were taken from the following literature sources: (Marlétaz et al. 2023; Yamaguchi et al. 2023). c) Total length numbers (log10-scaled) of the most abundant transposable element families in the Greenland shark (bars). Black circles indicate the respective total length of the elements in the great white shark.

High genomic synteny with evolutionary distant shark species

Next, we analyzed the genomic position of highly conserved vertebrate genes (BUSCO-genes, n=3,354) in the GLS and the GWS. This analysis reveals a striking synteny between the two

respective genomes (**Fig. 2a**) - despite a separation between these two lineages dating about 250 million years ago (Kumar et al. 2017). For several GLS chromosome-sized scaffolds (>10Mb), we find that all BUSCO genes of any one such scaffold map exclusively to a single scaffold of the GWS genome (**Fig. 2b**). This one-to-one correspondence is observed in 19 GLS scaffolds (44% of all scaffolds). For five additional scaffolds, at least 99% of BUSCO genes mapped to the same single GWS scaffold. Thus, our data reveal very high chromosome-level synteny of 25 GLS scaffolds (60% of scaffolds) covering approximately 3 billion base pairs (46% of scaffolded nucleotides). In several instances, our synteny map indicates potential chromosomal rearrangements, *e.g.*, involving scaffolds 11, 26, 7, and 3, as well as 29, 6, and 24. Furthermore, the map suggests potential metacentric splits of the great white shark sequences NC_04476.1 and NC_04479.1, leading to scaffolds 14 and 20 as well as 21 and 18, respectively. Three smaller scaffolds (42, 44, and 45) contained no BUSCO genes with correspondence in chromosome-scale GWS scaffolds.

An analysis of gene duplication events identified ancient, *i.e.*, at least 250 million years old, duplications conserved between GLS and GWS. An extreme example of these ancient duplications is an expansion of the ferritin light chain (*FTL*) and the ferritin heavy chain (*FTH1*): we detected 155 and 179 copies of these genes, respectively. Approximately 80% of the copies are located in three larger clusters. The synteny of these clusters is identical in the GWS and GLS genomes. The largest cluster is located at the end of scaffold 30, containing 104 copies of *FTL* and 118 copies of *FTH1*. More than 85% of all *FTH1* and *FTL* copies appear to be of full length and share at least 90% of amino acids with the human orthologs. RNAseq data analysis revealed a tissue-specific expression of this cluster in the ovary (see genome browser, HiC_scaffold_30:44,903,755-48,980,971).

Overall, despite an evolutionary distance of several hundred million years between them (**Fig. 1b**), GLS and GWS show striking similarities in their genomic structure and organization.
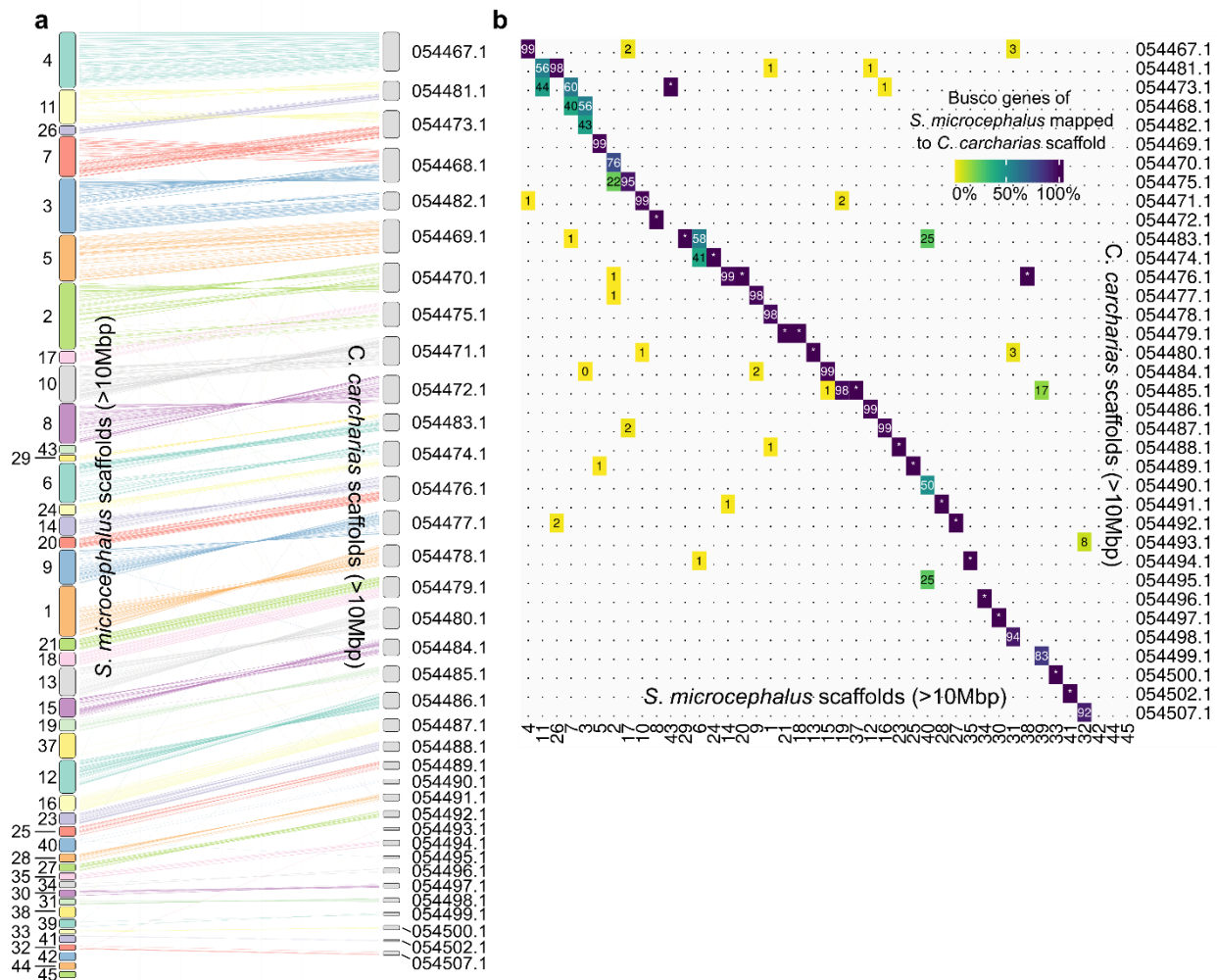
Figure 2. Synteny between the Greenland shark (*S. microcephalus*) and Great white shark (*C. carcharias*) genomes. a) Synteny graph of single-copy orthologous vertebrate genes found by BUSCO in both genomes. b) Heatmap quantifying for each Greenland shark scaffold (x-axis) the percentage of Busco genes mapping to the sequences of the Great White shark (y-axis). Dots indicate 0%, and stars indicate 100%.

Specific gene duplications in the Greenland shark

Next, we sought to identify unique characteristics of the GLS genome that may underlie its phenotypic traits, particularly its longevity. Regarding gene duplications, we identified 81 genes present as single copies in all other publicly available Elasmobranch genomes (n=10) but present in multiple copies in the GLS genome. When analyzed using the STRING database, these 81 genes show significant network connectivity (p = 0.029, **Fig. 3a**), indicating that the encoded proteins cooperate in executing specific biological functions. Furthermore, this network is enriched for the gene ontology (GO) term "site of double-strand break" (GO:0035861, FDR=0.037), which is also the only GO term for which a significant overrepresentation was detected. *TP53* appears as a hub of this network, although *TP53* itself is present as a single copy

in the GLS. Adding this gene manually increases the support for both network connectivity (p=0.018, **Fig. 3b**) and enrichment for double-strand break (FDR=0.002), potentially suggesting a functional link between the duplicated genes and *TP53*. Test for enrichments against the respective hallmark gene sets of the Molecular Signatures Database MSigDB (Liberzon et al. 2015) provided additional evidence for the functional relevance of the gene duplication network in DNA repair and the p53 pathway (one-sided exact Fisher-test; p=0.047 and p=0.005, respectively). Another GLS-specific gene duplication affects *NF2* coding for MERLIN, a vital tumor suppressor with known functional connections to p53 verified in mouse (Fig. 3b, (Kim et al. 2004)) and human cells (Li et al. 2017).

One of the further genes exclusively duplicated in the GLS is *RAD51*, a highly connected node in the network of duplicated genes (**Fig. 3c**), and, together with p53, its protein product plays a crucial role in DNA repair (Linke et al. 2003; Grundy et al. 2020; Laurini et al. 2020). We identified three *RAD51* copies in the GLS genome, all located on scaffold 12 in a region of approximately 30Mb. One copy of RAD51 has the same coding exons structure as in the GWS (exons 1 – 8 in the GLS; 2 – 9 in the GWS) and other Elasmobranch species displaying a high degree of conservation (**Fig. 3c,d**). Another similar copy consists of at least five exons. The third and shorter *RAD51* copy consists of a single exon (**Fig. 3e**). Of note, scaffold 12 shows a substantial level of synteny with the GWS chromosome 20 and potentially underwent an inversion of a chromosomal arm (**Fig. 3f**). Our data indicates the presence of a local duplication of a region containing *RAD51* (**Fig. 3f inset**). The single-exon copy is located about 30Mb downstream of the multi-exon copies. Interestingly, this single exon concatenates the genomic sequence of four exons of the longer copies and the corresponding four exons in the GWS and shows ubiquitous expression (exons 2 – 5, **Fig 3c**). Thus, the unique genomic organization of this *RAD51* copy may result from a retrotransposition, *i.e.*, reverse transcription of the spliced mRNA to cDNA and subsequent integration into the genome. Several other genes exclusively duplicated in the GLS appear to result from retrotransposition. For example, *OPTN*, *PLEKHA1*, *LRAT,* and *WAPAL* have copies of only one exon matching multiple exons of the ancestral copy in the genomes of the GLS and/or other shark species. Either LINEs or LTRs flank all of these retrogenes within 1,000 base pairs.

Taken together, duplications of genes associated with DNA repair and the p53 pathway appear to distinguish the extremely long-lived GLS from other Elasmobranch species, outlining the path for additional analyses and validations.
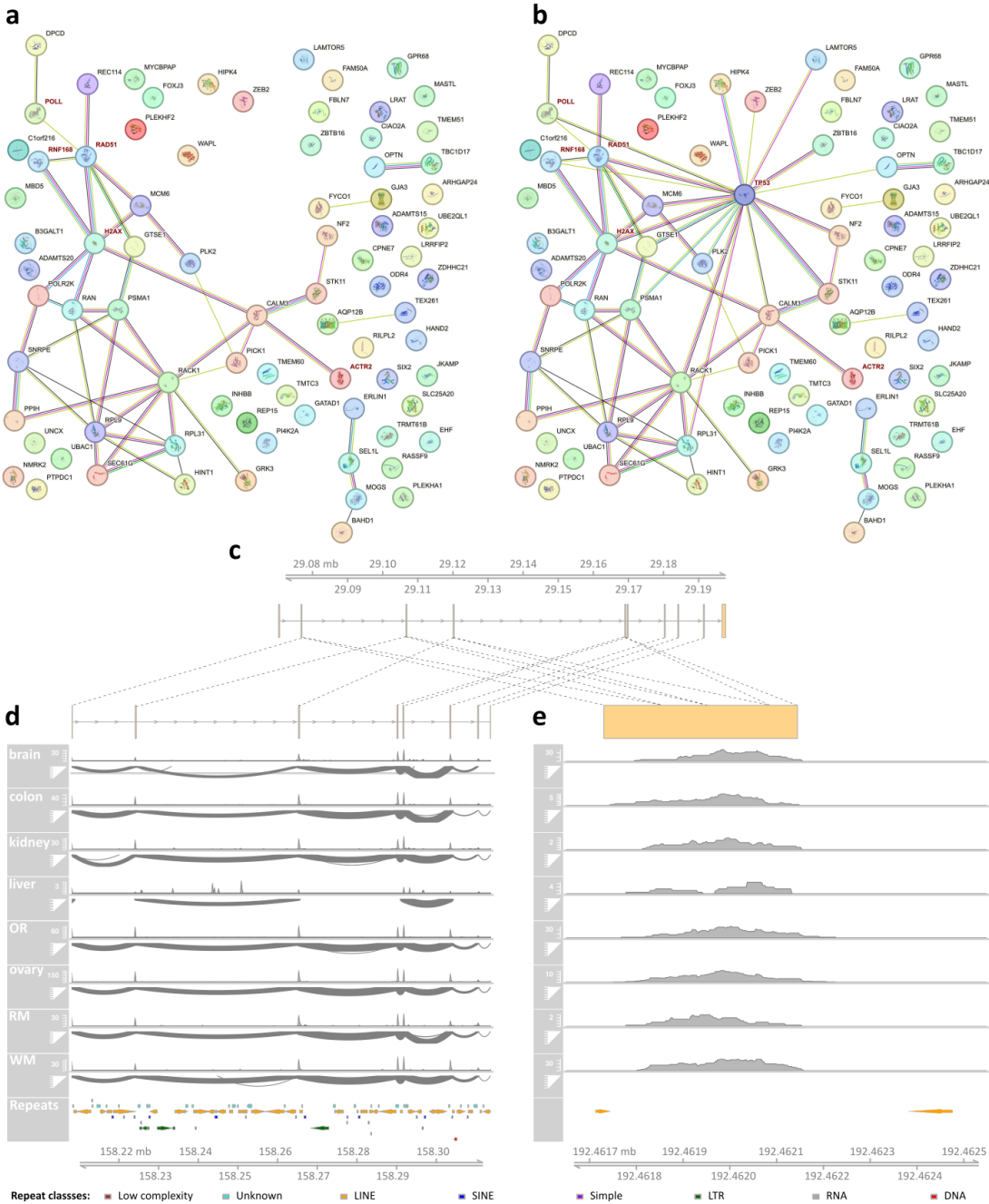
Fig. 3. Gene duplications distinguishing the Greenland shark from other Elasmobranch species. a) Uniquely duplicated genes with potential functional connections between them (STRING database, network connectivity: p = 0.029). Red labels indicate genes contributing to the enrichment found for "site of double-strand break" (GO:0035861, FDR=0.037). b) Same as in a) with *TP53* added manually to the network. c) *RAD51* gene structure in the great white shark and two of three gene copies in the Greenland shark (d and e). Note that introns were lost in the *RAD51* copy displayed in e). At the bottom of d) and e), gene expression coverage and sashimi plots for eight tissues and repeat annotation are shown. OR – olfactory receptor; RM – red muscle; WM – white muscle. f) Alignment-based dot plot of scaffold 12 and chromosome 20 (NC_054486.1) of the Great white shark. The gene regions of the three copies of the RAD51 (indicated by arrows) show similarity with the ortholog in the Great white shark (solid vertical line). The zoom-in suggests a local duplication of a region around RAD51 in the Greenland shark. The color intensity of the dots indicates the regional sequence identity.

Positive selection and unique changes in *TP53*

We conducted a genome-wide search for positively selected protein-coding genes by analyzing the ratio between synonymous and non-synonymous substitutions to further correlate molecular and organismal evolution. So far, our analysis highlights eight genes with evidence for positive selection in the GLS, *i.e.*, significant evolutionary changes compared to other sharks (FDR<0.1). Out of these eight positively selected genes, seven are part of pathways relevant for aging and aging-associated diseases: ZCCHC17 is a master regulator of synaptic gene expression in Alzheimer's disease (Tomljanovic et al. 2018). ACKR3 is a chemokine receptor with pro-proliferative and anti-senescence activity (Takaya et al. 2022; Gritsina and Yu 2023). TRAPPC8, DDX42, and DESI1 are involved in autophagy, assembly of the spliceosome, and deSUMOylation, respectively (Shin et al. 2012; Zhao et al. 2017; Yang et al. 2023). A Poly (ADP-ribose) polymerase family member, PARP14, and the mitochondrial ribosomal protein MRPS35 are particularly interesting. These two genes are involved in one of two fundamental mechanisms associated with the evolution of longevity: *PARP14* in DNA repair (Nicolae et al. 2015) and *MRPS35* in mitonuclear protein balance (Houtkooper et al. 2013).

Given the role of *TP53* in longevity and cancer resistance in other large animals and its connection to the DNA repair genes duplicated in the GLS, we examined its coding sequence more closely. We detected a specific amino acid insertion in the highly conserved C-terminal region of the protein, where the GLS protein is composed of three adjacent lysines. In comparison, all other Elasmobranchs examined have only two adjacent lysines at this site (**Fig. 4a**). Notably, with the sole exception of the GLS, the two-lysine pattern has been conserved for at least ~240 million years, which is the time when the last common ancestor of these Elasmobranchs lived (compare **Fig. 1b**). This pattern could not be detected by positive selection analysis because the probabilistic model used excludes positions that contain indels in some taxa. Structural modeling using AlphaFold3 suggests that the lysine insertion in the GLS leads to

a significantly increased helix propensity and, with that, to a more extended C-terminal helix of the protein (**Fig. 4b**). Furthermore, the modeling suggests that possible interactions of the HKKEKE motif (**Fig. 4a**) with other residues of the p53 tetramer lose stability and might not be formed upon the insertion of a third lysine (K) due to the shifted position of the interaction partner (compare **Fig. 4c** to **Fig. 4d**).

In conclusion, our analysis shows that alterations in several aging-relevant genes, including *TP53*, distinguish the GLS from shorter-lived shark species.
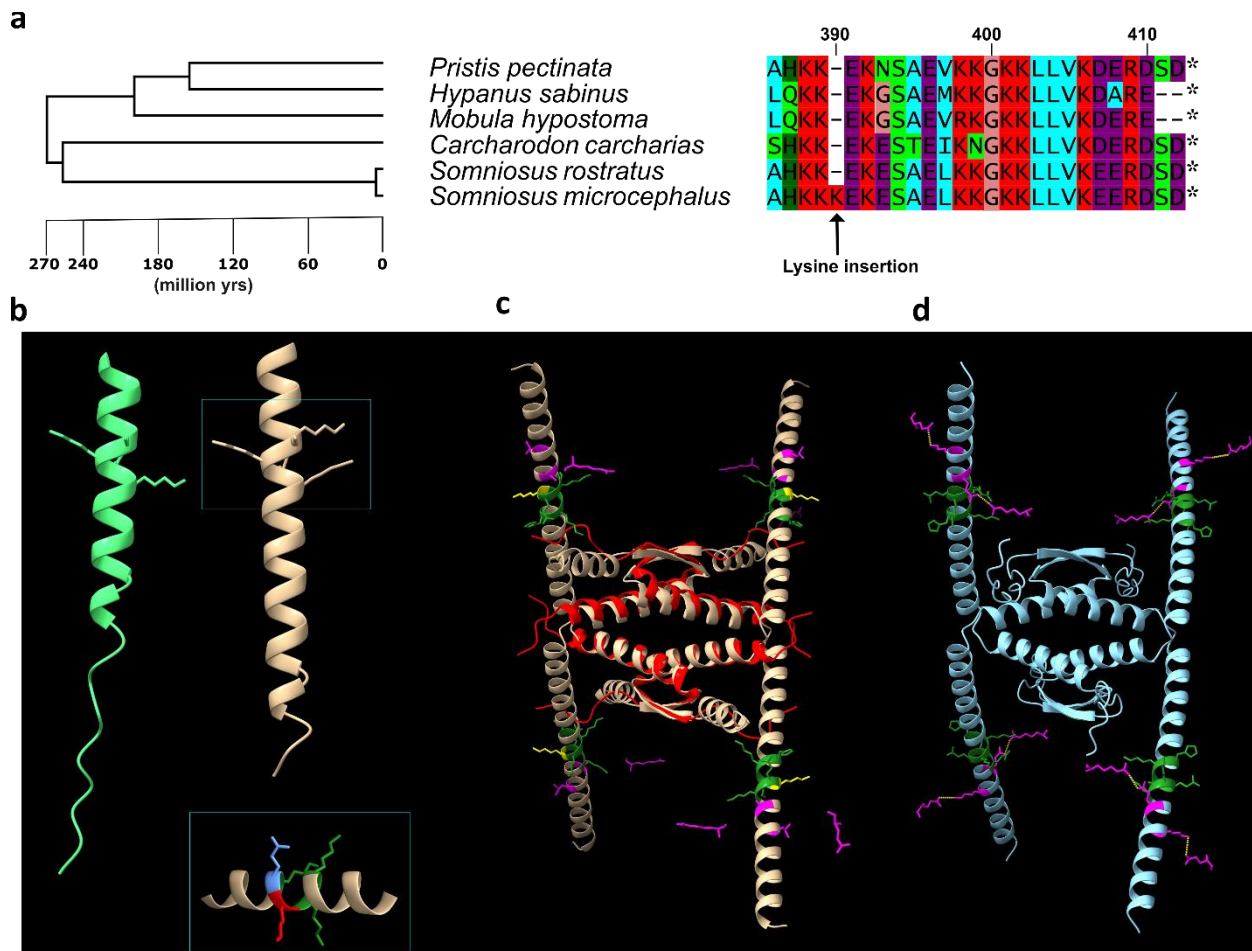

Fig. 4. Evolutionary changes in the Greenland shark's p53aminoacidic sequence. a) Phylogenetic tree of the protein and sequence alignment of p53's C-terminus that contains the Greenland Shark-specific Lysine insertion. b) Side-by-side comparison of the Alphafold3 structure predictions for the C-terminal fragments of monomeric p53 of the *S. rostratus* (green, little sleeper shark) and *S. microcephalus* (Greenland shark, brown) with the latter forming a longer alpha helix. The zoom-in is showing the insertion site with its surrounding amino acids. Alignment positions K388, K389, K382 green, K390 (insertion) red, E391 in blue. c) Alphafold3 predictions for a tetrameric arrangement of p53. Superposition of the NMR solution structure of the p53 oligomerization domain (human; PDB code 1OLG; red) with the predicted model of the GLS C-terminal fragment (residues 319-409; brown). The sequence stretch H387-K-K-K390-E-K-E393 (positions by alignment, compare with a) with heavy atoms of the side chain (green) is shown with the lysine insert K390 (yellow). d) as in c) but without the K390 insertion (residues 319-408; blue). In addition, possible

stabilizing interactions between the motif shown and other residues of p53 (located in different chains; E55, R310) are indicated by yellow dashes (respective residues marked in magenta).

Genome Browser

To visualize and access the GLS genome and annotation, we created a genome hub for the UCSC genome browser (Nassar et al. 2023). The hub also shows GC content, repeats, BUSCO hits, and RNA-seq data of different GLS tissues mapped to the GLS genome assembly. The genome annotation and BUSCO hit tracks are searchable and can be used to find the locations and structures of genes. BLAT (Kent 2002) and *in silico* PCR (Kuhn et al. 2013) services were also set up and made available for the GLS genome assembly through the tools tab on the UCSC genome browser page. BLAT can be used to find GLS genome sequences similar to the users' query sequence, while *in silico* PCR can be used to predict PCR product given the primer sequences and identify potential off-target sequences that may also be amplified with those primers. The hub link can be found at http://glshark.leibniz-fli.de.

Taken together, we provide an online resource for easy access to the GLS genome and its gene annotation.

Discussion

Our study provides a chromosome-level assembly of the 6.45 Gb Greenland shark genome with an annotation of 22,634 protein-coding genes and hundreds of thousands of repetitive elements. Given the limited availability of Methuselah genomes, the Greenland shark genome provides a cornerstone for studying the genetic adaptations enabling extreme longevity.

Compared to other large deep-sea animals, we, thus far, identified fewer positively selected genes: 8 in the GLS compared to 15 in the bowhead whale, 21 in the whale shark, and 61 in the great white shark (Keane et al. 2015; Marra et al. 2019). Although few positively selected genes were found in the GLS, they may represent adaptations to multiple critical hallmarks of aging, *i.e.*, genome instability, cellular senescence, loss of proteostasis, and disabled macroautophagy (López-Otín et al. 2023). Furthermore, we identified a network of functionally connected gene duplications in the GLS exclusively associated with double-strand break repair mechanisms and the p53 pathway, an essential regulator of DNA damage response (Williams and Schumacher 2016).

These results are in line with a large body of evidence in mammals linking double-strand break repair and longevity. For example, in elephants, the increased copy number of tumor suppressor genes, especially *TP53*, is thought to be a significant contributor to their ability to combine a considerable body size with a long lifespan (Vazquez and Lynch 2021).

Recent comparative transcriptomic studies also identified a positive correlation between the expression of DNA repair genes and the lifespan in mammals (Liu et al. 2023). For instance, the DNA repair gene XRCC6 carries a unique amino acid substitution in the Galapagos giant tortoise (Quesada et al. 2019) , and duplication of DNA repair genes was detected in long-lived turtles (Glaberman et al. 2021). In addition, DNA repair genes are enriched for positive selection when comparing the longest-lived rockfish species (maximum lifespan ≥ 105 years) with the shorter-lived species of the genus (Kolora et al. 2021). At a functional level, double-strand break repair is more efficient in long-lived mammals. The somatic mutation rate scales with mammalian lifespan (Cagan et al. 2022), highlighting a pattern of evolutionary convergence to protect the genomic integrity of long-lived species. Therefore, our preliminary data may emphasize the role of DNA repair for longevity.

Furthermore, p53 itself is affected by a unique amino acid insertion in the conserved C-terminal region of the protein. This is an interesting parallel to the other very distantly related deep-sea animals, as all of these species show adaptations in key effectors of DNA repair and tumor suppression, *e.g., PARP14 / NF2* (GLS); *ERCC1*/*ST13* (bowhead whale); *FIGL1 / CMTM7* (whale shark); *POLD3 / MDM4* (great white shark). Since these species can be assumed to have over a thousand times more cells than humans, it could be speculated that they are also more prone to cancer (Peto's paradox, (Tollis et al. 2017)). Therefore, adapting the above mechanisms may have been a critical step in evading the development of malignancies.

However, an analysis of the adaptations in the four genomes of large deep-sea animals also reveals critical differences between the shorter-lived species great white shark (70 years maximum life span) and whale shark (130 years) and the even longer-lived bowhead whale (211 years) and GLS (392 years). The authors of the Great White Shark genome paper explained that they "did not find copy number variants of genes associated with genome stability". They concluded that the positively selected genes represent "fine-tuning mechanisms related to the maintenance of genome integrity" (Marra et al. 2019). In contrast, the bowhead whale genome also contains gene duplications of individual essential DNA repair genes (Keane et al. 2015), such as *PCNA*.

Another interesting parallel concerns the shortest and longest-lived of the four species: compared to most other sharks, the great white shark (73 years) shows a substantial expansion of retrotransposon sequences (Marra et al. 2019), which is even more pronounced in the GLS (**Fig. 1c**). The expansion of retrotransposon sequences in both species contributed substantially to the increased genome sizes. In the case of the GLS, we have provided evidence that this mechanism may have contributed to the expansion of DNA repair gene sequences. Given that retrotransposons themselves are a source of double-strand breaks (Warkocki 2023), our findings suggest that the evolution of DNA repair genes and retrotransposons may be interlinked in the

GLS. Specifically, retrotransposon activity may have led to the expansion of DNA repair-associated retrogenes, which, in turn, allowed for the tolerance of higher retrotransposon activity. While our observations require additional analyses and validations, it is tempting to speculate that this cycle might have been a driving force in the evolution of the GLS genome and the shark's longevity.

Materials and Methods

Sample collection

Analyzed sharks were caught from 2021 - 2023 by scientific longlining from a depth between approximately 200 and 600 meters in the fjords in the vicinity of Narsaq and Qeqertoq, South Western Greenland from R/V Dana, and in Disko Bay in the area of Qeqertarsauq, North Western Greenland from R/V Porsild. All sampling was carried out following laws and regulations and with authorization from the Government of Greenland (Ministry of Fisheries, Hunting & Agriculture, permit Nr. 2020-26794 and 2023-6108).

The sharks were euthanized immediately after capture by an initial blow to the head with a bolt pistol for cattle, after which the spinal cord was transected. The tissue samples were collected as soon as possible, flash frozen in liquid nitrogen and then stored in a -80 freezer. Some samples were returned to Denmark on dry ice, and others were shipped in a freezing container.

Smaller sleeper sharks were collected in the vicinity of Genoa, Italy. One eye of 3 smaller sleeper sharks was subsequently prepared at the Aarhus AMS Centre (Department of Physics and Astronomy, Aarhus University, Denmark) by isolating the embryonic eye lens nucleus under light microscopy from concentrically arranged layers of secondary fiber cells. A 4-5 mg subsample of the innermost part of the embryonic nucleus was used for isotopic analyses with Accelerator Mass Spectrometry (AMS), as described in Nielsen et al., 2016.

Sequencing

DNA was extracted by directly adding a digest buffer (60 mM Tris pH 8.0, 100 mM EDTA pH 8.0, 0.5% SDS) to the frozen tissue. Proteinase K was mixed into the solution until reaching a final concentration of 500 µg/ml. The mixture was kept in a water bath at 54°C until the tissue was completely dissolved. Subsequently, precipitation with phenol-chloroform was performed three times. After incubation at room temperature (3 min) and subsequent centrifugation (5 min), the aqueous phase was transferred into a fresh tube. The interphase was washed with chloroform and precipitated with 0.8 parts isopropanol. After gently swirling the mixture, it was incubated for 15 min at room temperature and centrifuged for 30 minutes. After discarding the supernatant, the pellet was washed with 70% ethanol and centrifuged again. Subsequently, the supernatant was discarded. The pellet was eluted by adding water. The mixture was placed in the refrigerator until the DNA was completely dissolved.

For Illumina sequencing, isolated DNA was quantified and quality-checked using a TapeStation 4200 Instrument in combination with Genomic DNA ScreenTape (Agilent Technologies). According to the manufacturer's instructions, 1 µg of DNA was sheared on a Covaris M220 to a fragment size of 400 bp using the microTUBE–50 AFA Fiber Screw-Cap System. Illumina DNA libraries were prepared from 400 ng using NEBNext Ultra II DNA Library Prep Kit in combination with NEBNext Multiplex Oligos for Illumina (Unique Dual Index UMI Adaptors DNA) according to the manufacturer's instructions (New England Biolabs). Isolated total RNA was quantified and quality checked using a TapeStation 4200 Instrument in combination with RNA ScreenTape (Agilent Technologies). Illumina RNA libraries were prepared from 800 ng using NEBNext Ultra II Directional RNA Library Prep Kit in combination with NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Multiplex Oligos for Illumina (Unique Dual Index UMI Adaptors RNA) according to the manufacturer's instructions (New England Biolabs). Quantification and quality check of DNA and RNA libraries was performed using a TapeStation 4200 Instrument in combination with D5000 ScreenTape (Agilent Technologies).

Illumina DNA and RNA libraries were pooled and loaded on NovaSeq 6000 (Illumina). Sequencing was performed using a paired-end 2x101 bp configuration on S4 flow cell (Reagent Kit v1.5 200 cycles) and S2 flow cell (Reagent Kit v1.5 200 cycles). Sequencing data were converted in FASTQ format and demultiplexed using bcl2fastq software (v2.20.0.422).

HiC-libraries were prepared in part using the Arima HiC+ Kit and the Arima Library Prep Kit V2, and in part as published previously (Eremenko et al. 2021). Briefly, nuclei were isolated by homogenizing liver tissue with a Dounce Homogenizer in a fixation buffer containing 1% formaldehyde. Fixation was performed in a final volume of 10 ml of fixation buffer with moderate rotation. The reaction was quenched with 2.5 M glycine (to a final concentration of 0.2 M) for 5 minutes at room temperature with rotation. Fixed cells were washed twice, and nuclei were purified using centrifugation with a sucrose cushion. Chromatin digestion was performed by adding 200 units of DpnII restrictase per sample and incubating at 37°C overnight with constant agitation. The overhangs of the digested DNA were labeled by adding biotin-14-dATP using DNA Polymerase I, Large (Klenow) Fragment, and a mixture of other nucleotides. After pelleting the nuclei, a ligation mix (1x NEB T4 DNA ligase buffer, 1% Triton X-100, BSA, 400 U/µl T4 DNA Ligase) was added, and the DNA fragments were ligated with overnight incubation at room temperature with gentle rotation. DNA was reverse-crosslinked by incubation with Proteinase K and SDS (final concentration 2%) and then isolated using the phenol-chloroform method. DNA was sheared using a Bioruptor Pico (Diagenode) ultrasound device. MyOne Streptavidin C1 Dynabeads (Thermo Fisher Scientific) was used to pull down biotinylated fragments. Preparation for Illumina sequencing and final amplification were performed using reagents from the NEBNext Ultra II DNA Library Prep Kit for Illumina in combination with NEBNext Multiplex Oligos for Illumina

(Unique Dual Index UMI Adaptors DNA) according to the manufacturer's instructions. A mock PCR was done to finely tune the number of cycles needed for library amplification (7 cycles). Before sequencing, the size distribution and concentration of the libraries were assessed using the Agilent TapeStation System in combination with D5000 ScreenTape (Agilent Technologies). Libraries were pooled and sequenced using one NovaSeq 6000 run. System runs in SP/2x 51 cycle (paired-end)/standard-loading workflow mode. Sequence information was converted to FASTQ format using bcl2fastq v.2.20.0.422.

For PacBio sequencing, high molecular weight DNA was sheared to 7-12 kb using the MegaRuptor3 system (Speed 32, T=56min). We used the SMRTbell prep kit 3.0. Specifically, the sheared DNA was subjected to 1x MRTbell beads cleanup to reduce volume. Single-stranded DNA overhangs were removed in the first step. DNA strand breaks were repaired, and an A-overhang was added. Subsequently, non-barcoded SMRTbell adapters were ligated, followed by AmpurePB purification with two ethanol (80%) washing steps. A nuclease treatment was performed to remove residual adapters. Last, the high molecular weight fraction of the library was enriched by an AMPure PB beads size selection with a 3.1x ratio. For all runs, a total of 10 sample preps were done. All cells were sequenced on the PacBio Sequel II instrument with (10) SMRT cells using 30h video times using version 2.0 sequencing chemistry and 2h pre-extension. HiFi/CCS analysis was performed using CCS v6.3.0 with default parameters.

Genome assembly

Contigs were assembled using hifiasm (Cheng et al. 2021) with PacBio HiFi reads as input filtered for potential contaminations using Kraken2 (Wood et al. 2019). Scaffolding was conducted to feed the contig assembly and HiC-reads into the 3D-DNA pipeline, with subsequent manual curation using a juicebox (Durand et al. 2016; Dudchenko et al. 2017). Genome completeness was measured by applying BUSCO (Simao et al. 2015) and using single-copy vertebrate genes as a reference.

Genome annotation

Repeats were annotated using RepeatModeler and RepeatMasker (Tempel 2012; Flynn et al. 2020). Gene annotation was conducted using the braker3 pipeline (Brůna et al. 2021) with protein hints from the great white shark genome (Rhie et al. 2021) and aligned RNA-seq reads. Before this, RNA-seq reads were quality trimmed, adapter clipped, and mapped with Trimmomatic, Cutadapt, and STAR (Martin 2011; Dobin et al. 2013; Bolger et al. 2014). Functional annotation was done using eggNogMapper2 (Cantalapiedra et al. 2021). Additionally, we used BLAST (Altschul et al. 1990) to annotate unlabeled gene isoforms.

Transcriptome assembly of *Somniosus rostratus*

RNA-seq reads were quality trimmed and adapter clipped using Trimmomatic and Cutadapt (Martin 2011; Bolger et al. 2014). Transcriptome assembly and coding sequence extraction were conducted using Trinity and TransDecoder (Haas et al. 2013).

Comparative genomics

Analysis of transposable elements was repeated for the great white shark using RepeatModeler and RepeatMasker (Tempel 2012; Flynn et al. 2020).

Tree inference and positive selection analysis was conducted using PosiGene (Sahm et al. 2017). Coding sequences derived from our GLS genome assembly, a little sleeper shark transcriptome assembly, and publicly available coding sequences from 12 other sharks were used as input. Of the latter, ten were downloaded from NCBI (*Amblyraja radiata, Carcharodon carcharias, Chiloscyllium plagiosum, Hypanus sabinus, Leucoraja erinacea, Mobula hypostoma, Pristis pectinate, Rhincodon typus, Scyliorhinus canicula, Stegostoma tigrinum*) and two from the Squalomix webpage (*Scyliorhinus torazame, Chiloscyllium punctatum*) (Nishimura et al. 2022).

Genes duplications were analyzed by examining all gene symbols assigned more than once by the gene annotation (n=1,224). All protein sequences of all respective genes were compared against the protein sequences of ten other shark species using NCBI BLAST+. We checked for each of these ten species whether all GLS sequences associated with a specific gene symbol hit exactly one gene in the other shark species. If this was the case for each of the ten shark species for a specific gene symbol, we accepted the respective gene as exclusively duplicated in the GLS (n=81).

Dot plots and chromosome-level alignments were obtained using the D-GENIES server (Cabanettes and Klopp 2018), and visualized via ggplot using custom scripts.

Structure prediction

Structure predictions of p53 were performed using the AlphaFold3 web server with default options (Abramson et al. 2024). For graphical representations the program ChimeraX (Pettersen et al. 2021) was employed.

UCSC Genome Browser Hub

The make_hub.py script (Hoff 2019) was used to convert the genome assembly into 2bit file format. This script was also used to get the genome's GC content track and to make hub descriptions and hub, genome, and track database definition file drafts. The same script was used to convert RNA-seq BAM files into bigWig files for the hub.

The genome annotation file was adjusted by ensuring correct hierarchy and attributes for gene, transcript, exon and CDS rows using gtftk (Lopez et al. 2019) and a custom bash script. The

adjusted annotation was converted to USCS browser hub appropriate format using UCSC Kent Utilities gtfToGenePred, genePredToBigGenePred and bedToBigBed (Kent et al. 2010). To add BUSCO hits track, the scores in the bed file were divided by 10 to make the range of the scores compatible with the UCSC browser and bedToBigBed was used to convert the BUSCO hits bed file to USCS browser readable bigBed file format. ixIxx from USCS Kent tools was used to create indices for both the annotation and the BUSCO hits tracks to make them searchable on the hub. Repeats from RepeatMasker were converted into UCSC genome browser usable format using RepeatMasker tool rmToTrackHub.pl (Tempel 2012) followed by bedToBigBed. gfServerHuge from USCS Kent tools was used to start untranslated and translated BLAT (Kent 2002) and in silico PCR (Kuhn et al. 2013) services for the GLS genome assembly.

Acknowledgments

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493–500.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.

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

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* 3:lqaa108.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way.Tullius T, editor. *PeerJ* 6:e4958.

Cagan A, Baez-Ortega A, Brzozowska N, Abascal F, Coorens THH, Sanders MA, Lawson ARJ, Harvey LMR, Bhosle S, Jones D, et al. 2022. Somatic mutation rates scale with lifespan across mammals. *Nature* 604:517–524.

Cailliet GM, Andrews AH, Burton EJ, Watters DL, Kline DE, Ferry-Graham LA. 2001. Age determination and validation studies of marine fishes: do deep-dwellers live longer? *Experimental Gerontology* 36:739–764.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38:5825–5829.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170–175.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92–95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* 3:99–101.

Edrey YH, Hanes M, Pinto M, Mele J, Buffenstein R. 2011. Successful aging and sustained good health in the naked mole rat: a long-lived mammalian model for biogerontology and biomedical research. *ILAR J* 52:41–53.

Eremenko E, Golova A, Stein D, Einav M, Khrameeva E, Toiber D. 2021. FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C. *STAR Protocols* 2:100643.

Finch CE. 1998. Variations in Senescence and Longevity Include the Possibility of Negligible Senescence. *The Journals of Gerontology: Series A* 53A:B235–B239.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* 117:9451–9457.

Glaberman S, Bulls SE, Vazquez JM, Chiari Y, Lynch VJ. 2021. Concurrent Evolution of Antiaging Gene Duplications and Cellular Phenotypes in Long-Lived Turtles. *Genome Biology and Evolution* 13:evab244.

Gritsina G, Yu J. 2023. CXCR7 as a novel therapeutic target for advanced prostate cancer. *Oncogene* 42:785–792.

Grundy MK, Buckanovich RJ, Bernstein KA. 2020. Regulation and pharmacological targeting of RAD51 in cancer. *NAR Cancer* 2:zcaa024.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:1494–1512.

Hansen PM. 1963. Tagging Experiments with the Greenland shark (Somniosus microcephalus (Block and Schneider)) in Subarea 1. *International Comission for the Northwest Atlantic Fisheries.* Special Publication no. 4:172–175.

Hoff KJ. 2019. MakeHub: Fully Automated Generation of UCSC Genome Browser Assembly Hubs. *Genomics, Proteomics & Bioinformatics* 17:546–549.

Holtze S, Gorshkova E, Braude S, Cellerino A, Dammann P, Hildebrandt TB, Hoeflich A, Hoffmann S, Koch P, Terzibasi Tozzini E, et al. 2021. Alternative Animal Models of Aging Research. *Frontiers in Molecular Biosciences* 8.

Houtkooper RH, Mouchiroud L, Ryu D, Moullan N, Katsyuba E, Knott G, Williams RW, Auwerx J. 2013. Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* 497:451–457.

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI, et al. 2015. Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep* 10:112–122.

Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* 12:656–664.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26:2204–2207.

Kim H, Kwak N-J, Lee JY, Choi BH, Lim Y, Ko YJ, Kim Y-H, Huh P-W, Lee K-H, Rha HK, et al. 2004. Merlin Neutralizes the Inhibitory Effect of Mdm2 on p53*. *Journal of Biological Chemistry* 279:7812–7818.

Kolora SRR, Owens GL, Vazquez JM, Stubbs A, Chatla K, Jainese C, Seeto K, McCrea M, Sandel MW, Vianna JA, et al. 2021. Origins and evolution of extreme life span in Pacific Ocean rockfishes. *Science* 374:842–847.

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Briefings in Bioinformatics* 14:144–161.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* 34:1812–1819.

Laurini E, Marson D, Fermeglia A, Aulic S, Fermeglia M, Pricl S. 2020. Role of Rad51 and DNA repair in cancer: A molecular perspective. *Pharmacology & Therapeutics* 208:107492.

Li X, Chen H, Xue L, Pang X, Zhang X, Zhu Z, Zhu W, Wang Z, Wu H. 2017. p53 performs an essential role in mediating the oncogenic stimulus triggered by loss of expression of neurofibromatosis type 2 during in vitro tumor progression. *Oncol Lett* 14:2223–2231.

Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417–425.

Linke SP, Sengupta S, Khabie N, Jeffries BA, Buchhop S, Miska S, Henning W, Pedeux R, Wang XW, Hofseth LJ, et al. 2003. p53 Interacts with hRAD51 and hRAD54, and Directly Modulates Homologous Recombination12. *Cancer Research* 63:2596–2605.

Liu W, Zhu P, Li Meng, Li Z, Yu Y, Liu G, Du J, Wang X, Yang J, Tian R, et al. 2023. Large-scale across species transcriptomic analysis identifies genetic selection signatures associated with longevity in mammals. *The EMBO Journal* 42:e112740.

Lopez F, Charbonnier G, Kermezli Y, Belhocine M, Ferré Q, Zweig N, Aribi M, Gonzalez A, Spicuglia S, Puthier D. 2019. Explore, edit and leverage genomic annotations using Python GTF toolkit. *Bioinformatics* 35:3487–3488.

López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2023. Hallmarks of aging: An expanding universe. *Cell* 186:243–278.

Magalhães JP de, Costa J, Church GM. 2007. An Analysis of the Relationship Between Metabolism, Developmental Schedules, and Longevity Using Phylogenetic Independent Contrasts. *The Journals of Gerontology: Series A* 62:149–160.

Marlétaz F, de la Calle-Mustienes E, Acemel RD, Paliou C, Naranjo S, Martínez-García PM, Cases I, Sleight VA, Hirschberger C, Marcet-Houben M, et al. 2023. The little skate genome and the evolutionary emergence of wing-like fins. *Nature* 616:495–503.

Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Pavinski Bitar P, Richards VP, Komissarov A, Rayko M, Kliver S, et al. 2019. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proceedings of the National Academy of Sciences* 116:4446–4455.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 10.14806/ej.17.1.200*.

Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee BT, et al. 2023. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Research* 51:D1188–D1195.

Nicolae CM, Aho ER, Choe KN, Constantin D, Hu H-J, Lee D, Myung K, Moldovan G-L. 2015. A novel role for the mono-ADP-ribosyltransferase PARP14/ARTD8 in promoting homologous recombination and protecting against replication stress. *Nucleic Acids Research* 43:3143–3153.

Nielsen J. 2018. The Greenland shark (Somniosus microcephalus) - Diet, tracking and radiocarbon age estimates reveal the world's oldest vertebrate.

Nielsen J, Hedeholm RB, Heinemeier J, Bushnell PG, Christiansen JS, Olsen J, Ramsey CB, Brill RW, Simon M, Steffensen KF, et al. 2016. Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (Somniosus microcephalus). *Science* 353:702–704.

Nishimura O, Rozewicki J, Yamaguchi K, Tatsumi K, Ohishi Y, Ohta T, Yagura M, Niwa T, Tanegashima C, Teramura A, et al. 2022. Squalomix: shark and ray genome analysis consortium and its data sharing platform [version 1; peer review: 2 approved]. *F1000Research* [Internet] 11. Available from: https://f1000research.com/articles/11-1077/v1

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* 30:70–82.

Quesada V, Freitas-Rodríguez S, Miller J, Pérez-Silva JG, Jiang Z-F, Tapia W, Santiago-Fernández O, Campos-Iglesias D, Kuderna LFK, Quinzin M, et al. 2019. Giant tortoise genomes provide insights into longevity and age-related disease. *Nature Ecology & Evolution* 3:87–95.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592:737–746.

Ricklefs RE. 2010. Life-history connections to rates of aging in terrestrial vertebrates. *Proceedings of the National Academy of Sciences* 107:10314–10319.

Sahm A, Bens M, Platzer M, Szafranski K. 2017. PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Res* 45:e100.

Shin EJ, Shin HM, Nam E, Kim WS, Kim J, Oh B, Yun Y. 2012. DeSUMOylating isopeptidase: a second class of SUMO protease. *EMBO reports* 13:339–346.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

Smith TR, Bishop A, Guthridge J, Hocking R, Horning M, Lowe CG. 2022. Capture, husbandry, and oxygen consumption rate of juvenile Pacific sleeper sharks (Somniosus pacificus). *Environmental Biology of Fishes* 105:1519–1534.

Speakman JR. 2005. Body size, energy metabolism and lifespan. *Journal of Experimental Biology* 208:1717–1730.

Ste-Marie E, Watanabe YY, Semmens JM, Marcoux M, Hussey NE. 2020. A first look at the metabolic rate of Greenland sharks (Somniosus microcephalus) in the Canadian Arctic. *Scientific Reports* 10:19297.

Takaya K, Asou T, Kishi K. 2022. Selective Elimination of Senescent Fibroblasts by Targeting the Cell Surface Protein ACKR3. *International Journal of Molecular Sciences* 23.

Tempel S. 2012. Using and Understanding RepeatMasker. In: Bigot Y, editor. Mobile Genetic Elements: Protocols and Genomic Applications. Totowa, NJ: Humana Press. p. 29–51. Available from: https://doi.org/10.1007/978-1-61779-603-6_2

Tollis M, Boddy AM, Maley CC. 2017. Peto's Paradox: how has evolution solved the problem of cancer prevention? *BMC Biology* 15:60.

Tomljanovic Z, Patel M, Shin W, Califano A, Teich AF. 2018. ZCCHC17 is a master regulator of synaptic gene expression in Alzheimer's disease. *Bioinformatics* 34:367–371.

Vazquez JM, Lynch VJ. 2021. Pervasive duplication of tumor suppressors in Afrotherians during the evolution of large bodies and reduced cancer risk.Rokas A, Wittkopp PJ, Stearns SC, Gorbunova V, editors. *eLife* 10:e65041.

Voituron Y, de Fraipont M, Issartel J, Guillaume O, Clobert J. 2011. Extreme lifespan of the human fish (Proteus anguinus): a challenge for ageing mechanisms. *Biol Lett* 7:105–107.

Warkocki Z. 2023. An update on post-transcriptional regulation of retrotransposons. *FEBS Letters* 597:380–406.

Watanabe YY, Lydersen C, Fisk AT, Kovacs KM. 2012. The slowest fish: Swim speed and tail-beat frequency of Greenland sharks. *Journal of Experimental Marine Biology and Ecology* 426–427:5–11.

Williams AB, Schumacher B. 2016. p53 in the DNA-Damage-Repair Process. *Cold Spring Harbor Perspectives in Medicine* [Internet] 6. Available from: http://perspectivesinmedicine.cshlp.org/content/6/5/a026070.abstract

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology* 20:257.

Yamaguchi K, Uno Y, Kadota M, Nishimura O, Nozu R, Murakumo K, Matsumoto R, Sato K, Kuraku S. 2023. Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization. *Genome Research* 33:1527–1540.

Yang F, Bian T, Zhan X, Chen Z, Xing Z, Larsen NA, Zhang X, Shi Y. 2023. Mechanisms of the RNA helicases DDX42 and DDX46 in human U2 snRNP assembly. *Nature Communications* 14:897.

Zhao S, Li CM, Luo XM, Siu GKY, Gan WJ, Zhang L, Wu WKK, Chan HC, Yu S. 2017. Mammalian TRAPPIII Complex positively modulates the recruitment of Sec13/31 onto COPII vesicles. *Scientific Reports* 7:43207.