RNA analysis of the longest living vertebrate Greenland shark revealed an abundance of LINE-like elements in its transcriptome

Martin Bartas[1*], Jiří Červeň[1], Natália Valková[2], Adriana Volná[3], Michaela Dobrovolná[2,4], Lucie Šislerová[2,4], Hörður Baldvinsson[5], Petr Pečinka[1], Václav Brázda[2,4,**]

[1]*Department of Biology and Ecology, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic*
[2]*Institute of Biophysics of the Czech Academy of Sciences, 612 65 Brno, Czech Republic*
[3]*Department of Physics, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic*
[4]*Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 612 00, Brno, Czech Republic.*
[5]*Knowledge Center of Vestmannaeyjar, Ægisgata 2/900, Vestmannaeyjar, Iceland*

Abstract

The Greenland shark (*Somniosus microcephalus*) is an enigmatic species known for its exceptional longevity and extraordinary adaptations to the cold environment. This animal lives in the arctic and subarctic regions of the North Atlantic Ocean. Surprisingly, even though it is a vertebrate with the longest known lifespan, its transcriptome has not been studied yet. Therefore, we isolated and analyzed RNA in the Greenland shark samples. Our findings reveal some important information about the possible genetic mechanisms that could contribute to its longevity. We identified a highly expressed long interspersed nuclear element-like transcript (LINE-like) that is supposed to be associated with extended lifespan and resilience to age-related diseases, possibly through an improved telomere maintenance mechanism. This research not only contributes to our understanding of the biology and evolution of the Greenland shark but could also have implications for human longevity research.

Key words: Greenland shark, transcriptome, bioinformatics, LINE-1, longevity research, *Somniosus microcephalus*

DOI: 10.5817/CPR2023-2-17

Received November 28, 2023, *accepted* January 6, 2024.
[*]Corresponding author: [*]M. Bartas <martin.bartas@osu.cz>; [**]V. Brázda <vaclav@ibp.cz>
Acknowledgements: We express our sincere gratitude to the Czech Neuron Foundation for their generous financial support, which enabled us to conduct the inaugural expedition to Iceland in August 2022. This expedition not only facilitated valuable scientific connections and experiences but also laid the foundation for future research in this field. Our great thanks go to the Bjarnarhöfn Shark Museum for providing us sample of the Greenland shark and to Klara Jakobsdóttir (Marine and Freshwater Research Institute) for sample storage and valuable advices. In addition to the team of authors, we thank other important people who made the research expedition to Iceland possible: Photographer Tomas Vit, Research vessel captain Sigmundur Einarsson, and Guðjón Hildibrandsson from the Shark Museum. This work was supported by the European Union under the LERCO project number CZ.10.03.01/00/22_003/0000003 via the Operational Programme Just Transition. Computational resources were provided by the e-INFRA CZ project (ID:90254) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Introduction

The life of every living creature is limited; however, differences in the lifespan of various organisms are enormous from a few hours, days, weeks up to several thousand years for some tree species. In vertebrates, this range is much smaller from a few months or years to decades. Only exceptionally a few species of vertebrates could overcome hundreds of years, for example, the giant tortoise (*Aldabrachelys gigantea)* or the whale *Balaena mysticetus*. However, the record holder among vertebrates is the Greenland shark (*Somniosus microcephalus*), which can live more than 400 years.

The Greenland shark is known for its remarkable longevity. Radiocarbon dating of crystals within their eye lenses estimated the age of one about six-meter-long animal for 392±120 years, suggesting that it was born between 1504 and 1744 (Nielsen et al. 2016). Although it could be partly explained by its way of life and environment (slow motion, cold water, *etc.*), it is evident, that other organisms with similar modes of life, comprising evolutionarily closely related vertebrates, live much more shorter (De Magalhães and Costa 2009). The Greenland shark has long intrigued scientists and researchers due to its unique characteristics and remarkable adaptations to its cold environment (Nielsen et al. 2014, Chernova et al. 2015). One of the distinguishing features of the Greenland shark is its immense size. Growing to an average length of 4 to 6 meters (13 to 20 feet) and weighing up to 6,000 kilograms (13,200 pounds), it is one of the largest shark species in existence (Leclerc et al. 2012, Augustine et al. 2017). Its elongated body is covered in thick and rough skin that helps maneuver through icy waters. These sharks have a dark gray or blackish coloration, which helps to conceal themselves within the depths they inhabit. Their eyes are (similarly to other shark species) equipped with tapetum lucidum, which reflect light thanks to a layer of parallel cells with silver guanine crystals, and ´thus enhances shark vision in low light conditions, a crucial adaptation for their deep-sea lifestyle. Interestingly, many individuals (up to 85%) have eyes infected with the small parasitic crustacean *Ommatokoita elongata* attached to the cornea (Beck and Mansfield 1969, Yopak et al. 2019). Another intriguing aspect of the Greenland shark is the low metabolic rate (Ste-Marie et al. 2020), allowing them to adapt to the limited food resources of their environment. The diet of the Greenland sharks consists of carcasses and various animals, including fish, seals, and even the occasional polar bear (Nielsen et al. 2014). Their feeding habits are facilitated by rows of sharp teeth, capable of tearing flesh and crushing bones. For a long time, it was quite mysterious how these slow-moving sharks could catch the seals, until 2010, when a detailed report on this issue was published (Lucas and Natanson 2010). The authors focused on the unexpected deaths of seals on Sable Island, Canada between 1993 and 2001. Many seal carcasses bore traces of sharp teeth and wounds corresponding to the Greenland shark (Watanabe et al. 2012).

However, cold water and slow metabolisms are probably only one part of the reasons for the longevity of the Greenland shark, other animals, even other species of sharks, sharing the same environment do live usually less than 10-times shorter life. Therefore, it is clear that the exceptional longevity has some molecular and genetic background. Several hallmarks of longevity, including changes in the DNA repair system, which should protect genetic material in long-lived organisms before mutations and cancer development, have been described (Clarke and Mostoslavsky 2022, Bartas et al. 2021). One of the critical parts of vertebrate genomes are the ends of the chromosomes, and the shortening of the

telomers during life is considered a very crucial factor leading to cell senescence and aging (Kim Sh et al. 2002). Interestingly, special parts of the genome called LINE-1 (Long Interspersed Element 1) were shown to play an important role in telomere maintenance (Aschacher et al. 2020). LINE is a type of transposable element (also known as retrotransposon) without long terminal repeats (LTR) that is present in the genomes across the eukaryotic kingdom, including humans (Thawani 2023). Non-LTR retrotransposons typically contain poly(A) or simple repeats at their 3'terminus (Kojima 2019). It is a repetitive DNA sequence that can move or "transpose" within the genome, potentially causing genetic mutations or rearrangements. Although LINE elements are often considered "junk DNA" with no known function, recent research suggests that they may play important regulatory role in genome organization (Lu et al. 2021), transgenerational inheritance (Mirabello et al. 2010), and are involved in embryonic and placental development (Sharif et al. 2023). Although it was believed that LINEs are generally active in embryonic or cancerous cells, a recent study found, that the expression of the LINE-1 element inspected in various human and mouse tissues (McKerrow et al. 2023) reached up to 25 TPM (transcripts per million) in various noncancerous tissues.

As there is no available information about the genomic and transcriptomic data of Greenland sharks except for the sequence of their mitochondrial DNA (Santaquiteria et al. 2017, Edwards et al. 2019), we decided to analyze its genome and transcriptome. The first step was the acquisition of suitable the Greenland shark samples. This was possible thanks to the support of the Czech Foundation Neuron expedition to Iceland (August 2022). Here, we present our first analyses of the Greenland shark transcriptome, which revealed a strong abundance of LINE-like mRNA sug-gesting its important involvement in its longevity.

Material and Methods

Sample acquisition

The first attempt to catch the Greenland shark was made in August 2022. The Fridrik Jesson motor vessel (MMSI: 251772110) was rented, along with an experienced sailor, and fishing equipment. The longline (200 m) with a buoy was set 30 kilometers east of Heimaey, the largest island in the Icelandic Vestmannaeyjar archipelago (63.4617753 N, 19.5810081 W). Rotten chicken and horse meat were used as bait, however, without success. The shark sample was obtained the following year (March 2023) as a bycatch on a fishing trawler and provided in cooperation with the Bjarnarhöfn Shark Museum (Iceland). Samples were from a three-meter-long female. Muscle and liver tissue samples were frozen for several days, stored in RNAlater (ThermoFisher Scientific) upon retrieval, and then stored at -80° before analyses.

RNA isolation, sequencing, and assembly

Approximately 200 mg of shark tissue was homogenized in 4 ml of TriReagent (MRC, Cincinnati, USA) using soft tissue homogenization tubes in MinyLys homogenizer (both Bertin, Berlin, Germany). The samples were homogenized in three cycles of shaking at 1 min. 5000 rpm and 1 min. ice bath. A modified TriReagent protocol was used for RNA isolation (samples were centrifuged at 12 000 g for 10 min. at 4°C prior to the addition of chloroform, and RNA was precipitated using isopropanol with a high salt precipitation solution (0.8 M sodium citrate and 1.2 M NaCl)). In the case of liver tissue, the lipid layer that appeared after homogenization and during phase separation was discarded. The RNA integrity number (RIN) was estimated using Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). RIN was low, so the kit for low-quality RNA was used for library preparation. Total RNA sequencing was carried out by DNA Link company (Seoul, Republic of South Korea) using TrueSeq total RNA library kit with Ribo-Zero plus solution (library preparation) and sequenced to 50Gb depth using the Illumina NovaSeq 6000 platform.

The resulting files containing 101bp-long paired-end reads were post-processed, *i.e.* adapters were computationally removed and nucleotides with PHRED score < 20 were filtered out. Assembly was made using the Trinity tool (Grabherr et al. 2011). The adapters were removed at first, bases with PHRED quality below 20 and reads shorter than 20nt were filtered out using the cutadapt tool. Trinity-v2.9.1 ran with the default setting for paired-end sequencing. The found unique RNA sequences assigned as LINE-like element were deposited in the NCBI database, accession numbers: OR828930.1, OR828936.1[1].

Data analysis and visualization

The nucleotide and dinucleotide compositions were calculated using a 'DNA composition' tool accessed from Sequence Manipulation Suite (Stothard 2000). Open reading frames (ORFs) within the LINE-like transcripts were identified using the Expasy Translate Tool (Artimo et al. 2012) (*Supplementary material 1*). The domain composition of the identified ORFs was determined through the Web server of the Conserved Domain Database (Lu et al. 2020). A schematic diagram of domain composition was made using the DOG tool (v. 2.0) (Ren et al. 2009). Protein parameters were computed using the Expasy ProtParam tool (Gasteiger et al. 2005). Subcellular localization was predicted using CELLO v.2.5. tool for eukaryotic protein sequences (Yu et al. 2006) (*Supplementary material 2*). The secondary structure of identified protein sequences was predicted using a GOR4 method (Garnier et al. 1996). For modeling the LINE-1/ORF2 structure of the Greenland shark, the trRosetta web server (Du et al. 2021) was used. PDB files can be found in the *Supplementary material 3*. The superposition of the Greenland shark LINE-1/ORF2 and the human was made using FATCAT 2.0 web server (Li et al. 2020) in flexible alignment mode (*Supplementary material 4*). Structures were visualized in the UCSF Chimera standalone tool (Pettersen et al. 2004).

Phylogenetic tree construction

Protein sequences similar to the Greenland shark ORF2 (liver) were gathered using NCBI blastp search with Default parameters (Altschul et al. 1990). Then the final dataset was made containing 50 manually selected non-redundant representative protein sequences comprising both newly identified the Greenland shark ORF2 (*Supplementary material 1*). The phylogenetic analysis was performed on the Phylogeny.fr platform (Dereeper et al. 2008) and comprised the following steps. Protein sequences were aligned with MUSCLE (v3.8.31) (Edgar 2004) configured for the highest accuracy (MUSCLE with default settings). After alignment, ambiguous regions (*i.e.* containing gaps and/or poorly aligned) were removed with Gblocks™ (v0.91b) (Castresana 2000) using the these parameters: minimum length of a block after gap cleaning was 10; no gap positions were allowed in the final alignment; all segments with contiguous non-conserved positions bigger than 8 were rejected; minimum number of sequences for a flank position was 85%. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.1/3.0 aLRT) (Guindon and Gascuel 2003). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.028) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.100). Reliability for internal branch was assessed using the aLRT test (SH-Like) (Anisimova and Gascuel 2006). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3) (Chevenet et al. 2006).

Results

Research expedition to Iceland

The two-week research expedition to Iceland was carried out in August 2022. Six people in total were employed, two from the Institute of Biophysics of the Czech Academy of Sciences, three from the University of Ostrava and one photographer (Fig. 1A). The temporary base was settled at Vestmannaeyjar and daily cruises to Reynisfjara "Shark Bay" (Fig. 1B) were taken, using a local marine research boat called *Fridrik Jesson* (Fig. 1C). Rotten chicken and horse meat were used as bait (Fig. 1D). Using the longline fishing technique, we carried out an attempt to catch the living individual of the Greenland shark for fresh tissue sampling from the dorsal fin, however, unsuccessfully. The baits were partially consumed, but the hooks remained intact. The behavioral biology of Greenland sharks is still largely unexplored (Grant et al. 2018), but it may be that during the summer months they seek deeper places in the ocean with colder water. Nevertheless, valuable local contacts were gained, which allowed the acquisition of samples in the future. This occurred in March 2023, when the Greenland shark was caught by a commercial fishing trawler and delivered to the Shark Museum in Bjarnarhöfn, where various tissues, including muscles and liver, were collected and further transported to the Czech Republic.

Fig. 1. Photodocumentation of Expedition Neuron 2022 - Behind the Secret of the Greenland shark's longevity. (A) Research members of the expedition and local crew, from the left side: Martin Bartas, Jiří Červeň, Václav Brázda, Hörður Baldvinsson (Managing Director, Knowledge Center of Vestmannaeyjar), Michaela Dobrovolná, Adriana Volná, and the local sailor Mr. Sigmundur Einarsson. (B) Research boat *Fridrik Jesson*. (C) The location where the expedition took place, Heimaye island is marked by a red arrow, and the approximate location where the baits were placed is in an orange oval. (D) Rotten chicken and horse meat baits and chained hooks (longline technique, each had a size of 30 cm).

Global characterization of liver and muscle transcriptome assemblies

In the domestic institutions, the total RNA was isolated and sent for the whole transcriptome sequencing (*see* Materials and Methods section). Although the integrity of isolated RNA wasn't very high (RIN = 3), 34.8 millions of paired-end reads were obtained from liver tissue, and 23.5 millions of paired-end reads in the case of muscle tissue. Detailed parameters of the resulting transcriptome assemblies are depicted in Table 1 below. Complete transcriptome assemblies in FASTA format will be released in a follow-up study.

Parameter of transcriptome assembly	Liver tissue	Muscle tissue
Maximal transcript length (bp)	14,563	77,128
Minimal transcript length (bp)	173	179
Mean transcript length (bp)	335	402
Median transcript length (bp)	274	279
Number of adenines (A)	3,525,149	12,499,120
Number of thymines (T)	3,640,665	12,900,693
Number of cytosines (C)	4,521,263	11,312,013
Number of guanines (G)	3,365,236	10,393,089
Overall number of sequenced bp	15,052,313	47,104,915
Overall number of sequenced transcripts	44,894	117,084
Percentual GC content	52.39%	46.08%

Table 1. Detailed parameters of the Greenland shark liver and muscle tissue transcriptome assemblies. Values were calculated using a Fasta Statistics tool (Kyran 2021) on the Galaxy Europe web server (The Galaxy community 2022[3]).

Although most of the highly expressed transcripts have their known homologs, we have found highly abundant transcript without any clear homology in the NCBI database of RNAs. Therefore, we continue with a detailed characterization of this unique RNA transcript with 545 transcripts per million (TPM) in liver assembly and 937 TPM in muscle assembly (quantification of particular transcripts can be found in *Supplementary material 5*). This means that approximately every 2000th detected transcript in liver assembly (and 1000th in muscle assembly) was this unique sequence. The following analyses show that this sequence contains open reading frames and typical nucleotide sequence composition, which allows its characterization as the LINE-like sequence.

Characterization of LINE-like transcript within the liver transcriptome assembly

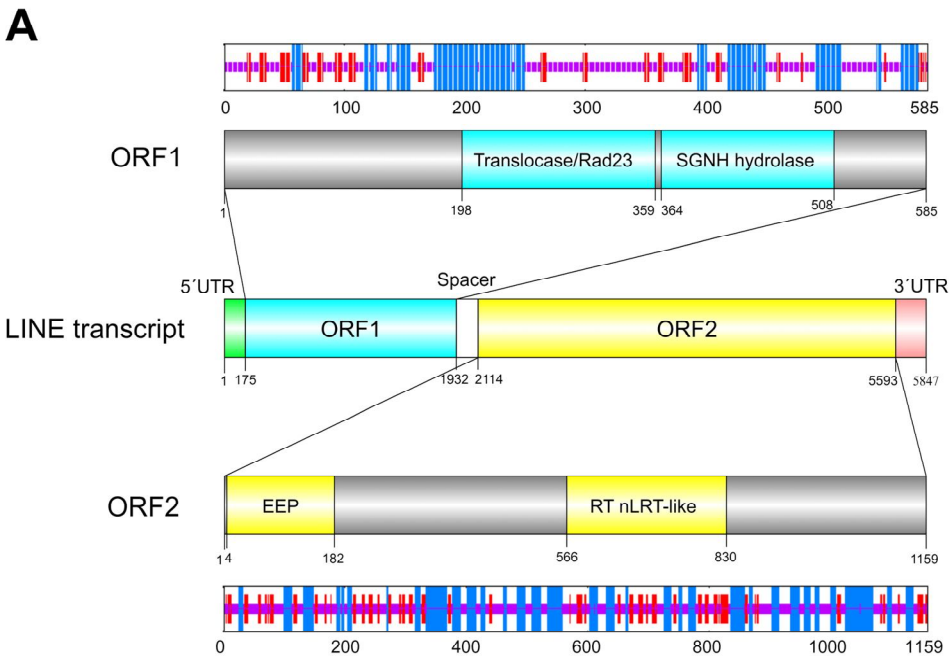
The LINE-like transposable element detected in the liver transcriptome of the Greenland shark has a length of 5,487 nt (the sequence is enclosed in *Supplementary material 1*). Because LINE elements are generally characterized by a high adenine (A) content (An et al. 2011, Dai et al. 2014), we computed nucleotide and dinucleotide frequencies in this putative LINE-like transposable element (Table 2). It is visible, that A is strongly represented in the whole transcript (37.40%) and especially within 5´-UTR (47.70%). Another interesting feature of LINEs is the very low content of CpG dinucleotides, only 2.12% for the whole transcript. The most frequent dinucleotide was AA (14.64%) (Table 2).

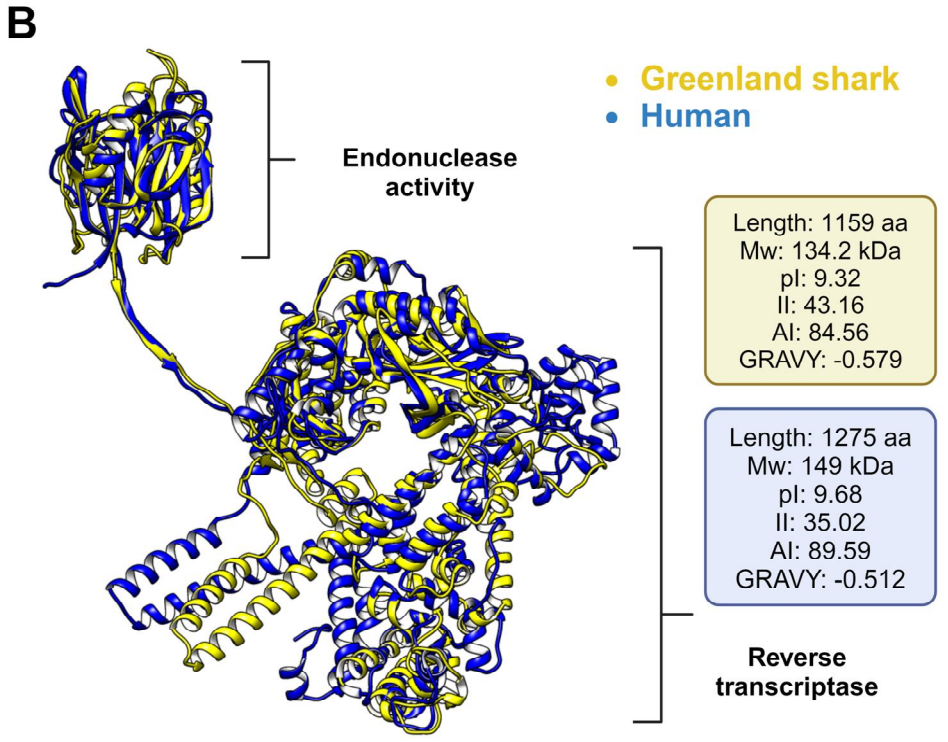
Expasy Translate Tool identified two non-overlapping open reading frames (ORFs) in LINE-like transposable element from liver transcriptome (Fig. 2). The shorter - ORF1, encodes 585 amino acid (aa) long protein. Domain homology search in the Conserved Domain Database revealed in this protein sequence a significantly homologous region to *sec-independent translocase* (aa 198-359; E-value of 2.51e-05) having also lower homology to the *UV excision repair protein Rad23* (aa 200-330; E-value of 8.18e-05).

	Whole transcript	5´-UTR	ORF1	Spacer	ORF2	3´-UTR
G	18.44	18.97	19.91	9.39	18.10	18.90
A	**37.40**	**47.70**	**36.69**	**33.15**	**37.39**	**38.58**
T	19.53	15.52	14.16	**43.09**	20.37	31.10
C	24.63	17.82	29.24	14.36	24.14	11.42
GG	4.14	**2.89**	4.61	**0.56**	4.37	**1.19**
GA	7.58	**11.56**	9.11	2.78	6.73	9.49
GT	3.01	**2.31**	**2.33**	4.44	3.10	5.93
GC	3.69	**2.31**	3.87	**1.67**	3.91	**1.98**
AG	7.18	9.83	8.25	3.89	6.67	7.51
AA	**14.64**	**25.43**	**13.32**	6.67	**15.00**	**16.21**
AT	7.01	4.05	4.61	**18.89**	7.53	9.88
AC	8.57	8.09	**10.47**	3.33	8.16	5.14
TG	4.98	4.62	4.15	4.44	5.12	9.49
TA	5.68	5.20	3.70	**16.11**	5.98	8.30
TT	4.11	3.47	**1.82**	**16.11**	4.14	**11.46**
TC	4.77	**2.31**	4.50	6.67	5.15	**1.98**
CG	**2.12**	**1.16**	**2.90**	**0.56**	**1.95**	**0.79**
CA	9.51	5.78	**10.53**	7.78	9.66	4.35
CT	5.41	5.78	5.41	3.33	5.61	3.95
CC	7.59	5.20	**10.42**	**2.78**	6.93	**2.37**

Table 2. Percentual nucleotide and dinucleotide composition of the whole LINE-like transcript (and its features) expressed in the Greenland shark. Enriched nucleotides (green) /dinucleotides (grey) are in bold. Depletion of CpG and other dinucleotides is in red. The composition was computed using a "DNA composition" tool accessed from Sequence Manipulation Suite (Stothard 2000).

Fig. 2. ► Characterization of the LINE-like element identified in the liver transcriptome assembly of the Greenland shark. *Top*: The identified liver LINE-like transcript has a length of 5,847 nucleotides and encodes two separate nonoverlapping ORFs (in different reading frames/fr.). ORF1 (cyan) has a length of 585 amino acid residues and contains 2 domains, one with homology to Translocase/Rad23 and a second similar to the SGNH hydrolase domain. ORF2 (yellow) has a length of 1,159 amino acid residues and contains an endonuclease (EEP) domain at its N-terminus and reverse transcriptase (RT) domain. For both ORF1 and ORF2, the prediction of secondary structure was made using a GOR4 method (Garnier et al. 1996), alpha helices are depicted as the biggest blue vertical segments, the beta strands as red vertical segments, and random coils as the smallest purple segments. *Bottom*: Structural superimposition of modeled the Greenland shark (yellow) and human (blue) ORF2 structures, together with several predicted protein parameters (Mw - molecular weight; pI - isoelectric point; II - instability index; AI - aliphatic index; GRAVY - grand average of hydropathicity index).


A
0
100
200
300
400
500
585
ORF1
Translocase/Rad23
SGNH hydrolase
1
198
359
364
508
585
5´UTR
Spacer
3´UTR
LINE transcript
ORF1
ORF2
1
175
1932
2114
5593
5847
ORF2
EEP
RT nLRT-like
14
182
566
830
1159
0
200
400
600
800
1000
1159


B
Greenland shark
Human
Endonuclease activity
Length: 1159 aa
Mw: 134.2 kDa
pI: 9.32
II: 43.16
AI: 84.56
GRAVY: -0.579
Length: 1275 aa
Mw: 149 kDa
pI: 9.68
II: 35.02
AI: 89.59
GRAVY: -0.512
Reverse transcriptase

The second identified domain in ORF1 (aa 364-508) belongs to the SGNH hydrolase superfamily and is homologous to the *sialate O-acetylesterase-like subfamily* (E-value of 5.57e-10). The longer ORF2 coding sequence for 1159 aa long protein has very significant domain homology to *Non-LTR (long terminal repeat) retrotransposon and non-LTR retrovirus reverse transcriptase* (aa 566-830; E-value of 1.00e-54) and homology to the *Exonuclease-Endonuclease-Phosphatase (EEP) domain superfamily* near the N-terminus (aa 4-182; E-value of 5.16e-08). Notably, the presence of non-truncated ORFs and the conservancy of all 13 critical amino acid residues forming the active site of reverse transcriptase suggest this element can be functional in terms of retroposition capability.

The theoretical ability of ORF1 and ORF2 to bind the LINE-like RNA (that is one of the characteristic features of LINE-1 (Naufer et al. 2019) was checked using an RNA-Protein Interaction Prediction (RPISeq) web server (Muppirala et al. 2011). Both ORF1 and ORF2 were predicted to interact with the Greenland shark LINE-like transcript. Further, RNA interaction aa residues were predicted using the PPRInt web server (Kumar et al. 2008). Detailed outputs of both predictions are enclosed in *Supplementary material 6*. Final subcellular localization was predicted as nuclear, for both ORFs (Detailed prediction can be found in *Supplementary material 2*). In addition, we found that the 3′-untranslated region (3′-UTR) of the LINE-like transcript harbors interesting repetitive elements: two copies of AAGACTGTTTAAAATTAATATTG, five copies of TTGAA, and the GAGAGAGAGAGAGAGAGAGAGAG tract (polyGA) at the very end of the transcript (*Supplementary material 7*).

To further characterize LINE-like ORF2 from the Greenland shark, the full-length 3D model was constructed using trRosetta (Du et al. 2021) and superimposed with human LINE-1 retrotransposable element ORF2 protein (LORF2; Uniprot ID: O00370) AlphaFold (Jumper et al. 2021) modeled structure (Fig. 2). For this purpose, the FATCAT flexible approach was applied, and the resulting structures were significantly similar with a P-value of 2.49e-11. In total, 986 structurally equivalent positions with an RMSD of 3.92 Å were found. Detailed FATCAT output can be found in *Supplementary material 4*.

Characterization of LINE-like transcript within the muscle transcriptome assembly

In the muscle transcriptome assembly, the LINE-like transcript identified had a length of 5,511 nt (*Supplementary material 1*). The similarity with liver LINE-like transcript was checked using pairwise global alignment (EMBOSS Needle algorithm). The results showed 85.4% identity and 13.5% of gaps (complete alignment is enclosed in *Supplementary material 4*). A dot plot of the longest LINE-like transcripts from liver and muscle tissues is shown in Fig. 3. From the global alignment and dot plot, it is evident that the differences are located mainly in the first 1,500 nt of both transcripts (2 longer gaps), nonetheless, several smaller changes (nucleotide substitutions) are present across the whole alignment. This suggests that the liver LINE-like and muscle LINE-like transcripts are transcribed from different/separate genes. On the contrary, there was no observed homology between shark and human LINE elements on the RNA sequence level (Fig. 3B).

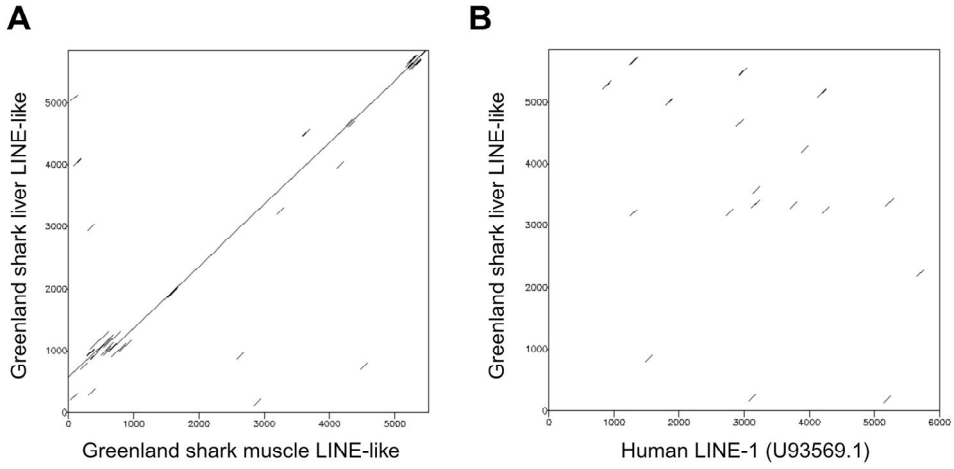


Fig. 3. Dot plot of the longest identified LINE-like transcripts in liver and muscle tissues of the Greenland shark (A), and liver LINE-like transcript from the Greenland shark and human LINE-1 (GenBank: U93569.1); (B). The dot plots were constructed using EMBOSS Dotmatcher (accessed from [2]) with the following parameters: window size = 100; threshold = 50.

Similarly as in the liver LINE-like transcript, several direct nucleotide repeats were found in untranslated regions (*Supplementary material 7*), and ORF1 and ORF2 were identified (*Supplementary material 1*). The ORF1 is 383 aa long and shows homology to large tegument protein UL36 (aa 11-162; E-value of 7.28e-07), and to the SGNH hydrolase superfamily (aa 163-306, E-value of 5.02e-09), as can be seen in Fig. 4 (pairwise ORF1 protein alignment can be found in the *Supplementary material 4*). ORF2 was very similar to that identified in liver LINE-like, with the same domain composition, only 5 amino acid substitutions were found in the whole 1159 long protein (therefore not shown in the Figure, pairwise ORF2 protein alignment can be found in the *Supplementary material 4*). Again, the theoretical ability of these ORFs to bind muscle LINE-like RNA was checked. Both ORF1 and ORF2 were predicted to interact with the Greenland shark muscle LINE-like transcript, and RNA interaction residues were predicted (*Supplementary material 6*). The final subcellular localization was predicted to be nuclear, for both ORFs (*Supplementary material 2*).

Except for that, the longest LINE-like muscle transcript contains an additional overlapping ORF, here called "ORF0" (Fig. 4). This ORF0 partially spans ORF1, but would be translated into a different reading frame (-1 shift with respect to ORF1). This transcript contains several possibilities of ORFs with different translation starts, which strengthens our suggestion, that this transcript belongs to the transposable element family, as this phenomenon is especially common in vertebrate transposons (Wright et al. 2022). The resulting putative protein would be 250 aa long with a predicted pI of 8.66. Conserved Domain Database domain homology search revealed significant homology to the *Chromosome segregation ATPase* domain (E-value = 2.83e-05; aa 3-83) and to the *superantigen-like protein SSL3* domain (E-value = 2.35e-03; aa 71-153).

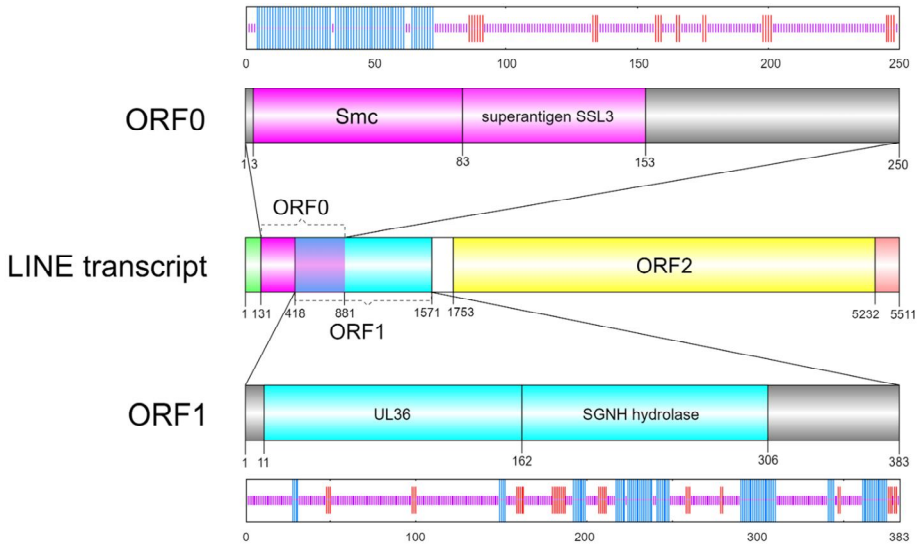


Fig. 4. Characterization of the longest LINE-like element identified in muscle transcriptome assembly of the Greenland shark. The longest identified liver LINE-like transcript has length of 5,511 nucleotides and encodes two separate non-overlapping ORFs (in different reading frames). ORF0 (purple) has a length of 383 aa residues and contains 2 domains, one with homology to Chromosome segregation ATPase domain (Smc), and a second similar to the superantigen-like protein SSL3 domain. ORF1 (yellow) has a length of 383 aa residues and contains a domain similar to the large tegument protein UL36 domain on its N-terminus and a region similar to the SGNH hydrolase domain. For both ORF0 and ORF1, the prediction of secondary structure was made using a GOR4 method (Garnier et al. 1996), alpha helices are depicted as the biggest blue vertical segments, beta strands as red vertical segments, and random coils as the smallest purple segments. As muscle LINE-like ORF2 domain composition is the same as liver LINE-like ORF2, it is not shown in this figure again.

Possible phylogenetic relationships

As transposable elements can generally undergo horizontal gene transfer (HGT), *i.e.* "jump" not only within the genome of an individual but also between genomes of different species (Zhang et al. 2020), it is difficult to reconstruct some meaningful evolutionary history. Here, we limit our effort to the construction of a simple protein tree of the most conserved LINE-like protein ORF2. At first sight, most LINE-like ORF2s are unannotated and designated as hypothetical or uncharacterized proteins in the NCBI database (Fig. 5). The closest protein homolog of a length of 1155 aa residues was found in the tropical freshwater fish *Prochilodus magdalenae*, an endemic species that inhabits Río Magdalena (Columbia). There was a protein sequence identity of 54% and a similarity of 71% with both ORF2 of the Greenland sharks, with only 2.7% gaps. Interestingly, significant hits were found also in invertebrates, like in soft coral *Paramuricea clavata* (*e.g.* CAB4009131.1 has a length of 1172 aa, identity of 40%, similarity of 59% and 4.6% gaps), or blue mussel *Mytilus edulis* (CAG2192112.1 with the length of 1406 aa, identity of 23%, similarity of 38% and 28.7% gaps).

Fig. 5. Protein tree of the Greenland shark LINE-like ORF2 homologs. Greenland shark ORFs are highlighted in red (liver variant) and orange (muscle variant). The closest vertebrate homolog (*Prochilodus magdalenae*) is highlighted in purple. The representative human ORF2 protein is shown in blue. Detailed tree construction parameters are described in the Materials and Methods section.

Discussion

The Greenland shark is a fascinating animal capable of living longer than any other vertebrate organism and excellently adapted to the cold waters of the Arctic and subarctic regions. Surprisingly, information about its DNA and RNA is limited. In our study, we sequenced the Greenland shark transcriptome from muscle and liver samples. Analysis of these data yielded interesting and unexpected results showing the abundance of LINE-like elements in the actively transcribed RNA of these tissues.

The evolution of LINEs is quite complicated, which results in a remarkable diversity of their sequences and types of domain arrangement (Ivancevic et al. 2016). Active LINE-1 has been described to have been lost during evolution in some species, *e.g.* rhinoceros, where LINE-1 was lost approximately 20 million years ago (Sookdeo et al. 2018). On the other hand, LINE elements were described as very active in ontogenesis (Protasova et al. 2021), and their high transcription is associated with some types of cancers in humans (Xiao-Jie et al.

2016). However, basic transcription has also been shown for some regular human tissues (McKerrow et al. 2023). Therefore, the significantly higher expression of the full-length LINE-like transcript in long-lived Greenland sharks is an unexpected result. However, increased LINE-1 expression is also associated with maintenance of the telomere (Mueller et al. 2018). Since telomere shortening is widely recognized as one of the hallmarks of aging (Aunan et al. 2016), the active transcription of this LINE-like RNA suggests its potential to preserve genome stability and to improve telomere maintenance. Alternative telomere lengthening in cancer cells can be suppressed with reverse transcriptase inhibitors (Bondarev and Khavinson 2016). Morover, it was shown that LINE-1 ribonucleoprotein particles (comprising the reverse transcriptase domain) can protect telomeric ends in human malignant cell lines in the absence of telomerase activity (Aschacher et al. 2020). Even if the nucleotide sequence homology between the human and the Greenland shark LINE is undetectable, and the similarities in the protein sequence between the Greenland shark and human ORF2 (which encodes LINE reverse transcriptase (Kopera et al. 2011)) is slightly above 30%, both ORF2 proteins are structurally very similar (*see* Fig. 2). In 2010, Georges St. Laurent and colleagues proposed a LINE*age* theory (Laurent et al. 2010), where the LINE elements represent an important component of organismal aging, although in a negative sense. LINEs, according to this theory, act as a double-edged sword, allowing evolutionary advantage at the cost of higher genomic instability leading to mutations and rearrangements, and therefore accelerated aging (Laurent et al. 2010, Kemp and Longworth 2015). In Greenland sharks, we know nothing about the overall number of copies of the *LINE* gene, as the genome (DNA) is still not available. It may be that the observed RNAs of LINE-like expression come from a single or a few copies of the *LINE* gene that lost the ability of retroposition and, instead, the main function of encoded proteins diversifies to improved telomere maintenance and/or another longevity-promoting molecular mechanism.

Even if we found the strong abundance of the LINE-like active element as a unique feature of both liver and muscle tissues of the Greenland shark, further research is necessary to validate their functions and investigate the specific mechanisms through which the LINE-like expression may contribute to extreme longevity in Greenland sharks.

References

AN, W., DAI, L., NIEWIADOMSKA, A. M., YETIL, A., O'DONNELL, K. A., HAN, J. S. and BOEKE, J. D. (2011): Characterization of a synthetic human LINE-1 retrotransposon ORFeus-Hs. *Mobile DNA*, 2(1): 2. doi: 10.1186/1759-8753-2-2

ANISIMOVA, M., GASCUEL, O. (2006): Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55: 539-552. doi: 10.1080/10635150600755453

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. and LIPMAN, D. J. (1990): Basic local alignment search tool. *Journal of Molecular Biology*, 215(3): 403-410. doi: 10.1016/S0022-2836(05)80360-2

ARTIMO, P., JONNALAGEDDA, M., ARNOLD, K., BARATIN, D., CSARDI, G., DE CASTRO, E., DUVAUD, S., FLEGEL, V., FORTIER, A., GASTEIGER, E., GROSDIDIER, A., HERNANDEZ, C., IOANNIDIS, V., KUZNETSOV, D., LIECHTI, R., MORETTI, S., MOSTAGUIR, K., REDASCHI, N., ROSSIER, G., XENARIOS, I. and STOCKINGER, H. (2012): ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40: W597-W603. doi: 10.1093/nar/gks400

ASCHACHER, T., WOLF, B., ASCHACHER, O., ENZMANN, F., LASZLO, V., MESSNER, B., TÜRKCAN, A., WEIS, S., SPIEGL-KREINECKER, S., HOLZMANN, K., LAUFER, G., EHRLICH, M. and BERGMANN, M. (2020): Long interspersed element-1 ribonucleoprotein particles protect telomeric ends in alternative lengthening of telomeres dependent cells. *Neoplasia*, 22(2): 61-75. doi: 10.1016/j.neo.2019.11.002

AUGUSTINE, S., LIKA, K. and KOOIJMAN, S. A. L. M. (2017): Comment on the ecophysiology of the Greenland shark, Somniosus microcephalus. *Polar Biology*, 40: 2429-2433. doi: 10.1007/s00300-017-2154-8

AUNAN, J. R., WATSON, M. M., HAGLAND, H. R. and SØREIDE, K. (2016): Molecular and biological hallmarks of ageing. *British Journal of Surgery*, 103: e29-e46. doi: 10.1002/bjs.10053

BARTAS, M., BRÁZDA, V., VOLNÁ, A., ČERVEŇ, J., PEČINKA, P. and ZAWACKA-PANKAU, J. E. (2021): The changes in the p53 protein across the animal kingdom point to its involvement in longevity. *International Journal of Molecular Sciences*, 22(16), 8512. doi: 10.3390/ijms22168512

BECK, B., MANSFIELD, A. W. (1969): Observations on the Greenland Shark, Somniosus microcephalus, in Northern Baffin Island. *Journal of the Fisheries Research Board of Canada*, 26: 143-145. doi: 10.1139/f69-013

BONDAREV, I. E., KHAVINSON, V. KH. (2016): Suppression of alternative telomere lengthening in cancer cells with reverse transcriptase inhibitors. *Advances in Gerontology*, 6: 272-274. doi: 10.1134/S2079057016040020

CASTRESANA, J. (2000): Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17: 540-552. doi: 10.1093/oxfordjournals.molbev.a026334

CHERNOVA, N. V., SMIRNOVA, E. V. and RASKHOZHEVA, E. V. (2015): First record of the Greenland shark Somniosus microcephalus (Squaliformes: Somniosidae) in the Siberian Arctic with notes on its distribution and biology. *Journal of Ichthyology*, 55: 827-835. doi: 10.1134/S0032945215060053

CHEVENET, F., BRUN, C., BAÑULS, A-L., JACQ, B. and CHRISTEN, R. (2006): TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, 7: 439. doi: 10.1186/1471-2105-7-439

CLARKE, T. L., MOSTOSLAVSKY, R. (2022): DNA repair as a shared hallmark in cancer and ageing. *Molecular Oncology*, 16(18): 3352-3379. doi: 10.1002/1878-0261.13285

DAI, L., LACAVA, J., TAYLOR, M. S. and BOEKE, J. D. (2014): Expression and detection of LINE-1 ORF-encoded proteins. *Mobile Genetic Elements*, 4: e29319. doi: 10.4161/mge.29319

DE MAGALHÃES, J. P., COSTA, J. (2009): A database of vertebrate longevity records and their relation to other life‐history traits. *Journal of Evolutionary Biology*, 22: 1770-1774. doi: 10.1111/j.1420-9101.2009.01783.x

DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J. F., GUINDON, S., LEFORT, V., LESCOT, M., CLAVERIE, J. M. and GASCUEL, O. (2008): Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36: W465-W469. doi: 10.1093/nar/gkn180

DU, Z., SU, H., WANG, W., YE, L., WEI, H., PENG, Z., ANISHCHENKO, I., BAKER, D. and YANG, J. (2021): The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols*, 16(12): 5634-5651. doi: 10.1038/s41596-021-00628-9

EDGAR, R. C. (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32: 1792-1797. doi: 10.1093/nar/gkh340

EDWARDS, J. E., HILTZ, E., BROELL, F., BUSHNELL, P. G., CAMPANA, S. E., SCHOU CHRISTIANSEN, J., DEVINE, B. M., GALLANT, J. J., HEDGES, K. J., AARON MACNEIL, M., MCMEANS, B. C., NIELSEN, J., PRÆBEL, K., SKOMAL, G. B., STEFFENSEN, J. F., WALTER, R. P., WATANABE, Y. Y., VANDERZWAAG, D. L. and HUSSEY, N. E. (2019): Advancing research for the management of long-lived species: A case study on the Greenland shark. *Frontiers in Marine Science*, 6. doi: 10.3389/fmars.2019.00087

GARNIER, J., GIBRAT, J-F. and ROBSON, B. (1996): GOR method for predicting protein secondary structure from amino acid sequence. *In*: Methods in enzymology. Elsevier, pp. 540–553.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., APPEL, R. D. and BAIROCH, A. (2005): Protein identification and analysis tools on the ExPASy server. *In*: J. M. Walker (ed.): The proteomics protocols handbook. Humana Press, New York, pp. 571–607. doi: 10.1385/1-59259-890-0:571

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. and REGEV, A. (2011): Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7): 644-652. doi: 10.1038/nbt.1883

GRANT, S. M., SULLIVAN, R. and HEDGES, K. J. (2018): Greenland shark (*Somniosus microcephalus*) feeding behavior on static fishing gear, effect of SMART (Selective Magnetic and Repellent-Treated) hook deterrent technology, and factors influencing entanglement in bottom longlines. *PeerJ*, 6: e4751. doi: 10.7717/peerj.4751

GUINDON, S., GASCUEL, O. (2003): A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52: 696-704. doi: 10.1080/ 10635150390235520

IVANCEVIC, A. M., KORTSCHAK, R. D., BERTOZZI, T. and ADELSON, D. L. (2016): LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. *Genome Biology and Evolution*, 8: 3301-3322.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T. and HASSABIS, D. (2021): Highly accurate protein structure prediction with AlphaFold. *Nature*, 596: 583-589. doi: 10.1038/s41586-021-03819-2

KEMP, J. R., LONGWORTH, M. S. (2015): Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer. *Frontiers in Chemistry*, 3: 68. doi: 10.3389/fchem.2015.00068

KIM SH, S. H., KAMINKER, P. and CAMPISI, J. (2002): Telomeres, aging and cancer: In search of a happy ending. *Oncogene*, 21(4): 503-511. doi: 10.1038/sj.onc.1205077

KOJIMA, K. K. (2019): Structural and sequence diversity of eukaryotic transposable elements. *Genes & Genetic Systems*, 94: 233-252.

KOPERA, H. C., MOLDOVAN, J. B., MORRISH, T. A., GARCIA-PEREZ, J. L. and MORAN, J. V. (2011): Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proceedings of the National Academy of Sciences*, 108(51): 20345-20350. doi: 10.1073/pnas.1100275108

KUMAR, M., GROMIHA, M. M. and RAGHAVA, G. P. S. (2008): Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins*, 71: 189-194. doi: 10.1002/prot.21677

KYRAN, A. (2021): Fasta Statistics: Display summary statistics for a fasta file. https://github.com/ galaxyproject/tools-iuc

LAURENT, G. ST., HAMMELL, N. and MCCAFFREY, T. A. (2010): A LINE-1 Component to Human Aging: Do LINE elements exact a longevity cost for evolutionary advantage? *Mechanisms of Ageing and Development*, 131: 299-305. doi: 10.1016/j.mad.2010.03.008

LECLERC, L-E., LYDERSEN, C., HAUG, T., BACHMANN, L., FISK, A. T. and KOVACS, K. M. (2012): A missing piece in the Arctic food web puzzle? Stomach contents of Greenland sharks sampled in Svalbard, Norway. *Polar Biology*, 35: 1197-1208. doi: 10.1007/s00300-012-1166-7

LI, Z., JAROSZEWSKI, L., IYER, M., SEDOVA, M. and GODZIK, A. (2020): FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, 48(W1): W60-W64. doi: 10.1093/nar/ gkaa443

LU, J. Y., CHANG, L., LI, T., WANG, T., YIN, Y., ZHAN, G., HAN, X., ZHANG, K., TAO, Y., PERCHARDE, M., WANG, L., PENG, Q., YAN, P., ZHANG, H., BI, X., SHAO, W., HONG, Y., WU, Z., MA, R., WANG, P. and SHEN, X. (2021): Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Research*, 31: 613-630. doi: 10.1038/s41422-020-00466-6

LU, S., WANG, J., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C., GONZALES, N. R., GWADZ, M., HURWITZ, D. I., MARCHLER, G. H., SONG, J. S., THANKI, N., YAMASHITA, R. A., YANG, M.,

ZHANG, D., ZHENG, C., LANCZYCKI, C. J. and MARCHLER-BAUER, A. (2020): CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research*, 48: D265-D268. doi: 10.1093/nar/gkz991

LUCAS, Z. N., NATANSON, L. J. (2010): Two shark species involved in predation on seals at Sable Island, Nova Scotia, Canada. *Proceedings of the Nova Scotian Institute of Science (NSIS)*, 45: 64-88.

MCKERROW, W., KAGERMAZOVA, L., DOUDICAN, N., FRAZZETTE, N., KAPAROS, E. I., EVANS, S. A., ROCHA, A., SEDIVY, J. M., NERETTI, N., CARUCCI, J., BOEKE, J. D. and FENYÖ, D. (2023): LINE-1 retrotransposon expression in cancerous, epithelial and neuronal cells revealed by 5′ single-cell RNA-Seq. *Nucleic Acids Research*, 51(5): 2033-2045. doi: 10.1093/nar/gkad049

MIRABELLO, L., SAVAGE, S. A., KORDE, L., GADALLA, S. M. and GREENE, M. H. (2010): LINE-1 methylation is inherited in familial testicular cancer kindreds. *BMC Medical Genetics*, 11: 77. doi: 10.1186/1471-2350-11-77

MUELLER, C., ASCHACHER, T., WOLF, B. and BERGMANN, M. (2018): A role of LINE-1 in telomere regulation. *Frontiers in Bioscience* (Landmark Ed), 23: 1310-1319. doi: 10.2741/4645

MUPPIRALA, U. K., HONAVAR, V. G. and DOBBS, D. (2011): Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics*, 12: 489. doi: 10.1186/1471-2105-12-489

NAUFER, M. N., FURANO, A. V. and WILLIAMS, M. C. (2019): Protein-nucleic acid interactions of LINE-1 ORF1p. *Seminars in Cell & Developmental Biology*, 86: 140-149. doi: 10.1016/j.semcdb.2018.03.019

NIELSEN, J., HEDEHOLM, R. B., HEINEMEIER, J., BUSHNELL, P. G., CHRISTIANSEN, J. S., OLSEN, J., RAMSEY, C. B., BRILL, R. W., SIMON, M., STEFFENSEN, K. F. and STEFFENSEN, J. F. (2016): Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (*Somniosus microcephalus*). *Science*, 353(6300): 702-704. doi: 10.1126/science.aaf1703

NIELSEN, J., HEDEHOLM, R. B., SIMON, M. and STEFFENSEN, J. F. (2014): Distribution and feeding ecology of the Greenland shark (*Somniosus microcephalus*) in Greenland waters. *Polar Biology*, 37: 37-46. doi: 10.1007/s00300-013-1408-3

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. and FERRIN, T. E. (2004): UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13): 1605-1612. doi: 10.1002/jcc.20084

PROTASOVA, M. S., ANDREEVA, T. V. and ROGAEV, E. I. (2021): Factors regulating the activity of LINE1 retrotransposons. *Genes* (Basel), 12: 1562. doi: 10.3390/genes12101562

REN, J., WEN, L., GAO, X., JIN, C., XUE, Y. and YAO, X. (2009): DOG 1.0: illustrator of protein domain structures. *Cell Research*, 19(2): 271-273. doi: 10.1038/cr.2009.6

SANTAQUITERIA, A., NIELSEN, J., KLEMETSEN, T., WILLASSEN, N. P. and PRÆBEL, K. (2017): The complete mitochondrial genome of the long-lived Greenland shark (*Somniosus microcephalus*): Characterization and phylogenetic position. *Conservation Genetics Resources*, 9: 351-355. doi: 10.1007/s12686-016-0676-y

SHARIF, J., KOSEKI, H. and PARRISH, N. F. (2023): Bridging multiple dimensions: roles of transposable elements in higher-order genome regulation. *Current Opinion in Genetics & Development*, 80: 102035. doi: 10.1016/j.gde.2023.102035

SOOKDEO, A., HEPP, C. M. and BOISSINOT, S. (2018): Contrasted patterns of evolution of the LINE-1 retrotransposon in perissodactyls: The history of a LINE-1 extinction. *Mobile DNA*, 9: 12. doi: 10.1186/s13100-018-0117-4

STE-MARIE, E., WATANABE, Y. Y., SEMMENS, J. M., MARCOUX, M. and HUSSEY, N. E. (2020): A first look at the metabolic rate of Greenland sharks (*Somniosus microcephalus*) in the Canadian Arctic. *Scientific Reports*, 10 (1): 19297. doi: 10.1038/s41598-020-76371-0

STOTHARD, P. (2000): The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28: 1102-1104.

THAWANI, A., ARIZA, A. J. F., NOGALES, E. and COLLINS, K. (2023): Template and target site recognition by human LINE-1 in retrotransposition. *Nature*, Advance online publication. doi: 10.1038/s41586-023-06933-5

WATANABE, Y.Y, LYDERSEN, C., FISK, A. T. and KOVACS, K. M. (2012): The slowest fish: Swim speed and tail-beat frequency of Greenland sharks. *Journal of Experimental Marine Biology and Ecology*, 426–427: 5-11. doi: 10.1016/j.jembe.2012.04.021

WRIGHT, B. W., MOLLOY, M. P. and JASCHKE, P. R. (2022): Overlapping genes in natural and engineered genomes. *Nature Reviews Genetics*, 23: 154-168.

XIAO-JIE, L., HUI-YING, X., QI, X., JIANG, X. and SHI-JIE, M. (2016): LINE-1 in cancer: multifaceted functions and potential clinical implications. *Genetics in Medicine*, 18(5): 431-439. doi: 10.1038/gim.2015.119

YOPAK, K. E., MCMEANS, B. C., MULL, C. G., FEINDEL, K. W., KOVACS, K. M., LYDERSEN, C., FISK, A. T. and COLLIN, S. P. (2019): Comparative brain morphology of the Greenland and Pacific sleeper sharks and its functional implications. *Scientific Reports*, 9(1): 10022. doi: 10.1038/s41598-019-46225-5

YU, C-S, CHEN, Y-C, LU, C-H and HWANG, J-K (2006): Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*, 64: 643-651. doi: 10.1002/prot.21018

ZHANG, H-H, PECCOUD, J., XU, M-R-X, ZHANG, X. G. and GILBERT, C. (2020): Horizontal transfer and evolution of transposable elements in vertebrates. *Nature Communications*, 11: 1362. doi: 10.1038/s41467-020-15149-4

Web sources / Other sources

[1] Links to sequences deposited in the NCBI database:
https://www.ncbi.nlm.nih.gov/nuccore/OR828930.1/
https://www.ncbi.nlm.nih.gov/nuccore/OR828936.1/

[2] https://www.ebi.ac.uk/jdispatcher/seqstats/emboss_dotmatcher

[3] The Galaxy Community (2022): The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50: W345-W351. doi: 10.1093/nar/gkac247

Supplementary materials

Supplementary material 1: Sequences of identified LINE-like transcripts and ORFs
Supplementary material 2: CELLO subcellular localisation prediction
Supplementary material 3: PDB files of modelled structures
Supplementary material 4: Sequence and structural alignments
Supplementary material 5: Quantification of transcripts (Salmon tool)
Supplementary material 6: RPISeq and PPRInt predictions
Supplementary material 7: LINE-like features and repeats