UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS
BACHARELADO EM CIÊNCIAS BIOLÓGICAS, ÊNFASE EM BIOLOGIA MARINHA E COSTEIRA

ALICE MORAES LINCK

Molecular Analysis of Cancer-Related Genes in South American Fur Seals (*Arctocephalus australis*)

IMBÉ

2026

Molecular Analysis of Cancer-Related Genes in South American Fur Seals (*Arctocephalus australis*)

Trabalho de Conclusão de Curso apresentado como requisito para a obtenção do título de Bacharel em Ciências Biológicas, ênfase em Biologia Marinha e Costeira na Universidade Federal do Rio Grande do Sul.

Artigo formatado para o repositório de preprints: *https://www.biorxiv.org*

Orientador: Prof. Dr. Enéas Ricardo Konzen

Orientadora: Profa. Dra. Rossana Colla Soletti

IMBÉ

2026

CIP - Catalogação na Publicação

Linck, Alice Moraes
Molecular Analysis of Cancer-Related Genes in South American Fur Seals (Arctocephalus australis) / Alice Moraes Linck. -- 2026.
37 f.
Orientador: Enéas Ricardo Konzen.

Coorientadora: Rossana Colla Soletti.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Campus Litoral Norte, Curso de Ciências Biológicas: Biologia Marinha e Costeira, Tramandaí, BR-RS, 2026.

1. Arctocephalus australis. 2. Genética de Mamíferos Marinhos. 3. Oncologia Comparativa. 4. Polimorfismos. 5. Bioinformática. I. Konzen, Enéas Ricardo, orient. II. Soletti, Rossana Colla, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Alice Moraes Linck

Molecular Analysis of Cancer-Related Genes in South American Fur Seals (*Arctocephalus australis*)

Trabalho de Conclusão de Curso apresentado como requisito para a obtenção do título de Bacharel em Ciências Biológicas, ênfase em Biologia Marinha e Costeira na Universidade Federal do Rio Grande do Sul.

Imbé, 1° de junho de 2026

_____

Profa. Dra. Elisabeth Cabral da Silva
Coordenadora do Curso

_____

Prof. Dr. Enéas Ricardo Konzen
Orientador
Universidade Federal do Rio Grande do Sul

_____

Profa. Dra. Rossana Colla Soletti
Coorientadora
Universidade Federal do Rio Grande do Sul

Banca Examinadora:

_____

Profa. Dra. Karen Luisa Haag
Universidade Federal do Rio Grande do Sul

_____

Prof. Dr. Fábio Ribeiro Queiroz
Instituto Mário Penna

Agradecimentos

A realização deste trabalho é o resultado da colaboração, do apoio e do carinho de diversas pessoas que cruzaram o meu caminho. A conclusão desta etapa não seria possível sem cada um de vocês.

Gostaria de iniciar agradecendo ao meu orientador, Prof. Dr. Enéas, por ter me acolhido desde o começo do curso. Muito obrigada por todas as oportunidades de projetos e bolsas que tive graças à sua confiança em mim. Você foi maravilhoso durante todo o processo, e o considero, com muito carinho, o meu “pai científico”. À minha coorientadora, Profa. Dra. Rossana, meu profundo agradecimento por ser uma mentora tão maravilhosa. Sempre me identifiquei muito com sua área focada em saúde e com as suas aulas. Você é uma verdadeira inspiração para mim, como cientista, mãe e mulher.

Agradeço ao Centro de Reabilitação de Animais Silvestres e Marinhos (CERAM) pelo fornecimento dos materiais de linfonodos. Em especial, agradeço ao veterinário Derek por toda a ajuda para adquirir os materiais, e também pelos ensinamentos que tive quando fui estagiária do centro, o que me ajudou a crescer como pessoa. Também à estudante de biologia Luana Beatriz, que, muito querida, me alcançou algumas das amostras que ela mesma retirou.

Agradeço também ao Laboratório do Departamento de Plantas de Lavoura da Faculdade de Agronomia e ao Laboratório de Sinalização e Plasticidade Celular (LabSinal) do Campus do Vale, por abrirem suas portas e me permitirem o uso de seus equipamentos em alguns momentos. Na Agronomia, agradeço à Profa. Carla Andréa Delatorre e ao Bruno Delatorre Salomon pelo acolhimento e ajuda, com um agradecimento muito especial ao Gustavo Leão, que dedicou tantos dias e muito do seu tempo para me auxiliar com enorme paciência. No LabSinal, minha eterna gratidão à Beatriz Müller, que desenvolveu a primeira etapa deste projeto no TCC dela e inspirou o meu trabalho. Bia, você foi uma amiga incrível, super querida e acolhedora, com uma didática fantástica que me ensinou muito durante o estágio. Devo muito a você.

Às minhas amigas do curso, muito obrigada por terem me apoiado em tantos momentos marcantes de nossa jornada. Meu coração se enche de alegria por estarmos nos formando juntas. A todos os meus outros amigos, sejam eles da biologia, os de fora ou mesmo aqueles com quem não tenho mais tanto contato, obrigada por sempre entenderem o meu lado, por me apoiarem e por trazerem leveza aos meus dias com jogos, artes maravilhosas (tenho sorte de ter muitos amigos artistas!) e estudos de idiomas.

À minha família, que é a minha base de tudo. Ao meu pai, Hamilton, agradeço por ter me inspirado a ser uma pessoa crítica e sempre interessada no conhecimento; levo comigo a sua essência, que surgiu com muito amor. À minha mãe, Maria, muito obrigada por todo o amor e apoio incondicional, não apenas durante o curso, mas em toda a minha vida. À minha irmã, Aline, obrigada por compartilharmos muitos momentos preciosos e ótimas conversas científicas, que tivemos por nossa área similar de estudo. Ao meu cunhado, Willian, deixo aqui um agradecimento muito especial; muito obrigada por ter sido um verdadeiro porto seguro em um momento em que eu muito precisei.

Por fim, ao meu namorado, Carlos, agradeço por ter sido um grande incentivador da minha entrada e permanência neste curso. Obrigada por me ajudar com tanta frequência, por compartilhar comigo tantos assuntos de interesses científicos em comum e, muito memorável, por ter me inspirado a fazer Cálculo 1 de extracurricular.

Tudo o que conquistei, nós conquistamos juntos.

Apresentação

Este trabalho dá continuidade ao estudo desenvolvido por Müller (2024), que iniciou a investigação de genes associados ao câncer em *Arctocephalus australis*, com foco em *TP53*, *NOTCH1* e *SF3B1*. Naquele trabalho, foram realizadas análises filogenéticas e de motivos proteicos, testes de extração de DNA, desenho de primers e testes de PCR, com a autora indicando caminhos possíveis para a continuidade desta investigação molecular, proposta que buscou ser atendida neste estudo.

Neste contexto, o atual trabalho buscou complementar as análises anteriores dentro das limitações materiais e técnicas disponíveis, focando nos genes *TP53* e *SF3B1* somente. Assim, foram feitas novas extrações de DNA de amostras não-tumorais, foi testada uma nova metodologia para extração genética da amostra tumoral em Tecido Fixado em Formalina e Embebido em Parafina (FFPE) e foram incorporadas análises comparativas com espécies congêneres de *Arctocephalus*, inclusive de proteínas geradas por *AlphaFold*.

Optou-se pela redação do trabalho em formato de *preprint*, dado que a metodologia e os resultados aqui descritos farão parte de um artigo científico em desenvolvimento junto aos resultados obtidos por Müller (2024). Assim, a versão atual serve como base para futura submissão, que ainda depende da finalização do sequenciamento de *SF3B1*, atualmente em processamento, e da subsequente análise de polimorfismos e seus efeitos deletérios.

Resumo

A oncologia comparativa permite investigar mecanismos de desenvolvimento tumoral compartilhados entre espécies, sendo especialmente relevante em mamíferos marinhos porque a aparente menor incidência de câncer em muitos desses grupos, discutida no contexto do Paradoxo de Peto, pode revelar mecanismos evolutivos de resistência ou susceptibilidade ao câncer. A descrição de um caso único de linfoma linfocítico de pequenas células B (B-SLL) em *A. australis* motivou a continuidade da investigação iniciada por Müller (2024), voltada à análise de genes associados ao câncer em pinípedes. O lobo-marinho-sul-americano (*Arctocephalus australis*), é uma espécie sem genoma de referência disponível em bancos públicos, o que dificulta investigações moleculares. Assim, este trabalho teve como objetivo geral extrair e investigar regiões dos genes *TP53* e *SF3B1* em amostras não-tumorais de linfonodo de *A. australis*, visando a identificação de polimorfismos e seus potenciais efeitos deletérios. Buscou-se reconstruir sequências codificantes desses genes em espécies congêneres de *Arctocephalus*, comparar padrões de conservação entre pinípedes e outros mamíferos, e avaliar a conservação estrutural das proteínas p53 e SF3B1. Partiu-se das hipóteses de que fragmentos de *TP53* e *SF3B1* poderiam apresentar polimorfismos entre indivíduos de *A. australis*, podendo ser diferentes das observadas em espécies comparativas, como *Homo sapiens* e *Zalophus californianus*, e que ocorreriam em regiões funcionalmente relevantes. Além disso, estas regiões-alvo seriam conservadas entre pinípedes e outros mamíferos, inclusive as estruturas tridimensionais das proteínas, e uma anotação comparativa de *Arctocephalus* spp. poderia ser reconstruída com base em genomas de espécies relacionadas, como o *Z. californianus*, para uso nas análises moleculares desenvolvidas neste trabalho. Foram seguidos protocolos de extração de DNA para amostras não-tumorais de 15 indivíduos de *A. australis*, uma amostra de *Lobodon carcinophaga* e uma amostra tumoral em Formalina Embebido em Parafina (FFPE) referente ao caso do B-SLL. As amostras não-tumorais apresentaram DNA adequado para PCR, enquanto a amostra tumoral, apesar da recuperação mensurável de DNA, não amplificou nos ensaios realizados, provavelmente devido à baixa integridade do material FFPE e ao tamanho dos fragmentos-alvo dos primers. Entre os genes testados, *SF3B1* foi o alvo mais viável, apresentando amplificação mais consistente em amostras não-tumorais e sendo selecionado para sequenciamento por Sanger. *TP53* apresentou amplificação inespecífica ou ausência de produtos adequados nas condições testadas. As análises *in silico* permitiram reconstruir sequências codificantes de *TP53* e *SF3B1* para *A. gazella* e *A. townsendi*, com resultados concordantes entre diferentes estratégias de anotação comparativa. Além disso, análises de alinhamento, filogenia, motivos conservados e estrutura proteica indicaram alta conservação de *SF3B1* e conservação relevante de *TP53* entre pinípedes e outros mamíferos, incluindo regiões correspondentes a hotspots associados ao câncer em humanos. Conclui-se que este trabalho estabeleceu uma base molecular e comparativa inicial para o estudo de genes associados ao câncer em *Arctocephalus* spp., identificou *SF3B1* como o alvo mais promissor para análises futuras em *A. australis* pelos procedimentos feitos neste trabalho e evidenciou a necessidade de protocolos específicos para material FFPE, desenho de primers espécie-específicos e abordagens baseadas em RNA ou amostras tumorais melhor preservadas para investigar possíveis alterações funcionais associadas ao B-SLL. Com os resultados dos sequenciamentos, ainda em processamento, será possível analisar os polimorfismos e seus potenciais efeitos deletérios para a finalização do artigo final.

Abstract

Comparative oncology allows the investigation of tumour development mechanisms shared among species, and is especially relevant in marine mammals because the apparent lower incidence of cancer in many of these groups, discussed in the context of Peto's Paradox, may reveal evolutionary mechanisms of cancer resistance or susceptibility. The description of a single case of B-cell small lymphocytic lymphoma (B-SLL) in *A. australis* motivated the continuation of the investigation initiated by Müller (2024), focused on the analysis of cancer-associated genes in pinnipeds. The South American fur seal (*Arctocephalus australis*) is a species without an available reference genome in public databases, which hinders molecular investigations. Thus, the general objective of this study was to extract and investigate regions of the *TP53* and *SF3B1* genes in non-tumoural lymph node samples from *A. australis*, aiming to identify polymorphisms and their potential deleterious effects. This study sought to reconstruct coding sequences of these genes in congeneric species of *Arctocephalus*, compare conservation patterns among pinnipeds and other mammals, and evaluate the structural conservation of the p53 and SF3B1 proteins. The study was based on the hypotheses that fragments of *TP53* and *SF3B1* could present polymorphisms among individuals of *A. australis*, that these could differ from those observed in comparative species, such as *Homo sapiens* and *Zalophus californianus*, and that they would occur in functionally relevant regions. In addition, these target regions were expected to be conserved among pinnipeds and other mammals, including the three-dimensional structures of the proteins, and a comparative annotation of *Arctocephalus* spp. could be reconstructed based on genomes from related species, such as *Z. californianus*, for use in the molecular analyses developed in this study. DNA extraction protocols were followed for non-tumoural samples from 15 individuals of *A. australis*, one sample of *Lobodon carcinophaga*, and one Formalin-Fixed Paraffin-Embedded (FFPE) tumour sample corresponding to the B-SLL case. The non-tumoural samples yielded DNA suitable for PCR, whereas the tumour sample, despite measurable DNA recovery, did not amplify in the assays performed, probably due to the low integrity of the FFPE material and the size of the primer target fragments. Among the genes tested, *SF3B1* was the most viable target, showing more consistent amplification in non-tumoural samples and being selected for Sanger sequencing. *TP53* showed non-specific amplification or absence of suitable products under the tested conditions. The *in silico* analyses allowed the reconstruction of *TP53* and *SF3B1* coding sequences for *A. gazella* and *A. townsendi*, with concordant results among different comparative annotation strategies. In addition, alignment, phylogenetic, conserved motif and protein structure analyses indicated high conservation of *SF3B1* and relevant conservation of *TP53* among pinnipeds and other mammals, including regions corresponding to cancer-associated hotspots in humans. It is concluded that this study established an initial molecular and comparative basis for the study of cancer-associated genes in *Arctocephalus* spp., identified *SF3B1* as the most promising target for future analyses in *A. australis* based on the procedures performed in this work, and highlighted the need for specific protocols for FFPE material, species-specific primer design, and RNA-based approaches or better-preserved tumour samples to investigate possible functional alterations associated with B-SLL. Once the sequencing results, which are still being processed, are available, it will be possible to analyse the polymorphisms and their potential deleterious effects for the completion of the final article.

MOLECULAR ANALYSIS OF CANCER-RELATED GENES IN SOUTH AMERICAN FUR SEALS (*Arctocephalus australis*)

A PREPRINT

Alice Moraes Linck [1], **Beatriz Cristine Müller** [1], **Rossana Colla Soletti** [1], **and Enéas Ricardo Konzen** [1]
[1]Center for Coastal, Limnological and Marine Studies (CECLIMAR), Federal University of Rio Grande do Sul, Imbé, RS, Brazil
[*]To whom correspondence should be addressed: erkonzen@gmail.com;

June 1st, 2026

ABSTRACT

The South American fur seal (*Arctocephalus australis*) is an otariid for which a single case of B-cell small lymphocytic lymphoma (B-SLL) has been described. However, genomic information for the species remains scarce, and no reference genome is available in public databases. Continuing the work of Müller (2024), this study aimed to analyse phylogenetic relationships, predict motifs and protein structure, and investigate nucleotide polymorphisms in genes possibly associated with cancers in *A. australis*, specifically *SF3B1* and *TP53*. It was hypothesised that fragments of these genes in *A. australis* samples would show nucleotide variability among the analysed individuals, although low due to the high degree of functional conservation of these genes. It was also considered that *TP53* and *SF3B1* sequences from congeneric *Arctocephalus* species could provide more suitable comparative references for interpreting variation in *A. australis* than sequences from other pinnipeds. Non-tumoural lymph node samples from 15 *A. australis* individuals and one *Lobodon carcinophaga* individual were used, as well as one formalin-fixed paraffin-embedded tumour sample for microscopy, corresponding to the only lymphoma record currently known for the species. DNA from non-tumoural and tumoural *A. australis* samples was extracted using different protocols, assessed by electrophoresis, spectrophotometry and fluorometry, with the non-tumoural samples amplified by PCR using primers designed from related pinnipeds. In parallel, the coding sequences of *TP53* and *SF3B1* were reconstructed for the still unannotated *A. gazella* and *A. townsendi* using three independent comparative annotation strategies, with concordant results. Genome annotations, multiple sequence alignments and maximum likelihood genomic and protein phylogenetic tree construction were all conducted on the `SentinelBio` web platform, developed during this project. Structural analyses of the p53 and SF3B1 proteins, including three-dimensional structure superpositions and hotspot mapping, performed in UCSF ChimeraX, allowed qualitative assessment of protein conservation and identification of identity and geometric differences among the mapped hotspot residues. For DNA extraction from non-tumoural samples, the manufacturer's protocol for the kit used was effective, whereas for formalin-fixed tumour tissue, a bench protocol was more effective for DNA recovery. Among the tested targets, only *SF3B1* amplified consistently in non-tumoural *A. australis* samples and was selected for Sanger sequencing. *TP53-3*, a small region of the target gene, showed non-specific amplification, and the remaining primers did not generate products of the expected size. The DNA sample derived from tumour tissue did not amplify in any assay using the employed primers, despite measurable DNA recovery, which may be due to poor DNA quality or even mutations in the gene binding sites. Structural analyses indicated high conservation of p53 and SF3B1 hotspot regions between human and pinniped proteins. It is concluded that this work generated comparative molecular resources for *Arctocephalus* spp., identified *SF3B1* as the most viable sequencing target under the tested conditions and established an easily reproducible methodology. The *SF3B1* sequencing, currently in progress, will allow subsequent characterisation of intraspecific polymorphisms.

Keywords *Arctocephalus australis* · *SF3B1* · Marine Mammals Genetics · FFPE · Comparative Oncology · Polymorphism

1 Introduction

Cancer comprises a diverse group of diseases characterised by genetic alterations that disrupt normal cellular regulation [1, 2]. Although widely studied in humans, cancer also affects wildlife and may have implications for conservation and comparative oncology [3]. One central question in comparative oncology is Peto's Paradox, which describes the absence of a direct relationship between body size, longevity and cancer incidence across species [4]. Despite having more cells and longer lifespans, larger animals do not consistently show higher cancer rates, suggesting the evolution of effective tumour suppression mechanisms [5]. Marine mammals are particularly relevant in this context because neoplasms appear to be relatively uncommon in several groups [6]. Additionally, as top predators capable of bioaccumulating lipophilic contaminants, marine mammals may serve as sentinel species for ecosystem and public health disturbances [7].

The South American fur seal, *Arctocephalus australis* Zimmermann, 1783, is an otariid pinniped distributed along the southern coast of South America, with frequent records on the coast of Rio Grande do Sul, Brazil [8, 9]. Although currently classified as Least Concern by the International Union for Conservation of Nature, genetic and pathological data for the species remain limited [10]. Reports of cancer in fur seals of the genus *Arctocephalus* are also rare, despite the relatively high cancer-related mortality reported for Carnivora in comparative studies [5]. In contrast, the California sea lion, *Zalophus californianus*, presents a high frequency of urogenital carcinoma associated with genetic predisposition, otarine herpesvirus-1 infection and exposure to persistent organic pollutants [11, 12, 13]. Given the phylogenetic proximity between *Z. californianus* and *A. australis*, comparative analyses may help to identify conserved genes and pathways involved in tumour susceptibility or resistance in pinnipeds.

Recently, the first reported case of B-cell small lymphocytic lymphoma (B-SLL) in a free-ranging South American fur seal was described in an adult male *A. australis* stranded in southern Brazil [14]. This case represents a relevant starting point for investigating whether genetic variation in cancer-associated genes may contribute to lymphomagenesis in this species, or whether the apparent rarity of cancer in *A. australis* reflects true biological resistance, low exposure to specific risk factors, or underreporting. Among the molecular alterations associated with B-SLL and CLL/SLL, changes affecting *TP53* and *SF3B1* are particularly relevant in the pathogenesis of these neoplasms [15]. These genes are involved in tumour suppression, cell proliferation and RNA splicing, and their alteration has been associated with several haematological malignancies [16, 17, 18].

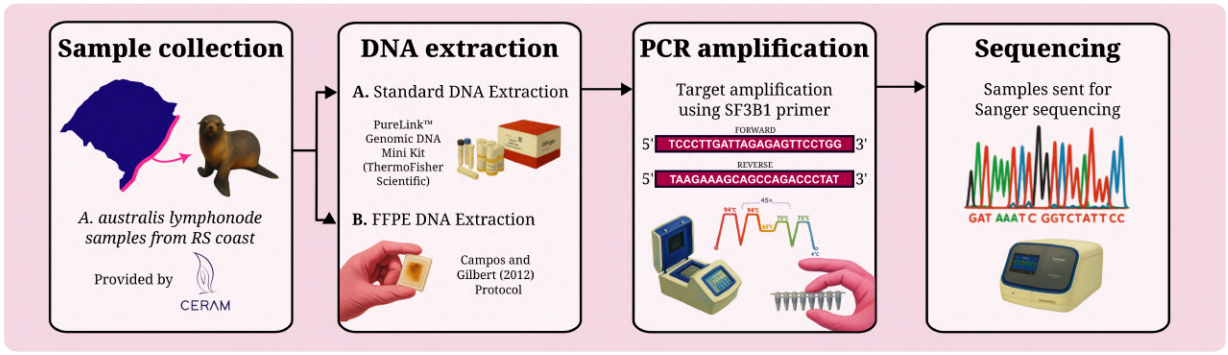
This study builds on the work of Müller [19], who established the initial molecular framework for investigating cancer-associated genes in *A. australis*. In that study, primers were designed for selected regions of *TP53*, *SF3B1* and *NOTCH1* using sequences from closely related pinnipeds, and preliminary PCR assays were performed to evaluate their amplification potential in *A. australis* samples. Although *NOTCH1* is relevant to B-SLL and was included in the previous study, it was excluded from the present study because it did not amplify reliably. In addition, phylogenetic analyses and conserved protein motif searches across representatives of Carnivora were conducted to assess the evolutionary conservation of these genes and to identify regions of possible functional relevance.

In this context, the present study continues this approach by expanding the molecular analysis of non-tumour lymph node samples, incorporating comparative annotation of *Arctocephalus* congeners and integrating structural analyses of cancer-associated hotspot regions for both *SF3B1* and *TP53*. This study aims to contribute to the understanding of cancer-associated genetic variation in South American fur seals and to support future studies in marine mammal health and comparative oncology.

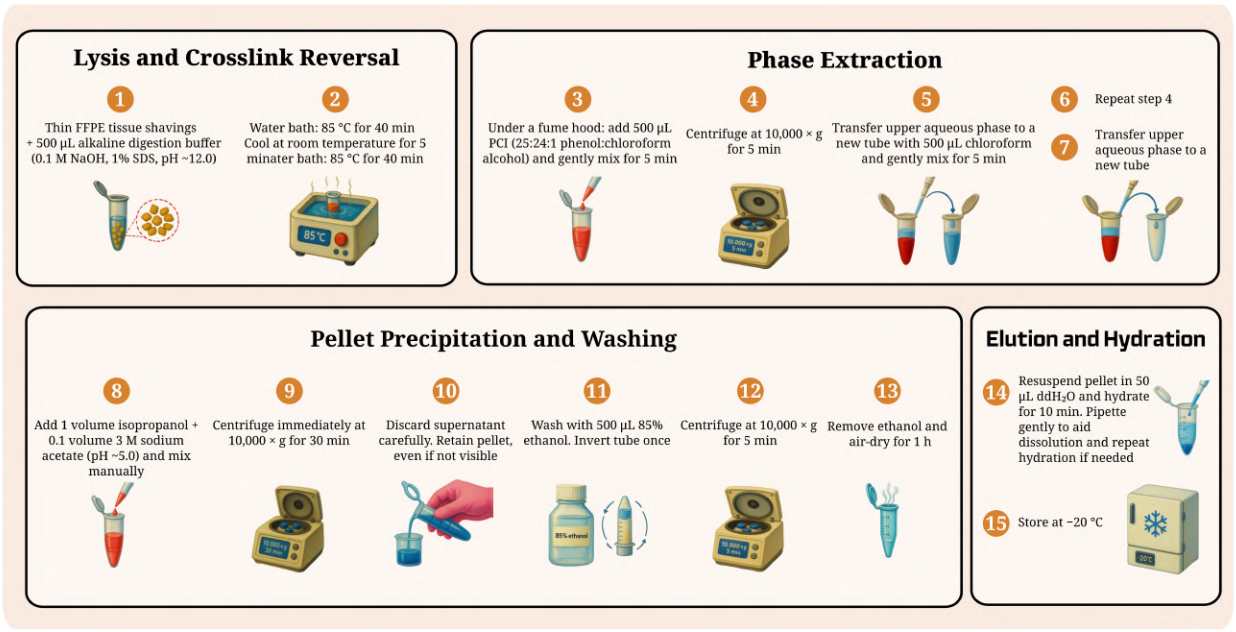
2 Materials and Methods

The overall methodological design of this study is summarised in Figure 1. This workflow integrates the main laboratory methods, the adapted Formalin-Fixed, Paraffin-Embedded (FFPE) DNA extraction protocol, and the *in silico* analyses performed through `SentinelBio`, a web-based platform made for this study, providing a general overview of the experimental and computational steps described in the following sections.

A)


Sample collection
A. australis lymphonode samples from RS coast
Provided by
CERAM
DNA extraction
A. Standard DNA Extraction
PureLink™ Genomic DNA Mini Kit (ThermoFisher Scientific)
B. FFPE DNA Extraction
Campos and Gilbert (2012) Protocol
PCR amplification
Target amplification using SF3B1 primer
FORWARD
5' TCCCTTGATTAGAGAGTTCCTGG 3'
REVERSE
5' TAAGAAAGCAGCCAGACCCTAT 3'
Sequencing
Samples sent for Sanger sequencing
GAT AAATC GGTCTATTCC

B)


Lysis and Crosslink Reversal
1 Thin FFPE tissue shavings + 500 µL alkaline digestion buffer (0.1 M NaOH, 1% SDS, pH ~12.0)
2 Water bath: 85 °C for 40 min Cool at room temperature for 5 minater bath: 85 °C for 40 min
Phase Extraction
3 Under a fume hood: add 500 µL PCI (25:24:1 phenol:chloroform alcohol) and gently mix for 5 min
4 Centrifuge at 10,000 × g for 5 min
5 Transfer upper aqueous phase to a new tube with 500 µL chloroform and gently mix for 5 min
6 Repeat step 4
7 Transfer upper aqueous phase to a new tube
Pellet Precipitation and Washing
8 Add 1 volume isopropanol + 0.1 volume 3 M sodium acetate (pH ~5.0) and mix manually
9 Centrifuge immediately at 10,000 × g for 30 min
10 Discard supernatant carefully. Retain pellet, even if not visible
11 Wash with 500 µL 85% ethanol. Invert tube once
12 Centrifuge at 10,000 × g for 5 min
13 Remove ethanol and air-dry for 1 h
Elution and Hydration
14 Resuspend pellet in 50 µL ddH2O and hydrate for 10 min. Pipette gently to aid dissolution and repeat hydration if needed
15 Store at −20 °C

C)

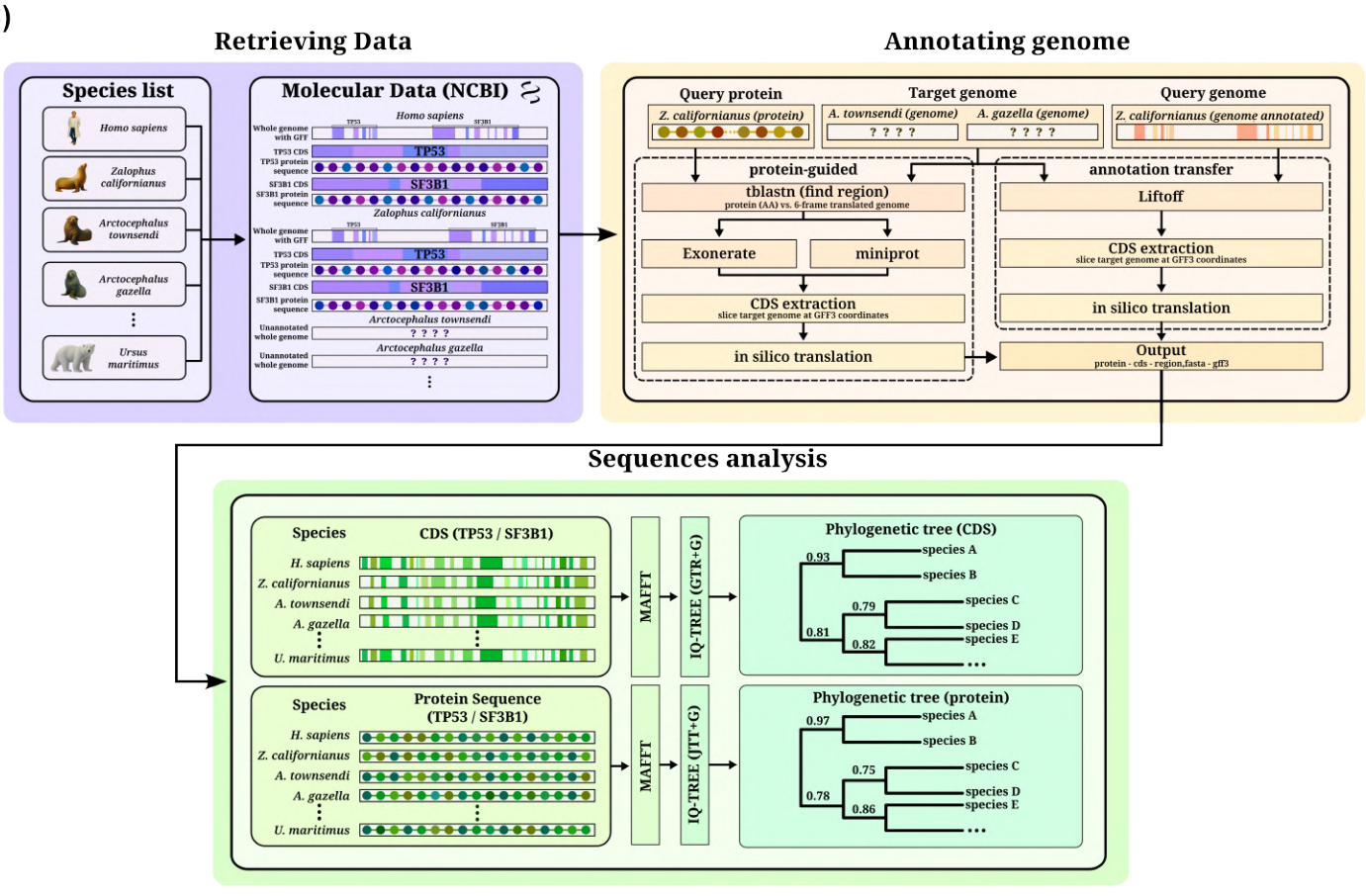

Retrieving Data
Species list
Homo sapiens
Zalophus californianus
Arctocephalus townsendi
Arctocephalus gazella
Ursus maritimus
Molecular Data (NCBI)
Homo sapiens
Zalophus californianus
Arctocephalus townsendi
Arctocephalus gazella
TP53
SF3B1
Annotating genome
Query protein
Z. californianus (protein)
Target genome
A. townsendi (genome)
A. gazella (genome)
Query genome
Z. californianus (genome annotated)
protein-guided
tblastn (find region)
protein (AA) vs. 6-frame translated genome
Exonerate
miniprot
CDS extraction
slice target genome at GFF3 coordinates
in silico translation
annotation transfer
Liftoff
CDS extraction
slice target genome at GFF3 coordinates
in silico translation
Output
protein - cds - region.fasta - gff3
Sequences analysis
Species
CDS (TP53 / SF3B1)
H. sapiens
Z. californianus
A. townsendi
A. gazella
U. maritimus
MAFFT
IQ-TREE (GTR+G)
Phylogenetic tree (CDS)
species A
species B
species C
species D
species E
Species
Protein Sequence (TP53 / SF3B1)
IQ-TREE (JTT+G)
Phylogenetic tree (protein)

Figure 1: Overview of the methodological workflow used in this study. (A) Wet-laboratory steps, including lymph node sample collection from *A. australis*, DNA extraction from non-tumoural and FFPE tumoural samples, PCR amplification targeting *SF3B1*, and sample submission for Sanger sequencing. (B) Adapted Campos and Gilbert (2012) protocol used for DNA extraction from FFPE tissue, including lysis and crosslink reversal, phase extraction, pellet precipitation and washing, and final elution and hydration. (C) *In silico* workflow performed in `SentinelBio`, including species selection, molecular data retrieval from NCBI, annotation transfer or protein-guided annotation for unannotated genomes, CDS and protein sequence extraction, multiple sequence alignment with MAFFT, and phylogenetic reconstruction with IQ-TREE.

2.1 Sampling and DNA extraction

Non-tumoral lymph node samples were obtained *post mortem* from a total of 15 individuals of *Arctocephalus australis* and 1 individual of *Lobodon carcinophaga*. All samples were collected from stranded animals along the coast of Rio Grande do Sul, Brazil, through the Centro de Reabilitação de Animais Silvestres e Marinhos (CERAM). Tissues were stored at $-20\,^{\circ}\mathrm{C}$ in microcentrifuge tubes until DNA extraction. The inclusion of *L. carcinophaga* aimed to provide a comparative reference from a phylogenetically related species. In addition, 1 rare tumoral lymph node sample from *A. australis*, previously diagnosed as B-cell small lymphocytic lymphoma, was included in the study. The animal was also collected from the coast of Rio Grande do Sul, but the material available for this study consisted only of a formalin-fixed paraffin-embedded (FFPE) block obtained from the pathology archive of the Veterinary Faculty at UFRGS.

DNA extraction from non-tumour tissues was performed using the PureLink™ Genomic DNA Mini Kit (Thermo Fisher Scientific), based on silica-column purification, using an adapted version of the manufacturer's protocol. Frozen tissue samples were initially macerated with liquid nitrogen until a fine powder was obtained. For each sample, approximately $25\,\mathrm{mg}$ of macerated tissue was transferred to a sterile $1.5\,\mathrm{mL}$ microcentrifuge tube, followed by the addition of $180\,\mu\mathrm{L}$ of digestion buffer. The samples were incubated in a water bath at $55\,^{\circ}\mathrm{C}$ for 1 h. For samples in which DNA extraction was not initially successful, the incubation time was increased up to 4 h. After incubation, the lysate was centrifuged for 3 min, and the supernatant was transferred to a new microcentrifuge tube. Then, $20\,\mu\mathrm{L}$ of RNase A were added, followed by vortexing and incubation at room temperature for 2 min. Subsequently, $200\,\mu\mathrm{L}$ of lysis buffer and $200\,\mu\mathrm{L}$ of 100% ethanol were added to the lysate, with vortexing after the addition of each reagent. The purification step was then performed using the silica columns provided in the kit. Approximately $640\,\mu\mathrm{L}$ of lysate were added to the column placed in a collection tube and centrifuged for 1 min. The collection tube was discarded and replaced, and $500\,\mu\mathrm{L}$ of Wash Buffer I, previously mixed with 100% ethanol, were added to the column, followed by centrifugation for 1 min. The collection tube was again discarded and replaced. Next, $500\,\mu\mathrm{L}$ of Wash Buffer II, previously mixed with 100% ethanol, were added to the column, which was centrifuged for 3 min. The collection tube was then discarded and replaced with a sterile $1.5\,\mathrm{mL}$ microcentrifuge tube. For the first elution, 30–50 $\mu\mathrm{L}$ of elution buffer were added to the column, followed by incubation for 1 min and centrifugation for 1 min. To increase DNA recovery, a second elution was performed, followed by centrifugation for 1.5 min. The extracted DNA was stored at $-20\,^{\circ}\mathrm{C}$ until further analysis.

For the FFPE-derived tumoral sample extraction, the protocol performed used an adapted alkaline digestion and organic extraction method based primarily on Campos and Gilbert [20], with additional modifications adapted from Oba *et al.* [21]. Thin FFPE shavings were placed in a microcentrifuge tube and incubated with $500\,\mu\mathrm{L}$ of alkaline digestion buffer (0.1 M NaOH with 1% SDS, pH approximately 12.0) in a water bath at $85\,^{\circ}\mathrm{C}$ for $40\,\mathrm{min}$, followed by cooling at room temperature for $5\,\mathrm{min}$. DNA was then extracted by adding $500\,\mu\mathrm{L}$ of phenol:chloroform:isoamyl alcohol (25:24:1), gently mixing for $5\,\mathrm{min}$, and centrifuging at 10,000 $\times g$ for $5\,\mathrm{min}$. The upper aqueous phase was transferred to a new tube containing $500\,\mu\mathrm{L}$ of chloroform, gently mixed for $5\,\mathrm{min}$, and centrifuged again under the same conditions. The recovered aqueous phase was transferred to a clean tube, and DNA was precipitated with one volume of isopropanol and 0.1 volume of 3 M sodium acetate (pH approximately 5.0), followed by centrifugation at 10,000 $\times g$ for $30\,\mathrm{min}$. After removal of the supernatant, the pellet was washed with $500\,\mu\mathrm{L}$ of 85% ethanol, centrifuged for $5\,\mathrm{min}$ at maximum speed, and air-dried for $1\,\mathrm{h}$. DNA was eluted in $50\,\mu\mathrm{L}$ of dd$\mathrm{H_2O}$, allowed to hydrate for at least $10\,\mathrm{min}$, and stored at $-20\,^{\circ}\mathrm{C}$.

The quality of the extracted DNA was assessed using two complementary approaches: agarose gel electrophoresis for visual evaluation of DNA integrity, and spectrophotometric analysis using a NanoDrop™ Lite Spectrophotometer to estimate DNA purity and concentration. DNA integrity was assessed by 1% agarose gel electrophoresis, while DNA concentration and purity were estimated by spectrophotometry using either a NanoDrop™ Lite or a NanoDrop™ One Spectrophotometer, depending on instrument availability.

For the FFPE-derived tumoral sample, DNA concentration was additionally assessed using a Qubit™ fluorometer with the Qubit™ dsDNA High Sensitivity (HS) Assay, due to the limited quantity and expected fragmentation of DNA obtained from archival FFPE material.

2.2 *In silico* sequence search and `SentinelBio` project framework

Conserved protein motif analyses for *TP53* and *SF3B1* across Carnivora were incorporated from the prior work of Müller (2024) [19], in which they were conducted as a complementary strategy for investigating gene conservation and identifying regions associated with functionally relevant variation. Those analyses employed the MEME tool from the MEME Suite [22] under the *anr* distribution, permitting any number of motif repetitions, with the search limited to 20 motifs per protein. Putative motif functions were inferred through comparison against entries in the NCBI Conserved Domain Database [23]. Protein sequences were obtained from NCBI for at least one representative of each Carnivora family with available data, spanning *Zalophus californianus*, *Eumetopias jubatus*, *Callorhinus ursinus*, *Phoca vitulina*, *Odobenus rosmarus divergens*, *Mustela putorius furo*, *Ursus americanus*, *Canis lupus familiaris*, *Felis catus*, *Hyaena hyaena*, and *Suricata suricatta*; corresponding Gene IDs are listed in Appendix 6. *NOTCH1* was also used in the motif results, but since amplification of this gene in *A. australis* was not achieved in the preceding experimental work of Müller [19], it was excluded from this study. All updated analyses described below therefore concern *TP53* and *SF3B1* exclusively.

To maintain full traceability of intermediate files and standardise parameter management across tools, all updated bioinformatic analyses were coordinated through `SentinelBio`, a web-based research workspace developed for this project and publicly available on GitHub. The system organises the workflow into *projects*, each associating species records which are assigned explicit analytical roles such as primary species, reference, or outgroup with biological targets (genes, proteins, or genomic regions), indexed input and output files, and computational logs. Every analytical step is captured as a *tool run* recording the tool identifier, parameter values, input and output file references, and execution timestamps, so that the full provenance of any intermediate product can be reconstructed without manual bookkeeping. Output files follow a structured naming convention encoding the producing tool, a species short code, the gene label, and the product type, which eliminates ambiguity when handling closely related species or multiple genes processed in parallel. In addition to the modules used here, the platform includes further tools for read quality control (FastQC [24]; cutadapt [25]) and positive selection analyses (HyPhy); these were not employed in the present study. The modules used here are summarised in Table 1.

Reference sequences for *TP53* and *SF3B1* from annotated species were retrieved through the NCBI Entrez Direct integration in `SentinelBio`, which queries the Gene database and links matching records to their corresponding RefSeq RNA entries via the `gene_nuccore_refseqrna` pathway. Each search combined the gene symbol with the NCBI taxonomic identifier for the target clade, following the structure `GENE[Gene Name] AND txid<taxon>[Organism:noexp]`; for example, Otariidae records for *SF3B1* were retrieved using `SF3B1[Gene Name] AND txid9704[Organism:noexp]`. The annotated protein sequences of *Z. californianus*, which is the closest fully annotated otariid available in NCBI RefSeq, were separately retrieved and designated as the reference source for downstream comparative annotation. Unannotated whole-genome assemblies for *Arctocephalus gazella* and *A. townsendi* were downloaded from NCBI via their assembly accessions using the genome assembly fetch module, which streams and decompresses each genomic FASTA directly to project storage.

2.3 Gene annotation

Because no annotated records were available in NCBI for *A. gazella* or *A. townsendi* at either locus, coding sequences for *TP53* and *SF3B1* in both species were reconstructed by comparative annotation, executed through `SentinelBio`. Three independent strategies were applied to each target genome and their outputs cross-compared, to reduce the risk of method-specific prediction artefacts.

In all three strategies, the first step consisted of a translated nucleotide BLAST search (`tblastn`; BLAST+ [26]) of the *Z. californianus* protein sequence against a local index of the target assembly, using an E-value cutoff of 1×10^{-20}. Because both genes span multiple exons, each typically producing a separate high-scoring segment pair (HSP), the candidate genomic locus was identified as the contig carrying the highest cumulative bit-score summed across all HSPs, rather than selected on the basis of a single best hit. The interval spanning the outermost HSPs on the best-scoring contig was extracted with $5\,\mathrm{kb}$ of flanking sequence on each side, extended to $20\,\mathrm{kb}$ for *SF3B1* to accommodate its longer introns.

In the first annotation strategy, the extracted genomic region was supplied to Exonerate [27] as the target, with the *Z. californianus* protein as query, under the `protein2genome` model (`-bestn 1`). Coding sequences were obtained by slicing the target FASTA at the CDS coordinates reported in Exonerate's GFF3 output, reverse-complementing minus-strand predictions where applicable, and translating *in silico* using the standard genetic code. In the second strategy, the same candidate region was supplied to miniprot [28], run with GFF output enabled (`-gff`) and `-N 1` to retain only the single best alignment per query protein. CDS extraction and translation followed the same coordinate-based procedure. In the third strategy, annotations were transferred directly from *Z. californianus* to each target genome using

Liftoff [29], which aligns reference gene sequences internally via minimap2 and projects coordinate annotations to the target assembly without protein-guided correction. The `-polish` option was enabled to improve alignment at gene boundaries. The longest transcript per gene was selected from the resulting annotation, and CDS bases were extracted at the lifted coordinates and translated as above. This strategy provides a methodologically independent cross-check: unlike the protein-driven aligners, Liftoff transfers coordinates without sequence correction and therefore captures only what is directly inferable from structural homology with the reference.

Predictions from all three methods were compared at the level of the reconstructed CDS and translated protein sequences for each gene and target species. All three strategies produced concordant annotations for both *TP53* and *SF3B1* in *A. gazella* and *A. townsendi*, supporting the reliability of the reconstructed coding sequences. The miniprot-derived sequences were retained for all downstream analyses, as miniprot represents the most recent development in protein-to-genome alignment methodology among the three tools applied, offering improved handling of frameshifts and insertion–deletion events relative to Exonerate.

2.4 Sequence alignment and phylogenetic inference

For each gene, CDS and translated protein sequences for the complete species set were compiled and aligned separately using MAFFT [30], executed through `SentinelBio`. The input comprised available pinniped RefSeq sequences, the newly annotated *A. gazella* and *A. townsendi* sequences, selected ursid representatives (*Ursus maritimus*, *U. arctos*, *U. americanus*, and *Ailuropoda melanoleuca*), and *Homo sapiens* as outgroup. The `-auto` strategy was specified, delegating algorithm selection to MAFFT based on the size and composition of the input. Default gap-opening and gap-extension penalties of 1.53 and 0.0 were applied. Both nucleotide CDS and protein alignments were retained: CDS alignments for analyses requiring codon-level resolution, and protein alignments for residue-level comparisons and phylogenetic inference.

Maximum-likelihood phylogenetic trees for *TP53* and *SF3B1* were inferred from the protein alignments using IQ-TREE 2 [31], executed through `SentinelBio`. The substitution model was selected automatically via ModelFinder Plus (`-m MFP`), which evaluates candidate amino acid substitution matrices and selects the one minimising the Bayesian Information Criterion, determining the appropriate model for each gene independently. Branch support was quantified using the ultrafast bootstrap approximation with 1000 replicates. *Homo sapiens* was specified as the outgroup for display rooting (`-o`), which affects only the visual representation of the topology and not the underlying maximum-likelihood optimisation. For publication, phylogenetic trees were rendered outside `SentinelBio` using a custom Python pipeline based on Biopython's `Bio.Phylo` module for parsing and matplotlib for drawing, distributed as a Jupyter notebook.

2.5 Protein structure prediction

To complement the sequence-level analyses, three-dimensional structural comparisons were conducted to assess whether cancer-associated mutation hotspots characterised in human proteins occupy spatially equivalent positions in pinniped models. Structural models for *H. sapiens* and *Z. californianus* SF3B1 and p53 were retrieved from the AlphaFold Protein Structure Database [32] through the platform's structure fetch module, which resolves UniProt accessions and downloads the corresponding predicted structures. For *A. gazella* and *A. townsendi*, for which no entries were available in the AlphaFold database, models were generated locally using ColabFold [33] via the local structure prediction module integrated in `SentinelBio`. A preliminary superposition of the *Arctocephalus* spp. models against the corresponding *Z. californianus* models confirmed highly similar overall conformations across species; accordingly, the final structural figures were produced using *Z. californianus* and *H. sapiens* as representative models for the pinniped–human comparison.

Pairwise structural superpositions were performed outside `SentinelBio` using a custom Jupyter notebook distributed alongside this study. For each gene, the notebook loads the AlphaFold and ColabFold models described above, extracts C_α tables, and globally aligns the amino acid sequences of each non-human model against the *H. sapiens* reference using pairwise Needleman–Wunsch alignment (`Bio.Align`), establishing a residue-number mapping that is independent of the coordinate numbering in each PDB file. Residues with a pLDDT score below 70 were excluded from the rigid-body fit to prevent disordered loop regions from distorting the superposition; positions below 50 were additionally flagged as disordered and omitted from structural comparisons. The rotation matrix and translation vector were computed by Kabsch singular value decomposition as implemented in Biopython's `Bio.PDB.Superimposer` [34]. After applying the transformation, per-residue C_α distances between each non-human model and the human reference were recorded across all aligned positions. The notebook also wrote the transformation directly into each mobile PDB file, producing pre-aligned coordinate sets, and automatically generated ChimeraX command scripts (`.cxc`) that loaded these pre-aligned structures, assigned a species-specific colour palette, rendered hotspot residues as spheres, displayed structural

neighbours within 8 Å, and saved publication-quality PNG images. These scripts were executed in UCSF ChimeraX for final figure generation.

Cancer-associated mutation hotspots for SF3B1 and p53 were selected from the Cancer Hotspots Database (cancerhotspots.org; Memorial Sloan Kettering Cancer Center [35, 36]) by ranking all recorded residues in ascending order of Q-value, where lower values indicate greater statistical significance under false discovery rate correction [37]. The seven residues with the smallest Q-values were retained per gene, a practical threshold chosen to focus the structural comparison on a representative set of statistically robust hotspots rather than to serve as a formal significance cutoff; all selected residues had $Q < 0.05$. The retained p53 hotspot positions were R175, R213, Y220, G245, R248, R273, and R282; the retained SF3B1 positions were residues 700, 625, 902, 666, 626, 741, and 742. These residues were mapped onto the superposed three-dimensional models and examined through visual inspection in UCSF ChimeraX, run externally to `SentinelBio`, with attention to their spatial distribution and local atomic environment.

Table 1: Bioinformatics tools integrated in `SentinelBio` and used in this study.

Function	**Tool**
Reference sequence retrieval	NCBI Entrez Direct
Similarity search	BLAST+ (`tblastn`)
Gene annotation, protein-to-genome	Exonerate, `protein2genome` model
Gene annotation, protein-to-genome	miniprot
Gene annotation, annotation lift-over	Liftoff
Multiple sequence alignment	MAFFT
Maximum-likelihood phylogeny	IQ-TREE 2
Protein structure retrieval	AlphaFold Protein Structure Database
Protein structure prediction	ColabFold (local)
Curated cancer mutation hotspots	cancerhotspots.org

2.6 Primers, amplification and sequencing

The primers used were the same from the previous work of Müller [19]. Because the *Arctocephalus australis* genome was not available in public databases, primers were designed using reference sequences from closely related pinniped species available in GenBank at the time of that study. Sequences of *TP53* and *SF3B1* were selected from species of Otariidae, including *Zalophus californianus*, *Eumetopias jubatus*, and *Callorhinus ursinus*, and from *Phoca vitulina* as an external reference from Phocidae.

In the previous primer design workflow, nucleotide sequences from the selected species were aligned in BioEdit to avoid divergent regions. The target regions showed high interspecific conservation, supporting their use for primer design. Primers were manually designed to amplify exons 14–16 of *SF3B1*, and exons 2–11 of *TP53*. Primer quality was evaluated *in silico* using NetPrimer (Premier Biosoft), prioritising sequences of 17–30 nucleotides, GC content between 40–60%, and melting temperature close to 57 °C. The primer sequences and expected amplicon sizes are shown in Table 2.

Table 2: Primer sequences used and expected amplicon sizes.

Primer	**F/R**	**Sequence** $5' \rightarrow 3'$	**Amplicon size**
SF3B1	F	TCCCTTGATTAGAGAGTTCCTGG	1190 nt
	R	TAAGAAAGCAGCCAGACCCTAT	
TP53-1	F	ATGCAGGACCCACAGTCAGA	1077 nt
	R	CCTCTGTCACGAACTCGGAC	
TP53-2	F	CTACAAGCTGTTTTGCCAG	1008 nt
	R	CGTCCAGCACATTACCACT	
TP53-3	F	TATGGTTGGAGTTAGGGGG	1244 nt
	R	CGGACAAGCGAGGAAACTG	
TP53-4	F	TATTGGCAATGCGAGGTTG	1190 nt
	R	AAAAGGCAAGAGAGGGAGGG	

Initial PCR screening and annealing-temperature optimisation for the primer pairs were performed by Müller (2024) [19]. In that previous screening, the primer pairs *SF3B1* and *TP53*-3 showed the most promising amplification profiles in non-tumoral *A. australis* samples. *NOTCH1* was also used in the previous work, but since it did not yield detectable amplification products, this gene was excluded from the molecular analyses in this study.

Based on the primer pairs that showed the most promising results in the previous PCR screenings, additional PCR tests were performed. The test reactions were prepared in a final volume of 25 µL, consisting of 23 µL of PCR mix and 2 µL of DNA template. The PCR mix contained 16.3 µL of Milli-Q water, 2.5 µL of 10× reaction buffer, 1.5 µL of 50 mmol L^{-1} $MgCl_2$, 0.5 µL of dNTP mix, 1.0 µL of forward primer, 1.0 µL of reverse primer, and 0.2 µL of Taq DNA polymerase. The final PCRs, prepared for sequencing, were carried out using the same proportions but with a final volume of 50 µL, consisting of 46 µL of PCR mix and 4 µL of DNA template.

The amplification conditions consisted of an initial denaturation at 94 °C for 2 min, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 63 °C for 30 s, and extension at 72 °C for 1 min. A final extension step was performed at 72 °C for 7 min, after which the reactions were held at 4 °C. The Taq DNA polymerase, $MgCl_2$, reaction buffer, and dNTPs used in these reactions were obtained from Ludwig Biotecnologia. The dNTP mix was prepared from the dNTP Set 100 mmol L^{-1} by combining 10 µL of each dNTP with 60 µL of Milli-Q water.

Final tests and final PCR products were analysed by electrophoresis on 1.5% agarose gels prepared in 1× TBE buffer and stained with ethidium bromide at a final concentration of approximately 0.4 µg mL^{-1}. Samples were loaded with loading dye, and fragment size was estimated by comparison with 100 bp DNA ladder (Ludwig Biotecnologia). Electrophoresis was performed at 80 V for approximately 50 min. Bands were visualised using a UV transilluminator.

Amplification products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega), following the manufacturer's instructions. Purified samples were quantified and submitted for bidirectional Sanger sequencing using the same forward and reverse primers used in the original amplifications. Sequencing was requested from ACT Gene (https://actgene.com.br/), using the ABI 3730 Capillary Electrophoresis Genetic Analyzer and the ABI BigDye Terminator v. 3.1 Cycle Sequencing kit. Sequence processing, including quality assessment, assembly of forward and reverse reads, and generation of consensus sequences, will be performed by the sequencing service provider. The processed sequencing files will be used for downstream comparative analyses after delivery.

After the processed sequencing data are received, consensus sequences from *A. australis* will be compared against reference *SF3B1* sequences to identify nucleotide polymorphisms. Annotated sequences from *A. gazella* and *A. townsendi* will also be included for comparison. Sequence variation will be inspected among individuals to detect conserved and variable positions. Identified variants will be classified according to their genomic position and predicted coding effect.

3 Results

3.1 DNA recovery and quality assessment

DNA extraction was evaluated using complementary approaches: NanoDrop analysis for concentration and purity profiles, Qubit quantification for the FFPE-derived tumour sample, agarose gel electrophoresis, and an amplification-based assay to assess whether the recovered DNA could be used in downstream reactions, as seen in Figure 2.

Representative NanoDrop absorbance spectra were obtained for both non-tumoral tissue samples and the FFPE-derived tumoral sample (Figure 2A). The non-tumoral DNA sample showed a typical nucleic acid absorbance profile, with a defined peak around 260 nm. The FFPE-derived tumoral sample also showed detectable nucleic acid absorbance, and its spectrum did not present marked distortion, despite the expected limitations associated with formalin fixation and archival storage.

A PCR assay was performed using samples from the extracted DNA to evaluate DNA quality and amplification suitability (Figure 2B). Both undiluted extracts and samples diluted to 20 ng µL^{-1} were tested, and the PCR products were visualised by agarose gel electrophoresis. Amplification was observed for the non-tumour DNA samples, indicating that DNA extraction was successful and that the recovered DNA was suitable for subsequent PCR amplification. In contrast, the FFPE-derived tumour sample did not show visible amplification in the gel, which was consistent with the expected limitations of archival FFPE material, including DNA fragmentation and reduced amplifiable template availability.

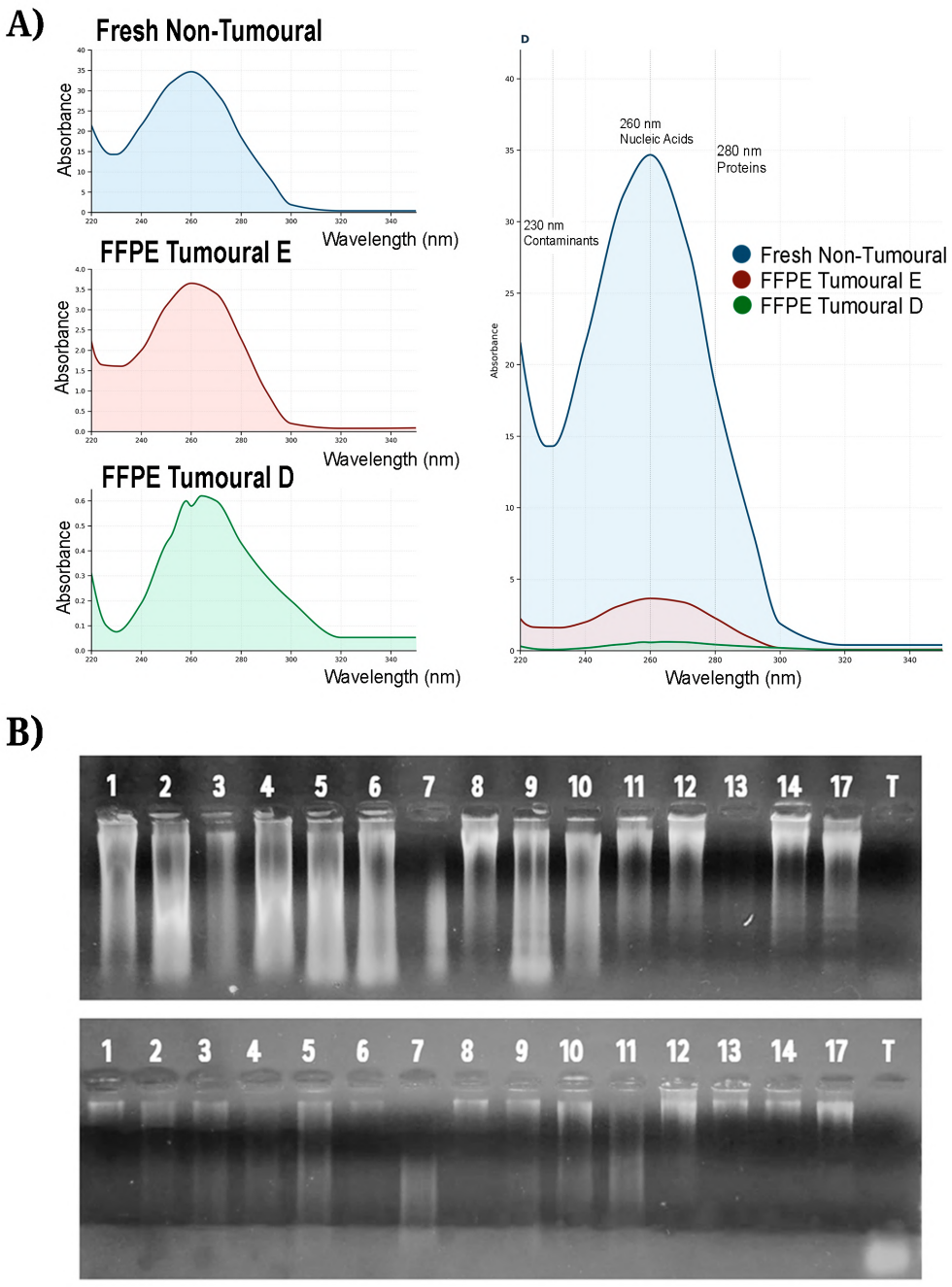


Figure 2: Quality assessments methods used for the DNA recovery. (A) Effect of tissue preservation method on nucleic acid yield and purity profiles. NanoDrop absorbance spectra show distinct quality differences between the fresh non-tumoural sample and FFPE tumoural extracts. Individual spectra are shown for the fresh non-tumoural sample, FFPE tumoural E and FFPE tumoural D, followed by all samples plotted on a unified absolute scale. The combined profile highlights the marked reduction in nucleic acid concentration in FFPE-derived samples, as well as baseline shifts and flattened peaks consistent with DNA degradation and chemical modification associated with FFPE preservation. Reference wavelengths associated with common contaminants at 230 nm, DNA/RNA at 260 nm and protein at 280 nm are indicated. (B) Agarose gel electrophoresis of DNA PCR amplification from *A. australis* samples. The first row shows undiluted DNA extracts, except for the tumoural sample (T), which was diluted. The second row shows diluted DNA extracts, except for the tumoural sample (T), which was undiluted. L indicates the DNA ladder, B indicates the blank control, numbered lanes correspond to *A. australis* non-tumoural samples, and T corresponds to the *A. australis* tumoural sample.

For the FFPE-derived tumoural sample, additional fluorometric quantification using the Qubit™ dsDNA High Sensitivity Assay resulted in a DNA concentration of $7.08\,\mathrm{ng\,\mu L^{-1}}$, corresponding to $354\,\mathrm{ng\,mL^{-1}}$ in the Qubit assay tube. This value was considered compatible with previously reported concentrations obtained from FFPE-derived DNA using Qubit-based fluorometric quantification, although such yields may vary substantially according to the extraction method and pre-analytical conditions [38]. Nevertheless, this result was interpreted as evidence of DNA recovery rather than as confirmation of downstream suitability. In partially degraded FFPE samples, Qubit-based quantification has been shown to provide a more reliable estimate of double-stranded DNA than NanoDrop measurements, but it does not directly assess whether the recovered DNA is amplifiable [39]. Moreover, the amount of FFPE-derived DNA measured in nanograms may not reflect the fraction of DNA available for PCR or sequencing, since formalin fixation, paraffin embedding and archival storage can lead to fragmentation and other forms of molecular damage [40]. Therefore, although DNA was successfully recovered from the archival tumoural material, the absence of PCR amplification was interpreted with caution and may reflect poor DNA integrity or a reduced amount of amplifiable template, rather than insufficient total DNA concentration alone.

3.2 *In silico* results

Conserved protein motif analysis was performed for *TP53* and *SF3B1* across representative species of the order Carnivora. Overall, both genes showed a high degree of motif conservation among the analysed taxa, with most motifs detected across all species included in each dataset. A total of 17 motifs were identified for *TP53* and 19 for *SF3B1*.

For *TP53*, several motifs were shared by all species, although some showed lineage-specific absence or restricted distribution. In *SF3B1*, most motifs were broadly conserved, supporting the conservation of recurrent regions within the protein across Carnivora. The distribution of conserved and lineage-restricted motifs for each gene is summarised in Tables 3 and 4.

Table 3: Conserved motifs identified in *TP53* protein sequences across Carnivora.

Motif	Amino acid sequence	Distribution among Carnivora
TP1	DDRNTFRHSVVVPYEPPEVGSDCTTVHYNYMCNSSCM GGMNRRPILTIIT	Present in all species analysed.
TP2	CVRAMAIYKKSEFVTEVVRRCPHHERCPDNSDGLAPP QHLIRVEGNLRAK	Present in all species analysed.
TP3	LQILGRESFEMFSEL	Present in all species analysed; multiple copies detected, with fewer copies in *Hyaena hyaena*.
TP4	FLQSGKTCPVTLTYS	Present in all species analysed; multiple copies detected, with fewer copies in *Ursus americanus*.
TP5	ALELKPAPSGPEPGGSRPLSS	Present in all species analysed.
TP6	CPGLDRRTEEE	Present in all species analysed.
TP7	HLKAKK	Absent only in *Hyaena hyaena*.
TP8	WNLLPE	Absent only in *Mustela putorius furo*.
TP9	MQDPQS	Present in all species analysed.
TP10	PSLNKL	Present in all species analysed.
TP11	PPPPDT	Present in all species analysed.
TP12	KTYPGTYGFR	Detected only in *Ursus americanus* and *Canis lupus*.
TP13	LLSEGV	Absent only in *Phoca vitulina* and *Mustela putorius furo*.
TP14	SRMSAA	Detected only in *Suricata suricatta* and *Hyaena hyaena*.
TP15	QKGKPL	Detected only in *Ursus americanus* and *Suricata suricatta*.
TP16	DGEYFV	Detected only in *Ursus americanus* and *Felis catus*.
TP17	LMFCRP	Detected only in *Ursus americanus*.

Table 4: Conserved motifs identified in *SF3B1* protein sequences across Carnivora.

Motif	Amino acid sequence	Distribution among Carnivora
SF1	TPGASKRKKRWDETPDRTPGG	Present in all species analysed.
SF2	LKBKEEKVREIAABLIAAJAEYAAEY	Present in all species analysed.
SF3	KAIAGADAVVANYIGIEVMLPFJKDFLKH	Present in all species analysed.
SF4	ELLKKLKKAIRQHRVKTLVAI	Present in all species analysed.
SF5	NALLNYVPPSLFQITPEVLQR	Present in all species analysed.
SF6	AKKAGKATKISKTRLDIDBMDEQ	Present in all species analysed.
SF7	LIKIEERIRILSNEALLAHLPKLVNDDEA	Present in all species analysed.
SF8	HEDIEAQIREIQGKKAALDEAQGVGLDSTGYYDQEIY GGSD	Present in all species analysed.
SF9	STPLLTPGKDSIGTP	Present in all species analysed.
SF10	KILIDEDEYTKEERTMIISPE	Present in all species analysed.
SF11	AJYARTYTPLDALPA	Present in all species analysed.
SF12	TAKELLRIITE	Present in all species analysed.
SF13	EEAKKPVDKVV	Present in all species analysed.
SF14	PSGGAGGA	Present in all species analysed.
SF15	GIESFD	Present in all species analysed.
SF16	LVKVID	Present in all species analysed.
SF17	MAKIAK	Absent only in *Hyaena hyaena* and *Zalophus californianus*.
SF18	IKGGSD	Detected only in *Suricata suricatta* and *Hyaena hyaena*.
SF19	CFEKJP	Detected only in *Suricata suricatta* and *Zalophus californianus*.

Comparative annotation successfully recovered *TP53* and *SF3B1* coding sequences for *A. gazella* and *A. townsendi*. In the multiple sequence alignments, both annotated *Arctocephalus* spp. proteins showed high similarity to the other pinniped sequences, with no evident large insertions, deletions, or disruptive changes in the analysed regions (Supplementary Material A.2). For *TP53*, amino acid differences were more frequent and distributed across several functional domains, including the transactivation domain, proline-rich region, and DNA-binding domain. Even so, *A. gazella* and *A. townsendi* remained highly similar to each other and to other otariids, especially *Z. californianus* and *C. ursinus*. In contrast, *SF3B1* showed very limited interspecific variation across the alignment, with only isolated divergent residues among the sampled taxa. Together, these results indicate that the reconstructed *Arctocephalus* sequences are consistent with the expected conservation pattern for pinnipeds.

To further evaluate the placement of the annotated *Arctocephalus* sequences, additional IQ-TREE phylogenies were constructed as seen in Figure 3 using a reduced dataset composed only of pinniped species, selected bear sequences for comparison, and Homo sapiens. In both *TP53* and *SF3B1*, the annotated A. gazella and *A. townsendi* sequences clustered close to other pinnipeds, including *Zalophus californianus* and *Callorhinus ursinus*. This placement is consistent with the expected proximity among otariid species, especially considering that *Callorhinus* has been discussed in relation to the same subfamilial known as Arctocephalinae.

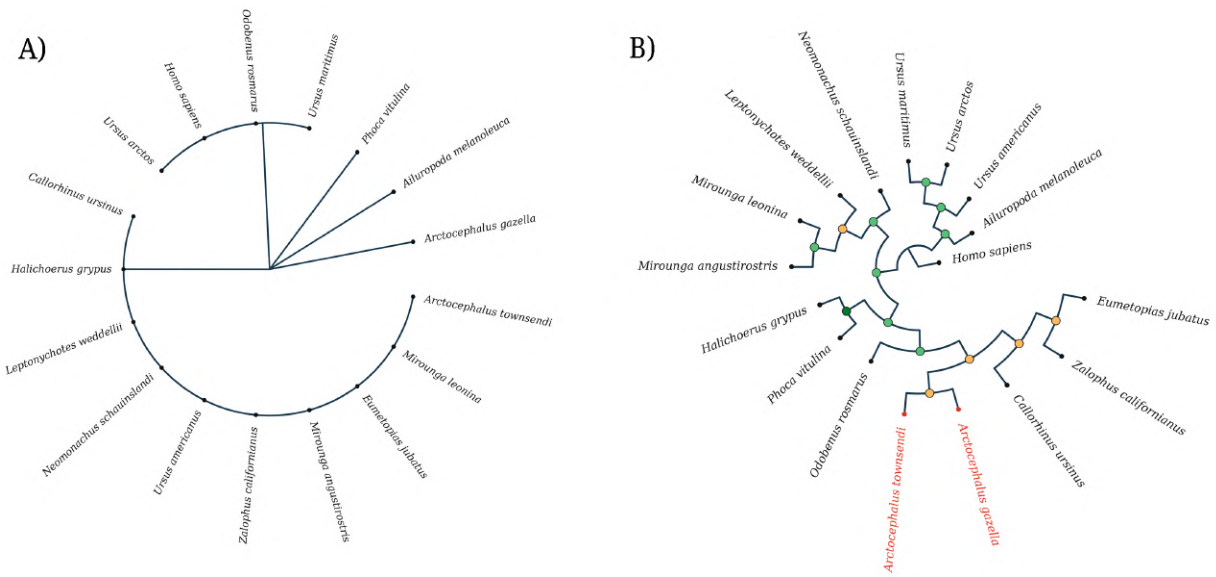


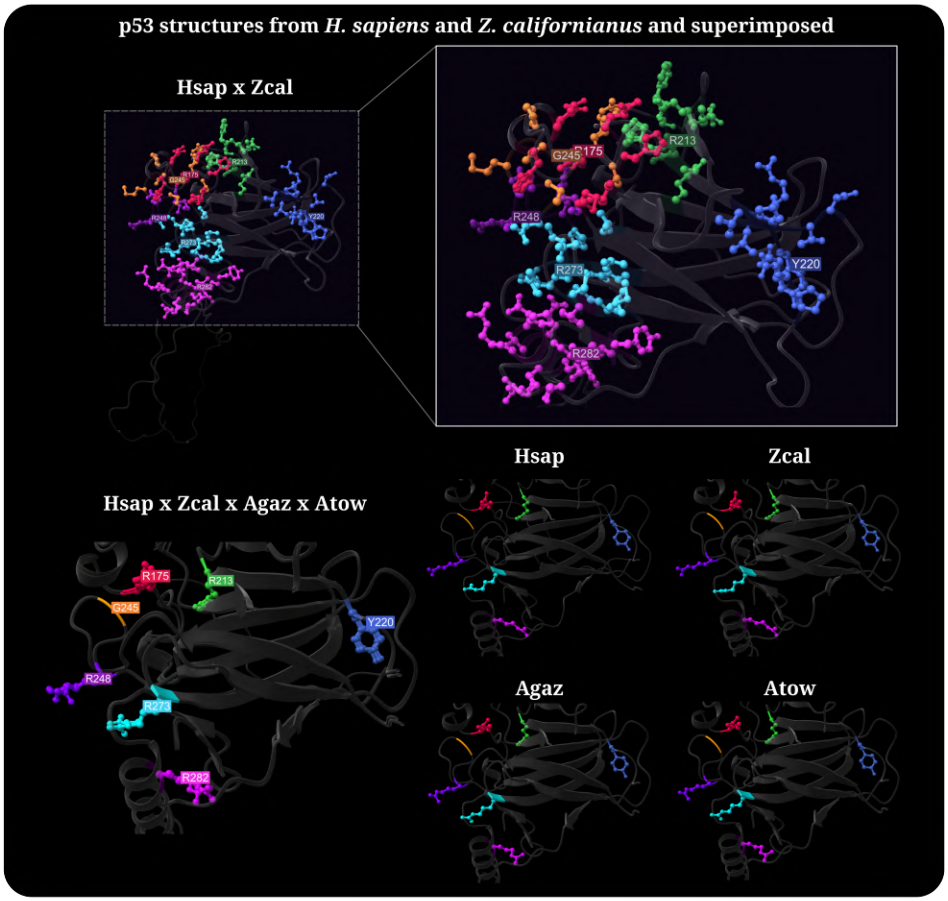
Figure 3: Maximum likelihood phylogenies of *SF3B1* (A) and *TP53* (B) inferred with IQ-TREE using reduced datasets composed of pinniped species, a representative group of bears and *H. sapiens*. Annotated *Arctocephalus* spp. sequences are highlighted in red. In (A), *SF3B1* showed limited interspecific variation among the sampled species, with the annotated *Arctocephalus* spp. sequences clustering similarly to the other pinnipeds. In (B), *TP53* showed clearer phylogenetic resolution, with the annotated *Arctocephalus* spp. clustering closer to *Z. californianus*, *C. ursinus* and *E. jubatus*, supporting the expected phylogenetic placement. Yellow indicates moderate branch support approximately above 50%, light green indicates high support above 90%, and dark green indicates very high support above 95%.

Following the phylogenetic analyses, structural comparisons were performed to evaluate whether this sequence-level conservation was also reflected at the protein level.

The independently generated *Arctocephalus* spp. models showed high structural similarity to the *Z. californianus* models. Based on this similarity, *Z. californianus* was used as the representative pinniped model in the final structural comparisons with *Homo sapiens*.

Structural superimposition indicated that the proteins encoded by *SF3B1* and *TP53* were highly conserved between *H. sapiens* and *Z. californianus* as seen in Figure 4. Several selected cancer-associated hotspot residues occupied comparable spatial positions and showed similar local atomic arrangements between the human and pinniped models.

A)


p53 structures from H. sapiens and Z. californianus and superimposed
Hsap x Zcal
R175
G245
R213
R248
R273
Y220
R282
Hsap
Zcal
Hsap x Zcal x Agaz x Atow
Agaz
Atow

B)

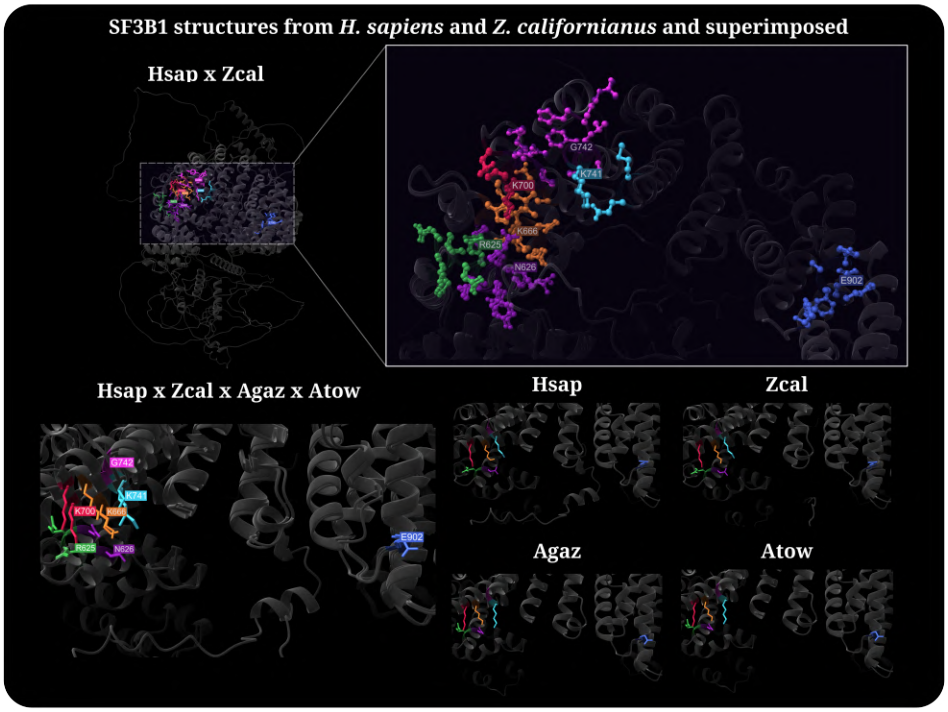

SF3B1 structures from H. sapiens and Z. californianus and superimposed
Hsap x Zcal
G742
K741
K700
K666
R625
N626
E902
Hsap x Zcal x Agaz x Atow
Hsap
Zcal
Agaz
Atow

Figure 4: Structural comparison of p53 (A) and SF3B1 (B) between *H. sapiens* and *Z. californianus*. Selected cancer-associated hotspot residues were mapped onto the superimposed three-dimensional protein models to assess their spatial correspondence between species. In p53, residues R175, R213, Y220, G245, R248, R273 and R282 were analysed. In SF3B1, residues R625, N626, K666, K700, K741, G742 and E902 were analysed. The selected positions showed high conservation and comparable local structural arrangements, supporting the conservation of major cancer-associated hotspot regions between human and pinniped proteins.

3.3 Molecular analyses

The gradient PCR assays performed by Müller [19] using non-tumour sample 1 identified the primer pairs with the most suitable amplification profiles and annealing temperatures, as summarised in Table 5. In these assays, the *SF3B1* and *TP53*-3 primer pairs amplified fragments of the expected size, approximately ~ 1200 bp. The optimal annealing temperatures were $63\,^{\circ}\mathrm{C}$ for *SF3B1* and $65\,^{\circ}\mathrm{C}$ for *TP53*-3. *TP53*-1, *TP53*-2 and *TP53*-4 primer pairs did not amplify fragments of the expected size in *A. australis* samples. In contrast, *TP53*-1 amplified only in the *L. carcinophaga* sample.

Based on the annealing temperatures established in the gradient PCR assays by Müller [19] (Table 5), subsequent PCRs were performed using all available non-tumour and tumour samples. Amplification of fragments within the expected size range was again confirmed in non-tumour *A. australis* samples using the *TP53*-3 and *SF3B1* primer pairs. The *TP53*-1 primer pair amplified only in the non-tumour *L. carcinophaga* sample, sample 7. None of the tumour samples showed amplification with the primers tested.

Table 5: Best annealing temperatures identified for each gene-specific primer pair.

Primer pair	**Best annealing temperature**	**Amplification result**
SF3B1	$63\,^{\circ}\mathrm{C}$	Expected-size fragment amplified in *A. australis*
TP53-3	$65\,^{\circ}\mathrm{C}$	Expected-size fragment amplified in *A. australis*
TP53-1	Not determined	Amplification only in *L. carcinophaga*
TP53-2	Not determined	No expected-size amplification
TP53-4	Not determined	No expected-size amplification

In the PCR assays for this study (Figure 5), some of the primers were tested based on the annealing temperatures that appeared most promising in the gradient PCR assay.

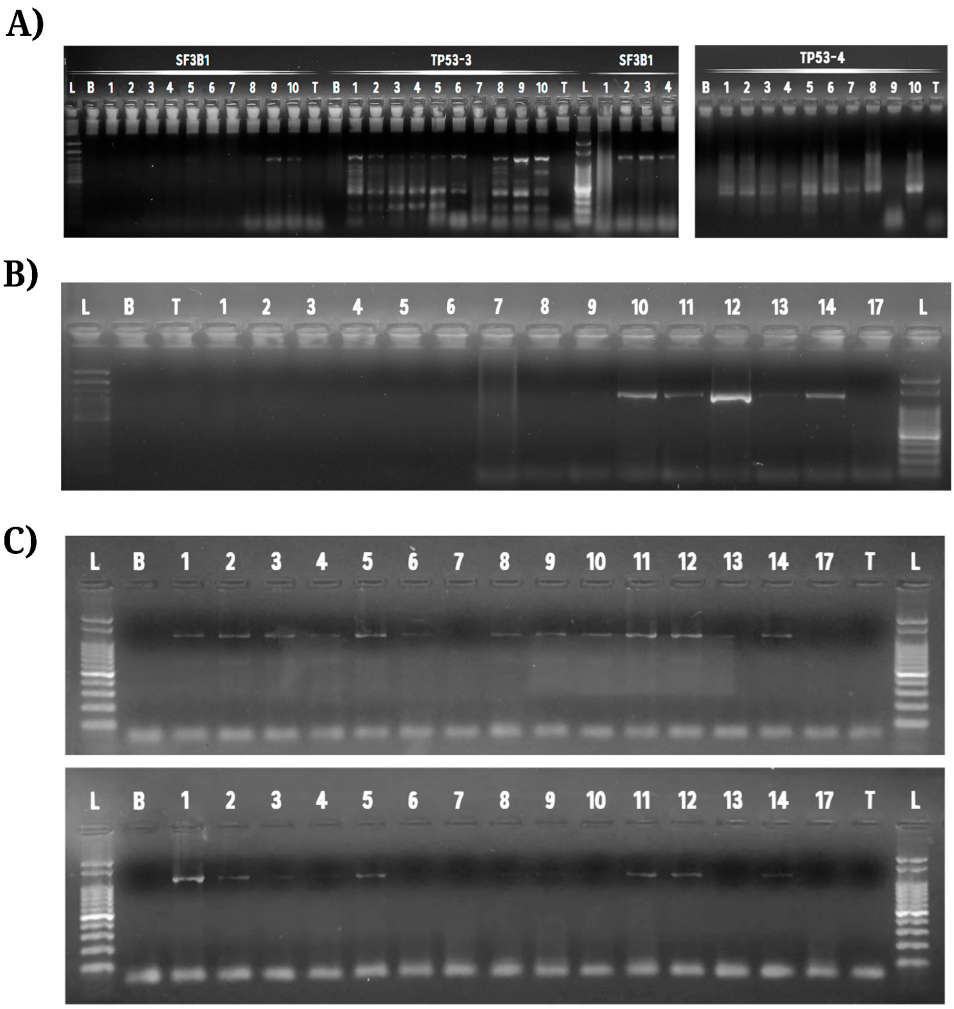


Figure 5: Agarose gel electrophoresis of PCR assays performed for primer evaluation and final sequencing submission. (A) PCR assay of *SF3B1*, *TP53*-3 and *TP53*-4. L: DNA ladder; B: blank control; numbered lanes: *A. australis* non-tumour samples; T: *A. australis* tumour sample. (B) PCR amplification using only the *SF3B1* primer pair. L: DNA ladder; B: blank control; samples 1–14: *A. australis* non-tumour samples; T: *A. australis* tumour sample; lane 17: *L. carcinophaga*. (C) Final *SF3B1* PCR amplification submitted for sequencing, shown in the first and second gel rows. L: DNA ladder; B: blank control; samples 1–14: *A. australis* non-tumour samples; T: *A. australis* tumour sample; sample 17: *L. carcinophaga*.

SF3B1 and *TP53*-3 primers were tested at 63 °C (Figure 5A) and *TP53*-4 was tested again (Figure 5A) at 56 °C. As the first *SF3B1* samples, numbered 1–4, had been loaded in substantially lower quantities in the first wells of the gel, these reactions were repeated in the final wells. As observed, *TP53*-3 showed excessive non-specific amplification and *TP53*-4 showed no proper amplification either. Meanwhile, *SF3B1* showed amplification at the expected amplicon size, approximately 1000 bp. The remaining primers were not tested due to practical limitations at that stage of the work, and priority was given to *SF3B1*, which appeared to be the most promising primer set based on the results obtained (Figure 5B).

Although *SF3B1* was initially selected as the most promising primer set, subsequent assays showed inconsistent amplification. After further optimisation of sample and reagent handling, amplification at the expected amplicon size was recovered, and the final PCR products were submitted for sequencing.

In the final PCR (Figure 5C), the blank control showed no amplification, indicating the absence of detectable contamination in the reaction, whereas the tumoural sample did not amplify in any of the tested assays. For sequencing, the *SF3B1* amplification products from the first row, corresponding to samples 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12 and 14, were

submitted together with products from the second row, corresponding to samples 1, 2, 3, 4, 11, 12 and 14, as a duplicate sequencing strategy for the overlapping samples. Since sample 13 was the only sample that amplified in the previous PCR assays, this amplification product was also submitted for sequencing.

4 Discussion

The results obtained in this study should be interpreted as a preliminary molecular and comparative framework for investigating cancer-associated genes in *Arctocephalus australis*, rather than as direct evidence of tumour-associated mutations in this species. This distinction is important because the original motivation of the work was linked to a rare case of B-cell small lymphocytic lymphoma (B-SLL) in *A. australis*, but the tumoural material available for molecular analysis was restricted to formalin-fixed paraffin-embedded (FFPE) tissue. Consequently, the study combined three complementary approaches: assessment of DNA quality and amplifiability, *in silico* characterisation of cancer-associated genes in related pinnipeds, and PCR-based selection of the most feasible target for sequencing in non-tumoural *A. australis* samples.

The DNA quality assessment showed a clear contrast between non-tumoural frozen tissue and FFPE-derived tumoural material. DNA extracted from non-tumoural lymph node samples showed suitable concentration and purity profiles and was successfully amplified in PCR assays, indicating that the extraction protocol produced DNA compatible with downstream molecular analyses. In contrast, although DNA was recovered from the FFPE-derived tumoural sample and quantified by Qubit fluorometry, this material failed to amplify in all PCR assays. This reinforces the interpretation that total DNA concentration alone is not a reliable indicator of amplifiable template quality in archival FFPE samples. Formalin fixation and paraffin embedding can cause DNA fragmentation, crosslinking and chemical modification, all of which reduce amplification efficiency even when measurable DNA is present [20, 21, 39, 40].

A likely contributor to the failure of the tumoural sample was the length of the targeted amplicons. The primer pairs used in this study were originally designed for DNA of higher integrity, and the expected PCR products were approximately 1 kb or longer. Although this size range was suitable for frozen non-tumoural tissue, it was probably suboptimal for FFPE-derived DNA, which is commonly fragmented. In retrospect, future assays using archival material should prioritise FFPE-specific primer design, with short overlapping amplicons across the regions of interest and an independent amplification-control assay to estimate the amount of amplifiable DNA before target sequencing [41]. This is particularly important for rare wildlife tumours, where material is limited and repeated extraction may not be possible.

The PCR results also clarified the practical scope of the present molecular analysis. The previous gradient PCR assays performed by Müller [19] indicated that the *SF3B1* and *TP53-3* primer pairs were the most promising targets in non-tumoural *A. australis* samples, whereas *NOTCH1*, *NOTCH2* and other *TP53* primer pairs did not produce fragments of the expected size. In the present study, *NOTCH1* was excluded from the tests, but additional PCR testing seem to have confirmed this pattern for *TP53*. *TP53-4* did not amplify adequately. *TP53-3* showed excessive non-specific amplification, limiting its suitability for direct sequencing under the tested conditions. By contrast, *SF3B1* produced fragments compatible with the expected size in several non-tumoural samples and was therefore selected as the most feasible sequencing target.

This selection should nevertheless be interpreted with caution. Amplification of *SF3B1* was not uniformly robust across all assays, and repeated optimisation of sample and reagent handling was required before usable products were obtained. Even in the final PCR, not all non-tumoural samples yielded visible amplification products. The clean blank control in the final assay supports the absence of detectable contamination, but the variability among samples indicates that amplification success was affected by template quality, reaction conditions, primer-template compatibility, or a combination of these factors. Because the primers were designed using *Z. californianus* and other related pinniped sequences, rather than an *A. australis* reference genome, primer-site mismatches caused by interspecific divergence remain a plausible explanation for inconsistent amplification. This reinforces the need for species-specific primer design once the *SF3B1* sequences generated here, or a complete reference genome for *A. australis*, become available.

The *in silico* analyses were essential for contextualising these molecular results. Conserved protein motif analysis and structural comparisons indicated that *TP53* and *SF3B1* are highly conserved across Carnivora, and that selected human cancer-associated hotspot residues in p53 and the *SF3B1*-encoded protein occupy comparable spatial positions in pinniped models. This conservation is not unexpected, since *SF3B1* encodes a core spliceosomal factor and *TP53* encodes a central tumour suppressor, both of which are expected to be under strong functional constraint in mammalian lineages. Therefore, the relevance of this conservation should not be interpreted as direct evidence of cancer susceptibility or resistance in *A. australis*. Rather, it provides a constraint-based argument: if variants are subsequently identified at or near these conserved positions in *A. australis*, they are more likely to warrant functional attention than variants occurring in poorly conserved regions.

The greater interspecific variation observed in *TP53* relative to *SF3B1* may reflect different levels of structural and functional constraint between these proteins. *SF3B1* showed stronger conservation across the sampled taxa, with limited amino acid variation in the alignments and apparent conservation of the selected hotspot-associated regions in the structural analyses. In contrast, *TP53* showed a comparatively higher number of amino acid differences among species, although the major cancer-associated hotspot residues analysed here remained spatially conserved between the human and pinniped models. These results support the use of conserved regions in both genes as biologically relevant targets for comparative cancer analysis, while not allowing direct inference about cancer incidence or tumour risk in pinnipeds.

A concrete output of the *in silico* component was the reconstruction of *TP53* and *SF3B1* coding sequences for *A. gazella* and *A. townsendi*, for which annotated records were not available in public databases. The concordance of three methodologically independent annotation strategies, *Exonerate*, *miniprot* and *Liftoff*, supports the reliability of these reconstructions. In addition, the placement of the annotated *Arctocephalus* spp. sequences close to other otariids in the phylogenetic analyses was consistent with the expected relationships among pinniped taxa. Together, these findings indicate that the reconstructed sequences are suitable for use as congeneric comparative references in downstream analyses involving *A. australis*. This is especially relevant because the absence of an available *A. australis* genome limits both primer design and variant interpretation.

The structural analyses further connected the sequence-level results with the biological rationale of the study. The observed spatial conservation of selected hotspot residues suggests that major cancer-associated regions described in humans are recognisable in pinniped proteins. However, the structural comparisons were based on predicted protein models rather than experimentally resolved structures. AlphaFold- and ColabFold-derived models are useful for comparative assessment of conserved folded regions, but they provide static predictions and are less informative for intrinsically disordered regions, flexible loops, protein complexes, ligand interactions and dynamic conformational changes [32, 33]. Consequently, the structural superpositions should be interpreted as qualitative evidence of spatial conservation, not as direct evidence that equivalent mutations would necessarily have identical biochemical or pathological effects in pinnipeds.

The comparison with *Z. californianus* is therefore useful mainly as evolutionary and methodological context. California sea lions represent the best-characterised otariid cancer model, particularly because urogenital carcinoma has been repeatedly reported in stranded animals and associated with a multifactorial model involving otarine herpesvirus infection, persistent organic pollutants and host susceptibility [42, 12]. In *A. australis*, by contrast, neoplasia is represented by isolated reports, including the B-SLL case that motivated this study and a captive case of otarine herpesvirus-1-associated urogenital carcinoma [43]. Thus, *Z. californianus* provides a useful comparative framework, but the same oncogenic mechanisms should not be assumed for *A. australis*, especially because the tumour type, exposure history and available molecular evidence differ.

The cross-species rationale adopted here is further supported by recent work in *Zalophus californianus* [44]. performed cross-species whole-exome sequencing of CSL urogenital carcinoma by leveraging the evolutionary conservation between the domestic dog and CSL genomes as a capture and analysis scaffold, in the absence of a high-quality species-specific reference. This approach recovered recurrent somatic single-nucleotide variants and copy-number aberrations, including a recurrent exon 4 C>T variant in *CD274*/PD-L1 identified in 54 of 68 (79.4%) animals with diagnosed urogenital carcinoma. Although that study addressed a different tumour type and gene set, and benefited from tumour–normal pairs and exome-wide data not available here, it provides a direct precedent for the strategy underlying the present work: using better-annotated related genomes as comparative references when the target otariid lacks an assembled genome. It demonstrates that cross-species capture and annotation can recover biologically recurrent alterations in otariids, supporting both the reconstruction of *TP53* and *SF3B1* for the unannotated *Arctocephalus* species and the use of related-pinniped sequences for primer design in *A. australis*. At the same time, the contrast in scope reinforces the limitations noted above: single-locus Sanger sequencing of non-tumoural material cannot capture the somatic and copy-number landscape accessible to exome-wide tumour sequencing.

These findings should also be interpreted in the context of uneven cancer surveillance among otariids. The high apparent cancer burden reported for *Z. californianus* is supported by long-standing stranding and necropsy series, whereas neoplasia in *Arctocephalus* spp. is still represented mainly by isolated case reports and limited pathological surveys [6, 42]. Therefore, the apparent rarity of cancer in *A. australis* should not be interpreted as evidence of tumour resistance. Instead, it may reflect a combination of biological differences, lower exposure to specific carcinogenic co-factors, and underdetection caused by the limited availability of systematic necropsy and molecular surveillance data.

Once the consensus *SF3B1* sequences from the *A. australis* individuals are received, several lines of analysis will be possible, although the scope of interpretation will remain limited by the use of Sanger sequencing of a single amplified region in a restricted number of individuals. First, alignment of the consensus sequences against one another will allow the identification of nucleotide polymorphisms segregating among the sampled non-tumoural individuals. Each variable position can be classified according to its location within the amplified region and, when located in coding sequence,

according to its predicted coding effect, such as synonymous, missense or nonsense substitution. This will provide a preliminary description of intraspecific *SF3B1* variation in *A. australis*.

Second, any variants identified in *A. australis* can be compared with the conserved hotspot-associated regions described in humans. The amplified region was selected because it overlaps a portion of *SF3B1* that includes recurrently mutated residues in human haematological malignancies. Therefore, variants detected within or near homologous positions would be of particular interest. Conversely, the absence of variation in these positions would also be informative, as it would support local conservation in a region already shown to be structurally conserved across the analysed models. However, even if a variant is detected near a known human hotspot, this would not be sufficient to classify it as pathogenic in *A. australis* without additional functional evidence.

Third, comparison of the *A. australis* sequences with the reconstructed *A. gazella* and *A. townsendi* sequences, as well as with the annotated *Z. californianus* reference, will place any intraspecific variants in an interspecific context. This comparison will help distinguish positions that are variable within *A. australis* from positions that represent fixed differences between otariid species. It may also indicate whether particular residues are broadly conserved across otariids or specific to the *Arctocephalus* lineage.

The sequencing strategy also limits the type of variation that can be detected. Because the present analysis is based on PCR amplification and Sanger sequencing of a single *SF3B1* region from non-tumoural material, the resulting data will primarily describe germline or constitutional polymorphism within the sampled individuals. This approach will not detect variants outside the amplified fragment, copy-number alterations, structural variants, regulatory changes, or low-frequency somatic subclones that would require tumour-derived material and deeper sequencing. Therefore, even if no variation is observed in the amplified region, this should be interpreted as local conservation of the assayed fragment rather than absence of cancer-associated variation in *SF3B1* or in other cancer-associated genes.

Because the tumour-derived FFPE sample failed to amplify, the present study cannot determine whether the reported B-SLL case carried alterations in *SF3B1* or *TP53*. This limitation changes the interpretative scope of the molecular results. Rather than providing direct evidence of tumour-associated mutations, the study establishes a comparative and methodological framework for future analyses of better-preserved tumoural material from *A. australis*. The rarity of the case still supports the relevance of reporting the data obtained here, but the absence of amplifiable tumour DNA prevents any direct conclusion regarding the molecular basis of the neoplasm.

Protein-level analysis was also initially considered as a complementary approach to evaluate the potential impact of genetic variants identified in *SF3B1*. In particular, immunohistochemistry could provide information on protein localisation within tumoural tissues. However, this approach was not pursued due to limitations in antibody availability, high cost and the absence of specific funding for this assay. In addition, mutations in *SF3B1* are primarily associated with altered RNA splicing rather than necessarily producing changes in total protein abundance or localisation detectable by immunohistochemistry [17, 45]. For this reason, immunohistochemistry would likely have limited interpretative value for the main question addressed here, especially in the absence of matched RNA or fresh tumoural material.

Future studies should therefore prioritise RNA-based approaches when suitable biological material is available. Reverse-transcription PCR or RNA sequencing would be more informative for evaluating the functional consequences of *SF3B1* variants, since they could detect aberrant splicing patterns directly [45]. Such analyses would be particularly relevant if future tumoural samples from *A. australis* or related pinnipeds become available in adequate preservation conditions. In parallel, species-specific primer design, broader sampling of non-tumoural individuals, sequencing of additional cancer-associated genes, and harmonised pathological surveillance would help clarify whether the patterns observed here are representative of the species or limited to the sampled individuals.

Overall, this study provides preliminary molecular resources for cancer-associated genes in *Arctocephalus* spp., identifies *SF3B1* as the most feasible sequencing target under the tested conditions, and places future *A. australis* variants in a comparative framework involving conserved hotspot-associated regions and closely related otariids. Although the absence of amplifiable tumour DNA prevents direct inference about the reported B-SLL case, the results establish a baseline for future studies using freshly preserved tumoural tissue, FFPE-adapted molecular assays, species-specific primers and RNA-based functional approaches.

5 Conclusion

This study established a preliminary molecular and comparative framework for the investigation of cancer-associated genes in *Arctocephalus australis*. DNA obtained from frozen non-tumoural lymph node samples showed sufficient quality for PCR amplification, whereas the FFPE-derived tumoural sample, despite measurable DNA recovery, failed to amplify in all tested assays. This indicates that the main limitation of the tumoural material was likely related to DNA integrity and amplifiable template availability rather than total DNA concentration alone.

Among the tested molecular targets, *SF3B1* was the only gene that amplified with sufficient consistency in non-tumoural *A. australis* samples and was therefore selected for Sanger sequencing. The remaining targets in *TP53* could not be reliably amplified under the tested conditions, probably due to a combination of primer-template incompatibility, target size and sample-dependent amplification efficiency. The sequencing results, once available, will allow a preliminary characterisation of intraspecific variation in the amplified region of *SF3B1*.

The *in silico* analyses complemented the molecular assays by reconstructing *TP53* and *SF3B1* coding sequences for *A. gazella* and *A. townsendi*, providing congeneric comparative references for future analyses of *A. australis*. Conserved motif analysis, phylogenetic reconstruction and structural comparisons showed high conservation of *SF3B1* and relevant conservation of *TP53* among pinnipeds and other carnivores. The spatial conservation of selected cancer-associated hotspot residues supports the biological relevance of these regions for comparative oncology, although it does not allow direct inference about cancer susceptibility or tumour development in *A. australis*.

Because the tumoural sample did not amplify, this study cannot determine whether the reported B-SLL case carried alterations in *SF3B1* or *TP53*. Nevertheless, the results provide useful methodological information for future work, particularly regarding the need for FFPE-adapted short amplicons, species-specific primer design, broader sampling and RNA-based approaches to assess the functional consequences of *SF3B1* variation. Overall, this study contributes initial molecular resources for *Arctocephalus* spp. and provides a basis for future investigations of cancer-associated genetic variation in South American fur seals.

6 Future perspectives

Future studies should prioritise three complementary directions: (i) broader sequencing of non-tumoural individuals from different regions to establish a more robust baseline of *SF3B1* polymorphism in *A. australis*; (ii) FFPE-adapted and fresh-tissue molecular protocols for rare tumoural cases, including short amplicons and matched tumour-normal comparisons; and (iii) RNA-based approaches, such as RT-PCR or RNA sequencing, to test whether candidate *SF3B1* variants affect splice-site usage or transcript structure. In parallel, harmonised pathological surveillance, virological testing and contaminant measurements would be necessary to distinguish true biological rarity of neoplasia in *Arctocephalus* spp. from underdetection.

7 Acknowledgements

The authors thank the Centro de Reabilitação de Animais Silvestres e Marinhos (CERAM) for providing access to the biological material of *Arctocephalus australis* used in this study. Additional thanks are extended to the Laboratório do Departamento de Plantas de Lavoura, Faculdade de Agronomia, and the Laboratório de Sinalização e Plasticidade Celular, Campus do Vale, both at the Universidade Federal do Rio Grande do Sul (UFRGS), for providing institutional and technical support during the development of this work.

8 Declaration on AI use

During the preparation of this work, GPT-5.5 Thinking, developed by OpenAI, was used to improve readability and language. Claude Opus 4.7, developed by Anthropic, was also used to assist with the development of specific Python scripts for the `SentinelBio` platform. After using these tools, the content and AI-assisted code were reviewed and edited. Full responsibility is assumed for the content of the publication.

9 Data Availability

The `SentinelBio` source code is publicly available at https://github.com/Sentinel-Bio/.

References

[1] L. C. U. Junqueira and J. Carneiro. *Biologia celular e Molecular*. Guanabara Koogan, Rio de Janeiro, 9 edition, 2012.

[2] D. Tarin. *Understanding Cancer: The Molecular Mechanisms, Biology, Pathology and Clinical Implications of Malignant Neoplasia*. Springer International Publishing, Cham, 2023.

[3] D. McAloose and A. L. Newton. Wildlife Cancer: A Conservation Perspective. *Nature Reviews Cancer*, 9(7):517–526, 2009.

[4] V. Callier. Solving Peto's Paradox To Better Understand Cancer. *Proceedings of the National Academy of Sciences*, 116(6):1825–1828, 2019.

[5] O. et al. Vincze. Cancer Risk Across Mammals. *Nature*, 601(7892):263–267, 2022.

[6] S. J. Newman and S. A. Smith. Marine Mammal Neoplasia: A Review. *Veterinary Pathology*, 43(6):865–880, 2006.

[7] G. D. Bossart. Marine Mammals as Sentinel Species For Oceans and Human Health. *Veterinary Pathology*, 2010.

[8] E. A. Crespo, A. C. M. Schiavini, N. A. García, V. Franco-Trecu, N. P. Goodall, D. Rodríguez, J. S. Morgante, and L. R. Oliveira. Status, Population Trend and Genetic Structure of South American Fur Seals (*Arctocephalus australis*) in Southwestern Atlantic Waters. *Marine Mammal Science*, 31:866–890, 2015.

[9] L. R. et al. de Oliveira. Morphological and Genetic Evidence for Two Evolutionarily Significant Units (ESUs) in the South American Fur Seal, *Arctocephalus gazella*. *Conservation Genetics*, 9(6):1451–1466, 2008.

[10] S. Cárdenas-Alayza, L. R. Oliveira, and E. A. Crespo. *Arctocephalus australis*, 2016. IUCN Red List of Threatened Species: e.T2055A45223529.

[11] H. M. Browning, F. M. D. Gulland, J. A. Hammond, K. M. Colegrove, and A. J. Hall. Common cancer in a Wild Animal: the California Sea Lion (*Zalophus californianus*) as an Emerging Model for Carcinogenesis. *EcoHealth*, 2015.

[12] F. M. D. et al. Gulland. Persistent Contaminants and Herpesvirus OtHV1 Are Positively Associated With Cancer in Wild California Sea Lions (*Zalophus californianus*). *Frontiers in Marine Science*, 7:602565, 2020.

[13] G. M. et al. Ylitalo. The Role of Organochlorines in Cancer-Associated Mortality in California Sea Lions (*Zalophus californianus*). *Marine Pollution Bulletin*, 50(1):30–39, 2005.

[14] D. B. Amorim, P. R. Ribeiro, M. B. Bandinelli, J. V. Z. Echenique, M. V. Bianchi, P. R. Almeida, F. R. Spilki, L. F. Baumbach, and L. Sonne. B-cell Small Lymphocytic Lymphoma in a Free-Ranging South American Fur Seal (*Arctocephalus australis*). *Diseases of Aquatic Organisms*, 156:1–6, 2023.

[15] Brunangelo Falini, Giovanni Martino, and Stefano Lazzi. A Comparison of the International Consensus and 5th World Health Organization Classifications of Mature B-cell Lymphomas. *Leukemia*, 37(1):18–34, 2023.

[16] A. M. Goh, C. R. Coffill, and D. P. Lane. The Role of Mutant p53 in Human Cancer. *The Journal of Pathology*, 223(2):116–126, 2011.

[17] D. et al. Cilloni. *SF3B1* Mutations in Hematological Malignancies. *Cancers*, 14(19):4927, 2022.

[18] J. D. Schiffman and M. Breen. Comparative Oncology: What Dogs and Other Species Can Teach Us About Humans with Cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1673):20140231, 2015.

[19] Beatriz Cristine Müller. Câncer em Mamíferos Marinhos: Investigando Genes Relacionados ao Linfoma em Lobos-marinhos-sul-americanos (*Arctocephalus australis* Zimmermann, 1783). Undergraduate thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, 2024.

[20] Paula F. Campos and Thomas M. P. Gilbert. DNA Extraction From Formalin-Fixed Material. In Beth Shapiro and Michael Hofreiter, editors, *Ancient DNA: Methods and Protocols*, volume 840 of *Methods in Molecular Biology*, pages 81–85. Humana Press, 2012.

[21] Utako Oba, Kenichi Kohashi, Yuhei Sangatsuda, Yoshinao Oda, Koh-Hei Sonoda, Shouichi Ohga, Koji Yoshimoto, Yasuhiro Arai, Shinichi Yachida, Tatsuhiro Shibata, Takashi Ito, and Fumihito Miura. An Efficient Procedure For The Recovery of DNA from Formalin-Fixed Paraffin-Embedded Tissue Sections. *Biology Methods and Protocols*, 7(1):bpac014, 2022.

[22] Timothy L. Bailey, Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Research*, 37(suppl. 2):W202–W208, jul 2009.

[23] Jiyao Wang, Aron Marchler-Bauer, Minghui Ge, Yuling Bo, Christopher J. Lanczycki, Lewis Y. Geer, Fascencies Marchler, Shenheng He, Jiansheng Song, Wannian Shi, Chanjuan Zheng, Farideh Chitsaz, Myra K. Derbyshire, J. Underhill, Shenzhi Lu, Stephen H. Bryant, R. Geer, and A. Marchler-Bauer. The Conserved Domain Database in 2023. *Nucleic Acids Research*, 51(D1):D384–D388, 2023.

[24] S. Andrews. FASTQC. A Quality Control Tool For High-Throughput Sequence Data, 2010.

[25] Marcel Martin. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. 17(1):10–12, 2011.

[26] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. BLAST+: Architecture and Applications. *BMC Bioinformatics*, 10:421, 2009.

[27] Guy S. C. Slater and Ewan Birney. Automated Generation of Heuristics for Biological Sequence Comparison. *BMC Bioinformatics*, 6(31):31, 2005.

[28] Heng Li. Protein-to-Genome Alignment with Miniprot. *Bioinformatics*, 39(1):btad014, 01 2023.

[29] Alaina Shumate and Steven L. Salzberg. Liftoff: Accurate Mapping of Gene Annotations. *Bioinformatics*, 37(12):1639–1643, 2021.

[30] Kazutaka Katoh and Daron M. Standley. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4):772–780, 04 2013.

[31] Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, 05 2020.

[32] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research*, 50(D1):D439–D444, 01 2022.

[33] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: Making Protein Folding Accessible to All. *Nature Methods*, 19(6):679–682, 2022.

[34] Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. de Hoon. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics*, 25(11):1422–1423, June 2009.

[35] Matthew T. Chang, Siddhartha Asthana, Sizhi P. Gao, et al. Identifying Recurrent Mutations in Cancer Reveals Widespread Lineage Diversity and Mutational Specificity. *Nature Biotechnology*, 34:155–163, 2016.

[36] Matthew T. Chang, Tripti Shrestha Bhattarai, Alison M. Schram, Craig M. Bielski, Mark T. A. Donoghue, Philip Jonsson, Debyani Chakravarty, Sarah Phillips, Cyriac Kandoth, Alexander Penson, Alexander Gorelick, Tambudzai Shamu, Swati Patel, Christopher Harris, JianJiong Gao, Selcuk Onur Sumer, Ritika Kundra, Pedram Razavi, Bob T. Li, Dalicia N. Reales, Nicholas D. Socci, Gowtham Jayakumaran, Ahmet Zehir, Ryma Benayed, Maria E. Arcila, Sarat Chandarlapaty, Marc Ladanyi, Nikolaus Schultz, José Baselga, Michael F. Berger, Neal Rosen, David B. Solit, David M. Hyman, and Barry S. Taylor. Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discovery*, 8(2):174–183, February 2018.

[37] John D. Storey and Robert Tibshirani. Statistical Significance for Genomewide Studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

[38] Agnieszka K. Sarnecka, Dominika Nawrat, Monika Piwowar, Janusz Ligęza, Jakub Swadźba, and Piotr Wójcik. DNA Extraction from FFPE Tissue Samples: A Comparison of Three Procedures. *Contemporary Oncology*, 23(1):52–58, 2019.

[39] Michele Simbolo, Marisa Gottardi, Vincenzo Corbo, Matteo Fassan, Andrea Mafficini, Giorgio Malpeli, Rita T. Lawlor, and Aldo Scarpa. DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS ONE*, 8(6):e62692, 2013.

[40] Anine Larsen Ottestad, Elisabeth F. Emdal, Bjørn H. Grønberg, Tarje O. Halvorsen, and Hong Yan Dai. Fragmentation Assessment of FFPE DNA Helps in Evaluating NGS Library Complexity and Interpretation of NGS Results. *Experimental and Molecular Pathology*, 126:104771, 2022.

[41] Barbara van Deventer, Lorraine du Toit-Prinsloo, and Chantal van Niekerk. Practical Tips to Using Formalin-Fixed Paraffin-Embedded Tissue Archives for Molecular Diagnostics in a South African Setting. *African Journal of Laboratory Medicine*, 11(1):6, 2022.

[42] Angela C. Deming, Kathleen M. Colegrove, Padraig J. Duignan, Ailsa J. Hall, James F. X. Wellehan, and Frances M. D. Gulland. Prevalence of Urogenital Carcinoma in Stranded California Sea Lions (*Zalophus californianus*) from 2005–15. *Journal of Wildlife Diseases*, 54(3):581–586, 2018.

[43] Mark P. Dagleish, Michelle Barrows, Mary Maley, Richard Killick, Jean Finlayson, Richard Goodchild, Anne Valentine, Richard Saunders, Kim Willoughby, Kenneth C. Smith, and Mark F. Stidworthy. The First Report of Otarine Herpesvirus-1-Associated Urogenital Carcinoma in a South American Fur Seal (*Arctocephalus australis*). *Journal of Comparative Pathology*, 149(1):119–125, 2013.

[44] Isabella G. Livingston, Catherine F. Wise, Allison N. Dickey, Rachael Thomas, Alissa C. Deming, Barbie Halaska, Frances M. D. Gulland, Kathleen M. Colegrove, Pádraig Duignan, and Matthew Breen. Cross-Species Exome Sequencing Reveals Recurrent Genomic Alterations in California Sea Lion (*Zalophus californianus*) Urogenital Carcinoma and Highlights a Recurrent PD-L1 Variant. *Genes*, 17(2), 2026.

[45] Alice D. Tang, Cameron M. Soulette, Marijke J. van Baren, Kyle Hart, Elizabeth Hrabeta-Robinson, Chia J. Wu, and Angela N. Brooks. Full-Length Transcript Characterization of *SF3B1* Mutation in Chronic Lymphocytic Leukaemia Reveals Downregulation of Retained Introns. *Nature Communications*, 11:1438, 2020.

A Supplementary Material

A.1 Gene IDs for preliminary phylogenetic analysis in Carnivora

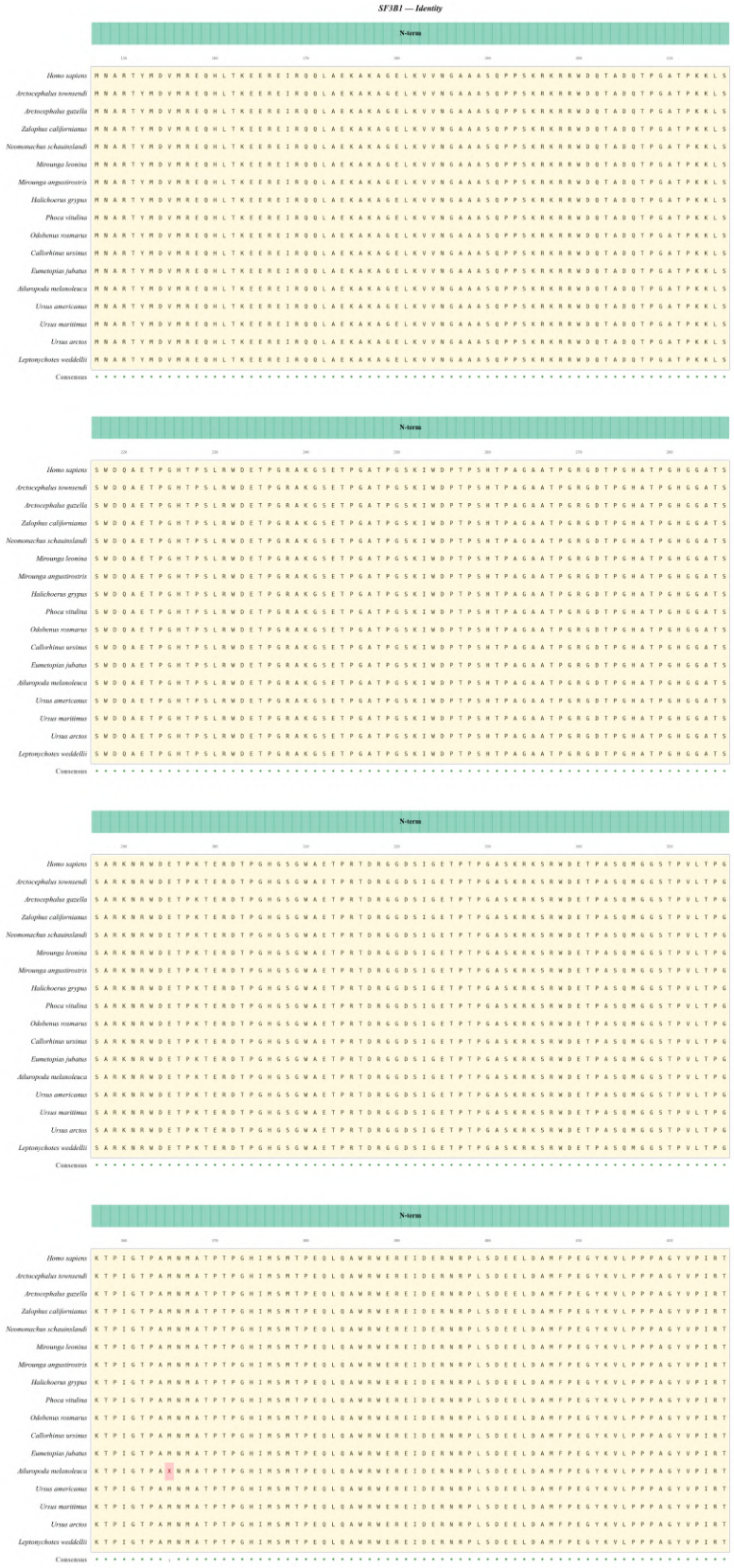
Table 6: Gene IDs used for the phylogenetic analysis of *TP53* and *SF3B1* in Carnivora.

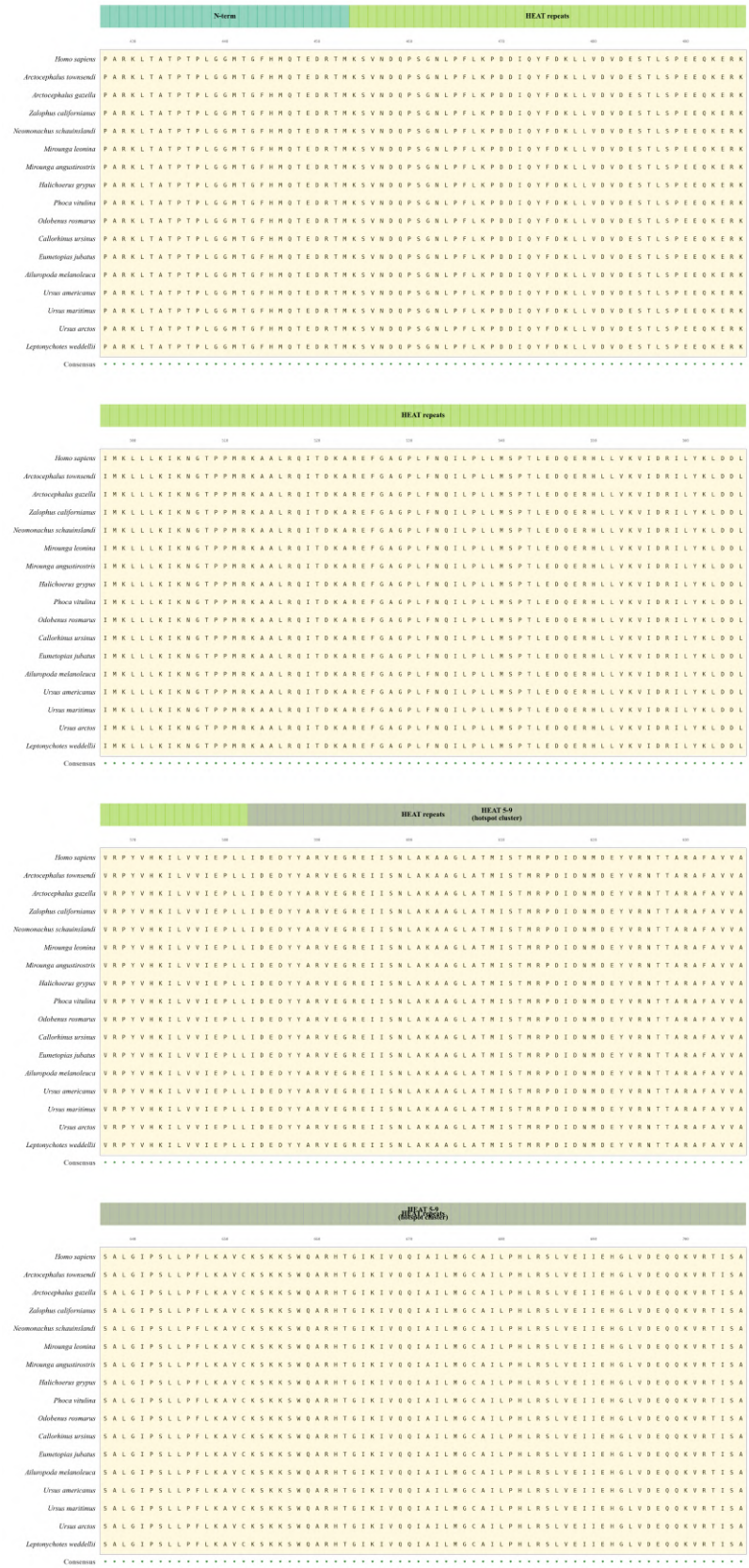
Species	***TP53***	***SF3B1***
Acinonyx jubatus	106972437	106799985
Ailuropoda melanoleuca	100468516	100477819
Callorhinus ursinus	112817082	112840713
Canis lupus dingo	112647438	112656476
Canis lupus familiaris	403869	488456
Enhydra lutris kenyoni	N/A	111145177
Eumetopias jubatus	114225325	114222822
Felis catus	493847	101080016
Halichoerus grypus	118526306	118545127
Hyaena hyaena	120231386	120222073
Leopardus geoffroyi	123604339	123599270
Leptonychotes weddellii	102726420	102734520
Lontra canadensis	116856570	116869170
Lutra lutra	125087037	125094969
Lynx canadensis	115500807	115521829
Lynx rufus	124519563	124516532
Meles meles	123929470	123950561
Mirounga angustirostris	123854560	123838986
Mirounga leonina	118000034	118012183
Mustela erminea	116578338	116597006
Mustela lutreola	131815809	131827622
Mustela nigripes	132003790	132013418
Mustela putorius furo	10167794	101682461
Neofelis nebulosa	131496929	131500801
Neomonachus schauinslandi	110584444	110592295
Neovison vison	122906719	122902150
Nyctereutes procyonoides	129497070	129522726
Odobenus rosmarus divergens	101362427	101376277
Panthera leo	122206426	122226441
Panthera onca	132688368	132671688
Panthera tigris	102962571	102960594
Panthera uncia	125927023	125911335
Phoca vitulina	116649291	116633155
Prionailurus bengalensis	122484900	122481607
Prionailurus viverrinus	125151405	125172120
Puma concolor	112856999	112848744
Puma yagouaroundi	121015982	121030870
Suricata suricatta	115281719	115287577
Ursus americanus	123792132	123798164
Ursus arctos	113271152	113250629
Ursus maritimus	103660368	103659709
Vulpes lagopus	121499929	121480539

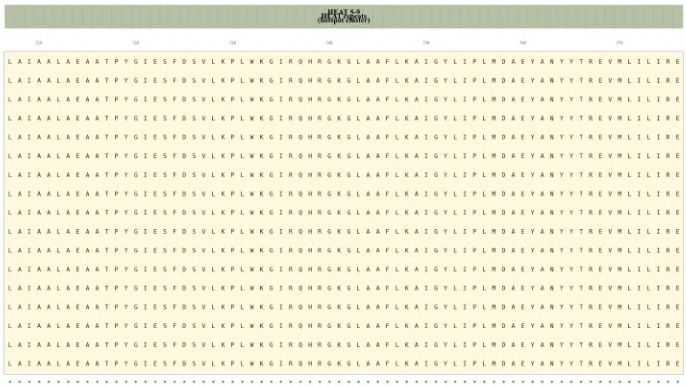
A.2 Sequence alignments

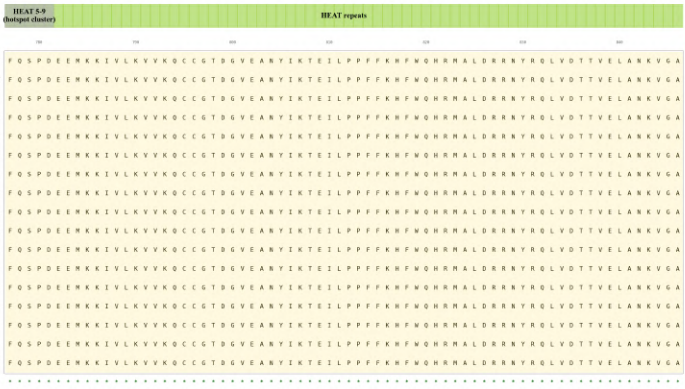


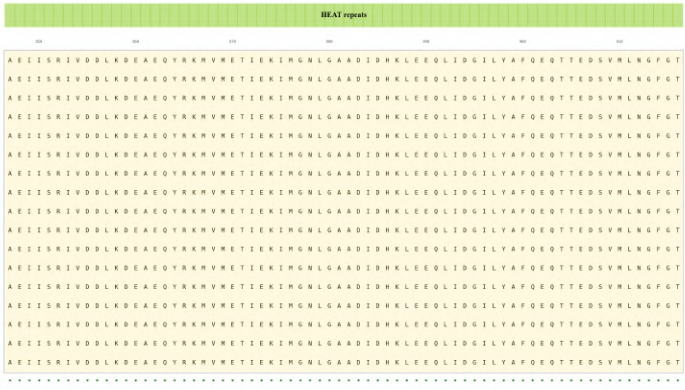
Multiple sequence alignment identity plot for *TP53*.

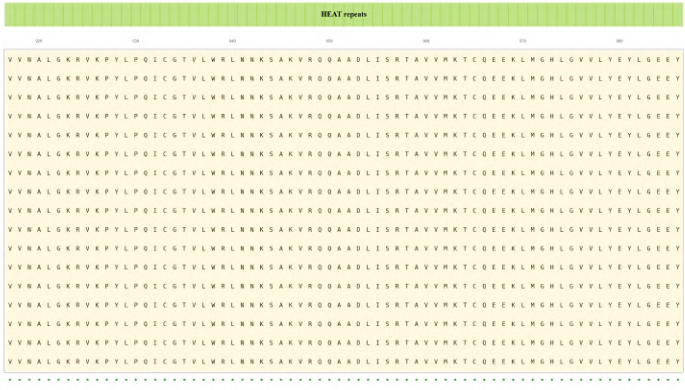

SF3B1 — Identity
N-term
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus

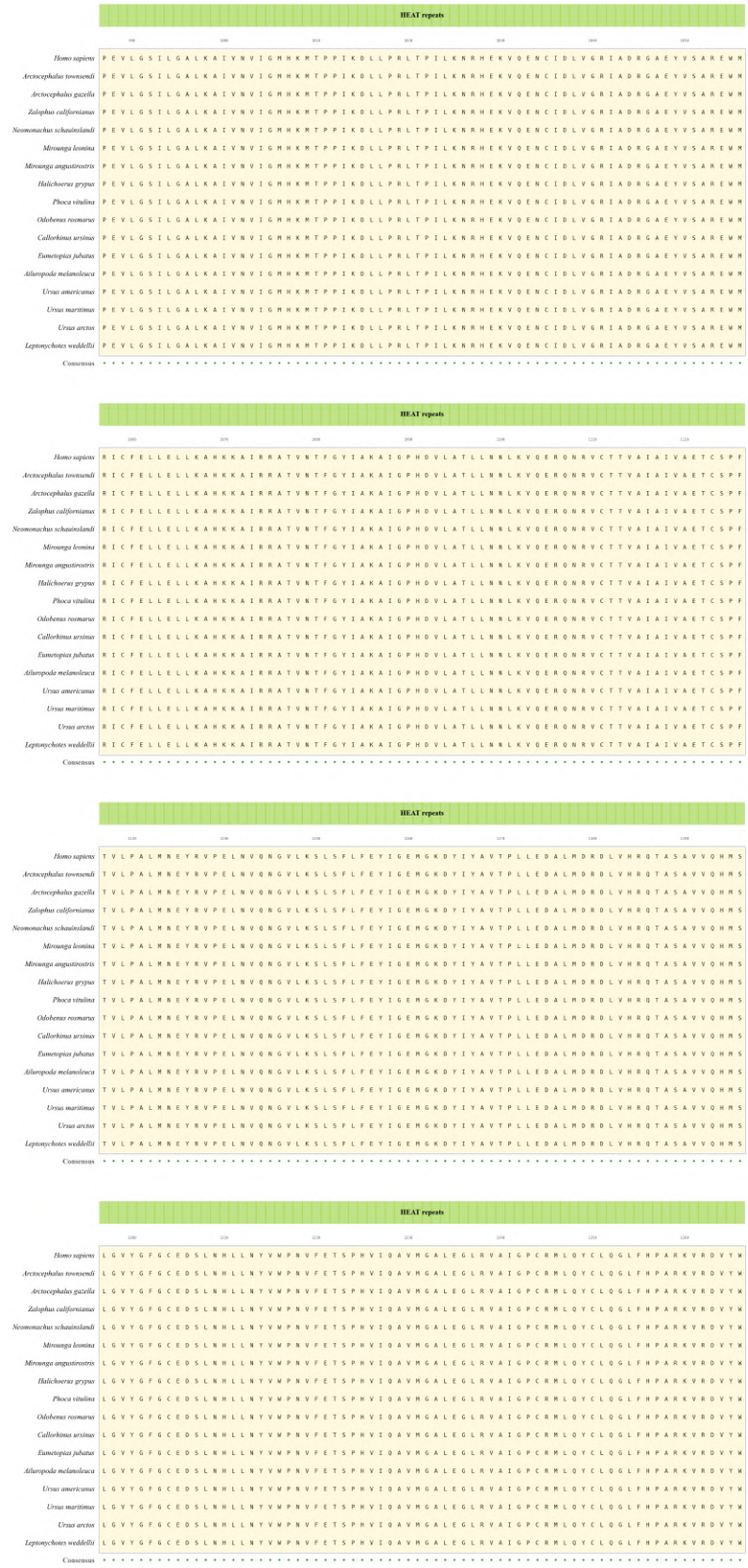

N-term
HEAT repeats
HEAT repeats
HEAT repeats
HEAT 5-9
(hotspot cluster)
HEAT 5-9
(hotspot cluster)
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus

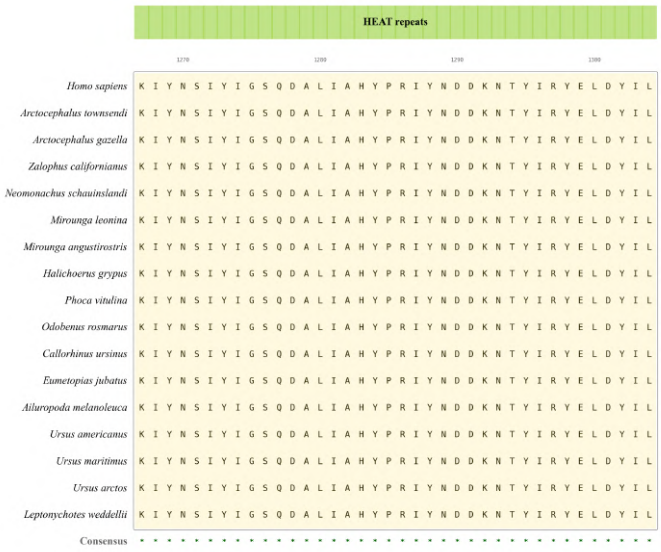

HEAT 5-9
(hotspot cluster)
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus


HEAT 5-9
(hotspot cluster)
HEAT repeats
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus


HEAT repeats
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus


HEAT repeats
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus


HEAT repeats
Homo sapiens
Arctocephalus townsendi
Arctocephalus gazella
Zalophus californianus
Neomonachus schauinslandi
Mirounga leonina
Mirounga angustirostris
Halichoerus grypus
Phoca vitulina
Odobenus rosmarus
Callorhinus ursinus
Eumetopias jubatus
Ailuropoda melanoleuca
Ursus americanus
Ursus maritimus
Ursus arctos
Leptonychotes weddellii
Consensus



Multiple sequence alignment identity plot for *SF3B1*.