UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Genomic characterisation of the thermoadaptation strategies in the penguin genus *Eudyptes*.

by

Debora Jackson

Student number: 1438116

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr Jean Mollett

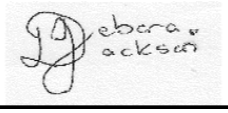
Co-supervisor: Prof. Pieter De Maayer

11/2022

DECLARATION

I, Debora Jackson (Student number: 1438116), am a student registered for Master of Science (Dissertation) in the academic year 2020 to 2022. I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise.
- I have not submitted this work before for any other degree or examination at any other university.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature:  **Date:** 02 November 2022

PREFACE

Penguins are scattered throughout the southern hemisphere and occupy a diverse range of habitats, ranging from polar, temperate to tropical environments. Concomitantly, penguins have had to develop an array of thermoadaptation strategies to survive and even thrive in these environments. While a lot is known about the behavioural and physiological thermoadaptation strategies of penguins living in cold environments, the molecular basis underlying these unique adaptations remains largely unknown, especially when compared to terrestrial and other aquatic birds. This study aimed to investigate the molecular determinants underlying thermoregulatory adaptation and their evolution among penguin species in the genus *Eudyptes* from different thermal environments.

Chapter 1 presents an introduction to the study along with a review of pertinent literature pertaining to penguin biology, ecology and thermoadaptation. In Chapter 2 the thermoadaptation strategies of two cold-adapted penguins, the macaroni penguin (*Eudyptes chrysolophus*) and the royal penguin (*Eudyptes schlegeli*), were analysed at the whole genome level, in comparison to six taxa from this genus which inhabit temperate environments. A core genome phylogeny along with a house-keeping phylogeny have revealed that two distinct clades of *Eudyptes* penguins emerge based on temperate and cold climates. A pan-genome analysis of the *Eudyptes* penguins was performed and a number of proteins unique to the cold adapted penguins were identified and functionally characterised. Similarly paralogous proteins overrepresented in the cold-adapted penguins were analysed. Thirty-five unique and thirteen paralogous proteins over-represented in the cold climate clade penguins were found to play a direct or indirect role in thermoregulation and adaptation within the cold-adapted *Eudyptes* penguins. Most were found to play roles in thermoregulatory mechanisms such as glucose metabolism, adipocyte differentiation, vascular development, feather development, skeletal muscle function and lipid metabolism.

Chapter 3 investigated thermoadaptation at the nucleotide and amino acid level for select proteins present both in temperate and cold-adapted penguins to look at microevolutionary signatures underlying cold adaptation. Fourteen single-copy orthologues that were found to function in thermoregulation showed an increased compositional bias towards the amino acids with small/neutral side chains (serine). An increase in a hydrophilic amino acid i.e. aspartic acid, was also observed as well as a decrease in the amino acids histidine and asparagine. Furthermore, a number of genes in the cold adapted penguins are uniquely under

purifying, neutral or positive selection, which suggests potential evolutionary drivers for life in the cold.

A summary is presented at the end of this dissertation that provides an overall description of the study and discusses limitations as well as future recommendations.

RESEARCH OUTPUTS

Conference outputs

Jackson D, Mollett J and De Maayer P. 2021. Genomic characterisation of the thermoadaptation strategies in the penguin genus *Eudyptes*. South African Society for Bioinformatics Student Council (SASBi-SC) and South African Genetics Society (SAGS) Virtual Student Symposium. 14-16 September 2021. *Oral presentation*.

Jackson D, Mollett J and De Maayer P. 2021. Genomic characterisation of the thermoadaptation strategies in the penguin genus *Eudyptes*. University of the Witwatersrand Molecular Biosciences Research Trust (MBRT) Virtual Research Day. 9 December 2021. *Oral presentation.*

DEDICATION

This dissertation is dedicated to the memory of my late grandparents, Bryan and Natalie Jackson.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following individuals and institution for helping me through this journey:

My supervisor Dr Jean Mollett for all your guidance, support and understanding.

My co-supervisor Prof. Pieter De Maayer for helping me get over my fear of bioinformatics and being patient with me. Also thank you for all the laughs.

My friend Alison Brigden for walking side by side with me through this experience and for being my stress reliever. Late night presentation preparation is always fun when you don't have to do it alone.

Most importantly my parents, brother and nonna. Thank you for all your unfaltering love, support, and constant encouragement. Thank you for always being there. I wouldn't be the person I am today, nor be where I am today, if it weren't for you.

Lastly, the National Research Foundation for financial support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2-AG	2-arachidonyl glycerol
aa	Amino acid
ACC	Antarctic circumpolar current
ACVRL1	Serine threonine-protein kinase receptor R3
AdipoR2	Adiponectin receptor protein 2
AK1	Adenylate kinase 1
AKR1B10	Aldo-keto reductase family 1 member 10
Akt	Protein kinase B
ALK1/5	Activin A receptor like type 1/5
ALOXE3	Epidermis-type lipoxygenase 3
AMPK	AMP-activated protein kinase
AMY2A	Pancreatic alpha-amylase
AQP7	Aquaporin 7
ASK1	Mitogen-activated protein kinase 5
ASMT	Acetylserotonin O-methyltransferase
ATP	Adenosine 5'-triphosphate
AvUCP	Avian uncoupling protein
B10K	The Bird 10,000 Genomes Consortium/Project
BaCoCa	Base composition calculator
BAT	Brown adipose tissue
BHE40	Basic helix-loop-helix family member E40
BlastKOALA	Basic Local Alignment Search Tool KEGG Orthology And Links Annotation
BlastP	Basic Local Alignment Search Tool: Protein
BUSCO	Benchmarking of Universal Single Copy Orthologues
cAMP	Cyclic adenosine monophosphate
CAR	Constitutive androstane receptor

CISD3	CDGSH iron sulphur domain 3
cm	Centimetres
CMKLR1	Chemokine-like receptor 1
COGs	Conserved Orthologous Groups
CO1	Cytochrome c oxidase subunit 1
COQ8B	Coenzyme Q protein 8B
COQ10B	Coenzyme Q protein 10B
COX2	Cytochrome c oxidase subunit 2/cyclooxygenase 2
CTSK	Cathepsin k
Cyt b	Cytochrome b
DGAT2	Diacylglycerol O-acyltransferase 2
DHODH	Dihydroorotate dehydrogenase
DLL4	Delta-like protein 4
DMH	Dorsomedial hypothalamus
DNA	Deoxyribose nucleic acids
DSG1	Desmoglein
ECH	Enoyl-CoA hydratase
EggNOG	Evolutionary genealogy of genes: Non-supervised Orthologous Groups
eLOX3	Epidermis-type lipoxygenase 3
ERK	Extracellular signal-regulated kinase
EvpL	Envoplakin
FADS1/*FADS1*	Fatty acid desaturase 1
FAM13A	Family with sequence similarity13 member A
FFAR3	Free fatty acid receptor 3
FASN	Fatty acid synthase
GABA	Gamma-aminobutyric acid
GABRA6	Gamma-aminobutyric acid receptor subunit alpha-6
Gb	Giga base pairs
GADD45b	Growth arrest and DNA damage inducible protein 45 beta

GLP-1	Glucagon-like peptide-1
GLUT4	Glucose transporter type 4
GO	Gene Ontology
GPCR	G protein-coupled receptor
GYS1	Glycogen synthase 1
HDAC5	Histone deacetylase 5
HVRI	Hypervariable region 1
IL	Interleukin
IL1	Interleukin 1
IL6	Interleukin 6
IL15	Interleukin 15
IL6R	Interleukin 6 receptor
IL6RA	Interleukin 6 receptor subunit alpha
IL6ST	Interleukin 6 cytokine family signal transducer
IRAK2	Interleukin-1 receptor-associated kinase 2
IUCN	International Union for Conservation of Nature
JAK	Janus kinase
JAK3	Janus kinase 3
KCNQ4/5	Potassium voltage-gated channel subfamily KQT member 4/5
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg	Kilogram
KIF5B	Kinesin-1 heavy chain
km	Kilometres
KRT13/15/75	Keratin 13/15/75
LC-PUFAs	Long chain polyunsaturated fatty acids
L-DOPA	L-dihydroxyphenylalanine
m	Meters
MAPK	Mitogen-activated protein kinase
MBOAT4	Ghrelin O-acyltransferase
Mbp	Mega base pairs

MDM2	Mouse double minute 2
MEC	Mechanistic empirical combination
ML	Maximum likelihood
mm	Millimetres
MMCT	Middle Miocene climate transition
MPK38	Murine protein serine-threonine kinase 38
mRNA	Messenger RNA
MtDNA	Mitochondrial DNA
mTORC1	Mammalian target of rapamycin complex 1
MYA	Million years ago
MYH1	Myosin heavy chain
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National centre for Biotechnology Information
ND6	NADH-ubiquinone oxidoreductase chain 6
NDUFA11	NADH dehydrogenase 1 alpha subcomplex subunit 11
NEB	Nebulin
NKX1-1	NK1 transcription factor related protein 1
NKX6-1	NK6 homeobox 1
nr	Non-redundant
NST	Non-shivering thermogenesis
ODC	Ornithine decarboxylase
OG	Orthogroup
PETF/*petf*	Photosynthetic ferredoxin
PGC-1α/1a	Peroxisome proliferator-activated receptor gamma coactivator 1 alpha
PI3K	Phosphatidylinositol 3-kinase
PIK3CG	Phosphatidylinositol 4,5-biphosphate 3-kinase catalytic subunit gamma isoform

POA	Preoptic area
PO/AH	Preoptic anterior hypothalamus
PPARγ	Peroxisome proliferator-activated receptor gamma
PPARγ2	Peroxisome proliferator-activated receptor gamma 2
PPARGC1B	Peroxisome proliferator-activated receptor gamma coactivator 1-beta
PRL	Prolactin
Q	Coenzyme Q
RAG-1	Recombination activating protein 1
RARRES2	Retinoic acid receptor responder protein 2
RCFV	Compositional heterogeneity and bias value
RNA	Ribonucleic acid
SAM	S-adenosyl-L-methionine
SCFAs	Short chain fatty acids
SCOs	Single-copy orthologues
SIC	Sea ice concentration
SLC17A6	Solute carrier family 17 member 6
SLC30A3	Zinc transporter 3
SLC38A2	Sodium-coupled neutral amino acid transporter 2
SMAD2/3/4	Suppressor of mothers against decapentaplegic 2/3/4
SOCS7	Suppressor of cytokine signalling 7
S1P	Sphinogosino-1-phosphate
SPHK1/2	Sphingosine kinase 1/2
sp.	Species
spp.	Several species
ST	Shivering thermogenesis
STAT3/5	Signal transducer and activator of transcription 3/5
TBK1	TANK-binding kinase 1-binding protein
TBX1	T-box transcription factor 1

TET2	Ten-eleven translocation 2 methyl cytosine dioxygenase
TGF-β	Transforming growth factor beta
TH	Tyrosine hydroxylase
TUBA1C	Tubulin alpha-1C chain
TUBA3E	Tubulin alpha-3 chain
UCP 1-3	Uncoupling protein 1-3
VDR	Vitamin D receptor
VGLUT2	Vesicular glutamate transporter 2
WAT	White adipose tissue
WEGO	Web Gene Ontology Annotation Plot

Chapter 1
Introduction

1.1. Penguins in general

Penguins are a group of flightless aquatic birds that fall into the avian order Sphenisciformes and comprise the family Spheniscidae (Li *et al.*, 2014; Pan *et al.*, 2019). Almost all extant species of penguin live exclusively in the southern hemisphere, where they enjoy a cosmopolitan distribution, with the exception of the Galápagos penguin (*Spheniscus mendiculus*), which is found north of the equator (World Wide Fund for Nature, 2020). There are currently twenty recognised extant penguin species (Pan *et al.*, 2019) and more than 50 fossil species have been documented to date that extend back 45-60 million years ago (MYA) (Baker *et al.*, 2006; Pan *et al.*, 2019). The extant penguin species are allocated to six genera (Figure 1.1), namely: *Aptenodytes*, *Eudyptes*, *Eudyptula*, *Megadyptes*, *Pygoscelis* and *Spheniscus* (Baker *et al.*, 2006; Pan *et al.*, 2019). *Aptenodytes* comprises the emperor (*A. forsteri*) and king (*A. patagonicus*) penguins that inhabit areas in and around the Antarctic. *Eudyptes* consists of eight species of crested penguins that inhabit areas in and around the sub-Antarctic region while *Eudyptula* comprises of the little blue penguin (*E. minor*) and the Australian fairy penguin (*E. novaehollandiae*) that inhabit areas in and around Australia and New Zealand (Grosser *et al.*, 2016). The yellow-eyed penguin (*Megadyptes antipodes*) inhabits areas in and around New Zealand (Baker *et al.*, 2006). The genus *Pygoscelis* is comprised of the Adélie (*P. adeliae*), chinstrap (*P. antarcticus*) and gentoo (*P. papua*) penguins that inhabit areas in and around Antarctica and the cool temperate waters of the Southern Ocean. Lastly, *Spheniscus* consists of the African (*S. demersus*), Humboldt's (*S. humboldti*), magellanic (*S. magellanicus*) and Galápagos (S. *mendiculus*) penguins that inhabit the western shores of southern Africa (African penguin) and the western shores of South America (the other three spheniscid penguins) (Figure 1.1; Baker *et al.*, 2006). The classification of penguins into these distinct genera is supported by morphological, behavioural, myological (i.e. the musculature), integumentary (i.e. the skin) and breeding data as well as molecular analysis of both nuclear and mitochondrial DNA sequences (Baker *et al.*, 2006; O'Hara, 1989; Jouventin, 1982; McKitrick, 1991; Giannini and Bertelli, 2004). However, there has been extensive debate regarding the species level delineation of penguins, in particular in the genera *Eudyptes* and *Eudyptula* (Cole *et al.*, 2019; Pan *et al.*, 2019). Taxonomic uncertainty limits conservation as inadequate understanding of both the limits between closely related species and the relationships among them has been a growing problem in the taxonomy field and as a result, it directly impacts the creation and

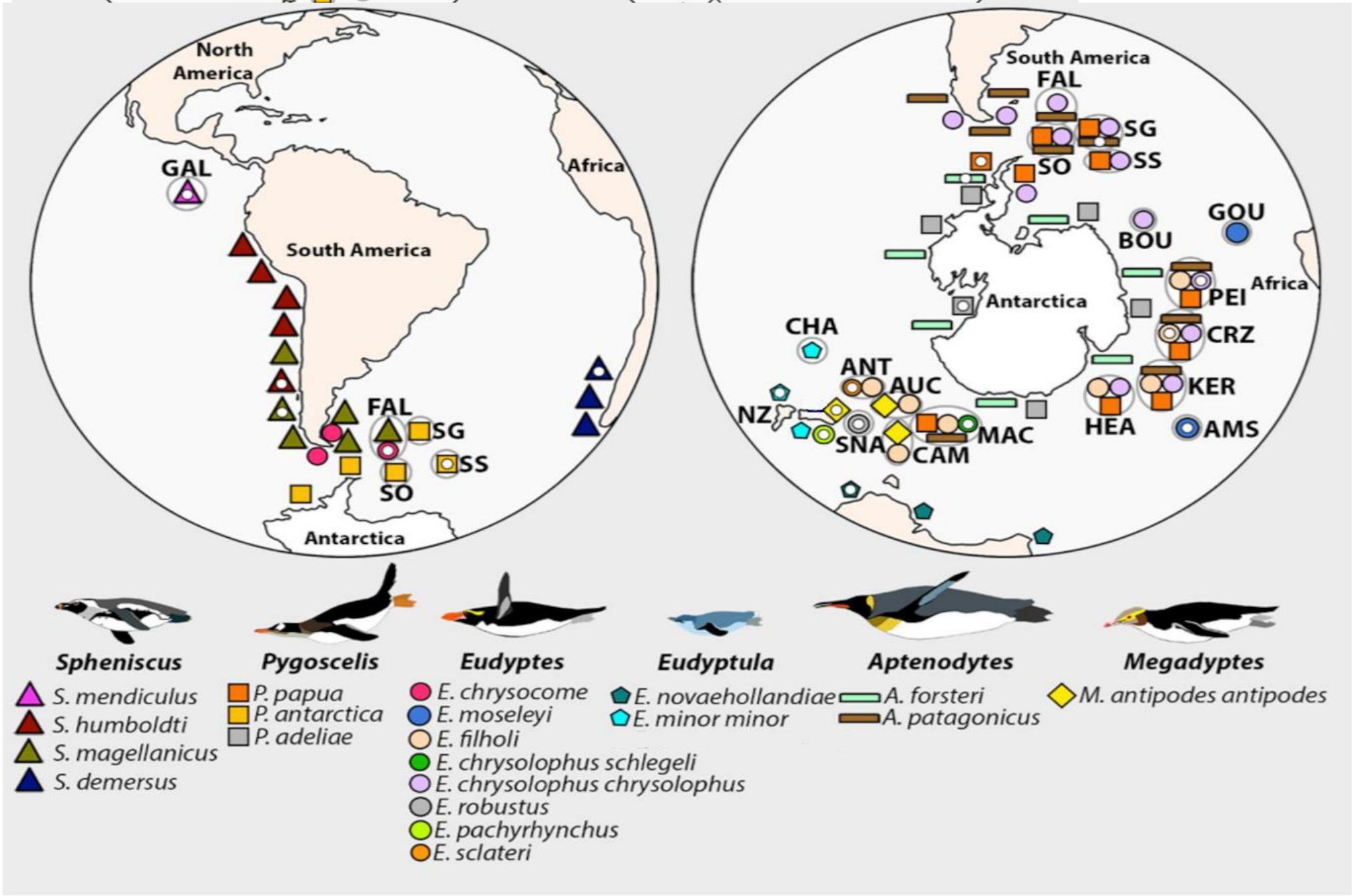


Figure 1.1: World map showing location of every extant penguin species along with their genera. Key symbols with a small white ellipse represent locations where penguins have been sampled. Note: The Humboldt's penguin (*Spheniscus humboldti*) sampling location is unclear as this individual was bred in the Copenhagen zoo (Taken from Pan *et al.*, 2019).

development of recommendations for conservation that are well-founded (Bacon and Bailey, 2006).

Penguins inhabit many unique ecological niches including the sea ice around Antarctica (e.g. the emperor penguin), the rocky shorelines of the sub-Antarctic islands primarily inhabited by the eastern rockhopper penguin (*E. filholi*), New Zealand's oceanic temperate forests largely inhabited by the Fiordland penguin (*E. pachyrhynchus*) and the tropical Galápagos Islands inhabited solely by the Galápagos penguin (Ropert-Coudert, 2014). Therefore, penguins basically occupy almost every coastline and island archipelago in the Southern Hemisphere and Southern Ocean (Boersma, 2008). Some species live in hostile environments that experience strong winds and extreme drops in temperature with seasonal fluctuations in the length of daylight across chick-rearing and breeding seasons (Goldsmith and Sladen, 1961; Li *et al.*, 2014; Pan *et al.*, 2019). By contrast, other penguin species occupy environments with tropical/temperate climates that experience little variation in daylight length (Pan *et al.*, 2019). The distinct environmental and climatic parameters to which the different species of penguins are exposed has led to the evolution of unique morphological and physiological adaptations, including distinct thermoadaptation strategies.

1.2. Penguin characteristics

Extant penguin species have a wide range of sizes, with *Eudyptula* spp. being the smallest (weighing approximately 1 kg and standing at an estimated 41 to 45 cm) and the emperor penguin being the largest (typically weighing up to 30 kg and standing at approximately 1.1 m) (Le Maho, 1977; Marchant and Higgins, 1990; Müller-Schwarze, 1984; Pan *et al.*, 2019, Stonehouse, 1967). Analysis of the fossil record has suggested that penguin species used to be larger, with extinct penguin species showing a body mass that may have surpassed 100 kg (Mayr *et al.*, 2017).

Since penguins spend most of their lives in the ocean, they have developed unique characteristics that allow them to be well-adapted for life in water (Thiébot *et al.*, 2012). As a result, they have evolved a substantive set of morphological characteristics that contribute towards their competence in and under the water such as their bodies being elongated, streamlined and tapered at both ends (fusiform) (Stonehouse, 1967). Penguins have lost the ability of aerial flight and, instead, have uniquely modified paddle-like flippers and immense breast muscles that are used for "underwater flight" or, more accurately, wing-propelled diving (Li *et al.*, 2014; Pan *et al.*, 2019). Their paddle-like flippers, which are a specialised

form of wing that have a unique structure and shape, (Stonehouse, 1967; Li *et al.*, 2014) have bones that are broad and flat with the elbow and wrist joint being almost fused which allows for a flat flipper that is adapted for swimming (Yellow-eyed Penguin Trust, 2017). Another evolutionary adaptation for life in water is their plumage of densely packed, scale-like feathers that allow for waterproofing and thermal insulation (Li *et al.*, 2014; Pan *et al.*, 2019). They have specially adapted eye lenses for acute visual sensitivity, adapted for visual sensitivity during deep dives and for efficient under water predation (Li *et al.*, 2014; Pan *et al.*, 2019). Furthermore, they have reduced distal wing musculature, stiff wing joints and dense bones that allow them to overcome the buoyancy force in water (Li *et al.*, 2014; Pan *et al.*, 2019). Although they spend the majority of their lives in water, penguins also spend time on land for breeding and nesting (Garcia and Boersma, 2013). Their upright stance on land is a result of their legs and webbed feet being set far back on the body as well as resting their body weight on the aforementioned appendages and their tail which is stiff, wedge-shaped and short (Stonehouse, 1967; Müller-Schwarze, 1984).

The diversity in penguin diet is reflected in bill morphology (Müller-Schwarze, 1984). Penguins that generally eat fish have a bill that tends to be long and thin (e.g. emperor and king penguins), whereas those that primarily eat krill have bills that are stouter and shorter (e.g. Adélie and chinstrap penguins) (del Hoyo *et al.*, 1992). To aid in swallowing live prey, rear-directed spines line the mouth (Gill, 1994). With regards to plumage, all penguins are counter shaded with largely pale ventral and dark dorsal colouration. Species differ largely in their head and facial markings, and in their overall size (del Hoyo *et al.*, 1992; Marchant and Higgins, 1990; Oliver *et al.*, 2006). Chicks, juveniles and immature penguins are duller than adults in colour and markings are slightly different when compared to adults (del Hoyo *et al.*, 1992). Most species of penguins are to some extent sexually dimorphic (males and females look different). For example, male crested penguins have larger bills and are more robust than the females (Müller-Schwarze, 1984). Moreover, the sexes with pairs of royal penguins are easily separated based on cheek colouration and bill morphology (Müller-Schwarze, 1984).

1.3. Penguins as bioindicators of ecosystem health

The importance of penguins as bioindicators of ecosystem health has long been appreciated. Penguins make excellent indicators as they have long life spans, a permanent ecological niche and are the dominant avifauna in Antarctica as well as some of the surrounding islands (Metcheva *et al.*, 2006). Penguins can also be considered keystone species given their

importance to soil fertility at their breeding colonies and that they are important mesopredators within marine food webs (Pan *et al.,* 2019; Zhu *et al*., 2014).

Given their biphasic biology, foraging at sea and returning to land to breed, penguins contribute significantly to soil fertility in and around their breeding colonies. For example, a study done by Zhu *et al*. (2014) showed that penguin activity greatly boosted soil phosphorus contribution as well as phosphine formation, which plays a key part in the phosphorous cycle in terrestrial ecosystems of maritime Antarctica. As such, soil richness will be detrimentally impacted if penguin populations continue to decline. A significant nutrition pathway from nearby marine environments to terrestrial roosting and breeding areas is provided by seabird guano (Havik, Catenazzi and Holmgren, 2014; Ellis, 2005). Although seabird guano can also alter soil pH, salinity and moisture, it is widely known that seabird guano increases soil nitrogen and phosphorus (Havik, Catenazzi and Holmgren, 2014; Ellis, 2005). For example, the composition of guano for White-capped Noddies (*Anous minutus*) and Wedge-tailed Shearwaters (*Puffinus pacificus*) is roughly 7.3% nitrogen and 1.5% phosphorus (Smith and Johnson, 1995), but for pelagic seabirds like penguins the nitrogen and phosphorus content of guano increases to approximately 15% - 20% and 10%, respectively (Lorrain *et al.*, 2017). As a result, seabird excrement can multiply the amount of nitrogen and phosphorus that is available in terrestrial ecosystems by 100 and 400 times, respectively (Mulder *et al.*, 2011). Additionally, marine-derived nutrient subsidies have an impact on a number of processes, such as primary production (Mulder and Keall, 2001; Rowe *et al.*, 2017), the composition of plant communities (Anderson *et al.*, 2008; Rowe *et al.*, 2017) and the population size of top predators (Rose and Polis, 1998; Rowe *et al.*, 2017). If seabird populations continue to decline there would be severe negative impacts on all trophic levels in the ecosystems where they breed and nest. In addition, penguin feathers are also used as a biomonitoring tool for assessing the concentration of toxic elements and heavy metals such as lead and mercury, which can be used to monitor contamination/pollution of the Antarctic sea zones (Metcheva *et al.*, 2006).

Penguins are mesopredators (i.e. predators that prey on smaller organisms and in turn are preyed upon by larger predators) (Ainley *et al.*, 2015) which are vital components of ecosystems. The removal of a mesopredator will ultimately cause a trophic cascade, however, the specific impacts this would have depends on the ecosystem in which penguins live (Ainley *et al.*, 2015). One potential outcome of a trophic cascade could be a significant increase in numbers of some species (Silliman and Angelini, 2012). For example, the prey of penguins (cephalopods, crustaceans and fish) would experience “predator release” and their populations would more than likely significantly increase, throwing ecosystems out of

balance. Competitors could also benefit as there would be no competition for food. Another possibility would be that the populations of other species would decline, such as the predators of penguins (Silliman and Angelini, 2012). The predators would need either to switch to different species of prey or face significant population declines. Both options would result in ecosystem imbalance.

1.3.1 Seabird life history traits

Seabirds have quite a few life history and ecological strategies that make them well adapted for life in the marine environment. These include low fecundity, long life spans and specialised foraging strategies (e.g. diving for prey underwater) (Richards *et al.,* 2021). Given the significant effort required to deliver food to young from the open ocean, it seems likely that these traits evolved to maximise adult survival (Velarde *et al.,* 2019). Seabirds are exposed to numerous and recurrent anthropogenic threats in both terrestrial and marine ecosystems since they need isolated terrestrial landmasses to breed and open oceans to feed. These threats include those that directly impair fecundity and survival (such as invasive species and bycatch), those that alter or destroy habitats (such as land use and energy production), and global change threats i.e. climate change (Richards *et al.,* 2021). In addition, these threats make seabirds the most threatened group of birds, and their conservation status is quickly deteriorating (Paleczny *et al.,* 2015).

Pelagic specialists with slow reproductive rates and varied diets are susceptible to threats that have a direct impact on survival and fecundity (e.g. invasive species and the exploitation of biological resources) and climate change (Richards *et al.,* 2021). Species with narrow habitat ranges, quick reproductive rates, and varied diets are more likely to be threatened by habitat-modifying processes (such as pollution and natural system modifications) (Richards *et al.,* 2021). Anthropogenic and climate change threats are more likely to affect species with ecologically identical traits, whereas ecologically flexible traits (such as broad habitat ranges and generalist feeding strategies) may be better protected against extinction (Cooke *et al.,* 2019; Richards *et al.,* 2021). Anthropogenic threats may be favouring fast-lived, wide-ranging generalists like gulls and terns at the expense of slow-living, specialized species like albatross and petrels. Due to the damage that climate change is causing to breeding and foraging habitats, species that are at risk by climate change will exhibit habitat specialization and slow reproductive speed traits (Richards *et al.,* 2021). Species that have narrow habitat breadths will be at risk from habitat-altering threats because they have fewer opportunities to change resource use or distribution in response to environmental change (Richards *et al.,* 2021). Pressures that directly impact survival and fecundity will have an effect on species with slow reproduction rates since these threats wipe out breeding sites and raise mortality

(Richards *et al.*, 2021). Identifying trends and factors that increase a species' risk of extinction will make it possible to create more informed and effective conservation plans (Ripple *et al*., 2017).

1.3.2 Penguin population trends

Many penguin populations are declining, and this is thought to be primarily due to anthropogenic activities like competition with fisheries, bycatch, overfishing, overexploitation (egg harvesting, penguin hunting), pollution (oil spills), habitat loss and degradation (BirdLife International, 2020; Pan *et al.,* 2019). Although anthropogenic activities play a large role in penguin population decline, climate change also plays a role (BirdLife International, 2020; Pan *et al.,* 2019). According to the International Union for Conservation of Nature (IUCN) red list of threatened species, 66% of penguin species have a decreasing population trend. The snares, gentoo and little blue penguin populations are stable, while the Adélie and king penguin populations are steadily increasing (IUCN Red List of Threatened Species, 2022). These increases in population trend for both Adélie and king penguins can be attributed to these species expanding their ranges into previously unoccupied areas due to environmental conditions brought about by climate change (Pütz *et al.,* 2021; Younger *et al.,* 2015). With regards to the red list categories, 42.1% of penguin species are of least concern, 10.5% are near threatened, 21.1% are vulnerable and 26.3% are endangered (IUCN Red List of Threatened Species, 2022). This makes penguins the second-most threatened bird group in the world, only second to the albatross. On the Antarctic Peninsula alone, penguin populations have decreased by more than 25% on average over the past twenty years (Oceanites, 2022). Antarctica is inhabited by 6.12 million breeding pairs which are made up of five penguin species i.e. the emperor, Adélie, gentoo, macaroni and chinstrap penguins. Both the Adélie and chinstrap penguin populations on Peterman Island and Deception Island, respectively, have been hit particularly hard where said populations have decreased by more than 50% (Oceanites, 2022). However, gentoo populations have actually increased, which may be due to generalist foraging strategies as well as a flexible trophic niche, plasticity in their breeding phenology and declines of sea ice (Herman *et al.,* 2020). In addition, some penguin colonies are not experiencing drastic loss and, instead, are growing such as the Adélie populations found in East Antarctica and the Ross Sea (Oceanites, 2022). This is due to ocean and wind currents shielding these regions from rising temperatures and in turn aiding the growth of these specific populations (Oceanites, 2022).

1.3.3 Penguins and climate change

Due to their essential role in marine, coastal and terrestrial ecosystems and their sensitivity to both anthropogenic and environmental change, penguins are one of the most studied groups

of organisms with regards to climate change research (Forcada and Trathan, 2009; Li *et al.*, 2014; Pan *et al.*, 2019).

The diversification of penguin taxa occurred during two abrupt cooling events in Earth's recent past (Baker *et al.*, 2006). The first cooling event happened approximately 35-24 MYA when *Eudyptes*, *Eudyptula* and *Spheniscus* diverged from the older Antarctic genera *Aptenodytes* and *Pygoscelis* (Baker *et al.*, 2006). These ancestral lineages then dispersed northward as a result of the newly formed circumpolar current. The second cooling event or the middle Miocene climate transition (MMCT) occurred between 14-12 MYA approximately and was accompanied by the cooling of the Southern Oceans' surface waters, the flow of the circumpolar current around the Antarctic increasing in intensity and an upsurge in the volume of ice in the Antarctic (Baker *et al.*, 2006; Sheveneli *et al.*, 2004). This MMCT event allowed for a second bout of diversification, which has led to the rise of many extant penguin species that we find today. The circumpolar current has allowed modern penguins to be dispersed to lower latitudes which include oceanic islands within the Antarctic Convergence to the tips of southern continents (Baker *et al.*, 2006). As such many unique adaptations have evolved within penguins in order to inhabit their various ecological environments, from the extreme cold Antarctic sea to temperate climates and tropical islands (Pan *et al.*, 2019). However, these adaptations may act as a double-edged sword with regards to climate change. Climate change might result in species being driven away from their preferred habitats and towards their ancestral distribution which may cause multiple species to go extinct (Baker *et al.*, 2006).

The species that inhabit extremely cold environments rely on sea ice for their food, nesting/breeding sites and to escape from predators (Sidder, 2016). Also, their thermoregulation systems (and behaviours) have adapted for cold temperatures. A rise in global temperatures would cause the ice to melt with concomitant catastrophic consequences for the penguins with regards to their environment and their physiology. Changes in the food web would occur through changes in sea ice cover and the Antarctic food web would be affected through the disappearance of key organisms like krill (Jenouvrier *et al.*, 2014; Loeb *et al.*, 1997). Low sea ice concentration (SIC) would cause food scarcity and high predation which would negatively impact adult penguin survival (Barbraud and Weimerskirch, 2001; Jenouvrier *et al.*, 2014). In addition, low SIC would have a negative impact not only on adult penguin survival, but on breeding success as well. This is because longer foraging trips would be needed and thus more energy will be expended by the parental penguins, ultimately leading to the chicks being fed less frequently (Jenouvrier *et al.*, 2014). Breeding sites would also be lost due to disappearing ice cover (Barbraud and Weimerskirch, 2001). By contrast,

penguin species that inhabit temperate/tropical environments can tolerate higher temperatures better than their Antarctic counterparts (BirdLife International, 2020). However, rapid cooling in these tropical environments would have the same effects as a rise in temperature would have in the Antarctic environment i.e. food web changes and loss of breeding sites which would negatively impact adult survival and breeding success. The rise in ocean temperatures could also negatively impact adult survival, as food may become scarce due to the penguin prey species seeking cooler ocean water (BirdLife International, 2020).

In order to protect the penguins from further decline caused by anthropogenic activities and climate change the underlying molecular basis of their thermoadaptation strategies need to be studied. Understanding the molecular basis for thermoregulatory abilities in penguins would help in determining how changes in temperature would affect the molecular elements involved in thermoregulation and thermoadaptation and in turn how this change would influence their thermoregulatory behaviour. This will then lead to helping penguin species survive and adapt to these changes brought about by climate change through the improvement of existing conservational strategies and the creation of new ones.

1.4. Thermoadaptation in penguins

Thermoregulation is the ability of a warm-blooded animal to regulate its internal temperature and keep it within certain boundaries even if the external/environmental temperature is completely different (Osilla and Sharma, 2020). Penguins have evolved multiple thermoregulatory adaptations as they inhabit a wide range of environments, from Antarctica where temperatures range from -60 to 10 °C, with high winds and seasonal fluctuations in the length of daylight (Goldsmith and Sladen, 1961), to the Galápagos Islands where temperatures range from 21 to 30 °C (Stonehouse, 1967). Enhanced thermoregulation has developed as a complicated system within the head, wing and legs of penguins to deal with temperature extremes (Frost *et al.,* 1975; Li *et al.*, 2014; Pan *et al.*, 2019).

1.4.1 Behavioural thermoregulation

Many penguin thermoregulatory adaptations are instinctual/behavioural. For example, for the retention of heat, penguins tuck their flippers in closer to their bodies which reduces heat loss, because this behaviour reduces the surface area exposed to the environment and therefore minimises heat loss (Stonehouse, 1967). Bills also play an important role in thermoregulation. As they are well vascularised and uninsulated, they serve as a 'thermal window' for heat exchange (Ryeland *et al.*, 2017). Bills thus allow for behaviour that leads to minimising heat loss, especially in cold climates. Penguins (and other birds that inhabit cold environments) attempt to insulate their bills in their plumage while roosting; this is known as 'back rest' behaviour and is another way to mitigate heat loss (Ryeland *et al.,* 2017). Small

bills have been selected to allow for the minimization of heat loss in cold environments, as seen in the emperor penguin, where their bills are small in proportion to their body size (McCafferty *et al.*, 2013). Some penguin species like the king and emperor penguins are able to rest their weight on their heels and tail by tipping up their feet and this reduces the amount of time in contact with the cold surface they are resting on, thus minimising heat loss through their feet (Le Maho, 1977; Stonehouse, 1967). The chicks and adults of some penguin species tend to huddle together in order to conserve heat. For example, emperor penguins huddle in groups of up to 6,000 individuals while also incubating their eggs during the Antarctic winter (Gilbert *et al*., 2006; Stonehouse, 1967). In contrast, behavioural/instinctual thermoregulation adaptations for the loss of heat in warmer temperatures include panting and moving to shaded areas (Müller-Schwarze, 1984), and ruffling their feathers which allows warm air to escape and increases loss of heat. They also hold their flippers away from their bodies increasing the surface area exposed to the environment to maximise heat loss through the flippers (Müller-Schwarze, 1984).

1.4.2 Physiological thermoregulatory approaches

In penguins, and all other birds, the thermoregulatory system that enables their body temperature to remain within certain boundaries i.e. relatively constant over a wide range of environmental temperatures, is composed of a sensory part, an integrating part and a command part (Ruuskanen *et al.*, 2019). The sensory part detects environmental changes and is comprised of baro-, osmo-, and thermo-receptors. The integrating part comprises the thermoregulatory centre of the preoptic anterior hypothalamus (PO/AH), in which neurons that are temperature sensitive detect changes in temperature and temperature information that is received from peripheral thermoreceptors (Ruuskanen *et al.*, 2019). The body temperature set point is then defended by heat production mechanisms or heat loss mechanisms that depend on the thermal status of the bird. The command part involves neurological and endocrine signals which lead to downstream body temperature control mechanisms. These include shivering thermogenesis (ST), non-shivering thermogenesis (NST), evaporative heat loss, peripheral vasodilation or vasoconstriction and behavioural changes (Ruuskanen *et al.*, 2019). ST and NST form part of the physiological process termed thermoregulatory thermogenesis (Mozo *et al.*, 2005). ST in birds involves increasing the contractile activity in skeletal muscles to ultimately increase the rate of heat production (most thermogenesis in birds relies on ST). NST, on the other hand, relies on metabolic energy transformations to increase heat production (Ruuskanen *et al.*, 2019).

A second thermogenesis, obligatory thermogenesis, (Mozo *et al.*, 2005) involves basal metabolic processes required for body consistency which includes diet-induced

thermogenesis that is a result of food absorption, digestion and storage of fat (Mozo *et al.*, 2005). Fat storage is vital for the survival of penguins as it allows penguins to withstand cold temperatures (thermoregulation) and survive through long fasting periods (Li *et al.*, 2014).

Penguins (like other warm-blooded birds) have circulatory systems that allow for counter-current blood flow (counter-current heat exchange systems) which enables the conservation or releasing of heat to keep their internal temperature constant through arterio-venous networks in the head, axillae and legs (McCafferty *et al.*, 2013). For heat to be conserved within the body, it is transferred from blood that is flowing to the flippers and legs to the blood that is leaving the flippers and legs. This then allows for heat to be trapped within the body rather than reaching the exposed parts (i.e. without feathers) of the flippers or legs and being lost to the external environment (McCafferty *et al.*, 2013). In the eye, nasal passages and jaw muscles the post-orbital *rete mirabile* (complex of arteries and veins) serves as a heat exchanger (McCafferty *et al.*, 2013).

Penguins have the highest density of feathers per unit area of any bird (Yong, 2015). The darker feathers on the dorsal side of penguins absorb heat which helps with thermoregulation in cold climates, the white feathers absorb little heat and as a result aid with heat loss in warmer climates (Yong, 2015; Müller-Schwarze, 1984). Feathers are good insulation on land, but in cold water a sub-cutaneous lipid layer provides better insulation. Approximately 30% of a penguin's body mass is blubber (Pond and Mattacks, 1985; Lewden *et al.*, 2017). Penguins found in colder climates generally have longer feathers and a thicker sub-cutaneous lipid layer, whereas penguins located in warmer climates tend to have shorter feathers and a thinner sub-cutaneous lipid layer (Pond and Mattacks, 1985; Kooyman *et al.*, 1976: Stonehouse, 1967). Penguin feathers are not structured the same as those found on birds that have the ability of aerial flight. Flighted birds possess large flat feathers arranged in tracts. By contrast penguin feathers are short (30-40 mm) and stiff with an under-layer of fine down that overlap and are evenly packed over the body surface (Li *et al.*, 2014; Stonehouse, 1967). This allows for waterproofing, insulation and a good streamline effect in water (Lewden *et al.*, 2017). On land, penguins can fluff out their feathers when cold to trap air closer to their bodies for better insulation. They do this by using small smooth muscles that are attached to feather shafts. This allows for the distortion of the surface of the skin, enabling the feather shaft to stand upright, which slows the movement of air across the skin (Lewden *et al.*, 2017; Stahel *et al.*, 1987).

1.4.3 Molecular thermoregulatory approaches

The proteins that make up feathers are alpha (α)-keratins (exist in all vertebrates) (Wu *et al.*, 2015) and beta (β)-keratins, which exist solely in birds and reptiles (Wu *et al.*, 2015). Diversification and duplication of β-keratin genes as well as the numbers of both α-keratin

and β-keratin genes are known to play vital roles in the diversification of avian feathers (Li *et al.*, 2014). The barbules and barbs of mature feathers are made up of β-keratins. Gene duplications of β- keratins in the Adélie and emperor penguins can be linked to their unique feather structure and thus their adaptation to the Antarctic climate (Li *et al.*, 2014).

Keratinocyte β-keratins are also expressed in the penguin epidermis (skin) (Li *et al.*, 2014). Penguin epidermis consists of a thick stratum corneum which is made up of flattened heavily keratinised cells and these cells lack basophilic nuclear remnants (Li *et al.*, 2014). The cornified envelope of keratinocytes is made up of an envoplakin protein component which is encoded by the *evpL* gene (Li *et al.*, 2014). The *DSG1* gene is believed to be involved in a dermatological disorder within humans characterised by skin on the palms and soles being thickened (Li *et al.*, 2014). These two genes along with variation of the α- and β-keratin genes are thought to have contributed to the unique feather structure and skin found in penguins (Li *et al.*, 2014). The elaborate epidermis with accompanying epidermal structures (feathers) is used for thermoregulation, among other things, and has thus allowed for penguins to successfully adapt to the various environments that they inhabit (Greenwold *et al.*, 2014).

Avian uncoupling protein (AvUCP) is a protein involved in NST. It is found in the mitochondria of birds and uncouples oxidative phosphorylation and dissipates the transmembrane mitochondrial gradient of protons into heat and thus plays a role in thermoregulation (Mozo *et al.*, 2005). There are quite a few genes that play a role in lipid metabolism and thus thermoregulation (through obligatory thermogenesis). One such gene is *FASN*, which codes for a fatty acid synthase and plays important roles in *de novo* lipogenesis and furthermore thermoregulation (Li *et al.*, 2014).

The molecular mechanisms of penguins' thermoregulatory adaptations need to be investigated in order to predict how temperature changes would impact the molecular components involved in thermoregulation and thermoadaptation, and how this change would affect the thermoregulatory behaviour of penguins. Understanding the molecular mechanisms of thermoregulatory adaptations will lead to improving current conservational strategies and to developing new ones, which will aid penguin species in surviving and adjusting to the changes brought about by climate change.

1.5. The genus *Eudyptes*

This study will focus on the thermoregulatory adaptation of the crested penguin species within the genus *Eudyptes*. This genus includes eight extant species, namely the southern rockhopper (*E. chrysocome*), eastern rockhopper (*E. chrysocome filholi*), macaroni (*E. chrysolophus chrysolophus*), northern rockhopper (*E. moseleyi*), Fiordland (*E.*

pachyrhynchus), snares (*E. robustus*), royal (*E. chrysolophus schlegeli*), and erect-crested (*E. sclateri*) penguins (Figure 1.2). According to the IUCN red list of threatened species, only one species of penguin in the genus is near threatened, the royal penguin, while the other seven are threatened. Among the seven threatened species, the snares, Fiordland, macaroni, southern rockhopper and eastern rockhopper penguins fall under the conservation status of vulnerable, whereas the erect-crested and northern rockhopper penguins are endangered (BirdLife International, 2020; IUCN Red List of Threatened Species, 2022).

The genus *Eudyptes* has been the subject of much taxonomic uncertainty, although most authorities now agree on eight species, the status of several subspecies/species remains a debate (Pan *et al.*, 2019). For decades, the only distinction made within *Eudyptes* was between the rockhopper penguin (the three rockhopper penguins were considered to belong to one species) and the macaroni penguin using only morphological data (del Hoyo *et al.*, 1992; Frugone *et al.*, 2018). Then a study by Jouventin *et al.* (2006) used genetic and behavioural data that lead to the rockhopper penguin being split into northern and southern rockhopper penguins This was followed by a study done by Ksepka *et al.* (2006) that determined the macaroni complex being split into the macaroni penguin and the royal penguin based on combined molecular and morphological data. Bertelli and Giannini (2005) also concluded that the macaroni and royal penguins should be seen as separate species, but this was on morphological and mitochondrial sequence data. Most recently, the eastern rockhopper penguin has been separated and this was determined by multiple studies. In a study by Frugone *et al.* (2018), two mitochondrial DNA (mtDNA: *HVRI, COI*) and two nuclear markers (*ODC, AK1*) were analysed in five *Eudyptes* penguins (the eastern rockhopper, northern rockhopper, southern rockhopper, macaroni and royal penguins) from thirteen different locations. It was determined that the rockhopper penguins comprise three separate species, the eastern rockhopper, northern rockhopper and southern rockhopper. However, due to the lack of genetic divergence observed between the macaroni and royal penguins, they are seen as subspecies according to Frugone *et al.* (2018). Cole *et al.* (2019) analysed 41 near completed mitochondrial genomes (mitogenomes) that represented all extant and recently extinct penguin taxa. They arrived at the same conclusions regarding the rockhopper penguins as well as the macaroni and royal penguins as Frugone *et al.* (2018). Most recently, a study by Pan *et al.* (2019) analysed twenty-one extant penguin genomes and they concluded that there are eight extant species that comprise the *Eudyptes* genus and this is the taxonomy that was followed for this study.

Penguins belonging to the genus *Eudyptes* have similar appearances with sharply delineated black and white plumage, prominent yellow crest ornamentation, red bills as well as red eyes

Figure 1.2: Penguin species within the genus *Eudyptes*. (A) *E. chrysocome* [western rockhopper penguin; Glenda Rees], (B) *E. chrysocome filholi* [eastern rockhopper penguin; Kyle Morrison], (C) *E. moseleyi* [northern rockhopper penguin; Wikipedia.org/wiki/Crested_penguin], (D) *E. chrysolophus chrysolophus* [macaroni penguin; wikipedia.org/wiki/Crested_penguin], (E) *E. pachyrhynchus* [fiordland penguin; wikipedia.org/wiki/Crested_penguin] (F) *E. robustus* [snares penguin; wikipedia.org/wiki/Crested_penguin], (G) *E. chrysolophus schlegeli* [royal penguin; wikipedia.org/wiki/Crested_penguin], (H) *E. sclateri* [erect-crested penguin; Tu De Roy].

(Figure 1.2; Williams, 1995). Some differences between species occur in the colouration of the face with royal penguins and a small percentage of macaroni penguins having a mostly white face, whereas the other crested penguins have black faces (Williams, 1995; Müller-Schwarze, 1984). Other key differences among species include the length of the ornamental head plumes, the size and colour of the fleshy margin of the gape, the colour pattern on the ventral side of the flipper and the variation of sizes of the superciliary stripe that is found in front of the eye (Banks *et al.*, 2006). Species further differ in their distributions (Figure 1.3; Frugone *et al.*, 2018), and in their behaviour including differing levels of natal philopatry and foraging site fidelity (Frugone *et al.*, 2018).

During the breeding season, *Eudyptes* penguins are constrained by the energy demands of their chicks, however, most crested penguins perform extensive nonbreeding migrations (for example, macaroni penguins have been found to have travelled approximately 3,425 km from their colonies in order to look for food) and vagrants are frequently reported well-outside of their normal distributions (Frugone *et al.*, 2018; Stonehouse, 1967). Breeding behaviour in eudyptids is more complex when compared to penguin species in other genera (Williams, 1995). Eudyptids display varying degrees of obligate brood reduction which is facilitated by laying a small first egg (extreme egg-size dimorphism), reversed hatching asynchrony, large hatching intervals and marked sibling-size asymmetry (Seddon, 1999). All species in this genus lay two eggs with the first egg being substantially smaller (15-45%) than the second egg, the most likely reasons being that the smaller egg is often lost due to intermale territorial fights during the early nesting stage and the smaller egg also represents the minimum investment that may result in viable offspring (Johnson *et al.*, 1987). The most extreme egg-size dimorphism is seen in the macaroni penguin, where the first egg averages only 60% the size of the second and only one chick is raised successfully per breeding season (Williams, 1995; Müller-Schwarze, 1984). If the first egg is kept (most likely by the Fiordland, southern rockhopper and snares penguins), it normally hatches after the second egg (Seddon, 1999). When both eggs hatch there is no overt sibling aggression, but the smaller chick always starves to death (Seddon, 1999). The purpose and adaptive significance for obligate brood reduction in *Eudyptes* has still eluded authorities to this day, however, several hypotheses have been proposed. Throughout the breeding cycle, both partners need to undergo fasting for extended periods of time (Williams, 1995; Müller-Schwarze, 1984). Males, for instance, frequently fast for more than 40 days from arriving to the colony to their first foraging expedition after egg laying (Williams, 1995; Müller-Schwarze, 1984). Both partners also take turns incubating the eggs. However, only the male broods the chick after it hatches for

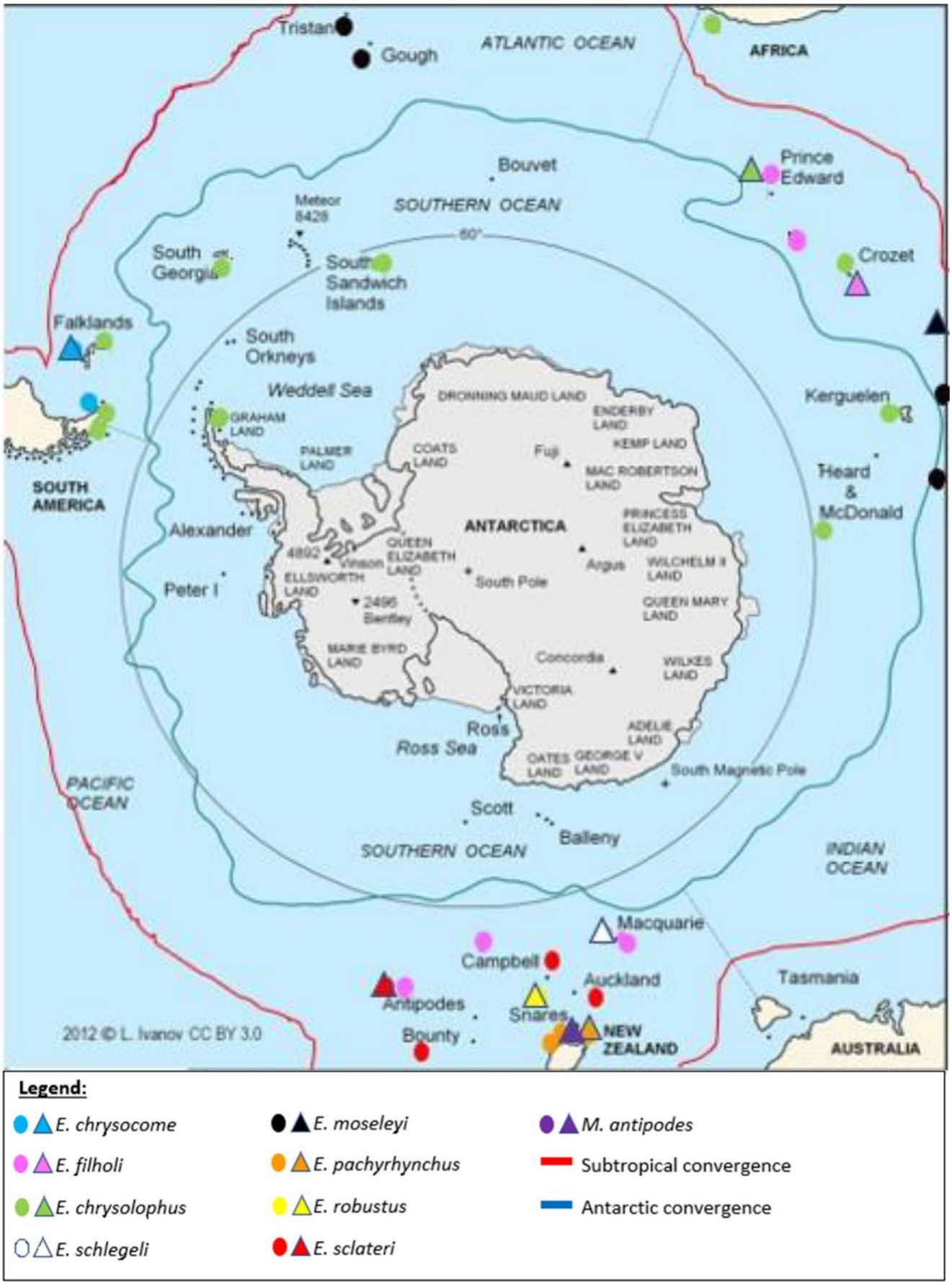


Figure 1.3: Location of each *Eudyptes* penguin species and the outgroup (*M. antipodes)* with regards to the Subtropical and Antarctic Convergence. Sampled individuals taken by Pan *et al.* (2019) for the final genomes are depicted by triangles, ovals represent other locations where breeding colonies of each species are found. Note that the sampled individual represented by a black triangle is found on Amsterdam Island which is off map (Based on Pan *et al.*, 2019).

approximately three weeks while the female feeds the chick (Williams, 1995; Müller-Schwarze, 1984).

As the yellow-eyed penguin is used as the outgroup in this study and this species is quite similar to the species that comprise *Eudyptes*, it is worthwhile to mention some differences. The yellow-eyed penguin was first thought to be closely related to the little penguin, however, through the use of mitochondrial and nuclear DNA it was discovered that the yellow-eyed penguin is more closely related to penguins of the genus *Eudyptes* (Pan *et al.*, 2019). The difference lies in the appearance of the yellow-eyed penguin where it is most easily identified by a band of bright yellow feathers that surrounds its eyes and encircles the back of the head (BirdLife International, 2022). The sides of the face, forehead and crown are pale grey and flecked with yellow. Its eyes are pale yellow and the foreneck is light brown. The bill and feet are light pink in colour. In addition, the bill also displays some orange colouration on the tip. Juveniles do not have a pale yellow band around their eyes and their heads are primarily grey (BirdLife International, 2022). While *Eudyptes* penguins are widely spread throughout the Southern Hemisphere, the yellow- eyed penguin is endemic to New Zealand and is the largest extant penguin that breeds on the mainland with a height that ranges between 62-79 cm (BirdLife International, 2022; IUCN Red List of Threatened Species, 2022). Breeding behaviour is also slightly different when compared to eudyptids. Yellow-eyed penguins are not colonial nesters. When coming ashore in groups of four to six or more individuals, they move along tracks to individual nest sites up to one kilometre inland. They lay two similar-sized eggs in September and these eggs hatch within a few hours of each other (synchronous hatching) and frequently fledge twins (BirdLife International, 2022; Seddon, 1999). The chicks do not compete with each other and parents do not favour one chick over the other, as a result, the growth rates are similar between the two (Seddon, 1999).

Compared to the other genera that comprise the family Spheniscidae, *Eudyptes* includes the most species of penguins, and these species live in a variety of habitats that experience different thermal conditions that are perfect for researching the molecular processes that underlie cold-adaptation.

1.6. Aim and objectives of research

This study aimed to investigate the molecular determinants underlying thermoregulatory adaptation and their evolution among *Eudyptes* penguin species from different thermal environments. This was achieved by 1) the functional annotation of the *Eudyptes* penguin genomes and identification of orthologous proteins/genes among the genomes, 2) the re-evaluation of the taxonomic status of *Eudyptes* species using whole-genome phylogenetic

approaches, 3) the identification of species- and group-specific genes that may be involved in thermoregulatory adaptation and, 4) phylogenetic, gene duplication, selection (positive, neutral and negative selection) and amino acid substitution analysis of thermoadaptation genes and proteins.

Genomic sequences from eight species of *Eudyptes* penguin were used. To re-evaluate the taxonomic status of the eight eudyptids (objective 2 and objective 4), a phylogenetic assessment was constructed using two approaches i.e. a core genome phylogeny and a house-keeping gene phylogeny. Orthologous and paralogous proteins were identified across the genomic sequences and classified according to their function and gene ontology (objective 1, objective 3 and objective 4). Proteins specifically related to cold-tolerant species or that were unique to a single species were identified, annotated, and classified by functionality using OrthoFinder and EggNOG-mapper, respectively (objective 3). Proteins identified as having a putative role in thermoregulation were then analysed at the amino acid level using BaCoCa (objective 4). Proteins that were found to have anomalous amino acid compositions were further subjected to selection analysis using the Selection Server (objective 4).

Chapter 2

Thermoregulation in *Eudyptes* penguins at the genome and pan-genome level

2.1. Introduction

The class Aves is the most species-rich of the tetrapod vertebrate classes, with over 10,500 extant bird species which are dispersed all over the world and found in diverse niches (Zhang *et al.*, 2014). Birds furthermore serve as model organisms when answering questions pertaining to animal conservation, developmental biology, neuroscience and population genetics (Zhang *et al.*, 2014). Despite their academic relevance, the genomes of only three bird taxa, the domestic chicken (*Gallus gallus*), domestic turkey (*Meleagris gallopavo*) and zebra finch (*Taeniopygia guttata*), had been sequenced by 2010 (Zhang *et al.*, 2014). However, thanks to the rapid advancement of sequencing technology, many large-scale bird genome sequencing initiatives have allowed for a better understanding of the genetic complexity of many bird taxa.

The first bird genome to be sequenced in 2004 was that of the domestic chicken. Its genome had a total length of 1.05 Gb (International Chicken Genome Sequencing Consortium, 2004). The next avian genomes, that of the zebra finch and the domestic turkey, were sequenced in 2010 (Dalloul *et al.*, 2010; Warren *et al.*, 2010). From 2013 there was a substantial increase in the number of genomes that were sequenced (National Centre for Biotechnology Information, 2022). According to the National Center for Biotechnology Information (NCBI), as of May 2022, 1,218 avian genomes have been sequenced. Birds have the smallest genomes with respect to other vertebrates. Their genomes range from 0.91 Gb (black-chinned hummingbird, *Archilochus alexandri*) to 1.3 Gb (common ostrich, *Struthio camelus*; Zhang *et al.*, 2014). Bird genomes are also less complex when compared to other vertebrate genomes (Zhang *et al.*, 2014).

Given the rapid advancement in genome sequencing technology there has been an exponential increase in the creation of large-scale bird genome sequencing initiatives. The Bird 10,000 Genomes (B10K) Project aims to produce representative genome sequences of all extant bird species (Feng *et al*., 2020; Jarvis *et al*., 2014; Zhang *et al*., 2014). This project was based on the success of a previous ordinal level project, the Avian Phylogenomics Project. The Avian Phylogenomics Project was the first proof of concept for performing large-scale sequencing of numerous representative species across a vertebrate class. The genomes of a total of 48 birds were sequenced which represented all 32 neognath (a clade of

birds that include majority of all extant species) and two of five paleognath (a clade of birds that include the tinamous and flightless ratites) orders (Feng *et al*., 2020; Jarvis *et al*., 2014; Zhang *et al*., 2014). The resultant genome sequences were used to investigate the bird family tree by performing large-scale phylogenomic analyses and applied various comparative genomic methods to solve avian-related questions such as avian genome evolution, loss of teeth, the molecular basis of flight, sex chromosome evolution and vocal learning. This project also led to the development of several new bioinformatic methodologies and software tools (Jarvis *et al*., 2014; Zhang *et al*., 2014).

The B10K Project's first milestone (phase 1) allowed much of the contested phylogenetic history of modern avian orders to be uncovered as well as address many questions related to avian evolution, diversity and behaviour (Feng *et al*., 2020; Jarvis *et al*., 2014; Zhang *et al*., 2014). The project's second milestone (phase 2) was achieved in 2020, with the release of 363 genomes that represent 92.4% of all avian families (Feng *et al*., 2020), with 267 being newly released genome assemblies. Phase 2 established a new pipeline that used a reference-free whole-genome alignment to identify an unprecedented number of orthologous sequences which allowed genomic novelties to be recognised within certain avian lineages (Feng *et al*., 2020). This densely sampled alignment allowed for the detection of signals of natural selection down to the individual bases. This demonstrates how increasing genomic diversity in comparative studies will uncover additional shared and lineage-specific variation as well as improve the investigation of genomic characteristics (Feng *et al*., 2020).

Phases 3 and 4 of the B10K project still need to be completed as sample collection is still underway. The objective is to acquire samples from 2,500 genera and subgenera, representing 2,400 distinct species in total. In phase 4 the B10K Project aims to collect samples for the remaining 8,000 species (Large Phylogenomics Projects, 2022). With the completion of this project the entire extant avian class will have a completed genomic level tree of life. The link between genetic and phenotypic variation will be uncovered. The correlation of biodiversity, biogeographical, evolutionary and genetic patterns across a broad range of species will be exposed. The demographic history of the entire class Aves will be uncovered as well as how the impact of ecological factors and human influence will affect species evolution (Zhang *et al*., 2014).

Penguin science has likewise been revolutionised through the availability of genome sequences. To date the genomes of twenty-one extant penguins have been sequenced. The Antarctic origin of penguins was hypothesised from non-molecular characters (i.e.

morphological, behavioural, myological, integumentary and breeding) as well as molecular data (i.e. DNA-DNA hybridisation, mtDNA and nuclear DNA markers) (Baker *et al.*, 2006). Vianna *et al.* (2020) used twenty-two whole genome sequences belonging to eighteen distinct penguin species and one outgroup. The study showed that crown-group penguins diverged in the early Miocene in Australia and New Zealand, and not in Antarctica as previously thought. It further showed how changing climatic conditions, the opening of the Drake Passage and associated increased intensity of the Antarctic Circumpolar Current (ACC) drove lineage diversification in penguins (Vianna *et al.*, 2020). The ACC not only promoted penguin diversification, but also allowed for geographic expansion. Through this study of the twenty-one penguin genomes, introgression was discovered among extant penguin species such as the crested penguins (*Eudyptes*). Throughout the majority of their evolutionary history, species within *Eudyptes* have been shown to exchange genes (Vianna *et al.*, 2020). This is seen between the eastern and southern rockhopper penguins transferring genes to the erect crested penguin (Vianna *et al.*, 2020). The directionality of introgression has been linked to the clockwise direction of the ACC that connects to the sub-Antarctic islands, furthermore, allowing for admixture (Vianna *et al.*, 2020). As such genes acting on diving capacity, oxygen metabolism and thermoregulation were studied and found to undergo adaptive evolution as penguins occupied more diverse thermal niches (Vianna *et al.*, 2020). *Eudyptes* penguins are widely spread and inhabit a variety of habitats that experience different thermal conditions making them excellent candidates to start uncovering the mystery that is cold-adaptation. In this study, a core genome phylogeny was constructed to resolve phylogenetic relationships between the *Eudyptes* penguins and the pan-genome of this genus was elucidated. The different pan-genome components were analysed and compared between cold-adapted *Eudyptes* species and the temperate *Eudyptes* species in order to uncover unique molecular determinants of thermoadaptation and thermoregulation.

2.2. Methods and materials

2.2.1 Genome sequences

The genome sequences of eight distinct species of *Eudyptes* were analysed in this study (Table 2.1). The genome sequence of the closely related yellow-eyed penguin was incorporated as outgroup. The latest genome assemblies along with the predicted protein datasets were obtained from the NCBI Genome Assembly database (National Centre for Biotechnology Information, 2022). These protein datasets were previously structurally and functionally annotated by Pan *et al.*, (2019) using the annotation pipeline developed by The B10K Genomes consortium. Genome metadata (G+C content, size, number of scaffolds and contigs) were also obtained from the NCBI database. Genome completeness for each of the

species was assessed using the BUSCO v. 3.0.2 tool (Benchmarking of Universal Single Copy Orthologues; Simão *et al.*, 2015) using all default settings, benchmarking the genome protein datasets against the Aves_odb9 database, comprised of 4,915 single-copy orthologous proteins conserved among birds.

Table 2.1: Penguin genomes that were utilised in this study, adapted from Pan *et al.*, 2019.

Scientific name	Common name	Sampling location	Accession no.	Genome size (Mbp)	GC content (%)	Number of scaffolds	Number of contigs	Genome coverage	Number of proteins per genome	Climate
Eudyptes chrysocome	Southern rockhopper	Falkland Islands/Malvinas	GCA_010085355.1	1231.07	40.9	3,371	31,096	163X	12,871	**Temperate**
Eudyptes chrysocome filholi	Eastern rockhopper	Crozet Island	GCA_010085365.1	1223.98	41	2,312	29,016	164X	12,955	**Temperate**
Eudyptes chrysolophus chrysolophus	Macaroni	Marion Island, Prince Edward Islands	GCA_010084205.1	1368.66	40.9	15,753	26,016	109X	15,079	**Cold**
Eudyptes chrysolophus schlegeli	Royal	Green Gorge, Macquarie Island	GCA_010080425.1	1310.61	40.5	17,083	86,483	320X	15,790	**Cold**
Eudyptes moseleyi	Northern rockhopper	Amsterdam Island	GCA_010082375.1	1306.70	40.2	20,381	78,721	140X	15,514	**Temperate**
Eudyptes pachyrhynchus	Fiordland	Harrison Cove, Milford Sound, New Zealand South Island	GCA_010085335.1	1310.92	41	52,188	97,297	116X	17,571	**Temperate**
Eudyptes robustus	Snares	The Snares, New Zealand sub-Antarctic	GCA_010085315.1	1248.62	41.6	62,719	89,364	139X	15,559	**Temperate**
Eudyptes sclateri	Erect-crested	Antipodes Island, New Zealand sub-Antarctic	GCA_010078445.1	1211.74	41.1	3,484	30,226	132X	12,811	**Temperate**
Megadyptes antipodes antipodes	Yellow-eyed	Otago Peninsula, New Zealand South Island	GCA_010078485.1	1317.73	40.3	17,379	37,645	89X	14,596	**Temperate**

2.2.2. Core genome and multi-gene phylogenies

To address some of the concerns relating to the taxonomy and species delineation in the genus *Eudyptes* (Cole *et al.*, 2019; Pan *et al.*, 2019), a core genome phylogeny was constructed. First, the protein datasets encoded by each *Eudyptes* genome and the outgroup were compared using OrthoFinder v. 2.2.7, which identifies orthologous proteins shared among genome datasets using a reciprocal best BlastP hit approach (Emms and Kelly, 2015). Orthologous proteins occurring in single copies (SCOs) among the *Eudyptes* penguins and the outgroup were selected. The proteins were individually aligned using the m-Coffee implementation of T-coffee (Notredame, *et al.*, 2000; Wallace *et al.*, 2006). The aligned proteins were concatenated and non-conserved or poorly aligned regions were removed using GBlocks v. 0.91b (Castresana, 2002). Subsequently, a maximum likelihood (ML) phylogeny was generated based on the curated concatenated alignment using IQ-Tree v. 1.6.12 (Nguyen *et al.*, 2015). The phylogeny was constructed using the best fit evolutionary model predicted using ModelTest (Prosada and Crandil, 1998) and the tree was evaluated with bootstrap support using the Ultrafast bootstrap approximation (n = 1,000 replicates) (Hoang *et al.*, 2018)

Similarly, a phylogeny was constructed using the nucleotide sequences of the house-keeping marker genes as previously used for phylogenetic analysis of all penguin species in the genus *Eudyptes* (Baker *et al.*, 2006). The latter phylogeny incorporated partial gene sequences for the mitochondrial ribosomal RNA genes 12S (668 bp) and 16S (905 bp) which are needed for messenger RNAs (mRNAs) to be translated into mitochondrial proteins (Yang *et al.*, 2014), cytochrome *b* (*cyt b*: 1,014 bp; main subunit of transmembrane cytochrome b6f and bc1 complex; Howell, 1989), cytochrome c oxidase subunit 1 (*COI*: 462 bp; encodes respiratory complex IV subunits for mitochondrial oxidative phosphorylation; Susanti *et al.*, 2018) and a nuclear exon of recombination activating protein 1 (*RAG-1*: 2,802 bp; plays a vital role in the immune systems of higher vertebrates; Baker *et al.*, 2006; Chiari *et al.*, 2009). The nucleotide sequences for each gene were obtained from the NCBI nucleotide database. The same workflow was used as for the core genome phylogeny; however, the IQ-Tree web server v. 1.6.12 (Trifinopoulos *et al.*, 2016) was used to generate a ML phylogeny. The phylogeny used the best fit evolutionary model predicted using ModelFinder (Kalyaanamoorthy *et al.*, 2017).

2.2.3. Identification and functional classification of orthologous protein datasets

The output from the OrthoFinder analysis described above was further used to perform a pan-genome analysis of the *Eudyptes* penguin consortium. The different pan-genome components: proteins which are uniquely encoded on a specific penguin genome or are only shared by a subset of the penguin genomes (collectively termed the accessory genome), as well as those encoded on all the analysed penguin genomes (the core genome), were determined. The accessory and core pan-genome components were classified according to function into their Conserved Orthologous Groups (COGs), using EggNOG-mapper v. 2 (Huerta-Cepas *et al.*, 2017) against the EggNOG v. 5.0 pre-clustered database (Huerta-Cepas *et al.*, 2019). The protein datasets were further classified according to their Gene Ontology (GO; that is according to their molecular function, biological process/es in which they are involved and their cellular location) using Blast2GO (Götz *et al.*, 2008) and WEGO (Web Gene Ontology Annotation Plot; Ye *et al.*, 2006), as well as according to their pathways using BlastKOALA (Basic Local Alignment Search Tool KEGG Orthology And Links Annotation; Kanehisa, Sato and Morishima, 2016) and KEGG Mapper (Kanehisa and Sato, 2020). Functions (or functional classes) which were over-represented (when proteins that belong to a specific GO term or KEGG pathway are present more than would be expected in data subsets) in penguins from colder climates, or which are present in single penguin taxa were identified and further functionally annotated by BlastP (Altschul *et al.*, 1990) analysis against the NCBI non-redundant (nr) protein database (Pruitt, *et al.*, 2005) and the Conserved Domain Database (Lu *et al.*, 2020) to identify orthologues in other cold adapted organisms.

Paralogous proteins are those that are derived through gene duplication events within the genome of one particular organism or species (Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine, 2005). Paralogues are significant as they can evolve new functions due to the lack of selective pressures acting on one copy (Ho-Huu *et al.*, 2012). As a result, this copy can mutate and acquire new functions (Ho-Huu *et al.*, 2012). Studies have shown that gene duplication is consistently identified as a means for adaptation to cold environments (Chen *et al.*, 2008; Zhang *et al.*, 2020). As such the number of paralogues (multiple copies per orthogroup) in each orthogroup for each *Eudyptes* penguin genome and the outgroup were evaluated from the Orthofinder output above. The average copy number was calculated for paralogous proteins found in the *Eudyptes* penguins to determine proteins that had high copy numbers as well as to cut out any extremely low outlying values represented by SCOs. Paralogous proteins that were over-represented by a two-fold difference in copy number in the cold climate clade penguins relevant to the temperate

climate clade penguins were then functionally annotated by BlastP (Altschul *et al.*, 1990) analysis against the NCBI nr protein database (Pruitt, *et al.*, 2005) and the Conserved Domain Database (Lu *et al.*, 2020).

2.3 Results and Discussion

2.3.1 Phylogenomic resolution of the genus *Eudyptes*

The genomes of penguins in the genus *Eudyptes* range in size between 1,211.74 Mbp (the erect-crested penguin) and 1,368.66 Mbp (the macaroni penguin) (Table 2.1). These genomes are relatively large among the bird genomes, especially, when compared to the smallest bird genome (910 Mbp) that belongs to the black-chinned hummingbird (*Archilochus alexandri*) (Zhang *et al.*, 2014). The penguins that make up the cold climate penguin clade (the macaroni and royal penguins) have two of the largest genome sizes among the eudyptid penguins, 1,368.66 Mbp and 1,310.61 respectively (Table 2.1). An average of 14,750 proteins ± 1,524.69 (average number of proteins ± standard deviation) are encoded on the genomes. The protein complement encoded on the genomes of eight *Eudyptes* penguins and the yellow-eyed penguin outgroup were compared using OrthoFinder v. 2.2.7 (Emms and Kelly, 2015). A total of 7,644 single-copy orthologues (SCOs) were identified to be conserved among the nine compared taxa. This contributes 43.5% - 59.66% of the total proteins encoded on the genomes. The SCOs were individually aligned using m-Coffee (Wallace *et al.*, 2006) and the alignments were concatenated. The original concatenated alignment comprised of 3,619,901 amino acid (aa) positions. After removal of poorly aligned regions in the alignment using GBlocks v. 0.91b (Castresana, 2002), the alignment comprised of 3,304,235 aa sites, with 6,938 distinct patterns, 26,325 singleton sites and 3,269,488 constant sites (98.95% of the total alignment). A ML phylogeny was constructed on the basis of this trimmed alignment using IQ-Tree v. 1.6.12 (Nguyen *et al.*, 2015). The best fit evolutionary model was determined to be JTT+F+I+G4.

In the core genome phylogeny, the *Eudyptes* penguins can be observed to form two distinct clades with strong bootstrap support (100%) (Figure 2.1). The delineation of the clades is in line with the biogeography of the taxa. As such, there is a clearly discernible cold climate clade, comprising of the macaroni and royal penguins (highlighted in blue in Figure 2.1) and a temperate clade comprising of six species, namely the southern rockhopper, eastern rockhopper, northern rockhopper, Fiordland, snares and erect-crested penguins (highlighted in red in Figure 2.1). The temperate clade can further be subdivided into two clades, namely clade 1 comprised of the southern rockhopper, eastern rockhopper and northern rockhopper

penguins and clade 2 consisting of the Fiordland, snares and erect-crested penguins.

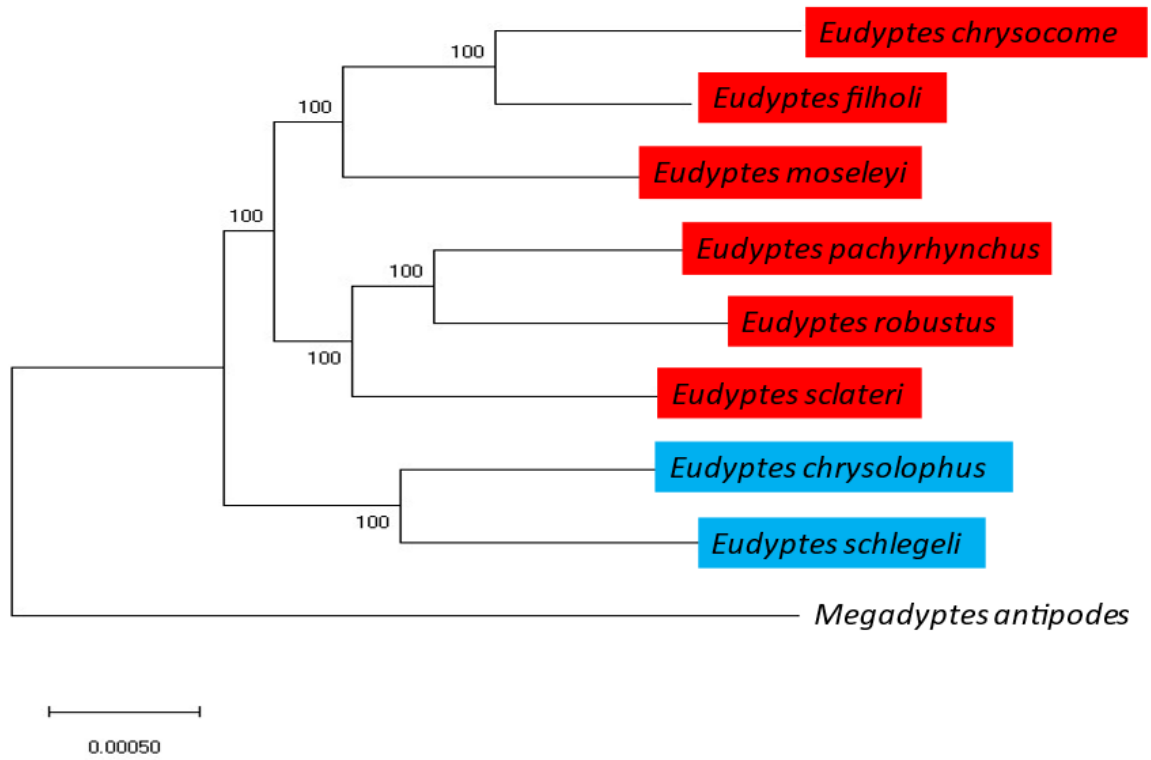


Figure 2.1: ML core genome phylogeny of the *Eudyptes* penguin species, rooted using the outgroup *Megadyptes antipodes*, created by IQ-Tree with the topology of all clades being strongly supported (bootstrap support: 100%). The red highlights represent the temperate climate penguins, the blue represents the cold climate penguins and no highlight represents the outgroup.

Similarly, a ML phylogeny was constructed on the basis of five distinct house-keeping genes. The trimmed concatenated gene alignment comprised 5,790 nucleotide positions, with 127 distinct patterns, 138 single sites and 5,483 constant sites (94.70% of total alignment). The DNA evolutionary model HKY+F+I was used to construct the phylogeny. The topology of the house-keeping ML phylogeny (Figure 2.2) is largely congruent with that of the core genome phylogeny (Figure 2.1), with one notable exception. In the house-keeping ML phylogeny (Figure 2.2) *E. sclateri* forms a branch separate from the temperate clade. Due to the greater sequence divergence (fewer conserved sites) in the nucleotide alignment (variable sites: 5.3% of total alignment), as compared with the amino acid-based core protein phylogeny (variable sites: 1.05% of total alignment), the house-keeping gene phylogeny may

be considered to provide the greater taxonomic resolution. This is expected given the concept of codon redundancy in which the higher variability in the few genes selected for the house-keeping gene ML phylogeny implies that a higher mutation rate of even a single gene would have a drastic effect on the resultant phylogeny (Spencer and Barral, 2012). The far larger set of amino acid sites analysed in the ML core genome phylogeny can thus be expected to provide a much better reflection of the true evolutionary relationships of the *Eudyptes* penguins at the whole genome scale. As such, the erect-crested penguin is more closely related to the Fiordland and snares penguins than indicated in the house-keeping gene phylogeny, which is in agreement with recent studies done by Pan *et al.* (2019) and Vianna *et al.* (2020) that used genomic data as well as in a study done by Ksepka *et al.* (2006) that used a combination of molecular and morphological data.

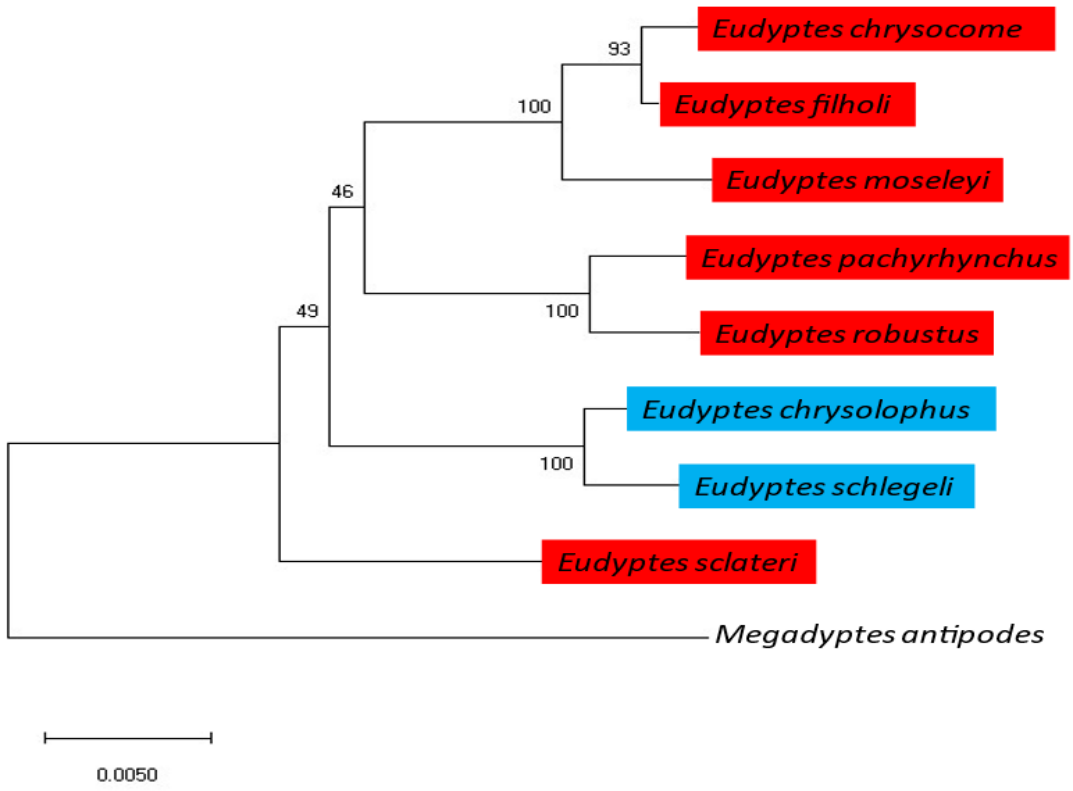


Figure 2.2: House-keeping gene ML phylogeny of the *Eudyptes* penguin species, the tree was rooted using *M. antipodes* and bootstrap values (n = 1,000 replicates) are shown. The red highlights represent the temperate climate penguins, the blue represents the cold climate penguins and no highlight represents the outgroup.

2.3.2. Pan-genome analyses of the genus *Eudyptes*

The pan-genome complement of the eight *Eudyptes* penguins and the yellow-eyed penguin outgroup was identified using OrthoFinder v. 2.2.7 (Emms and Kelly, 2015). The pan-genome complement totalled 16,511 orthogroups. Of these orthogroups 59.88% were shared among the nine compared taxa (core proteins). When taking into account core proteins that are not shared by all the temperate climate clade penguins or both the penguins from cold environments, the core proteins increase to 78.24% (Figure 2.3). With the exclusion of the yellow-eyed penguin, 61.32% of proteins were shared between all temperate climate *Eudyptes* penguins and the cold climate penguins. Considering the proteins that are only shared by the yellow-eyed penguin and the cold climate penguins, 8.15% of proteins are shared (accessory proteins). By contrast, only 0.72% and 0.23% are shared between the yellow-eyed penguin and the temperate clade penguins and the yellow-eyed penguin and the cold climate penguins, respectively (Figure 2.3). Only 1.28% of the proteins are unique to the penguins from cold environments, whereas in the temperate climate penguin clade, 10.89% of proteins are found to be unique (Figure 2.3). This shows that there is greater variability in the larger dataset of temperate climate penguin genomes. While relatively large unique fractions were observed both at the clade and individual penguin level, it cannot be excluded that these are the result of the genome sequencing, or the annotation pipeline used. With regards to genome sequencing, the analytical validity may be affected by the coverage depth, which in turn, is affected by the quantity and quality of DNA available for sequencing as well as the computational capabilities of the software used to assemble reads into the final genome assemblies (Raszek, Tisler and Hernandez, 2016). Also, depending on the annotation pipeline used, it may not be accurate enough with regards to orthogroup inference (Emms and Kelly, 2015). This may affect the relationships inferred between different proteins i.e. the annotation pipeline may identify a protein to be unique to a particular species, but due to the lower accuracy of protein annotation, orthology or paralogy may be incorrectly inferred.

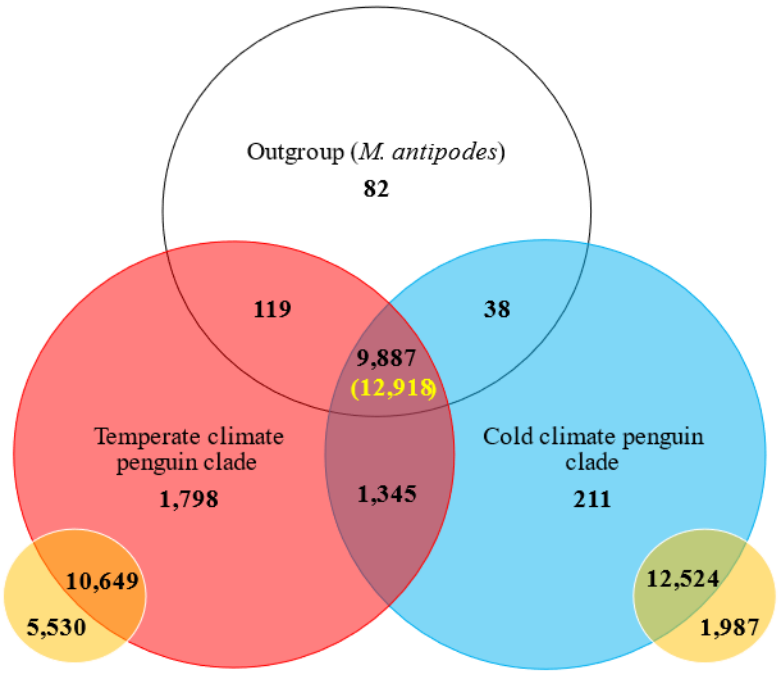


Figure 2.3: Venn diagram encompassing the pan-genome protein complement for the *Eudyptes* penguins and the outgroup. The black value in the center of the venn diagram represents core proteins shared by all nine compared taxa. The value in yellow and brackets represents the core proteins that are not shared by all the temperate climate clade penguins or the cold climate clade penguins. The lower values in the orange circles represent those proteins which are not shared by all taxa in the clade, while the upper values are core to all penguins in the clade.

2.3.3. Functional analysis of the *Eudyptes* pan-genome fractions

The unique and core protein elements of the pan-genome were assigned to COG categories using EggNOG-mapper v. 2 (Huerta-Cepas *et al.*, 2017). Of the unique proteins in the cold climate penguin clade, 82% could be assigned to COG categories, while 95% of those unique to the temperate climate penguin clade could be assigned to COG categories (Figure 2.4). Of the proteins shared by all nine taxa (core proteins), 99.2% were assigned to COG categories (Figure 2.4). For the 9,887 core proteins that were identified, 9,812 were assigned COG categories (Figure 2.4).

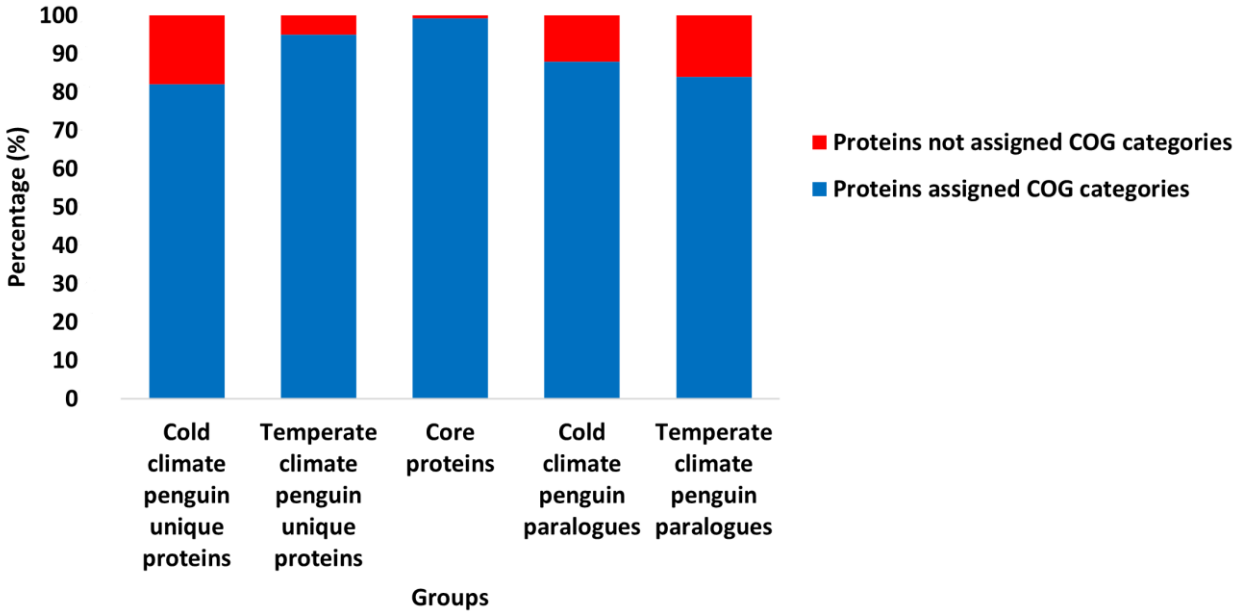


Figure 2.4: The percentage of COG categories assigned to the pan-genome complement within the cold climate penguin and the temperate climate penguin clades. The proteins that were assigned COG categories (represented by the colour blue) did not fall below 82%.

The majority of the core proteins fell within the Cellular Processes & Signalling supra-functional COG category (41%; 4,069 proteins), the Poorly Categorised category comprised 2,607 core proteins (27%), while a similar proportion of core proteins (1,593 and 1,542 respectively; Figure 2.5) belonged to the COG supra-functional categories Information Storage & Processing and Metabolism (16% each).

The cold climate penguins had a total of 174 unique proteins that were assigned to COG categories (Figure 2.4), with 86 of these proteins in the Cellular Processes & signalling supra-functional COG category (Figure 2.5). The Information Storage & Processing category consisted of 33 proteins, while 31 fell in the Poorly Categorised category (Figure 2.5) and 25 proteins belonged to the COG category Metabolism (Figure 2.5). For the temperate climate penguin clade, 1,707 out of 1,798 unique proteins were assigned COG categories in which 638 belong to the supra-functional category Cellular Processes & Signalling (Figure 2.5), 394 proteins were assigned to the Information Storage & Processing, 365 in the Poorly Categorised and 310 in the Metabolism categories (Figure 2.5), respectively.

The core proteins and the fractions unique to the cold climate penguins and the temperate climate penguins are dominated by the proteins involved in Cellular Processes & Signalling ($\geq$37%; Figure 2.5), however, the relative proportion of unique proteins with roles in metabolism are lower in the cold climate penguins (14%) than in the temperate climate penguin clade (18%) and the core proteins (16%; Figure 2.5).

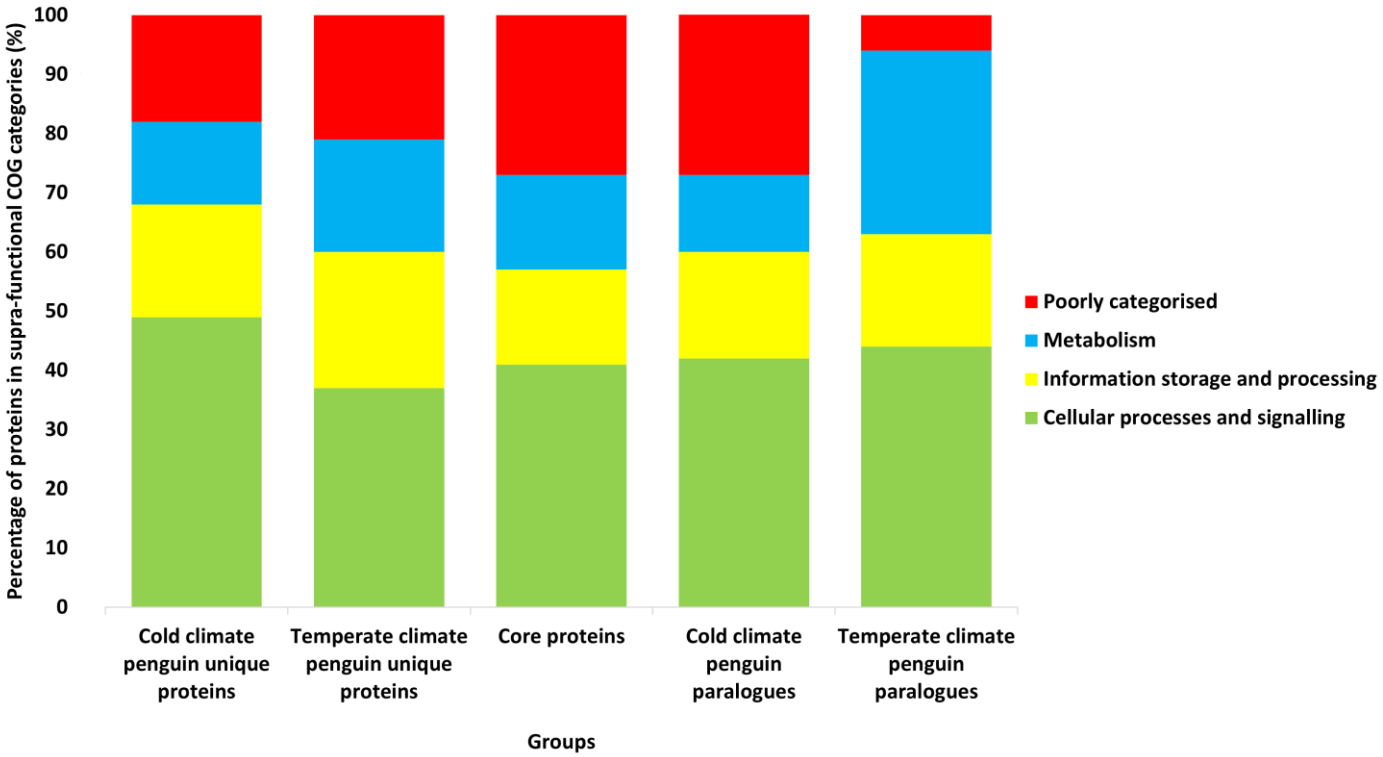


Figure 2.5: Supra-functional COG categories assigned to the pan-genome complement within the cold climate penguin and the temperate climate penguin clades. The Cellular Processes & Signalling supra-functional COG category was the category that was mostly assigned to proteins (37 – 49%).

With regards to the specific COG categories in each group (cold climate penguin unique proteins, temperate climate penguin unique proteins and core proteins; Figure 2.6), proteins in the COG categories Function Unknown (S; 20% - 30%), Signal Transduction Mechanisms (T; 15% - 22%), Transcription (K; 8% - 10%), Post-translational Modification, Protein Turnover, Chaperones (O; 7% - 10%) and Cytoskeleton (Z; 3% - 9%) are predominant (Figure 2.6). However, proteins involved in some categories are overrepresented in the cold climate penguins such as the Post-Translational Modification, Protein Turnover, Chaperones (10%), Cytoskeleton (9%), Replication, Recombination & Repair (L; 3%), Extracellular Structures (W; 3%), Nucleotide Transport & Metabolism (F; 2%), Chromatin Structure and Dynamics (B; 2%) and Nuclear Structure (Y; 0.6%) categories (Figure 2.6) relative to the proteins unique to the temperate climate penguins (Figure 2.6).

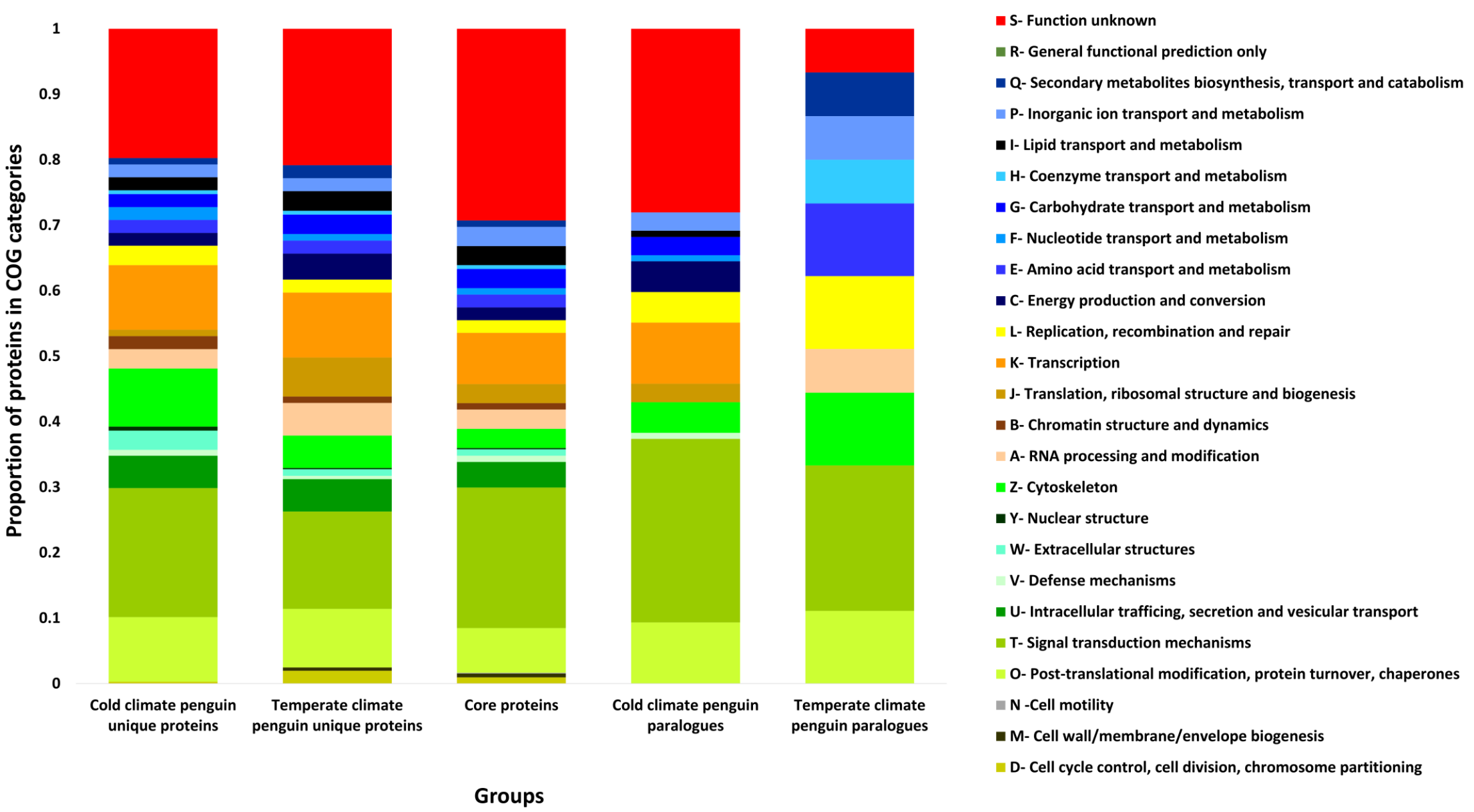


Figure 2.6: COG categories assigned to the pan-genome complement within the cold climate penguin and the temperate climate penguin clades. The Signal Transduction Mechanisms and Function Unknown COG categories were mostly assigned to proteins (15 – 30% and 6 – 30%, respectively).

2.3.4. The role of cold penguin-unique proteins in thermoregulation and adaptation

The protein complement unique to the cold clade penguins was further functionally annotated by BlastP analysis (Altschul *et al.*, 1990) against the NCBI nr protein database (Pruitt *et al.*, 2005) and the Conserved Domain Database (Lu *et al.*, 2020) to identify proteins with a potential role in thermoregulation. A total of 35 proteins unique to the cold clade penguins may play a potential direct or indirect role in thermoregulation within the cold-adapted *Eudyptes* penguins (Table 2.2).

Table 2.2: The 35 unique proteins that may play a potential role in thermoregulation within the cold-adapted *Eudyptes* penguins.

Protein short name	Preferred protein name	COG category	Putative function	Protein size (aa)		Locus Tag	
				E. chrysolophus	***E. schlegeli***	***E. chrysolophus***	***E. schlegeli***
ACVRL1	Serine threonine-protein kinase receptor R3	T	Mediates cell growth and differentiation, especially in developing arteries.	135	246	FQV09_0012827	FQV10_0002877
ADCK4	AarF Domain-Containing Protein Kinase 4	S	Coenzyme Q (ubiquinone; Q), a crucial lipid component in respiratory electron and proton transport, is biosynthesized by ADCK1. Coenzyme Q synthesis is modulated by this component, which is also necessary for the creation of a multi-subunit Q-biosynthetic complex.	_	349	_	FQV10_0010214
AKR1B10	Aldo-keto reductase family 1 member B10	S	This protein reduces aliphatic and aromatic aldehydes effectively, although it is less active on hexoses. It functions as an all-trans-retinaldehyde reductase (*in vitro*). Prior to the body's other organs receiving nutrients from digested food, it is in charge of detoxifying reactive aldehydes.	316	104	FQV09_0015800	FQV10_0016136
ALOXE3	Hydroperoxide Isomerase ALOXE3	E	Generates fatty acid hydroperoxides by adding oxygen to specific fatty acids.	171	92	FQV09_0014365	FQV10_0011004
ASMT	Acetylserotonin O-methyltransferase	S	Utilizing S-adenosyl-L-methionine (SAM or AdoMet) as the methyl donor, this enzyme catalyses the methylation of one or more particular substrates.	193	327	FQV09_0015520	FQV10_0001933
COQ10B	Coenzyme Q10B	I	Coenzyme Q, a vital electron carrier for the respiratory chain, is bound by COQ10B.	_	116	_	FQV10_0008137
CTSK	Cathepsin K	O	A cysteine proteinase that is mainly expressed in osteoclasts which cleaves bone matrix proteins.	95	124	FQV09_0014054	FQV10_0000959

Table 2.2 continued.

Protein short name	Preferred protein name	COG category	Putative function	Protein size (aa)		Locus Tag	
				E. chrysolophus	*E. schlegeli*	*E. chrysolophus*	*E. schlegeli*
			During bone resorption this protein plays a vital role in degrading the organic phase of bone.				
FFAR3	Free fatty acid receptor 2	T	A class A G-protein coupled receptor that binds free fatty acids.	55	55	FQV09_0012111	FQV10_0007026
GABRA6	Gamma-aminobutyric acid receptor subunit alpha-6	T	An anionic channel, mediating fast inhibitory synaptic transmission. Upon gamma-aminobutyric acid (GABA) binding to the ligand binding site on the extracellular domain, chlorine ions are selectively conducted through the GABRA pore, resulting in hyperpolarization of the neuron.	150	77	FQV09_0003462	FQV10_0010926
GYS1	Glycogen [Starch] Synthase, Muscle	G	Controls the production of 1,4-linked glucose chains in glycogen. It is the enzyme that restricts the rate of polysaccharide production.	_	150	_	FQV10_0012779
HDAC5	Histone deacetylase 5	B	Produces a deacetylated histone by catalysing the hydrolysis of a histone's N (6)-acetyl-lysine residue (EC 3.5.1.98). The process of histone acetylation and deacetylation is crucial for mediating the transcriptional control of several genes.	_	64	_	FQV10_0003698
IL6ST	Interleukin 6 Cytokine Family Signal Transducer	T	Mediates signals that regulate pain control, hematopoeisis, bone metabolism and immune response. Plays a role in embryonic development and is critical for the survival of sensory and motor neurons as well as the differentiation of astrocytes.	919	290	FQV09_0003049	FQV10_0008525
JAK3	Janus kinase 3	T	It is crucial for the signalling of cytokines that bind to the shared receptor subunit common gamma chain, including interleukin (IL)-2, IL-4, IL-7, IL-9, IL-15, and IL-21. The myeloid cell differentiation process and lymphoid development both depend on Jak3.	370	1136	FQV09_0013536	FQV10_0016040

Table 2.2 continued.

Protein short name	Preferred protein name	COG category	Putative function	Protein size (aa)		Locus tag	
				E. chrysolophus	*E. schlegeli*	*E. chrysolophus*	*E. schlegeli*
KCNQ4	Potassium voltage-gated channel subfamily KQT member 4	P	Creates a potassium channel that is thought to be essential for controlling neuronal excitability, notably in the cochlea's sensory cells.	162	611	FQV09_0014145	FQV10_0005597
KIF5B	Kinesin-1 heavy chain	Z	Suppresses both mitotic cell division and chromosomal instability during the early phases of female meiotic cell development. Allows for the localisation of alpha-sarcomeric actin, nestin and desmin in differentiating myoblast cells.	212	965	FQV09_0001785	FQV10_0000139
KRT13	Keratin, type I cytoskeletal 13	Z	A type I cytokeratin that is expressed in non-cornified stratified epithelia's suprabasal layers.	341	94	FQV09_0009751	FQV10_0006075
KRT75	Keratin, type II cytoskeletal 75	S	The development of hair and nails is greatly influenced by this intermediate filament protein.	541	225	FQV09_0014681	FQV10_0011979
MBOAT4	Ghrelin O-acyltransferase	S	An integral transmembrane enzyme that modifies protective cell-surface polymers, catalysing the transfer of an acetyl group from acetyl-CoA to its target substrates.	_	426	_	FQV10_0000282
MDM2	E3 Ubiquitin-Protein Ligase Mouse Double Minute 2	O	Forms homo-oligomers and displays E3 ubiquitin ligase activity, catalysing the attachment of ubiquitin to p53 as a vital step in the regulation of its expression levels in cells.	461	259	FQV09_0001601	FQV10_0005891
MYH1	Myosin heavy chain, skeletal muscle, adult	Z	Responsible for producing actomyosin contraction in metazoan muscle and non-muscle cells.	1945	574	FQV12_0015905	FQV10_0007550
NDUFA11	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	C	Transfers electrons in the mitochondrial respiratory chain from NADH to ubiquinone.	_	63	_	FQV10_0012821
NKX1-1	NK1 transcription factor-related protein 1	K	Is essential for the coordinated crosstalk of elements involved in maintaining energy homeostasis and is expressed primarily in the brainstem.	_	384	_	FQV10_0001847
NKX6-1	NK6 Homeobox 1	K	Transcription factor that is essential for the growth and operation of pancreatic beta cells.	_	262	_	FQV10_0011756
PIK3CG	Phosphatidylinositol 4,5-bisphosphate 3-kinase	T	Numerous immunological and vascular processes, such as cell recruitment, mast cell activation,	1106	91	FQV09_0001441	FQV10_0016266

Table 2.2 continued.

Protein short name	Preferred protein name	COG category	Putative function	Protein size (aa)		Locus tag	
				E. chrysolophus	*E. schlegeli*	*E. chrysolophus*	*E. schlegeli*
	catalytic subunit gamma isoform		platelet aggregation, and smooth muscle contraction, are regulated by phosphoinositol 3-kinase (PI3K) gamma signalling.				
PPARGC1B	Peroxisome proliferator-activated receptor gamma coactivator 1-beta	A	Has a non-redundant role in regulating the mitochondrial oxidative energy metabolism, has an impact on both mitochondrial biogenesis and insulin sensitivity, and performs a variety of tasks in oxidative tissues.	_	984	_	FQV10_0010906
RARRES2	Retinoic acid receptor responder protein 2	T	Functions in inflammation and metabolism; exhibits antimicrobial function in the skin.	_	78	_	FQV10_0013504
SLC17A6	Vesicular glutamate transporter 2	G	Enables the transport of inorganic phosphate as well as the absorption of glutamate into synaptic vesicles in excitatory neuronal cells' presynaptic nerve terminals.	429	125	FQV09_0006415	FQV10_0009662
SLC38A2	Sodium-coupled neutral amino acid transporter 2	E	Is a transporter of amino acids that is dependent on sodium. Mediates the 1:1 stoichiometric, pH-sensitive, and electrogenic cotransport of sodium ions with neutral amino acids. Possibly involved in the blood-brain barrier's transfer of amino acids.	106	504	FQV09_0001566	FQV10_0000533
SMAD4	Small mothers against decapentaplegic homolog 4	K/T	Multiple signalling pathways are mediated by this transcriptional regulator and signal transducer. SMAD4, a member of the Dwarfin protein family, is essential for a variety of cellular processes, including differentiation, apoptosis, gastrulation, embryonic development, and the cell cycle.	551	543	FQV09_0003796	FQV10_0017019
SPHK2	Sphingosine kinase 1	I/T	A secondary messenger that acts as a protein kinase C activator.	377	382	FQV09_0009094	FQV10_0001705
TBX1	T-Box Transcription Factor 1	K	TBX1 is a T-box transcription factor that is crucial for the formation of the second heart field's extracellular matrix-cell interactions, which are essential for the development of the cardiac outflow tract.	260	398	FQV09_0013579	FQV10_0010460

Table 2.2 continued.

Protein short name	Preferred protein name	COG category	Putative function	Protein size (aa)		Locus tag	
				E. chrysolophus	*E. schlegeli*	*E. chrysolophus*	*E. schlegeli*
TET2	Ten-eleven translocation (TET) 2 methylcytosine dioxygenase	S	Through the repeated oxidation of 5-methylcytosine into 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine, TET2 participates in DNA demethylation.	1996	_	FQV09_0000424	_
TH	Tyrosine Hydroxylase	E	Ensures that tyrosine is converted into L-dihydroxyphenylalanine (L-DOPA), the rate-limiting step in the production of the catecholamines dopamine, noradrenaline, and adrenaline.	130	130	FQV09_0016083	FQV10_0012259
TUBA3E	Tubulin alpha-3 chain	Z	An alpha-tubulin that makes up (along with beta-tubulin) microtubules. Plays roles in development Slit-Robo signalling and cooperation of prefoldin as well as triC/CCT in actin and tubulin folding.	450	450	FQV09_0004750	FQV10_0012630
VDR	Vitamin D (1,25-dihydroxyvitamin D3) receptor	K	VDR is a high-affinity receptor for 1alpha,25-dihydroxyvitamin D3, the biologically most active Vitamin D metabolite. It regulates a wide range of biological processes, including calcium metabolism, cell proliferation and differentiation, and immunomodulation.	86	441	FQV09_0014667	FQV10_0008511

2.3.4.1 Cold penguin-unique proteins with a role in the brain and thermoregulation

Thermoregulation is controlled by the preoptic area (POA) of the hypothalamus in the brain. The POA exerts control over all thermoefferent mechanisms (signals that carry temperature information away from the central nervous system) as it receives thermoafferent signals (signals that provide temperature information to the central nervous system from outside stimuli) from the skin as well as other parts of the body (Osaka, 2004). The POA contains GABAergic neurons that release GABA to the extracellular space and express GABRA6 (GABAA) receptors (Osaka, 2004). This GABAergic mechanism in the POA inhibit warm-sensitive and thermally insensitive neurons. The extracellular GABA level in the POA has been shown to increase when organisms have been exposed to acute cold, furthermore, heat exposure has resulted in GABA decreasing (Osaka, 2004). As a result, GABA and by extension GABRA6 receptors may be involved in the mechanism of thermoregulation in the POA through cold-induced thermogenesis (Osaka, 2004). Heat-activated GABAergic neurons in the POA reduce the activity of cold-activated neurons in the dorsomedial hypothalamus (DMH). This leads to an increase in thermogenesis (Zhao *et al.,* 2017). The SLC17A6 (VGLUT2) protein is required for the synthesis of postsynaptic neurons found in the dorsal part of the DMH, which are activated by cooling (Zhao *et al.,* 2017). Both GABRA6 and VGLUT2 are unique to the cold-adapted penguin clade.

ASMT is an enzyme that converts normelatonin to melatonin (catalyses the final reaction in melatonin biosynthesis). Melatonin has a vital role in circadian thermoregulation in which it sends signals to the POA of the anterior hypothalamus. Body temperature is then adjusted to be consistent with the metabolic rate of the organism (Saarela and Reiter, 1994). ASMT is found to be present in both cold-adapted *Eudyptes* penguins, but absent in the temperate clade penguins. Another protein unique to the cold climate clade is tyrosine hydroxylase (TH). TH catalyses the first step in catecholamine synthesis (Doan *et al.,* 2016). Catecholaminergic mechanisms are employed by the central and peripheral systems for temperature regulation (Myers, 1980). Uncoupling proteins (UCP1-3 in mammals and avUCP in birds) uncouple oxidative phosphorylation, thereby converting the energy required to synthesise ATP into heat (Rajagopal *et al.*, 2019). Proteins that take part in this mechanism include, PPARGC1B which is a regulator of fatty acid beta-oxidation and oxidative phosphorylation in mitochondria (Shao *et al.,* 2010). It also allows for the modulation of energy expenditure. NDUFA11 is an accessory subunit of the mitochondrial membrane respiratory chain NADH

dehydrogenase (Complex I). The respiratory chain receives electrons from NADH through the action of complex I. Ubiquinone serves as the enzyme's primary electron acceptor. COQ8B is an atypical kinase involved in the production of ubiquinone (coenzyme Q10) which is important for oxidative phosphorylation. Coenzyme Q's role in the respiratory chain depends on COQ10B. (Rajagopal *et al.,* 2019). NDUFA11, COQ8B and COQ10B play a role in NST through participating in the mechanism of mitochondrial proton uncoupling through oxidative phosphorylation (Rajagopal *et al.,* 2019). These proteins are restricted in distribution to a single cold climate penguin, namely the royal penguin.

2.3.4.2 Cold penguin-unique proteins involved in glucose metabolism and non-shivering thermogenesis

Glucose metabolism and homeostasis is needed in order for NST to take place. NST occurs in brown adipose tissue (mammals; BAT) and beige adipose tissue (birds; Xu *et al.*,2021) which allows for the maintenance of stable body temperature when exposed to low temperatures. These adipose tissues metabolise glucose (and glycogen) as well as fatty acids (lipids) via oxidation and results in dissipating energy as heat (Xu *et al.*,2021). Among the cold climate penguin unique protein complement, AKR1B1 is a key enzyme in the polyol pathway, which catalyses the reduction of glucose to sorbitol consuming NADPH during hyperglycaemia (in humans). Sorbitol is later converted to fructose by sorbitol dehydrogenase (Khayami *et al.,* 2020). Fructose is used to produce energy (and heat as a by-product) through glycolysis. Another protein, MBOAT4 (unique to the royal penguin), is essential for the activation of ghrelin (28 aa hormone) which has been shown to play a role in cellular energy homeostasis as well as in glucose metabolism and lipid homeostasis (Churm *et al.,* 2016). PIK3CG is implicated in the control of metabolic functions in multiple places. It plays a role in controlling glucose and lipid homeostasis (Nürnberg and Beer-Hammer, 2019). The *GYS1* gene (unique to the royal penguin) codes for muscle glycogen synthase that is produced in most cells, but is abundant in cardiac muscle and skeletal muscle. Muscle glycogen synthase forms glycogen by linking together glucose molecules which is then stored in the muscles and used as an energy source when needed (Kollberg *et al.,* 2007). In addition, this process is essential for glucose homeostasis as well as energy homeostasis.

Asparagine has been shown to drastically upregulate thermogenesis, glucose influx and lipogenesis in adipocytes, which allows for better cold tolerance (Xu *et al.,* 2021). Asparagine also stimulates the mTORC1 pathway, which promoted expression of

thermogenic genes and key enzymes in glycolysis via transcriptional and post-transcriptional mechanisms. Asparagine bioavailability thus affects thermogenic and glycolytic activities in adipose tissue. SLC38A2 is an asparagine transporter that allows adipose tissue to acquire asparagine (Xu *et al.*, 2021) and is unique to the cold climate penguin clade.

Interleukin 15 (IL15) is a specialised cytokine (myokine) that regulates glucose metabolism in skeletal muscle through the JAK3 - signal transducer and activator of transcription 3 (STAT3) signalling pathway. This signalling pathway along with the intracellular JAK – mitogen-activated protein kinase (MAPK) signalling pathway is activated by IL6ST homodimerization as well as the formation of a high-affinity receptor complex induced by the binding of interleukin 6 (IL6) to the interleukin 6 receptor (IL6R) i.e. the receptor systems for interleukin 6 can utilise IL6ST to initialise signal transmission (Hibi *et al.*, 1990). These two pathways (JAK3 – STAT3 and JAK – MAPK) have been implicated to have a function in adipocyte differentiation and function, lipid metabolism and glucose metabolism (Burrell *et al.*, 2020). JAK3 as well as IL6ST may play an indirect role in adaptive thermogenesis (Krolopp *et al.*, 2016; Richard and Stephens, 2014). CTSK, a cysteine protease involved in extracellular matrix remodelling is one of the determinants of adipocyte differentiation (Xiao *et al.*, 2006). Additionally, it has been demonstrated that the ribosomal protein-MDM2-p53 axis controls lipid synthesis and storage in response to nutrient availability (Krstic et al., 2018). Furthermore, it has been shown that MDM2 recruits cAMP response element binding protein to begin adipocyte differentiation in a p53-independent manner. (Creb1; Krstic *et al.*, 2018). JAK3 IL6ST, CTSK as well as MDM2 are unique to the cold-adapted penguins.

SMAD4 (along with SMAD2 and SMAD3) differentially regulates glucose and lipid metabolism by modulating murine protein serine-threonine kinase 38 (MPK38)-dependent mitogen-activated protein kinase 5 (ASK1)/TGF-β/p53 signalling (Seong *et al.*, 2018). Smad class and MPK38-mediated Smad phosphorylation are required for this. (Seong *et al.*, 2018). Through activation of the chemokine-like receptor 1 (CMKLR1), the adipocyte-secreted protein (adipokine), RARRES2 (unique to the royal penguin) regulates adipogenesis, metabolism, and inflammation (Rourke et al., 2014). It affects the expression of adipocyte genes involved in lipid and glucose metabolism, favorably influences adipocyte differentiation, and may contribute to angiogenesis (vascular development), a procedure necessary for the growth of white adipose tissue (WAT; Rourke *et al.*, 2014). KIF5B regulates glucose transporter type 4 (GLUT4) translocation and adiponectin secretion in adipocytes and has been shown to be involved in metabolism homeostasis. KIF5B also affects

mitochondrial biogenesis as well as key thermogenic gene expression. KIF5B therefore plays a vital role in glucose and lipid metabolism as well as thermoregulation (Cui *et al.,* 2017). SMAD4, RARRES2 and KIF5B are only present in the cold-adapted penguins.

2.3.4.3 Cold penguin-unique proteins implicated in adipose tissue metabolism

Adipose tissue is a major metabolic organ that stores substrates (white adipose tissue; WAT), produces heat (BAT and beige adipose tissue) as well as conserves heat (WAT) and as such plays a vital role in thermoregulation (Choe *et al.,* 2016) in conjunction with glucose metabolism. TET2 (unique to *E. chrysolophus*) interacts with PPARγ, a master transcription factor of adipogenesis and is primarily expressed in adipose tissue. Numerous genes that are critical for adipocyte processes like lipid metabolism, adipokine secretion, and insulin sensitivity are regulated by this protein interaction in a key way (Bian *et al.,* 2018). Insulin directly activates body heat production by stimulating glucose uptake. TET2 (unique to the macaroni penguin) is an epigenetic regulator of PPARγ ligand-mediated insulin sensitisation and transcriptional regulation in adipocytes. Through insulin sensitisation, TET2 can also be linked to glucose metabolism and homeostasis.

The G protein-coupled receptor (GPCR) FFAR3 is activated by short chain fatty acids (SCFAs), a product of dietary fibre digestion (Fuller *et al.,* 2015). This protein plays a role in regulating adipose insulin signalling by sensing SCFAs that are provided by gut microbes and as a result regulates fat accumulation and whole-body energy homeostasis (Ichimura *et al.,* 2014). FFAR3 may also play a role in glucose homeostasis by triggering the release of glucagon-like peptide-1 (GLP-1) as a result of SCFAs accumulating in the intestine (Fuller *et al.,* 2015). In response to cold exposure, heat needs to be produced rapidly and sustainably. Major thermogenic tissues such as adipose tissue and skeletal muscle increase heat production to preserve a constant body temperature through a mechanism that is heavily reliant on glucose oxidation (Brown *et al.,* 2019). One other protein, SPHK2 (along with SPHK1) regulates angiogenesis, glucose and lipid homeostasis through sphingosino-1-phosphate (S1P) regulation (Juchnicka *et al.,* 2021). The activation of protein kinase B (Akt) in muscles, through S1P, improves insulin responsiveness by increasing glucose uptake and glycogen synthesis. However, S1P has been shown to inhibit Akt activation in adipose tissue after insulin stimulation (Juchnicka *et al.,* 2021).

NKX6-1 (present in *E. schlegeli*) is involved in the development of insulin-producing beta cells in the pancreas, which may play a role in glucose homeostasis (Iype *et al.,* 2004).

Biological factors in the hypothalamus and brainstem play a major role in mediating the regulation of energy balance. To allow the coordinated crosstalk of factors involved in the maintenance of energy homeostasis, NKX1-1 (unique to the royal penguin) expression may be necessary, possibly in controlling the transcription of particular factors involved in energy balance (Simon *et al.,* 2010). A genetic indicator of beige adipose tissue is TBX1. It is necessary for maintaining beta-adrenergic sensitivity as well as appropriate insulin signalling in adipose tissue. The regulation of beige adipocyte function, energy homeostasis, and adipocyte formation are all mediated by adipose TBX1. (Markan *et al.,* 2020). Vitamin D's VDR-mediated activity affects adipocyte biology and energy metabolism. Vitamin D controls energy metabolism through participating in the control of beta oxidation in WAT (Wong *et al.,* 2009). *ALOXE3* codes for the enzyme eLOX3 which has functions in epidermal tissues. *ALOXE3* plays a role in adipocyte differentiation via PPARγ agonism as well as whole-body insulin sensitivity through the *ALOXE3*-PPARγ axis (Higgins *et al.,* 2018). NKX6-1, NKX1-1, TBX1, VDR and *ALOXE3* are uniquely present in the cold climate *Eudyptes* penguins.

2.3.4.4 Cold penguin-unique proteins that function in skeletal muscle and shivering thermogenesis

Skeletal muscle is one of the main thermogenic tissues in birds and has specific and precise intracellular organization (Sébastien *et al.,* 2018). Tubulin is the major constituent of microtubules. Skeletal muscle fibres are mostly made up of microtubules which are composed of a heterodimer of alpha and beta tubulin (one such alpha tubulin is TUBA3E). Each skeletal muscle fibre contains many myofibrils. Myofibrils are bundles of actin and myosin filaments organised into a chain of repeating units (sarcomeres). This therefore results in efficient muscle contraction which is vital for ST especially in penguins. MYH1 is a myosin protein that is uniquely expressed in striated muscle (skeletal and cardiac muscle) and helps with ST through efficient muscle contraction.

HDAC5 (found only in *E. schlegeli*) is highly expressed in skeletal muscle as well as in the heart and brain. HDAC5 is a crucial protein needed for controlling the pancreatic beta/delta-cell lineages (Sharma *et al.,* 2021) and furthermore, regulates insulin production. Additionally, it has been demonstrated to be a detrimental epigenetic regulator of IL6 production and release in skeletal muscle (Klymenko *et al.,* 2020). HDAC5 may have positive effects through two separate pathways, including transcriptional regulation of genes necessary for the consumption and disposal of glucose and HDAC5-dependent IL-6 signalling crosstalk to enhance muscle uptake of glucose (Klymenko *et al.,* 2020). Therefore,

HDAC5 may be needed for the normal functioning of skeletal muscle. TUBA3E, MYH1 and HDAC5 are unique to the cold climate clade penguins.

2.3.4.5 Cold penguin-unique proteins that play a role in vascular development

KCNQ channels (KCNQ4 and KCNQ5) are essential for regulating smooth muscle contractility (Greenwood and Ohya, 2009). Most involuntary activity is driven by the contraction of smooth muscle, which is found in the urinary bladder, gastrointestinal system, airways, and reproductive organs. This mechanism is needed to redistribute blood to areas where it is needed and may play an essential role in the penguins' counter-current circulatory system which allows them to regulate their internal temperature (McCafferty *et al.*, 2013). KCNQ4 is unique to the cold-adapted penguins. ACVRL1 (ALK1) is a type 1 receptor that has been found to be expressed in blood vessels and plays an important role in vascular development (Oh *et al.*, 1999). ALK1, unique to the cold climate penguins, along with ALK5 signalling pathways in endothelial cells play a vital role in determining vascular endothelial properties during angiogenesis (Oh *et al.*, 1999). This protein allows for endothelial cells to release substances that control vascular relaxation as well as contraction.

2.3.4.6 Cold penguin-unique proteins that operate in feather development

Feathers play a critical role in heat retention and as such play a role in thermoregulation. Regarding the vertebrate integument, the feather is the most complex keratinised structure. Feathers and other avian integumentary appendages like claws, beaks and scales are mainly composed of α- and β-keratins. Alpha-keratins are found in all vertebrates (Wu *et al.*, 2015). Beta-keratins, on the other hand, are unique to birds and reptiles. Different combinations of α- and β-keratins have been shown to contribute to the structural diversity of feathers (Wu *et al.*, 2015). Feather structure has been shown to be dependent on α-keratins such as KRT75 (Ng and Li, 2018) and KRT13, which are unique to the cold-clade penguins. KRT75 is expressed in the rachis (feather backbone) and ramus (branches) in normal regenerating adult feathers. KRT13 is a type I α-keratin that has been shown to be expressed at a higher level in feathers than in other skin appendages (Wu *et al.*, 2015).

2.3.5. Gene duplication in the cold penguin clade

Studies have shown that gene duplication represents a key means for adaptation to cold environments. One study by Zhang *et al.* (2020) showed that the Antarctic sea ice green alga *Chlamydomonas* sp. UWO241 harboured 901 duplicated genes which were involved in a range of cellular processes, some of which were concluded to aid in its survival in the Antarctic as a result of gene dosage (higher protein accumulation). The duplication of the nuclear *petf* gene specifically, resulted in two similar copies of the photosynthetic ferredoxin (PETF) protein whose expression and retention was concluded to be an adaptation to the cold environment (Zhang *et al.,* 2020). Another study conducted by Chen *et al.* (2008) identified 118 Antarctic-specific duplicated protein-coding genes within Antarctic notothenioid fish. It was concluded that the dramatic genomic expansion of certain protein gene families, augmented gene expression as well as gene functions were contributing factors that resulted in the physiological fitness of Antarctic notothenioid fish to the extreme polar climate (Chen *et al.,* 2008).

The number of paralogues (multiple copies per orthogroup) in each orthogroup for each *Eudyptes* penguin genome were evaluated. A total of 3,203 proteins were identified to have paralogous copies in the *Eudyptes* and the yellow-eyed penguins combined, contributing 19.40% of the total pan-genome complement. However, when excluding the outgroup (yellow-eyed penguin), 3,138 proteins were found to incorporate paralogues. Considering only those proteins conserved in all *Eudyptes* penguin genomes (core proteins), a total of 2,242 proteins (Figure 2.4) with paralogues were present (13.58% of the total pan-genome complement). The 3,138 proteins with paralogous copies (both core and accessory) found in the *Eudyptes* penguins have an average copy number of 1.5 copies. In the temperate climate and cold climate clade penguins the average copy number per orthogroup totalled 1.46 and 1.61, respectively. The higher number of relative paralogous copies observed for the cold climate penguins provide a primary indication that gene duplication may serve as a mechanism for coping with cold stress among penguins.

The average copy numbers across each protein orthogroup was calculated for the temperate climate and cold climate penguins, respectively and compared. A small proportion (283 protein orthogroups – 9%) displayed no difference in the average protein copy number. By contrast, 1,382 protein orthogroups (44%) comprised more (average: 1.23x) paralogous copies in the temperate climate clade penguins. The largest proportion of proteins (1,473

protein orthogroups; 47% of total) had more copies per protein in the cold climate penguins (average 1.48-fold as many copies when compared to the temperate climate penguin clade).

Using a two-fold difference in copy number as threshold, nineteen orthogroups were observed to meet this criterion for the temperate climate penguins, while 90 orthogroups had more than two-fold as many paralogues for the cold climate penguins relative to the temperate penguins. Of the 90 proteins over-represented in copy number in the cold clade penguins, fifteen are found in only one of the cold clade penguins (nine in the macaroni penguin and six in the royal penguin, respectively), while the other 75 are found in both cold clade penguins. In 59 of the orthogroups, there are paralogues for only one of the two cold clade penguins, suggesting that gene duplication is frequently specific to a single penguin.

Greater than three-fold difference in the number of copies in the cold-adapted penguins relative to the temperate clade penguins was observed for eighteen of the proteins (Table 2.3). The most noticeable cases represented by over four-fold in the cold-adapted penguins relative to the temperate climate penguins belong to the orthogroups OG0000009, OG0000012, OG0000670, OG0002474 and OG0010318 (highlighted in blue in Table 2.3). The greatest number of protein copies belong to the cold climate penguins i.e. twenty-eight identified in OG0000009 and OG0000012 within the macaroni penguin alone compared to between two to seven and two to six, respectively in the temperate climate penguins.

Table 2.3: Orthogroups that display a greater than three-fold difference in protein copy number in the cold climate penguins relative to the temperate climate penguins.

Orthogroup no.	Copy no.								Average copy no.		Fold difference in copy no. (x)
	E. sclateri	*E. robustus*	*E. pachyrhynchus*	*E. moseleyi*	*E. filholi*	*E. chrysocome*	*E. chrysolophus*	*E. schlegeli*	Temperate climate penguin clade	Cold climate penguin clade	
OG0000009	4	6	7	2	3	7	28	14	4.83	21	4.34
OG0000012	3	6	5	2	3	3	28	6	3.67	17	4.64
OG0000205	2	1	1	2	2	1	8	2	1.50	5	3.33
OG0000231	1	1	2	1	1	1	7	1	1.17	4	3.43
OG0000641	2	1	2	2	1	1	8	1	1.50	4.50	3
OG0000670	1	1	2	2	1	1	10	1	1.33	5.50	4.13
OG0000671	1	1	2	1	1	1	2	5	1.17	3.50	3
OG0000694	1	1	2	1	1	1	1	6	1.17	3.50	3
OG0000735	1	2	1	1	1	1	2	5	1.17	3.50	3
OG0000773	1	1	2	2	1	1	5	3	1.33	4	3
OG0000775	1	1	1	1	1	1	5	2	1	3.50	3.5
OG0000781	1	2	1	1	1	1	2	6	1.17	4	3.43
OG0000803	1	4	1	1	1	1	3	6	1.50	4.50	3
OG0000813	1	1	1	1	2	1	6	2	1.17	4	3.43
OG0002474	1	1	1	1	1	1	1	10	1	5.50	5.5
OG0008631	1	1	1	1	1	1	6	1	1	3.50	3.5
OG0010318	1	1	1	1	1	1	1	9	1	5	5
OG0013323	1	1	1	1	1	1	5	1	1	3	3

2.3.6. Functional analysis of the over-represented paralogues in cold clade penguins

The orthogroups with over-represented copy numbers in the cold climate penguins (90 proteins) and temperate climate penguins (19 proteins) were assigned COG categories using EggNOG-mapper v. 2 (Huerta-Cepas *et al.*, 2017). Overall, 88% of paralogues were categorised into a COG category within the cold climate penguins and 84% were assigned a COG category within the temperate climate penguins (Figure 2.4). The majority 33/90 of the orthogroups with paralogues overrepresented in the cold clade penguins could be assigned to the Cellular Processes & Signalling supra-functional category, while twenty-two, fourteen and ten were assigned to the categories Poorly Characterised, Information Processing & Storage and Metabolism, respectively (Figure 2.5). There are particular COG categories that were assigned to the paralogous proteins only present within the cold climate penguins relative to the temperate climate penguins (Figure 2.6). These include Energy Production & Conversion (C), Nucleotide Transport & Metabolism (F), Carbohydrate Transport & Metabolism (G), Lipid Transport & Metabolism (I), Translation, Ribosomal Structure & Biogenesis (J), Transcription (K) and Defence Mechanisms (V; Figure 2.6).

The 90 (> 2x as many copies in cold climate versus temperate climate penguins) cold clade penguin specific paralogous proteins were further functionally annotated using BlastP (Altschul *et al.*, 1990) analysis against the NCBI non-redundant (nr) protein database (Pruitt, *et al.*, 2005) and the Conserved Domain Database (Lu *et al.*, 2020). Thirteen paralogous protein groups showed evidence for a direct or indirect role in thermoadaptation (Table 2.4).

Table 2.4: Thirteen paralogous proteins over-represented in the cold climate penguin clade that play a role in thermoregulation.

Protein short name	Preferred protein name	COG category	Putative function	Average copy no.	
				Cold climate penguin clade	Temperate climate penguin clade
AdipoR2	Adiponectin receptor protein 2	T	Maintains membrane fluidity in most cell types.	2	1
AMY2A	Pancreatic alpha-amylase	G	Enables the hydrolysis of certain oligosaccharides, similar polysaccharides, and the alpha-(1,4) glycosidic bonds of starch, glycogen, and related polysaccharides.	2.5	1.17
AQP7	Aquaporin 7	G	Facilitates the transport of water, glycerol and urea.	2	1
COX2	Cytochrome c oxidase subunit 2	C	Contributes to the activity of cytochrome-c oxidase. Also involved in the positive regulation of vasoconstriction, cytochrome c to oxygen transport, and mitochondrial electron transport.	2.5	1
DGAT2	Diacylglycerol O-acyltransferase 2	I	Responsible for incorporating endogenously synthesised fatty acids into triglycerides.	2.5	1.17
FAM13A	Family With Sequence Similarity 13 Member A	T	Functions in airway epithelial barrier integrity.	2.5	1
KRT15	Keratin type I cytoskeletal 15	S	A cytokeratin that functions in the structural integrity of epithelial cells.	4	1.83
MYH1	Myosin heavy chain, skeletal muscle, adult	Z	A fundamental contractile protein found in all eukaryote cell types.	5	1

Table 2.4 continued.

Protein short name	Preferred protein name	COG category	Putative function	Average copy no.	
				Cold climate penguin clade	Temperate climate penguin clade
ND6	NADH-ubiquinone oxidoreductase chain 6	C	This protein is part of complex I in the mitochondrial respiratory chain. It functions in the transfer of electrons from NADH to the respiratory chain.	2.5	1
NEB	Nebulin	T	An actin-binding protein that is involved in sarcomere organisation, muscle strength and development.	3	1.17
TUBA3E	Tubulin Alpha-3 Chain	Z	An alpha-tubulin that is one of the main components that make up microtubules which are vital for cell support.	5.5	1
TUBA1C	Tubulin Alpha-1C Chain	Z	An alpha-tubulin that is one of the main components that make up microtubules. Is a structural constituent of the cytoskeleton.	6.5	2.33
-	Feather keratin 2	S	Makes up feathers, hair, claws, scales, horns and hooves.	5	2.17

2.3.6.1 Cold penguin-paralogous proteins that play a role in the brain and thermoregulation

The preoptic area (POA) of the anterior hypothalamus is the region of the brain that controls thermoregulation. AdipoR2 is an adiponectin receptor that has been found to be expressed in warm sensitive neurons and in temperature insensitive neurons of the POA as well as other brain regions involved in nutrient homeostasis (Klein *et al.,* 2011). Through adiponectin, AdipoR2 allows for the oxidation of triglycerides (fatty acids) over carbohydrates. As a result, it functions in energy metabolism, an important component of thermoregulation (Klein *et al.,* 2011). Warm sensitive neurons (where AdipoR2 is detected) can activate both NST and ST through BAT (beige adipose tissue in birds) and vasoconstriction. These neurons inhibit heat production and excite heat loss effector neurons, while temperature insensitive neurons do the opposite (Klein *et al.,* 2011).

2.3.6.2 Cold penguin-paralogous proteins that function in non-shivering thermogenesis through mitochondrial proton uncoupling

Proteins that take part in NST through mitochondrial proton uncoupling include, a core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I), NADH ubiquinone oxidoreductase core subunit 6 (ND6). This protein catalyses the transport of electrons from NADH via the respiratory chain using ubiquinone as an electron acceptor (Vries et al., 1996). This is critical for complex I assembly as well as its catalytic activity (Vries et al., 1996). Cytochrome c oxidase includes the enzyme COX2 that is encoded by the mitochondria. This enzyme powers oxidative phosphorylation as the final component of the mitochondrial electron transport chain (Cerqua et al., 2018). The respiratory chain's cytochrome c oxidase enzyme catalyses the reduction of oxygen to water (Cerqua et al., 2018). Both ND6 and COX2 have a greater average copy number in the cold climate penguins relative to the temperate climate penguins (2.5 and 1, respectively).

2.3.6.3 Cold penguin-paralogous proteins that operate in glucose and lipid metabolism

Amylase alpha 2a (AMY2A) catalyses the first step in the digestion of dietary starch and glycogen (Kaczmarek and Rosenmund, 1977) and may therefore play a role in glucose homeostasis and energy homeostasis. AQP7 is an aquaglyceroporin that can be found in adipocytes (Upadhyay *et al.,* 2021). It transports water as well as glycerol (an energy storage molecule). AQP7 plays roles in thermoregulation through sweating and energy mobilisation from fat to muscle (glycerol transport; Upadhyay *et al.,* 2021). Glycerol is a key link between glucose and lipid metabolism (Blötz and Stülke, 2017). Diacylglycerol o-acyltransferase 2

(DGAT2) is an enzyme that is highly expressed in adipose tissues and plays a role in adipocyte differentiation (Gesta and Kahn, 2012). DGAT2 is required for the synthesis and storage of intracellular triglycerides (Gesta and Kahn, 2012) and thus plays a role in energy homeostasis and lipid metabolism/homeostasis. FAM13A plays a role in adipocyte differentiation and function as it regulates fat distribution and metabolic traits through its action on adipose tissue (Fathzadeh *et al.,* 2020).

2.3.6.4 Cold penguin-paralogous proteins with a role in skeletal muscle and shivering thermogenesis

Nebulin (NEB) is required for the normal functioning of skeletal muscle (Chu *et al.,* 2016). Myofibrils make up each skeletal muscle fibre and myofibrils consist of actin and myosin filaments. NEB stabilises actin filaments and therefore regulates filament length (Chu *et al.,* 2016). Tubulin alpha 1c (TUBA1C) and tubulin alpha 3 chain (TUBA3E) are alpha tubulins. These alpha tubulins, in conjunction with beta tubulin, form heterodimers that make up microtubules which eventually result in skeletal muscle fibres. MYH1 is a myosin protein that is found in skeletal muscle (and cardiac muscle) and plays a role in muscle contraction which helps with ST. The proteins NEB, TUBA1C, TUBA3E and MYH1 all have a greater average copy number in the cold-adapted penguins in comparison to the temperate climate penguins with the most copies being present in TUBA1C (average copy number of 6.5 in cold climate penguins).

2.3.6.5 Cold penguin-paralogous proteins implicated in feather development

The structural diversity of feathers can be linked to variations in the number of alpha- and beta-keratins (Hughes *et al.,* 2011). KRT15 is an alpha cytokeratin that has been found to be up-regulated in bird body feathers (Hughes *et al.,* 2011). These feathers help birds retain heat when exposed to low temperatures and dissipate heat when exposed to high temperatures. Feather keratin 2 has not been fully functionally annotated, however, due to its sequence similarity to other keratins it can be determined to serve some type of function in feathers (UniProt, 2002). Feather keratin 2 and KRT15 have more copies present in the cold climate penguins relative to the temperate climate penguins (average copy number of 5 and 4, respectively in the cold-adapted penguins compared to 2.17 and 1.83 in the temperate climate penguins).

2.4 Conclusion

The genus *Eudyptes* is comprised of eight species which form two distinct clades with regards to climate. The cold-adapted penguin clade includes the macaroni and royal penguins, while the temperate climate penguin clade is comprised of the southern rockhopper, eastern rockhopper, northern rockhopper, Fiordland, snares and erect-crested penguins. The pan-genome complement of the eight *Eudyptes* penguins and the outgroup was analysed. The pan-genome comprises 16,511 orthogroups, with 59.9% of the proteins conserved in all penguins (core proteins), whereas ~1.3% and 11% of the proteins were unique to the cold climate and temperate penguins, respectively. Nearly 20% of the pan-genome orthogroup complement comprised of proteins with at least one paralogous copy. Thirty-five unique proteins and thirteen proteins with paralogues over-represented by two-fold in the cold climate penguins (relative to the temperate climate penguins) were found to play a direct or indirect role in thermoregulation within the cold-adapted *Eudyptes* penguins. These proteins played roles in mechanisms such as glucose metabolism, adipocyte differentiation, vascular development, feather development, skeletal muscle functioning and lipid metabolism. All these mechanisms have been shown to help in adapting to the cold. For example, lipid metabolism could be implicated in blubber formation which provides good insulation against the cold, especially in water. As a result, blubber is essential for a penguins' survival in environments that exhibit low temperatures.

Chapter 3

Cold adaptation in the genus *Eudyptes* at the gene and protein level

3.1. Introduction

Cold adaptation can be defined as all aspects of an organism's physiology that allow it to inhabit regions that experience a cold climate i.e. anything that ranges from moderate to extreme cold (Clarke, 1991). Adaptation to cold may occur through acclimation or acclimatisation and includes behavioural, genetic, morphological and physiological responses (Makinen, 2010). Cold adaptation is distinctive because there are particular adaptations that are extremely specific to low temperatures. Inhabiting regions with temperate climate appears to be easier to accomplish than to inhabit environments with polar climate as there are physiological challenges that are difficult to overcome in cold environments such as the need to withstand freezing (Clarke, 1991). It is vital to understand the mechanisms of how organisms manage to perform certain functions at extremely cold temperatures. It is also important to see how effectively these organisms cope with the limitations imposed by temperature on enzymatic reactions through hormonal and nervous control to digestion, from respiration and osmoregulation to all aspects of an organism's performance and behaviour, especially when compared to their counterparts that inhabit environments with warmer temperatures (Clarke, 1991). As a result, it will be possible to see how climate change will affect certain species and populations in the years to come. For example, teleost (ray-finned) fish have tissue fluids that are hypoosmotic to seawater and would freeze in seawater below approximately -0.7 °C, however, their polar counterparts have been able to partly avoid this by producing antifreeze proteins/glycoproteins (Clarke, 1991).

Exposure to extremely cold temperatures is a fundamental problem in temperature biology as it can be lethal not only at the cellular level, but to organisms as a whole. Cold temperatures severely impact cell integrity, enzyme kinetics, macromolecular interactions, membrane fluidity, solute diffusion rates and water viscosity by putting enormous pressure on normal cell function (De Maayer *et al.*, 2014; Giordano *et al.*, 2020; Raymond-Bouchard *et al.*, 2018). In order to improve functionality under extremely cold temperatures these biological systems and mechanisms need to be modified to be able to tolerate this stress (Zhang *et al.*, 2018).

Biological molecules such as DNA and proteins need to be modified in order to properly function under extreme temperatures (Zhang *et al.*, 2018). DNA can be altered to function under extreme temperature by the modification of GC content, whereas proteins are altered via a change in amino acid sequence. Although the association between genomic GC content and temperature adaptation has been investigated for many years, there is still disagreement. There have been a number of theories offered to explain how GC content might influence environmental adaptability. However, none of these theories address whether temperature could shape the GC content in DNA and amino acid preferences of proteins (Galtier *et al.*, 2001; Vinogradove, 2003; Wada and Suyama, 1986; Zhang *et al.*, 2018).

Protein building blocks constitute vital structures within living cells. These building blocks include enzymes that allow for the modulation of cellular structures and responses by controlling biochemical and biological reactions (Tang *et al.*, 2019; Yusof *et al.*, 2021). Based on this, proteins are placed into two categories i.e. they are either active or structural proteins. Enzymes are biological catalysts that control hundreds upon thousands of chemical reactions within cells and are thus classified as active proteins (Tang *et al.*, 2019; Yusof *et al.*, 2021). Structural proteins would include keratin and collagen as they are proteins that mainly contribute to the structural properties of cells (Tang *et al.*, 2019; Yusof *et al.*, 2021).

Protein adaptations are shaped by drivers such as temperature. According to studies on protein function it has been shown that higher catalytic efficiency (kcat /KM) is displayed by the majority of cold-adapted proteins, however this is at the expense of thermal stability when compared to proteins that are adapted for a temperate climate (Lee *et al.*, 2017; Yusof *et al.*, 2021). This lower stability has been linked to greater conformational flexibility in protein architecture in either a part of the structure that plays a vital role in protein conformational changes or areas around the active site (Lee *et al.*, 2017; Yusof *et al.*, 2021). This was shown in comparative studies done on microtubule (Detrich, 1997), malate dehydrogenase (Kim *et al.*, 1999), RTX lipase (Mohamad Ali *et al.*, 2013) and phosphoglycerate kinase (Bentahir *et al.*, 2000) proteins from polar species that demonstrated higher structural flexibility. Amino acid substitutions are also a key factor in protein flexibility, especially where glycines are substituted at the loop regions of proteins (Yusof *et al.*, 2021). In addition, amino acids with small/tiny or neutral side chains (serine, threonine, aspartic acid and alanine) were seen to be preferred in psychrophilic prokaryotes. This could be a result of the amino acids contributing to higher protein flexibility by having more coils and less helices in the secondary structure (Metpally and Reddy, 2009). Additionally, protein structure stabilising interactions such as

salt bridges, hydrogen bonds, ionic and aromatic interactions in intra- or inter-protein interactions, are shown to decrease in cold-adapted proteins (Yusof *et al.*, 2021). Genomic studies on amino acid bias in cold adaptation in vertebrates are still scarce (Zhang *et al.*, 2018). However, as more genomic data becomes available it will be possible to investigate the effects of cold exposure on codon and amino acid evolution in more depth. In Chapter 2 we analysed thermoadaptation at the whole genome level within the *Eudyptes* penguins. In this chapter we investigate thermal adaptation in the penguins at the level of the individual proteins and genes that code for them.

3.2. Methods and materials

3.2.1. Evolutionary analysis of molecular determinants of thermoadaptation

Single-copy orthologues (SCOs) from the protein datasets encoded on each *Eudyptes* genome and the outgroup were identified using OrthoFinder v. 2.2.7 (Emms and Kelly, 2015) in Chapter 2. The amino acid sequences of these proteins were individually aligned using the m-Coffee implementation of T-coffee (Notredame, *et al.*, 2000; Wallace *et al.*, 2006). The SCO amino acid sequence alignments were then compared using the BaCoCa v. 1.1 (Base Composition Calculator) software package (Kück and Struck, 2013). BaCoCa was used for multiple statistical analyses such as the chi-square test of homogeneity and the compositional heterogeneity and bias value (RCFV) to identify those proteins with anomalous base compositions between the cold and temperate climate clade *Eudyptes* penguins along with the yellow-eyed penguin (outgroup). These SCOs were subsequently functionally annotated by BlastP (Altschul *et al.*, 1990) analysis against the NCBI non-redundant (nr) protein database (Pruitt, *et al.*, 2005) and the Conserved Domain Database (Lu *et al.*, 2020). Proteins with a putative role in thermoregulation and thermoadaptation as identified in Chapter 2 were further investigated using BaCoCa v. 1.1 (Kück and Struck, 2013). This was done in order to see if there were certain biases towards specific amino acids within these proteins, as a common strategy of protein adaptation from temperate to cold environments (or vice versa) is for modest changes in amino acid composition to occur without changing the overall structure of the macromolecule.

The candidate thermoadaptation proteins chosen from the output above were tested for natural selection i.e. for positive, neutral or purifying selection. Firstly, the protein sequences were compared to their matching genomic DNA sequences using a pairwise sequence alignment tool, GeneWise v. 2.4.1 (Birney *et al.*, 2004), which considers sequencing errors. The protein sequences and corresponding DNA sequences were uploaded on the PAL2NAL

v. 14 web server (Suyama *et al.,* 2006) using the universal code codon table to remove in-frame stop codons. The Selection Server v. 2.4 (Doron-Faigenboim *et al.,* 2005) was used to identify positive, neutral and purifying selection using a Bayesian inference approach by aligning the DNA coding-sequences using MAFFT-L-INS-i (Katoh *et al.,* 2005) and using the Mechanistic Empirical Combination (MEC) evolutionary model which uses the JTT matrix (Goldman and Yang, 1994; Doron-Faigenboim and Pupko, 2007). The type and strength of evolutionary selection is determined by the Ka/Ks ratio (or $\omega = N/S$) with $\omega > 1$ indicating positive selection, $\omega = 1$ indicating neutral evolution, and $\omega < 1$ indicating negative (purifying) selection. If positive selection was detected, a likelihood ratio test was performed using the M8a null evolutionary model (M8a, beta + w = 1; Swanson *et al.,* 2003) to verify if the positive selection was significant.

3.3. Results and discussion

3.3.1. Amino acid bias in individual proteins conserved among *Eudyptes* penguins

A total of 7,644 SCOs were identified to be conserved among the cold climate and temperate climate *Eudyptes* penguins as well as the outgroup (the yellow-eyed penguin) using OrthoFinder v. 2.2.7 (Emms and Kelly, 2015) These SCOs make up 46.30% of the total *Eudyptes* pan-genome complement. BaCoCa v. 1.1 (Kück and Struck, 2013) was used to analyse the 7,644 SCOs to identify biases in the aligned aa sequence data. Average RCFV and chi-squared values of 0.0135 $\pm$ 0,0223 and 4.7016 $\pm$ 24.5267 (average $\pm$ standard deviation) were determined, respectively. Proteins with RCFV and chi-squared values above and below the average $\pm$ standard deviation values were further analysed along with those with p-values < 0.05. A total of 847 SCOs that met at least one of the cut-off criteria were determined. Only nine proteins met all the cut-off criteria, while the proteins that met the cut-off criteria of the RCFV and chi-square values totalled 186. Further, 644 SCOs solely satisfied the RCFV cut-off criterion and eight met only the chi-square criterion.

3.3.2. Functional annotation of single-copy orthologues and their role in thermoregulation and adaptation

The 847 SCOs that met at least one of the cut-off criteria were functionally annotated. Fourteen were found to play a potential role in thermoregulation and/or thermoadaptation (Table 3.1). None of these fourteen proteins met all the cut-off criteria. Three proteins met both the RCFV and chi-square cut-off criteria, while eleven SCOs met at least one of the cut-off criteria (Table 3.1).

Table 3.1: Single-copy orthologs identified in the genus *Eudyptes* using the RCFV, chi-square and p-values that may play a potential role in thermoregulation.

Protein short name	Protein preferred name	Predicted function	RCFV value	Chi-square value	p-value	Selection statistic (overall)	Type of selection (overall)	Positive selection significance
BHE40	Class E basic helix-loop-helix protein 40	Transcriptional repressor that regulates circadian rhythm.	0.0441	11.1504	1	0.16	Negative	-
CISD3	CDGSH iron-sulphur domain-containing protein 3	Needed for the normal functioning of the mitochondria.	0.1579	48.4728	1	0.24	Negative	-
DHODH	Dihydroorotate dehydrogenase	This enzyme plays a role in the production of pyrimidines, which are building blocks for various biological molecules such as DNA and RNA.	0.0455	11.9462	1	0.75	Negative	Non-significant
DLL4	Delta-like protein 4	Part of NOTCH signalling pathway needed for normal development of many tissues.	0.0444	18.8753	1	1.76	Positive	Significant
ECH	Enoyl-CoA hydratase	Functions part of lipid transport and metabolism.	0.0432	6.4792	1	0.14	Negative	-
FADS1	Fatty acid desaturase 1	Plays a role in lipid transport and metabolism.	0.0416	8.5865	1	3.90	Positive	Significant
GADD45b	Growth arrest and DNA damage inducible protein 45 beta	Plays a role in the site-specific modification of RNA, ribosome structure and spliceosome assembly.	0.0543	10.1385	1	0.16	Negative	-

Table 3.1 continued.

Protein short name	Protein preferred name	Predicted function	RCFV value	Chi-square value	p-value	Selection statistic (overall)	Type of selection (overall)	Positive selection significance
IL6RA	Interleukin-6 receptor subunit alpha	Multi-functional cytokine that regulates inflammation, hemopoiesis, the immune response and acute phase response.	0.0361	2.5786	1	0.19	Negative	Non-significant
IRAK2	Interleukin-1 receptor-associated kinase-like 2	Functions in innate immunity and inflammation in vertebrates; signal transduction mechanisms.	0.0702	39.4400	1	0.53	Negative	Non-significant
PRL	Prolactin	Plays roles in multiple processes including growth control.	0.0703	11.0344	1	0.64	Negative	Non-significant
SLC30A3	Zinc transporter 3	Functions in the regulation of sequestering zinc ions and enables zinc ion transmembrane transporter activity.	0.0470	10.0226	1	0.20	Negative	Non-significant
SOCS7	Suppressor of cytokine signalling 7	Critical for the functioning of neuronal cells.	0.0395	10.4603	1	0.22	Negative	Non-significant
TBK1	TANK-binding kinase 1-binding protein 1	Plays a role in cell cycle control, cell division and chromosome partitioning.	0.1015	37.1974	1	0.36	Negative	-
-	Feather keratin 4	Avian keratin protein that makes up feathers, claws and scales	0.0385	2.2458	1	0.34	Negative	-

Note: The values highlighted in blue met the cut-off criteria of > average ± standard deviation for the RCFV, chi-square and p-values, for each protein that is linked to thermoregulation.

3.3.2.1 Single-copy orthologous proteins that function in vascular development

DLL4 regulates angiogenesis through the DLL4-Notch signalling pathway. This pathway prevents uncontrolled vascular sprouting as well as undirected vascular growth (Cao, 2013). This could play a potential role in developing the counter-current circulatory system that is found in penguins. Vascular development is also important for adipose tissues. The adipose vasculature allows for the modulation of many adipocyte functions such as removing metabolic products from adipose tissue, providing oxygen and essential nutrients to maintain adipocyte survival and transporting adipokines, cytokines, growth factors and hormones that regulate adipocyte growth and function (Cao, 2013). The adipose vasculature aids in the maintenance and survival of adipose tissue and furthermore, allows it to perform its functions that relate to thermoadaptation such as energy regulation and lipid metabolism.

3.3.2.2 Single-copy orthologous hormones implicated in thermoregulation

Prolactin (PRL) is a hormone that is multi-functional and plays roles in various biological mechanisms (Alamer, 2011). PRL is involved in acclimatory responses to high temperatures. During heat exposure PRL has been shown to affect body fluid regulation by controlling extracellular fluid volume which helps in heat dissipation (Alamer, 2011). SOCS7 represses PRL, growth hormone and leptin signalling through the interaction with STAT3/5 (Martens *et al.*, 2005). PRL plays a role in adapting to high temperatures, whereas leptin prevents the fall of the core body temperature when exposed to cold temperatures (Deem *et al.*, 2018). SOCS7 therefore has a negative effect on thermoregulation through the repression of key thermogenic hormones.

3.3.2.3 Single-copy orthologous proteins that are involved in mitochondrial thermogenesis

Mitochondria generate most of the heat in endotherms and is an organelle that is critical for thermogenesis and, concomitantly, for thermoregulation, especially in endotherms that are exposed to low temperatures. This heat generation is achieved through processes such as electron transport, oxidative phosphorylation and fatty acid oxidation. CISD3 is part of the CDGSH domain-containing family and may play a role in mediating electron transport and oxidative phosphorylation in the mitochondria (Wiley *et al.*, 2007). It is further needed for normal functioning of the mitochondria, which is critical for thermoregulatory mechanisms like NST. Interleukin 1 Receptor Associated Kinase 2 (IRAK2) is an adaptor molecule that takes part in a signalling axis, which directly impacts adipocyte metabolism (Zhou *et al.*, 2020). Interleukin-1 (IL-1) stimulation activates the translocation of the IRAK2 Myddosome

(oligomeric complex that is needed to transmit inflammatory signals) to the mitochondrial outer membrane. IRAK2 then interacts with two mitochondrial proteins that results in the suppression of oxidative phosphorylation and fatty acid oxidation, thereby having a negative effect on thermoregulation (Zhou *et al.,* 2020).

Enoyl-CoA hydratase (ECH) catalyses the second step in the physiologically important beta-oxidation pathway of fatty acid metabolism which takes place in the mitochondria (Agnihotri and Liu, 2003). Beta-oxidation allows for fatty-acids to be broken down to produce energy and heat and thus contributes to thermoregulation. BHE40 is involved in the regulation of mitochondrial and peroxisomal biogenesis and function, which play vital roles in energy supply and adaptation (Chang *et al.,* 2019). BHE40 also plays a role in skeletal muscle protection (Chang *et al.,* 2019). Skeletal muscle relies on *PGC-1α* (transcriptional coactivator) to aid in metabolic thermogenesis adaptation, fatty acid oxidation (adapting to high energy demands), mitochondria biogenesis and oxidative metabolism. Adipocyte differentiation is modulated by BHE40 which represses PPARγ2, a key adipogenic factor i.e. BHE40 negatively regulates fatty oxidation in adipocytes (Chang *et al.,* 2019). DHODH (mitochondrial, respiratory chain-coupled enzyme) is a protein found on the outer surface of the mitochondrial inner membrane that links the biosynthesis of pyrimidine to mitochondrial energy metabolism (Zhang *et al.,* 2021). This enzyme catalyses the fourth step in pyrimidine biosynthesis to generate orotate. Orotate accumulation results in intracellular lipid accumulation and affects lipid metabolism, which will impact energy homeostasis (Zhang *et al.,* 2021). DHODH may also play a potential role (through orotate) in blubber formation, which provides penguins with good insulation when exposed to low temperatures.

3.3.2.4 Single-copy orthologous proteins that play a role in lipid and glucose metabolism

FADS1 codes for the rate-limiting delta 5 desaturase enzyme that functions in the long chain polyunsaturated fatty acids (LC-PUFAs) biosynthesis pathway (Reynolds *et al.*, 2020). This pathway is linked to complex lipids that play roles in energy homeostasis, innate immunity, neurocognitive function and brain development. An endocannabinoid, 2-arachidonyl glycerol (2-AG), that is regulated in this pathway (and by extension *FADS1*) is needed in almost all physiological processes including thermoregulation and energy metabolism and homeostasis (Reynolds *et a.,* 2020). TBK1 plays a critical role in both inflammation and metabolism in adipose tissue. Through its metabolic function this protein affects energy and glucose

metabolism. This is achieved by TBK1 inhibition of AMP-activated protein kinase (AMPK) (Zhao *et al.,* 2018). Due to its role in glucose metabolism, TBK1 may play a role in NST.

GADD45b functions as a coactivator for constitutive androstane receptor (CAR). CAR plays a role in energy metabolism and the activation of CAR has been shown to improve insulin sensitivity as well as inhibit gluconeogenesis (Cai *et al.,* 2021). The inhibition of gluconeogenesis is due to the degradation (by CAR) of an important coactivator that controls glucose metabolism (peroxisome proliferator-activated receptor gamma coactivator 1-alpha; PGC1-a). GADD45b is in part needed for CAR-mediated improvement of insulin sensitivity and the inhibition of gluconeogenesis and lipogenesis (Cai *et al.,* 2021). In pancreatic beta-cells zinc transporter proteins respond to variations in zinc as well as glucose levels. SLC30A3 is a zinc transport protein linked to insulin secretion and furthermore to glucose metabolism (Schumann *et al.,* 2020). IL6RA is an exclusive IL6 receptor with ligand recognition, but no signalling properties and is only expressed in a few cell types (García *et al.,* 2018). The IL6/IL6RA complex takes part in classical signalling when bound to membrane anchored gp130 through downstream phosphorylation of the JAK/STAT pathway, ERK and PI3 kinase. IL6 classical signalling mediates insulin sensitivity through facilitating glucose disposal in adipose tissue and muscle (García *et al.,* 2018). Another function of IL6 binding to IL6RA allows for STAT3-induced hypothalamic expression of cyclooxygenase 2 (COX2) in brain endothelial cells which results in an increase in body temperature (García *et al.,* 2018).

3.3.2.5 Single-copy orthologous proteins with a role in feather development

Feathers can be linked to thermoregulation through heat retention and heat loss. The structural diversity of feathers is due to different combinations of both alpha and beta keratins. Feather keratin 4 may play a role in the structural organisation of feathers, however, no studies have been done on this specific keratin as of June 2022.

3.3.3. Natural selection analysis of the thermoregulatory single-copy orthologues

The fourteen SCOs which were found to have anomalous amino acid compositions between the cold and temperate clade penguins were further subjected to selection analysis at both the nucleotide and amino acid level. Of these, six proteins were found to be under the overall influence of purifying selection (Table 3.1). Another six SCOs were also under overall negative selection. However, there were individual amino acid sites that displayed positive selection, although these sites were found to be non-significant (Table 3,1) on the basis of a

likelihood ratio test that used the M8a null evolutionary model (M8a, beta + w = 1; Swanson *et al.*, 2003) to verify if the positive selection was significant in the overall individual proteins. When comparing amino acid selection of all the *Eudyptes* penguins (temperate climate and cold climate clade penguins combined), two proteins i.e. DLL4 and FADS1 were found to be under overall positive selection i.e. site-specific positive selection was found to be significant (Table 3.1).

Feather keratin 4 is under overall negative selection with no site-specific positive selection, but when only looking at the temperate climate clade penguins, some amino acids start displaying positive selection (Figure 3.1). IL6RA displays the opposite to feather keratin 4 in which site-specific positive selection influences a few amino acid sites in *Eudyptes* overall, but in the temperate climate clade only the entire protein is under purifying selection (Figure 3.2). TBK1 is under negative selection overall in the *Eudyptes* penguins, however, in the temperate climate penguin clade the protein is under overall positive selection (Figure 3.3). As such, distinct selection patterns are evident in proteins from the cold clade and temperate clade penguins which may indicate their roles in thermoadaptation.

Figure 3.1: Site-specific selection of the SCO feather keratin 4 in A) both the temperate and cold climate *Eudyptes* clade and, B) the temperate climate penguin clade only.

Figure 3.2: Site-specific selection of the SCO IL6RA in A) both the temperate and cold climate *Eudyptes* clade and, B) the temperate climate penguin clade only.

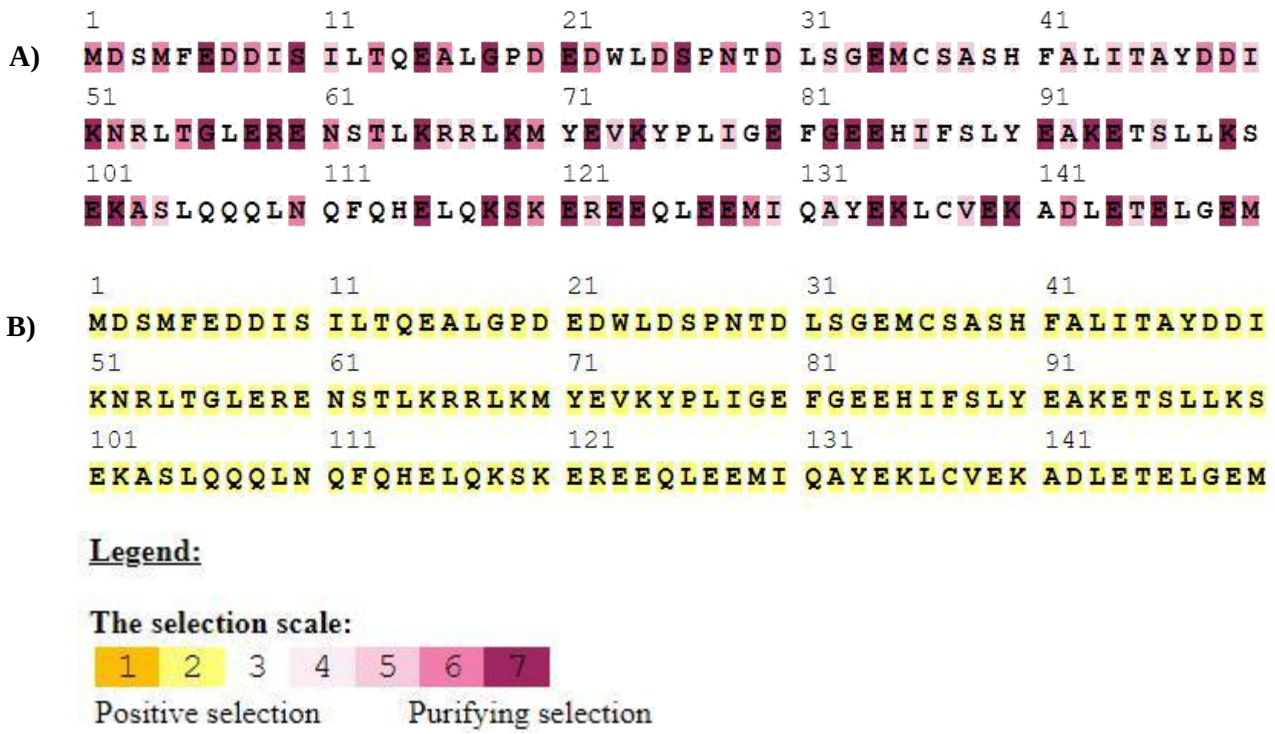


Figure 3.3: Site-specific selection of the SCO TBK1 in A) both the temperate and cold climate *Eudyptes* clade and, B) the temperate climate penguin clade only.

3.3.4. Amino acid composition analysis of the *Eudyptes* penguins single-copy orthologues

Temperature has been shown to be one of the biggest drivers in determining protein adaptations (Giordano *et al.*, 2020). In cold adapted organisms, proteins have been found to be less stable however, to make up for this short coming, they have been shown to exhibit greater conformational flexibility as well as higher specific activity when compared to high-temperature adapted proteins (De Maayer *et al.*, 2014; Giordano *et al.*, 2020). The choice of amino acids, according to their usage in nature, is essential in the evolutionary adaptation to temperature (Giordano *et al.*, 2020).

The amino acid (aa) composition of all 7,644 SCOs was analysed. The global aa composition between the cold-adapted penguins and the temperate climate clade penguins does not show any bias towards specific amino acids (Figure 3.4). However, when looking at the amino acid composition of the fourteen individual proteins that have a potential link to thermoadaptation, the cold-adapted proteins showed an increased compositional bias towards aspartic acid (D), cysteine (C) isoleucine (I), leucine (L), methionine (M), serine (S) and valine (V) (Figure 3.5). Mechanisms used to increase protein flexibility includes increased asparagine, methionine and glycine amino acids (De Maayer *et al.*, 2014). The cold climate penguins had 0.12%

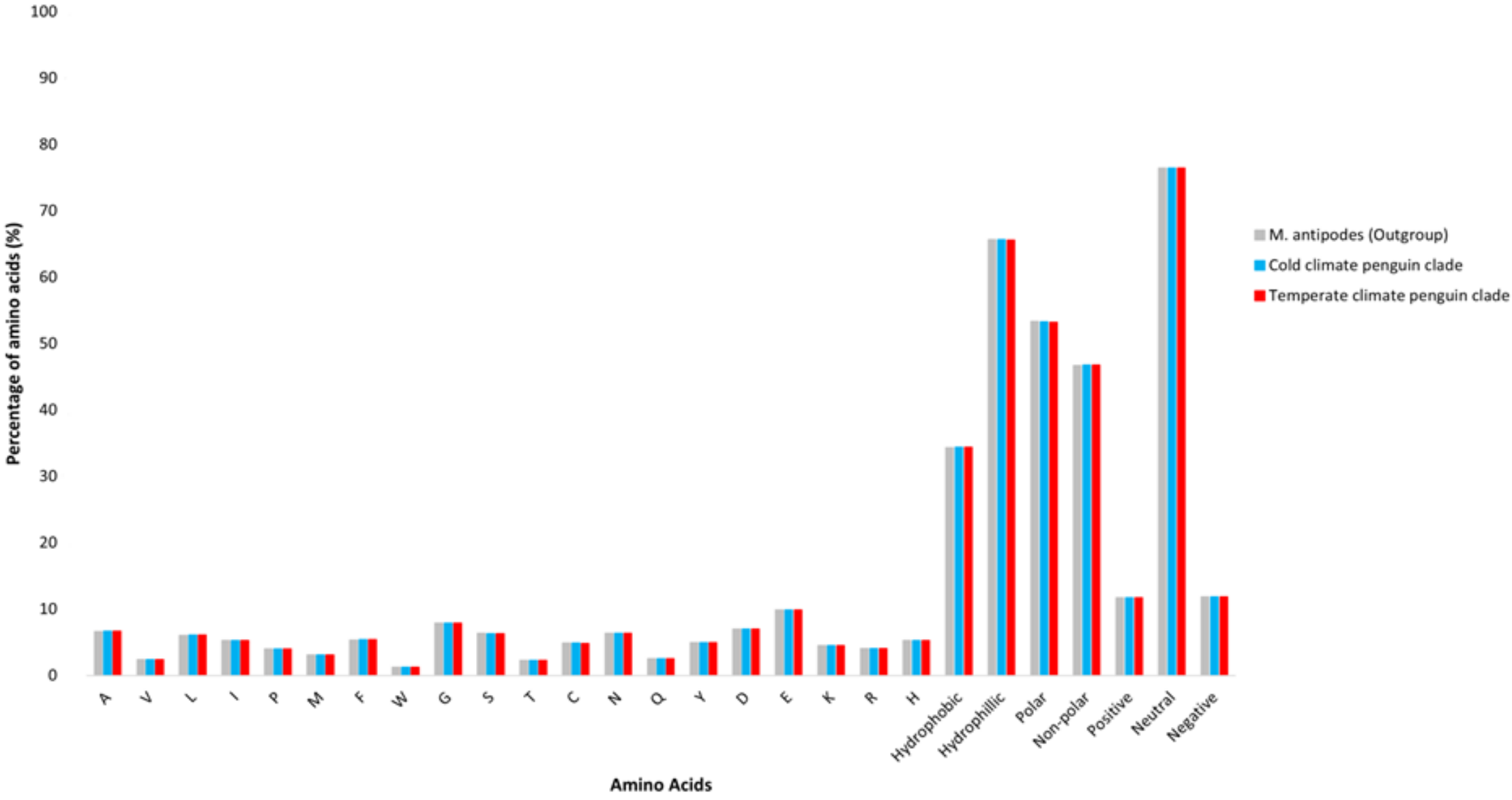


Figure 3.4: Global amino acid composition of SCOs among the cold climate and temperate climate *Eudyptes* penguin clades. Amino acid key**:** alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), methionine (M), phenylalanine (F), tryptophan (W), glycine (G), serine (S), threonine (T), cysteine (C), asparagine (N), glutamine (Q), tyrosine (Y), aspartic acid (D), glutamic acid (E), lysine (K), arginine (R) and histidine (H).

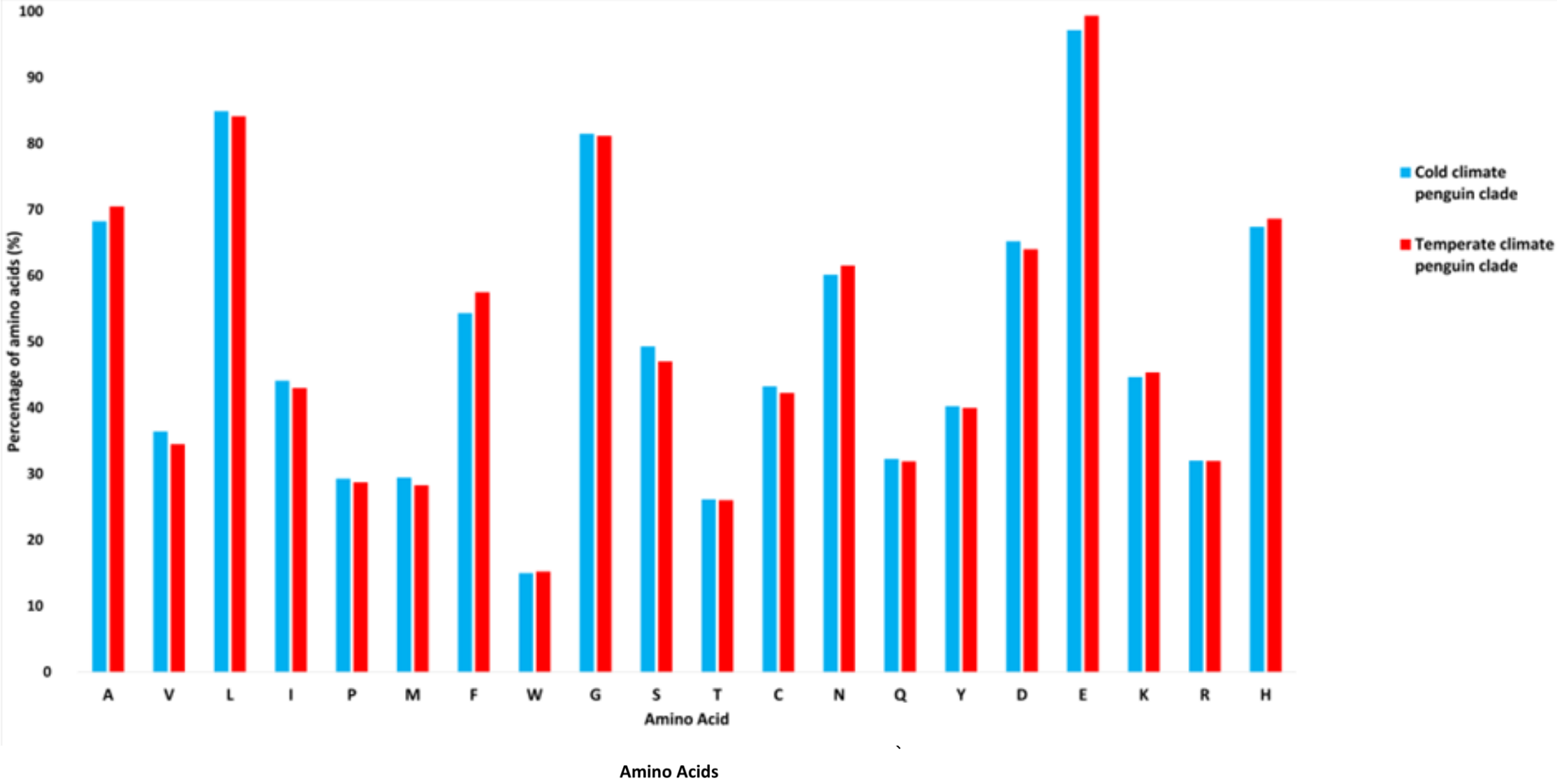


Figure 3.5: Amino acid composition of fourteen SCOs that play a potential role in thermoregulation among the cold climate and temperate climate *Eudyptes* penguin clades.

more methionines in their proteins than the proteins in the temperate climate penguins. Methionine is a non-polar, hydrophobic amino acid with an unbranched side chain which provides ample flexibility (Aledo, 2019). It has been predicted to amplify protein folding as well as reduce cold denaturation of the protein in cold-adapted organisms (Berthelot *et al.*, 2018). Serine and aspartic acid were found to be more common (by 0.23% and 0.12%, respectively) in the cold-adapted proteins relative to their temperate counterparts. Serine, as a polar uncharged amino acid, may compensate partly for the general reduction in charged residues in cold-adapted proteins (Raymond-Bouchard *et al.*, 2018). It has been shown to be over-represented in the coil regions of protein secondary structures (along with aspartic acid) which leads to an increase in flexibility of the proteins (Metpally and Reddy, 2009). Aspartic acid is hydrophilic and accumulation of this amino acid may lead to binding more water molecules, increasing the internal solvation of proteins. This will therefore result in increased protein conformational flexibility (De Maayer *et al.*, 2014). These cold-adapted proteins have also displayed a decrease in the amino acids alanine (A: -0.23%), asparagine (N: -0.14%), glutamic acid (E: -0.22%), histidine (H: -0.12%), lysine (K: -0.07%) and phenylalanine (F; -0.31%) relative to temperate climate penguin clade proteins (Figure 14). Histidine (0.12% less common in cold-adapted proteins when compared to their temperate counterparts) is an amino acid that is localised in active sites in approximately 50% of all enzymes and this contributes to the stability of a protein rather than flexibility (Hansen and Kay, 2014).

3.4. Conclusion

SCOs identified in the genus *Eudyptes* and the outgroup contribute to approximately 46% of the total pan-genome complement. The SCOs that played a role in thermoregulation and thermoadaptation in both the cold climate and temperate climate penguin clades had functions in a wide range of processes many of which involve the mitochondria, glucose metabolism, lipid metabolism and vascular development. All these mechanisms have been shown to be critical in adapting to life in the cold and are imperative for the survival of cold-adapted organisms like penguins. Selection has been shown to favour adaptations that allow for an organism to survive and reproduce. Any unfavourable adaptation that does more harm than good to an organism eventually comes under a deleterious effect, while an adaptation that grants an advantage to the organism comes under a positive effect (Bell, 2008). Most of the thermoregulatory proteins that have been discussed above are under purifying selection, however there are a few under positive selection. One such protein, FADS1 can be linked back to blubber formation that helps in providing insulation against cold exposure. The

protein DLL4 that helps in vascular development can be linked back to the formation of the counter-current circulatory system which enables the conservation or releasing of heat to keep internal temperature constant, is also under positive selection. These proteins aid in the survival of penguins in both temperate and cold climates. They are therefore advantageous and will help populations avoid extinction in the coming years by helping penguins tolerate various climatic shifts as fitness for many species will become increasingly linked to variation in traits vital for performance, especially in colder climates. Cold-adapted proteins tend to favour conformational flexibility over stability which is brought about through the increase of small, neutral amino acids such as serine and a decrease in the amino acid histidine. These are not set rules, however, as these amino acid cold adaptive strategies may vary depending on taxa and various factors they experience within these extreme environments (Raymond-Bouchard *et al.,* 2018). Furthermore, these nuances are only visible in the fourteen cold-adapted SCOs present within the cold climate penguin clade.

Summary

Penguins are some of the most well adapted vertebrates with regards to cold. They have managed to achieve this through maintaining thermal balance by reducing the amount of heat lost to the environment and by increasing heat production within the body through the use of various behavioural, physiological and molecular mechanisms. The genus *Eudyptes* is the most specious group within the family Spheniscidae and these species inhabit a range of different thermal environments ideal for studying the molecular mechanisms that drive cold-adaptation. This study aimed to investigate the molecular determinants underlying thermoregulatory adaptation and their evolution among *Eudyptes* penguin species from different thermal environments. A phylogenetic assessment was undertaken using two approaches. Orthologous and paralogous proteins were identified, annotated and classified to their function and gene ontology using OrthoFinder and EggNOG-mapper. Amino acid analysis and selection analysis were then undertaken on proteins that were found to have a putative role in thermoregulation using BaCoCa and the Selection Server, respectively. The penguins that inhabited sub-Antarctic environments were shown to form a clade in both core genome ML phylogeny and a house-keeping gene ML phylogeny that consisted of the macaroni and royal penguins. A clade was also formed with the penguins that occupy temperate environments i.e. the temperate climate penguin clade, composed of the southern rockhopper, eastern rockhopper, northern rockhopper, Fiordland, snares and erect-crested penguins. This temperate climate clade was further divided into two sub-clades.

When the *Eudyptes* pan-genome was compared with the yellow-eyed penguin (outgroup), 16,511 orthogroups were identified and of these approximately 60% were shared among the nine compared taxa (core proteins). The unique proteins in the cold climate penguin clade only made up 1.28% of the total pan-genome complement compared to the 10.9% that were found in the temperate climate penguin clade. Thirty-five of these unique proteins found in the cold climate *Eudyptes* penguins play potential roles in various thermoadaptation and thermoregulation strategies. Paralogous proteins were then investigated among the temperate climate and cold climate penguin clades to discover whether they play a potential role in cold-adaptation as gene duplication may serve as a mechanism for tolerating cold stress, especially in penguins. Approximately 19.4% of the total pan-genome protein complement comprised paralogous copies, of which 90 were found to be over-represented in the cold clade penguins by > two-fold relative to the temperate climate penguins. Of these, thirteen were linked to a potential role in thermoregulation with TUBA1C having the most copies present in the macaroni penguin compared to the temperate climate penguins.

Single copy orthologous proteins conserved in both cold and temperate clade penguins were subjected to microevolutionary scale analysis at the individual protein and gene level. SCOs contributed nearly 47% of the *Eudyptes* pan-genome complement. Through amino acid bias analysis, biases in the SCO aligned aa sequence data were identified using various statistical tests (RCFV, chi-square and p-value). Fourteen SCOs met at least one of the statistical cut-off criteria and were found to be linked to thermal regulation. Furthermore, two were discovered to be under overall positive selection (DLL4 and FADS1). Amino acid composition of the fourteen candidate thermoregulation proteins (as is supported in the literature) revealed that cold-adapted proteins prefer conformational flexibility over conformational stability. This is then achieved through an increase in the number of small, neutral amino acids such as serine as well as amino acids that increase the solubility of a protein i.e. aspartic acid. These observed changes are modest but may have a large potential impact on protein function in cold conditions regardless. All these analyses have shown that various molecular mechanisms can be linked to main thermoregulatory mechanisms such as glucose metabolism, lipid metabolism, adipocyte differentiation, vascular development, feather development and skeletal muscle functioning.

There are a few limitations to this study. Without looking at other factors such as morphology, integumentary data and behavioural data we cannot definitively say that all eight *Eudyptes* species are separate species, but our molecular data does suggest this is the most likely case. Another limitation is that many proteins still need to be fully functionally evaluated and annotated and many more proteins may play a role in thermoregulation and thermoadaptation. As a result, they may provide more information on mechanisms that help in cold adaptation. Future studies should incorporate all extant penguin species, especially the species localised in Antarctica i.e. the emperor and king penguins, and by the equator (the Galápagos penguin) to get an even more in-depth account of cold adaptation in penguins. Studies could also look at molecular mechanisms in their temperate counterparts in order to see how climate change may affect the different taxa. In conclusion, this study provides primary information in order to start uncovering the molecular mechanisms that allow penguins to inhabit various environments as well as to uncover the molecular mechanisms that drive cold adaptation in penguins.

References:

Ainley, D. G., Ballard, G., Jones, R. M., Jongsomjit, D., Pierce, S. D., Smith, W. O., Veloz, S. 2015. Trophic cascades in the western Ross Sea, Antarctica: revisited. Marine Ecology Progress Series, 534: 1-16.

Agnihotri, G. & Liu, H. 2003. Enoyl-CoA hydratase: reaction, mechanism, and inhibition. Bioorganic & Medicinal Chemistry, 11: 9–20. https://doi.org/10.1016/s0968-0896(02)00333-4.

Alamer, M. 2011. The Role of Prolactin in Thermoregulation and Water Balance During Heat Stress in Domestic Ruminants. Asian Journal of Animal and Veterinary Advances, 6: 1153-1169.

Aledo, J. C. 2019. Methionine in proteins: The Cinderella of the proteinogenic amino acids. Protein Science, 28: 1785–1796. https://doi.org/10.1002/pro.3698.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology, 215: 403-410.

Anderson, W. B., Wait, D. A. & Stapp, P. 2008. Resources from another place and time: Responses to pulses in a spatially subsidized system. Ecology, 89: 660–670.

Bacon, C. D. & Bailey, C. D. 2006. Taxonomy and Conservation: A Case Study from Chamaedorea alternans. Annals of Botany, 98: 755–763. https://doi.org/10.1093/aob/mcl158

Baker, A. J., Pereira, S. L., Haddrath, O. P., Edge, K. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proceedings of the Royal Society B, 273: 11-17.

Banks, J., Van Buren, A., Cherel, Y., Whitfield, J. B. 2006. Genetic evidence for three species of rockhopper penguins, Eudyptes chrysocome. Polar Biology, 30: 61-67.

Barbraud, C. & Weimerskirch, H. 2001. Emperor penguins and climate change. Nature, 411: 183-186.

Bell, G. 2008. Selection: The Mechanism of Evolution (2nd ed.). Oxford; New York: Oxford University Press. ISBN 978-0-19-856972-5.

Bentahir, M., Feller, G., Aittaleb, M., Lamotte-Brasseur, J., Himri, T., Chessa, J. P., Gerday, C. 2000. Structural, kinetic, and calorimetric characterization of the cold-active phosphoglycerate kinase from the antarctic Pseudomonas sp. TACII18. Journal of Biological Chemistry, 275: 11147–11153. https://doi.org/10.1074/jbc.275.15.11147.

Bertelli, S. & Giannini, N. P. A. 2005. A phylogeny of extant penguins (Aves: Sphenisiformes) combining morphology and mitochondrial sequences. Cladistics, 21: 209-239.

Berthelot, C., Clarke, J., Desvignes, T., William Detrich, H., Flicek, P., Peck, L. S., Peters, M., Postlethwait, J. H., Clark, M. S., 2019. Adaptation of Proteins to the Cold in Antarctic Fish: A Role for Methionine? Genome Biology and Evolution, 11: 220–231. https://doi.org/10.1093/gbe/evy262.

Bian, F., Ma, X., Villivalam, S. D., You, D., Choy, L. R., Paladugu, A., Fung, S., Kang, S. 2018. TET2 facilitates PPARγ agonist–mediated gene regulation and insulin sensitization in adipocytes. Metabolism, 89: 39–47. https://doi.org/10.1016/j.metabol.2018.08.006.

BirdLife International. 2020. IUCN red list for birds. https://www.birdlife.org/list-penguin-species (accessed 04/06/2022).

BirdLife International. 2022. IUCN red list for birds. https://www.birdlife.org/list-penguin-species (accessed 27/10/2022).

Birney, E., Clamp, M., Durbin, R. 2004. GeneWise and Genomewise. Genome Research, 14: 988– 995. https://doi.org/10.1101/gr.1865504.

Blötz, C. & Stülke, J. 2017. Glycerol metabolism and ts implication in virulence in Mycoplasma. FEMS Microbiology Reviews, 41: 640–652. https://doi.org/10.1093/femsre/fux033.

Boersma, P. D. 2008. Penguins as marine sentinels. Bioscience, 58: 597–607.

Brown, J. M., Scarlett, J. M., Schwartz, M. W. 2019. Rethinking the role of the brain in glucose homeostasis and diabetes pathogenesis. Journal of Clinical Investigation, 129: 3035– 3037. https://doi.org/10.1172/JCI130904.

Burrell, J. A., Boudreau, A., Stephens, J. M. 2020. Latest advances in STAT signalling and function in adipocytes. Clinical Science, 134: 629–639. https://doi.org/10.1042/CS20190522.

Cai, X., Feng, Y., Xu, M., Yu, C., Xie, W. 2021. Gadd45b is required in part for the anti-obesity effect of constitutive androstane receptor (CAR). Acta Pharmaceutica Sinica B, 11: 434–441. https://doi.org/10.1016/j.apsb.2020.08.015.

Cao, Z., Shang, B., Zhang, G., Miele, L., Sarkar, F. H., Wang, Z., Zhou, Q. 2013. Tumor cell-mediated neovascularization and lymphangiogenesis contrive tumor progression and cancer metastasis. Biochimica et Biophysica Acta (BBA), 1836: 273–286. https://doi.org/10.1016/j.bbcan.2013.08.001.

Cerqua, C., Morbidoni, V., Desbats, M. A., Doimo, M., Frasson, C., Sacconi, S., Baldoin, M. C., Sartori, G., Basso, G., Salviati, L., Trevisson, E. 2018. COX16 is required for assembly of cytochrome c oxidase in human cells and is involved in copper delivery to COX2. Biochimica et Biophysica Acta (BBA) – Bioenergetics, 1859: 244–252. https://doi.org/10.1016/j.bbabio.2018.01.004.

Castresana, J. 2002. GBlocks Server. Castresana Lab Animal Biodiversity and Evolution Program. http://molevol.cmima.csic.es/castresana/Gblocks_server.html (accessed 05/04/2022).

Chang, H. C., Kao, C. H., Chung, S. Y., Chen, W. C., Aninda, L. P., Chen, Y. H., Juan, Y. A., Chen, S. L. 2019. Bhlhe40 differentially regulates the function and number of peroxisomes and mitochondria in myogenic cells. Redox Biology, 20: 321–333. https://doi.org/10.1016/j.redox.2018.10.009.

Chen, Z., Cheng, C. H. C., Zhang, J., Cao, L., Chen, Lei, Zhou, L., Jin, Y., Ye, H., Deng, C., Dai, Z., Xu, Q., Hu, P., Sun, S., Shen, Y., Chen, Liangbiao. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. Proceedings of the National Academy of Sciences of the United States of America, 105: 12944–12949. https://doi.org/10.1073/pnas.0802432105.

Choe, S. S., Huh, J. Y., Hwang, I. J., Kim, J. I., Kim, J. B. 2016. Adipose Tissue Remodeling: Its Role in Energy Metabolism and Metabolic Disorders. Frontiers of Endocrinology, 7. https://doi.org/10.3389/fendo.2016.00030.

Chiari, Y., Van Der Meijden, A., Madsen, O., Vences, M., Meyer, A. 2009. Base composition, selection, and phylogenetic significance of indels in the recombination activating gene-1 in vertebrates. Frontiers in Zoology, 6: 32.

Chu, M., Gregorio, C. C., Pappas, C. T. 2016. Nebulin, a multi-functional giant. Journal of Experimental Biology, 219: 146–152. https://doi.org/10.1242/jeb.126383.

Churm, R., Davies, J., Stephens, J., Prior, S. 2017. Ghrelin function in human obesity and type 2 diabetes: a concise review: Ghrelin in human obesity and diabetes. Obesity Reviews, 18: 140–148. https://doi.org/10.1111/obr.12474.

Clarke, A., 1991. What Is Cold Adaptation and How Should We Measure It? American Zoologist, 31: 81– 92. https://doi.org/10.1093/icb/31.1.81.

Cole, T. L., Ksepka, D. T., Mitchell, K. J., Tennyson, A. J. D., Thomas, D. B., Pan, H., Zhang, G., Rawlence, N. J., Wood, J. R., Bover, P., Bouzat, J. L., Cooper, A., Fiddaman, S. R., Hart, T., Miller, G., Ryan, P. G., Shepherd, L. D., Wilmshurst, J. M., Waters, J. M. 2019. Mitogenomes uncover extinct penguin taxa and reveal island formation as a key driver of speciation. Molecular Biology and Evolution, 36: 784-797.

Cooke, R. S. C., Eigenbrod, F., & Bates, A. E. 2019. Projected losses of global mammal and bird ecological strategies. Nature Communications, 10 (1): 2279. https://doi.org/10.1038/s41467-019-10284-z.

Cui, J., Pang, J., Lin, Y. J., Gong, H., Wang, Z. H., Li, Y. X., Li, Jin, Wang, Z., Jiang, P., Dai, D. P., Li, Jian, Cai, J. P., Huang, J. D., Zhang, T. M. 2017. Adipose-specific deletion of Kif5b exacerbates obesity and insulin resistance in a mouse model of diet-induced obesity. FASEB J, 31: 2533–2547. https://doi.org/10.1096/fj.201601103R.

Dalloul, R. A., Long, J. A., Zimin, A. V., Aslam, L., Beal, K., Blomberg, L. A., Bouffard, P., Burt, D. W., Crasta, O., Crooijmans, R. P. M. A., Cooper, K., Coulombe, R. A., De, S., Delany, M. E., Dodgson, J. B., Dong, J. J., Evans, C., Frederickson, K. M., Flicek, P., Florea, L., Folkerts, O., Groenen, M. A. M., Harkins, T. T., Herrero, J., Hoffmann, S., Megens, H. J., Jiang, A., Jong, P. de, Kaiser, P., Kim, H., Kim, K. W., Kim, S., Langenberger, D., Lee, M. K., Lee, T., Mane, S., Marcais, G., Marz, M., McElroy, A. P., Modise, T., Nefedov, M., Notredame, C., Paton, I. R., Payne, W. S., Pertea, G., Prickett, D., Puiu, D., Qioa, D., Raineri, E., Ruffier, M., Salzberg, S. L., Schatz, M. C., Scheuring, C., Schmidt, C. J., Schroeder, S., Searle, S. M. J., Smith, E. J., Smith, J., Sonstegard, T. S., Stadler, P. F., Tafer, H., Tu, Z. (Jake), Tassell, C. P. V., Vilella, A. J., Williams, K. P., Yorke, J. A., Zhang, L., Zhang, H. B., Zhang, X., Zhang, Y., Reed, K. M. 2010. Multi-Platform Next-Generation Sequencing of the Domestic Turkey (Meleagris gallopavo): Genome Assembly and Analysis. PLOS Biology, 8: e1000475. https://doi.org/10.1371/journal.pbio.1000475

Deem, J. D., Muta, K., Ogimoto, K., Nelson, J. T., Velasco, K. R., Kaiyala, K. J., Morton, G. J. 2018. Leptin regulation of core body temperature involves mechanisms independent of the thyroid axis. American Journal of Physiology-Endocrinology and Metabolism, 315: E552–E564. https://doi.org/10.1152/ajpendo.00462.2017.

del Hoyo, J., Elliot, A., Sargatal, J. 1992. Handbook of the birds of the world, vol. 1: Ostrich to Ducks. Lynx Edicions, Barcelona, Spain.

De Maayer, P., Anderson, D., Cary, C., Cowan, D.A. 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO Reports, 15: 508–517. https://doi.org/10.1002/embr.201338170.

Detrich, H. W. 1997. Microtubule assembly in cold-adapted organisms: functional properties and structural adaptations of tubulins from Antarctic fishes. Comparative Biochemistry & Physiology, 118: 501–513. https://doi.org/10.1016/s0300-9629(97)00012-1.

De Vries, D. D., Went, L. N., Bruyn, G. W., Scholte, H. R., Hofstra, R. M., Bolhuis, P. A., van Oost, B. A. 1996. Genetic and biochemical impairment of mitochondrial complex I activity in a family with Leber hereditary optic neuropathy and hereditary spastic dystonia. American Journal of Human Genetics, 58: 703–711.

Doan, K. V., Kinyua, A. W., Yang, D. J., Ko, C. M., Moh, S. H., Shong, K. E., Kim, H., Park, S. K., Kim, D. H., Kim, I., Paik, J. H., DePinho, R. A., Yoon, S. G., Kim, I. Y., Seong, J. K., Choi, Y. H., Kim, K. W. 2016. FoxO1 in dopaminergic neurons regulates energy homeostasis and targets tyrosine hydroxylase. Nature Communications, 7: 12733. https://doi.org/10.1038/ncomms12733.

Doron-Faigenboim, A. & Pupko, T. 2006. A Combined Empirical and Mechanistic Codon Model. Molecular Biology & Evolution, 24: 388-397.

Doron-Faigenboim, A., Stern, A., Mayrose, I., Bacharach, E., Pupko, T. 2005. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics, 21: 2101–2103. https://doi.org/10.1093/bioinformatics/bti259.

Ellis, J. C. 2005. Marine Birds on Land: A Review of Plant Biomass, Species Richness, and Community Composition in Seabird Colonies. Plant Ecology, 181: 227–241. https://doi.org/10.1007/s11258-005-7147-y

Emms, D. M. & Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology, 16:157.

Fathzadeh, M., Li, J., Rao, A., Cook, N., Chennamsetty, I., Seldin, M., Zhou, X., Sangwung, P., Gloudemans, M. J., Keller, M., Attie, A., Yang, J., Wabitsch, M., Carcamo-Orive, I., Tada, Y., Lusis, A. J., Shin, M. K., Molony, C. M., McLaughlin, T., Reaven, G., Montgomery, S. B., Reilly, D., Quertermous, T., Ingelsson, E., Knowles, J. W. 2020. FAM13A affects body fat distribution and adipocyte function. Nature Communications, 11: 1465. https://doi.org/10.1038/s41467-020-15291-z.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., Pacheco, G., Sinding, M. H .S., Puetz, L., Cavill, E., Ribeiro, Â. M., Eckhart, L., Fjeldså, J., Hosner, P. A., Brumfield, R. T., Christidis, L., Bertelsen, M. F., Sicheritz-Ponten, T., Tietze, D. T., Robertson, B. C., Song, G., Borgia, G., Claramunt, S., Lovette, I. J., Cowen, S. J., Njoroge, P., Dumbacher, J. P., Ryder, O. A., Fuchs, J., Bunce, M., Burt, D. W., Cracraft, J., Meng, G., Hackett, S. J., Ryan, P. G., Jønsson, K. A., Jamieson, I. G., da Fonseca, R. R., Braun, E. L., Houde, P., Mirarab, S., Suh, A., Hansson, B., Ponnikas, S., Sigeman, H., Stervander, M., Frandsen, P. B., van der Zwan, H., van der Sluis, R., Visser, C., Balakrishnan, C. N., Clark, A. G., Fitzpatrick, J. W., Bowman, R., Chen, N., Cloutier, A., Sackton, T. B., Edwards, S. V., Foote, D. J., Shakya, S. B., Sheldon, F. H., Vignal, A., Soares, A. E. R., Shapiro, B., González-Solís, J., Ferrer-Obiol, J., Rozas, J., Riutort, M., Tigano, A., Friesen, V., Dalén, L., Urrutia, A. O., Székely, T., Liu, Y., Campana, M. G., Corvelo, A., Fleischer, R. C., Rutherford, K. M., Gemmell, N. J., Dussex, N., Mouritsen, H., Thiele, N., Delmore, K., Liedvogel, M., Franke, A., Hoeppner, M. P., Krone, O., Fudickar, A. M., Milá, B., Ketterson, E. D., Fidler, A. E., Friis, G., Parody-Merino, Á. M., Battley, P. F., Cox, M. P., Lima, N. C. B., Prosdocimi, F., Parchman, T. L., Schlinger, B. A., Loiselle, B. A., Blake, J. G., Lim, H. C., Day, L. B., Fuxjager, M. J., Baldwin, M. W., Braun, M. J., Wirthlin, M., Dikow, R. B., Ryder, T. B., Camenisch, G., Keller, L. F., DaCosta, J. M., Hauber, M. E., Louder, M. I. M., Witt, C. C., McGuire, J. A., Mudge, J., Megna, L. C., Carling, M. D., Wang, B., Taylor, S. A., Del-Rio, G., Aleixo, A., Vasconcelos, A. T. R., Mello, C. V., Weir, J. T., Haussler, D., Li, Q., Yang, H., Wang, J., Lei, F., Rahbek, C., Gilbert, M. T. P., Graves, G. R., Jarvis, E. D., Paten, B., Zhang, G. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature, 587: 252–257. https://doi.org/10.1038/s41586-020-2873-9.

Forcada, J. & Trathan, P. N. 2009. Penguin responses to climate change in the Southern Ocean. Global Change Biology, 15: 1618-1630.

Frost, P. G. H., Siegfried, W. R. & Greenwood, P. J. 1975. Arterio-venous heat exchange systems in the Jackass penguin Spheniscus demersus. Journal of Zoology, 175: 231–41.

Frugone, M. J., Lowther, A., Noll, D., Ramos, B., Pistorius, P., Dantas, G. P. M., Petry, M. V., Bonadonna, F., Steinfurth, A., Polanowski, A., Raya Rey, A., Lois, N. A., Pütz, K., Trathan, P., Wienecke, B., Poulin, E., Vianna, J. A. 2018. Contrasting phylogeographic pattern among Eudyptes penguins around the Southern Ocean. Scientific Reports, 8: 17481.

Fuller, M., Priyadarshini, M., Gibbons, S. M., Angueira, A. R., Brodsky, M., Hayes, M. G., Kovatcheva-Datchary, P., Bäckhed, F., Gilbert, J. A., Lowe, W. L., Layden, B. T. 2015. The short-chain fatty acid receptor, FFA2, contributes to gestational glucose homeostasis. American Journal of Physiology-Endocrinology and Metabolism, 309: E840–E851. https://doi.org/10.1152/ajpendo.00171.2015.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. Genetics, 159: 907–911. https://doi.org/10.1093/genetics/159.2.907.

García, M., Pazos, P., Lima, L., Diéguez, C. 2018. Regulation of Energy Expenditure and Brown/Beige Thermogenic Activity by Interleukins: New Roles for Old Actors. IJMS, 19: 2569. https://doi.org/10.3390/ijms19092569.

Garcia, B. P. & Boersma, P. D. 2013. Penguins: Natural History and Conservation. Seattle, WA, USA: University of Washington Press, 328.

Gesta, S. & Kahn, C.R. 2012. White Adipose Tissue, in: Symonds, M.E. (Ed.), Adipose Tissue Biology. Springer, New York, NY, pp. 71–121. https://doi.org/10.1007/978-1-4614-0965-6_4.

Giannini, N. P. & Bertelli, S. 2004. Phylogeny of extant penguins based on integumentary and breeding characters. The Auk, 121: 422-434.

Gilbert, C., Robertson, G., Le Maho, Y., Naito, Y., Ancel, A. 2006. Huddling behavior in emperor penguins: dynamics of huddling. Physiology and Behavior, 88: 479-488.

Gill, F.B. 1994. Ornithology. 2nd edition. W.H. Freeman and Company, New York.

Giordano, D., Boubeta, F. M., di Prisco, G., Estrin, D. A., Smulevich, G., Viappiani, C., Verde, C. 2020. Conformational Flexibility Drives Cold Adaptation in Pseudoalteromonas haloplanktis TAC125 Globins. Antioxidants & Redox Signalling, 32: 396–411. https://doi.org/10.1089/ars.2019.7887.

Goldman, N. & Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution, 11: 725-736.

Goldsmith, R. & Sladen, W. J. 1961. Temperature regulation of some Antarctic penguins. Journal of Physiology, 157: 251–62.

Götz, S., Garcia-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research, 36: 3420-3435.

Greenwood, I.A. & Ohya, S. 2009. New tricks for old dogs: KCNQ expression and role in smooth muscle: KCNQ genes in smooth muscle. British Journal of Pharmacology, 156: 1196–1203. https://doi.org/10.1111/j.1476-5381.2009.00131.x.

Greenwold, M. J., Bao, W., Jarvis, E. D., Hu, H., Li, C., Gilbert, M. T. P., Zhang, G., Sawyer, R. H. 2014. Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. BMC Evolutionary Biology, 14: 249.

Grosser, S., Rawlence, N. J., Anderson, C. N. K., Smith, I. W. G., Scofield, R. P., Waters, J. M. 2016. Invader or resident? Ancient-DNA reveals rapid species turnover in New Zealand little penguins. Proceedings of the Royal Society B: Biological Sciences, 283: 20152879. https://doi.org/10.1098/rspb.2015.2879

Hansen, A.L. & Kay, L.E. 2014. Measurement of histidine pK a values and tautomer populations in invisible protein states. Proceedings of the National Academy of Sciences of the United States of America, 111. https://doi.org/10.1073/pnas.1400577111.

Havik, G., Catenazzi, A. & Holmgren, M. 2014. Seabird Nutrient Subsidies Benefit Non-Nitrogen Fixing Trees and Alter Species Composition in South American Coastal Dry Forests. PLOS ONE, 9: e86381. https://doi.org/10.1371/journal.pone.0086381

Herman, R., Borowicz, A., Lynch, M., Trathan, P., Hart, T., Lynch, H. 2020. Update on the global abundance and distribution of breeding Gentoo Penguins (*Pygoscelis papua*). Polar Biology, 43: 1947–1956. https://doi.org/10.1007/s00300-020-02759-3

Hibi, M., Murakami, M., Saito, M., Hirano, T., Taga, T., Kishimoto, T. 1990. Molecular cloning and expression of an IL-6 signal transducer, gp130. Cell, 63: 1149–1157. https://doi.org/10.1016/0092-8674(90)90411-7.

Higgins, C. B., Zhang, Y., Mayer, A. L., Fujiwara, H., Stothard, A. I., Graham, M. J., Swarts, B. M., DeBosch, B. J. 2018. Hepatocyte ALOXE3 is induced during adaptive fasting and enhances insulin sensitivity by activating hepatic PPARγ. Journal of Insight, 3: e120794. https://doi.org/10.1172/jci.insight.120794.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Vinh, L. S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35: 518–522.

Ho-Huu, J., Ronfort, J., De Mita, S., Bataillon, T., Hochu, I., Weber, A., Chantret, N. 2012. Contrasted patterns of selective pressure in three recent paralogous gene pairs in the Medicagogenus (L.). BMC Evolutionary Biology, 12: 195. https://doi.org/10.1186/1471-2148-12-195.

Howell, N. 1989. Evolutionary conservation of protein regions in the proton motive cytochrome b and their possible roles in redox catalysis. Journal of Molecular Evolution, 29: 157-169.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggnog-mapper. Molecular Biology and Evolution, doi: 10.1093/molbev/msx148.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., Bork, P. 2019. EggNog 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research, doi: 10.1093/nar/gky1085.

Hughes, M. W., Wu, P., Jiang, T. X., Lin, S. J., Dong, C. Y., Li, A., Hsieh, F. J., Widelitz, R. B., Chuong, C. M. 2011. In search of the Golden Fleece: unraveling principles of morphogenesis by studying the integrative biology of skin appendages. Integrated Biology, 3: 388. https://doi.org/10.1039/c0ib00108b.

Ichimura, A., Hasegawa, S., Kasubuchi, M., Kimura, I. 2014. Free fatty acid receptors as therapeutic targets for the treatment of diabetes. Frontiers of Pharmacology, 5. https://doi.org/10.3389/fphar.2014.00236.

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432: 695–716. https://doi.org/10.1038/nature03154.

IUCN 2022. The IUCN Red List of Threatened Species. Version 2022-1. https://www.iucnredlist.org

Iype, T., Taylor, D. G., Ziesmann, S. M., Garmey, J. C., Watada, H., Mirmira, R. G. 2004. The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6.1. Molecular Endocrinology, 18: 1363–1375. https://doi.org/10.1210/me.2004-0006.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, Md. S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W. C., Ray, D., Green, R. E., Bruford, M. W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E. P., Bertelsen, M. F., Sheldon, F. H., Brumfield, R. T., Mello, C. V., Lovell, P. V., Wirthlin, M., Schneider, M. P. C., Prosdocimi, F., Samaniego, J. A., Velazquez, A. M. V., Alfaro-Núñez, A., Campos, P. F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D. M., Zhou, Q., Perelman, P., Driskell, A. C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F. E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F. K., Jønsson, K. A., Johnson, W., Koepfli, K. P., O'Brien, S., Haussler, D., Ryder, O. A., Rahbek, C., Willerslev, E., Graves, G. R., Glenn, T. C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S. V., Stamatakis, A., Mindell, D. P., Cracraft, J., Braun, E .L., Warnow, T., Jun, W., Gilbert, M. T. P., Zhang, G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science, 346: 1320–1331. https://doi.org/10.1126/science.1253451.

Jenouvrier, S., Holland, M., Stroeve, J., Serreze, M., Barbraud, C., Weimerskirch, H., Caswell, H. 2014. Projected continent-wide declines of the emperor penguin under climate change. Nature Climate Change: 4: 715-718.

Johnson, K., Bednarz, J. C. & Zack, S. 1987. Crested penguins: Why are the first eggs smaller? Oikos, 49 (3): 347-349.

Jouventin, P. 1982. Visual and vocal signals in penguins, their evolution and adaptive characters. Advances in Ethology, 24: 1-149.

Jouventin, P., Cuthbert, R. J. & Ottvall, R. 2006. Genetic isolation and divergence in sexual traits: evidence for the northern rockhopper penguin Eudyptes moseleyi being a sibling species. Molecular Ecology, 15: 3413–3423. https://doi.org/10.1111/j.1365-294X.2006.03028.x

Juchnicka, I., Kuźmicki, M., Szamatowicz, J. 2021. Ceramides and Sphingosino-1-Phosphate in Obesity. Frontiers of Endocrinology (Lausanne),12: 635995. https://doi.org/10.3389/fendo.2021.635995.

Kaczmarek, M. J. & Rosenmund, H. 1977. The action of human pancreatic and salivary isoamylases on starch and glycogen. Clinica Chimica Acta, 79: 69–73. https://doi.org/10.1016/0009-8981(77)90462-4.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., Jermiin, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Natural Methods, 14: 587–589. https://doi.org/10.1038/nmeth.4285.

Kanehisa, M. & Sato, Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Science, 29: 28-35.

Kanehisa, M., Sato, Y. & Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology, 428: 726-731.

Katoh, K., Kuma, K., Toh, H., Miyata, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research, 33: 511–518. https://doi.org/10.1093/nar/gki198.

Khayami, R., Hashemi, S. R., Kerachian, M. A. 2020. Role of aldo-keto reductase family 1 member B1 (AKR1B1) in the cancer process and its therapeutic potential. Journal of Cell & Molecular Medicine, 24: 8890–8902. https://doi.org/10.1111/jcmm.15581.

Kim, S. Y., Hwang, K. Y., Kim, S. H., Sung, H. C., Han, Y. S., Cho, Y. 1999. Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile Aquaspirillium arcticum. Journal of Biological Chemistry, 274: 11761– 11767. https://doi.org/10.1074/jbc.274.17.11761.

Klein, I., Sanchez-Alavez, M., Tabarean, I., Schaefer, J., Holmberg, K. H., Klaus, J., Xia, F., Marcondes, M. C. G., Dubins, J. S., Morrison, B., Zhukov, V., Sanchez-Gonzalez, A., Mitsukawa, K., Hadcock, J. R., Bartfai, T., Conti, B. 2011. AdipoR1 and 2 are expressed on warm sensitive neurons of the hypothalamic preoptic area and contribute to central hyperthermic effects of adiponectin. Brain Research, 1423: 1–9. https://doi.org/10.1016/j.brainres.2011.09.019.

Klymenko, O., Brecklinghaus, T., Dille, M., Springer, C., de Wendt, C., Altenhofen, D., Binsch, C., Knebel, B., Scheller, J., Hardt, C., Herwig, R., Chadt, A., Pfluger, P. T., Al-Hasani, H., Kabra, D. G. 2020. Histone deacetylase 5 regulates interleukin 6 secretion and insulin action in skeletal muscle. Molecular Metabolism, 42: 101062. https://doi.org/10.1016/j.molmet.2020.101062.

Kollberg, G., Tulinius, M., Gilljam, T., Ostman-Smith, I., Forsander, G., Jotorp, P., Oldfors, A., Holme, E. 2007. Cardiomyopathy and exercise intolerance in muscle glycogen storage disease 0. The New England Journal of Medicine, 357: 1507–1514. https://doi.org/10.1056/NEJMoa066691.

Kooyman, G. L., Gentry, R. L., Bergman, W. P., Hammel, H. T. 1976. Heat loss in penguins during immersion and compression. Comparative Biochemistry and Physiology A, 54: 75-80.

Krstic, J., Reinisch, I., Schupp, M., Schulz, T., Prokesch, A. 2018. p53 Functions in Adipose Tissue Metabolism and Homeostasis. IJMS, 19: 2622. https://doi.org/10.3390/ijms19092622.

Krolopp, J. E., Thornton, S. M., Abbott, M. J. 2016. IL-15 Activates the Jak3/STAT3 Signaling Pathway to Mediate Glucose Uptake in Skeletal Muscle Cells. Frontiers in Physiology, 7. https://doi.org/10.3389/fphys.2016.00626.

Ksepka, D. T., Bertelli, S & Giannini, N. P. 2006. The phlogeny of the living and fossil Sphenisciformes (penguins). Cladistics, 22(5): 412-441.

Kück, P., Struck, T. H. 2014. BaCoCa--a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Molecular Phylogenetics and Evolution, 70: 94– 98. https://doi.org/10.1016/j.ympev.2013.09.011.

Large Phylogenomics Projects, 2019. URL http://zhanggjlab.org/pages/project.html (accessed 7.7.22).

Lee, C., Jang, S. H., Chung, H. S., 2017. Improving the Stability of Cold-Adapted Enzymes by Immobilization. Catalysts 7, 112. https://doi.org/10.3390/catal7040112

Le Maho, Y. 1977. The emperor penguin: A strategy to live and breed in the cold: Morphology, physiology, ecology, and behaviour distinguish the polar emperor penguin from other penguin species, particularly from its close relative, the king penguin. American Scientist, 65: 680.

Lewden, A., Enstipp, M. R., Picard, B., van Walsum, T., Handrich, Y. 2017. High peripheral temperatures in king penguins while resting at sea: thermoregulation versus fat deposition. Journal of Experimental Biology, 220: 3084-3094.

Li, C., Zhang, Y., Li, J., Kong, L., Hu, H., Pan, H., Xu, L., Deng, Y., Li, Q., Jin, L., Yu, H., Chen, Y., Liu, B., Yang, L., Liu, S., Zhang, Y., Lang, Y., Xia, J., He, W., Shi, Q., Subramanian, S., Millar, C. D., Meader, S., Rands, C. M., Fujita, M. K., Greenwold, M. J., Castoe, T. A., Pollock, D. D., Gu, W., Nam, K., Ellegren, H., Ho, S. Y. W., Burt, D. W., Ponting, C. P., Jarvis, E. D., Gibert, M. T. P., Yang, H., Wang, J., Lambert, D. M., Wang, J., Zhang, G. 2014. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. GigaScience, 3: 3-27.

Loeb, V., Siegel, V., Holm-Hansen, O., Hewitt, R., Fraser, W., Trivelpiece, W., Trivelpiece, S. 1997. Effects of sea-ice extent and krill or salp dominance on the Antarctic food web. Nature: 387: 897-900.

Lorrain, A., Houlbrèque, F., Benzoni, F., Barjon, L., Tremblay-Boyer, L., Menkes, C., Gillikin, D. P., Payri, C., Jourdan, H., Boussarie, G., Verheyden, A., Vidal, E. 2017. Seabirds supply nitrogen to reef-building corals on remote Pacific islets. Scientific Reports, 7: 3721. https://doi.org/10.1038/s41598-017-03781-y

Lu, S., Wang, J., Chitsar, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., Marchler-Bauer, A. 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research, 48: D265-D268.

Makinen, T. M., 2010. Different types of cold adaptation in humans. Frontiers in Bioscience, (Schol Ed) 2: 1047–1067. https://doi.org/10.2741/s117.

Marchant, S. & Higgins, P. J. 1990. Handbook of Australian, New Zealand and Antarctic Birds. Vol. 1, Pt. B. Melbourne, Australia: Oxford University Press.

Markan, K. R., Boland, L. K., King-McAlpin, A. Q., Claflin, K. E., Leaman, M. P., Kemerling, M. K., Stonewall, M. M., Amendt, B. A., Ankrum, J. A., Potthoff, M. J. 2020. Adipose TBX1 regulates β-adrenergic sensitivity in subcutaneous adipose tissue and thermogenic capacity in vivo. Molecular Metabolism, 36: 100965. https://doi.org/10.1016/j.molmet.2020.02.008.

Martens, N., Uzan, G., Wery, M., Hooghe, R., Hooghe-Peters, E. L., Gertler, A. 2005. Suppressor of cytokine signaling 7 inhibits prolactin, growth hormone, and leptin signaling by interacting with STAT5 or STAT3 and attenuating their nuclear translocation. Journal of Biological Chemistry, 280: 13817–13823. https://doi.org/10.1074/jbc.M411596200.

Mayr, G., Scofield, R. P., De Pietri, V. L., Tennyson, A. J. D. 2017. A Paleocene penguin from New Zealand substantiates multiple origins of gigantism in fossil Sphenisciformes. Nature Communications, 8(1): 1927-1936.

McCafferty, D. J., Gilbert, C., Thierry, A. M., Currie, J., Le Maho, Y., Ancel, A. 2013. Emperor penguin body surfaces cool below air temperature. Biology Letters, 9: 20121192.

McKitrick, M. C. 1991. Phylogenetic analysis of avian hindlimb musculature. Miscellaneous Publications Museum of Zoology, University of Michigan, 179: 1-89.

Metcheva, R., Yurukova, L., Teodorova, S., Nikolova, E. 2006. The penguin feathers as bioindicator of Antarctica environmental state. Science of the Total Environment, 362: 259–265. https://doi.org/10.1016/j.scitotenv.2005.05.008

Metpally, R. & Reddy, B., 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. BMC Genomics 10, 11. https://doi.org/10.1186/1471-2164-10-11.

Mohamad Ali, M. S., Mohd Fuzi, S. F., Ganasen, M., Abdul Rahman, R. N. Z. R., Basri, M., Salleh, A. B. 2013. Structural adaptation of cold-active RTX lipase from Pseudomonas sp. strain AMS8 revealed via homology and molecular dynamics simulation approaches. Biomedical Research International, 2013: 925373. https://doi.org/10.1155/2013/925373.

Mozo, J., Emre, Y., Bouillaud, F., Ricquier, D., Criscuolo, F. 2005. Thermoregulation: What role for UCPs in mammals and birds?. Bioscience Reports, 25: 227-249.

Mulder, C. P. H. & Keall, S. N. 2001. Burrowing seabirds and reptiles: Impacts on seeds, seedlings, and soils in an island forest in New Zealand. Oecologia (Berl.) 127:350–360.

Mulder, C. P. H., Anderson, W. B., Towns, D. R., Bellingham, P. J., Mulder, C. P. H., Anderson, W. B., Towns, D. R., Bellingham, P. J. (Eds.). 2011. Seabird Islands: Ecology, Invasion, and Restoration. Oxford University Press, Oxford, New York.

Müller-Schwarze, D. 1984. The behavior of penguins: Adapted toice and tropics SUNY series in animal behavior. State University of New York Press, New York.

Myers, R. D. 1980. Catecholamines and the Regulation of Body Temperature, in: Szekeres, L. (Ed.), Adrenergic Activators and Inhibitors: Part I, Handbook of Experimental Pharmacology. Springer, Berlin, Heidelberg, pp. 549–567. https://doi.org/10.1007/978-3-642-67505-8_14.

National Centre for Biotechnology Information (NCBI). AVES - Genome, 1988. URL: https://www.ncbi.nlm.nih.gov/genome/?term=AVES (accessed 07/07/2022).

National Centre for Biotechnology Information (NCBI). Assembly – NCBI. 2022. URL https://www.ncbi.nlm.nih.gov/assembly (accessed 10/29/22).

Ng, C. S. & Li, W. H. 2018. Genetic and Molecular Basis of Feather Diversity in Birds. Genome Biology and Evolution, 10: 2572–2586. https://doi.org/10.1093/gbe/evy180.

Nguyen, L., Schmidt, H. A., von Haeseler, A., Minh, B. Q. 2015. IQ-Tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32: 268-274.

Notredame, C., Higgins, D. G. & Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology, 302: 205-217.

Nürnberg, B., Beer-Hammer, S. 2019. Function, Regulation and Biological Roles of PI3Kγ Variants. Biomolecules, 9: 427. https://doi.org/10.3390/biom9090427.

Oh, S. P., Seki, T., Goss, K. A., Imamura, T., Yi, Y., Donahoe, P. K., Li, L., Miyazono, K. 1999. Activin receptor-like kinase 1 modulates transforming growth factor-□1 signalling in the regulation of angiogenesis. Proceedings of the National Academy of Sciences of the United States of America, 97 (6): 2626-2631.

O'Hara, R. J. 1989. An estimate of the phylogeny of the living penguins (Aves: Spheniscidae). American Journal of Zoology, 273: 11-17.

Oliver, C., Parker, S. & Riley, P. 2006. 500 things you should know about science. Bardfield Press, Great Bardfield, Essex, pg. 118-119.

Osaka, T. 2004. Cold-induced thermogenesis mediated by GABA in the preoptic area of anesthetized rats. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 287: R306–R313. https://doi.org/10.1152/ajpregu.00003.2004.

Osilla, E. V. & Sharma, S. 2020. Physiology, temperature regulation. Treasure Island (FL): StatPearls Publishing.

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. 2015. Population Trend of the World's Monitored Seabirds, 1950-2010. PLOS ONE, 10(6): e0129342. https://doi.org/10.1371/journal.pone.0129342

Pan, H., Cole, T. L., Bi, X., Fang, M., Zhou, C., Yang, Z., Ksepka, D.T., Hart, T., Bouzat, J. L., Argilla, L. S., Bertelsen, M. F., Boersma, P. D., Bost, C., Cherel, Y., Dann, P., Fiddaman, S. R., Howard, P., Labuschagne, K., Mattern, T., Miller, G., Parker, P., Phillips, R. A., Quillfeldt, P., Ryan, P. G., Taylor, H., Thompson, D. R., Young, M. J., Ellegaard, M. R., Gilbert, M. T. P., Sinding, M. S., Pacheco, G., Shepherd, L. D., Tennyson, A. J. D., Grosser, S., Kay, E., Nupen, L. J., Ellenberg, U., Houston, D. M., Reeve, A. H., Johnson, K., Masello, J. F., Stracke, T., McKinlay, B., Borboroglu, P. G., Zhang, D., Zhang, G. 2019. High-coverage genomes to elucidate the evolution of penguins. GigaScience, 8: 1-17.

Pond, C. M. & Mattacks, C. A. 1985. Cellular structure of adipose tissue in birds. Journal of Morphology, 185: 195-202.

Posada, D. & Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics, 14: 817–818. https://doi.org/10.1093/bioinformatics/14.9.817.

Pruitt, K. D., Tatusova, T. & Maglott, D. R. 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 33: D501-D504.

Pütz, K., Gherardi-Fuentes, C., García-Borboroglu, P., Godoy, C., Flagg, M., Pedrana, J., Vianna, J. A., Simeone, A., Lüthi, B. 2021. Exceptional foraging plasticity in King Penguins (*Aptenodytes patagonicus*) from a recently established breeding site in Tierra del Fuego, Chile. Global Ecology and Conservation, 28: e01669. https://doi.org/10.1016/j.gecco.2021.e01669

Rajagopal, M. C., Brown, J. W., Gelda, D., Valavala, K. V., Wang, H., Llano, D. A., Gillette, R., Sinha, S. 2019. Transient heat release during induced mitochondrial proton uncoupling. Communications Biology, 2: 279. https://doi.org/10.1038/s42003-019-0535-y.

Raymond-Bouchard, I., Goordial, J., Zolotarov, Y., Ronholm, J., Stromvik, M., Bakermans, C., Whyte, L.G. 2018. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. FEMS Microbiology Ecology, 94. https://doi.org/10.1093/femsec/fiy023.

Reynolds, L. M., Dutta, R., Seeds, M. C., Lake, K. N., Hallmark, B., Mathias, R. A., Howard, T. D., hilton, F. H. 2020. FADS genetic and metabolomic analyses identify the Δ5 desaturase (FADS1) step as a critical control point in the formation of biologically important lipids. Scientific Reports, 10: 15873. https://doi.org/10.1038/s41598-020-71948-1.

Raszek, M. & Tisler, I. 2016. Advantages and limitations of genome sequencing. URL https://merogenomics.ca/en/advantages-and-limitations-of-genome-sequencing (accessed 07/06/2022).

Richard, A.J. & Stephens, J.M. 2014. The role of JAK-STAT signalling in adipose tissue function. Biochimica et Biophysica Acta, 1842: 431–439. https://doi.org/10.1016/j.bbadis.2013.05.030.

Richards, C., Cooke, R. S. C. & Bates, A. E. 2021. Biological traits of seabirds predict extinction risk and vulnerability to anthropogenic threats. Global Ecology and Biogeography, 30: 973–986. https://doi.org/10.1111/geb.13279

Ripple, W. J., Wolf, C., Newsome, T. M., Hoffmann, M., Wirsing, A. J., & McCauley, D. J. 2017. Extinction risk is most acute for the world's largest and smallest vertebrates. Proceedings of the National Academy of Sciences, 114 (40): 10678–10683. https://doi.org/10.1073/pnas.1702078114

Ropert-Coudert, Y., Hindell, M. A., Phillips, R., De Broyer, C., Koubbi, P., Griffiths, H. J., Raymond, B., Udekem d'Acoz d', Van de Putte, A. P., Danis, B., David, B., Grant, S., Gutt, J., Held, C., Hosie, G., Huettmann, F., Post, A., Ropert-Coudert, Y. 2014. Biogeographic patterns of birds and mammals. In: The Biogeographic Atlas of the Southern Ocean. Cambridge, Scientific Committee on Antarctic Research *et al.* Scientific Committee on Antarctic Research, 364–87.

Rose, M. D., & Polis, G. A. 1998. The distribution and abundance of coyotes: The effects of allochthonous food subsidies from the sea. Ecology, 79:998–1007.

Rourke, J. L., Muruganandan, S., Dranse, H. J., McMullen, N. M., Sinal, C. J. 2014. Gpr1 is an active chemerin receptor influencing glucose homeostasis in obese mice. Journal of Endocrinology, 222: 201–215. https://doi.org/10.1530/JOE-14-0069.

Rowe, J. A., Litton, C. M., Lepczyk, C. A., Popp, B. N. 2017. Impacts of Endangered Seabirds on Nutrient Cycling in Montane Forest Ecosystems of Hawai'i. Pacific Science, 71: 495–509.

Ruuskanen, S., Hsu, B, & Nord, A. 2019. Endocrinology of thermoregulation in birds in a changing climate. Evolutionary Ecology, 1-48.

Ryeland, J., Weston, M. A. & Symonds, M. R. E. 2017. Bill size mediates behavioural thermoregulation in birds. Functional Ecology, 31: 885-893.

Saarela, S., Reiter, R.J. 1994. Function of melatonin in thermoregulatory processes. Life Sciences, 54: 295–311. https://doi.org/10.1016/0024-3205(94)00786-1.

Schumann, T., König, J., Henke, C., Willmes, D. M., Bornstein, S. R., Jordan, J., Fromm, M. F., Birkenfeld, A. L. 2020. Solute Carrier Transporters as Potential Targets for the Treatment of Metabolic Disease. Pharmacology Reviews, 72: 343–379. https://doi.org/10.1124/pr.118.015735.

Sébastien, M., Giannesini, B., Aubin, P., Brocard, Julie, Chivet, M., Pietrangelo, L., Boncompagni, S., Bosc, C., Brocard, Jacques, Rendu, J., Gory-Fauré, S., Andrieux, A., Fourest-Lieuvin, A., Fauré, J., Marty, I. 2018. Deletion of the microtubule-associated protein 6 (MAP6) results in skeletal muscle dysfunction. Skeletal Muscle, 8: 30. https://doi.org/10.1186/s13395-018-0176-8.

Seddon, P. J. 1999. Brood size determination in penguins. In: Adams, N. J. and Slotow, R. H. (eds) Proc. 22 Int. Ornithol. Congr., Durban: 1162-1177. Johannesburg: BirdLife South Africa.

Seong, H. A., Manoharan, R., Ha, H. 2018. Smad proteins differentially regulate obesity-induced glucose and lipid abnormalities and inflammation via class-specific control of AMPK-related kinase MPK38/MELK activity. Cell Death & Disease, 9: 471. https://doi.org/10.1038/s41419-018-0489-x.

Sharma, D., Sharma, S., Chauhan, P. 2021. Acetylation of Histone and Modification of Gene Expression via HDAC Inhibitors Affects the Obesity. Biomedical and Pharmacology Journal, 14: 153–161.

Shao, D., Liu, Yang, Liu, X., Zhu, L., Cui, Y., Cui, A., Qiao, A., Kong, X., Liu, Yong, Chen, Q., Gupta, N., Fang, F., Chang, Y., 2010. PGC-1 beta-regulated mitochondrial biogenesis and function in myotubes is mediated by NRF-1 and ERR alpha. Mitochondrion, 10: 516–527. https://doi.org/10.1016/j.mito.2010.05.012.

Sheveneli, A. E., Kennett, J. P. & Lea, D. W. 2004. Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. Science, 305 (5691): 1766-70.

Sidder, A. 2016. National Geographic: Antarctica could lose most of its penguins to climate change. https://www.nationalgeographic.com/news/2016/06/adelie-penguins-antarctica-climate-change-population-decline-refugia/ (accessed 01/05/2020).

Silliman, B. R. & Angelini, C. 2012. Trophic cascades across diverse plant ecosystems. Nature Education Knowledge, 3 (10): 44.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31: 3210-3212.

Simon, R., Britsch, S., Bergemann, A. 2011. Ablation of Sax2 gene expression prevents diet-induced obesity. FEBS Journal, 278: 371–382. https://doi.org/10.1111/j.1742-4658.2010.07960.x

Smith, J.S. & Johnson, C.R. 1995. Nutrient inputs from seabirds and humans on a populated coral cay. Marine Ecology Progress Series, 124: 189–200.

Spencer, P.S. & Barral, J. M. 2012. Genetic code redundancy and its influence on the encoded polypeptides. Computational and Structural Biotechnology Journal, 1: e201204006. https://doi.org/10.5936/csbj.201204006.

Stahel, C. D., Nicol, S. C. & Walker, G. J. 1987. Heat production and thermal resistance in the little penguin Eudytula minor in relation to wind speed. Physiological Zoology, 60: 413-423.

State of Antarctic Penguin Report 2022. 2022. Oceanites - Penguins in Antarctica. URL https://oceanites.org/state-of-antarctic-penguins-reports/ (accessed 25/10/2022).

Stonehouse, B. 1967. The general biology and thermal balances of penguins. Advances in ecological research vol. 4: 131-196 (J. B. Cragg, Ed.). London, Academic Press.

Susanti, R., Iswari, R. S., Fibriana, F., Wati, I. 2018. The duck cytochrome oxidase I (COI) gene: Sequence and patterns analysis for potential barcoding tool. Biodiversitas, 19: 997-1003.

Suyama, M., Torrents, D., Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research, 34: W609-612. https://doi.org/10.1093/nar/gkl315.

Swanson, W.J., Nielsen, R. & Yang, Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Moecular Biology & Evolution, 20: 18-20.

Tang, Y., Wu, P., Jiang, S., Selvaraj, J. N., Yang, S., Zhang, G. 2019. A new cold-active and alkaline pectate lyase from Antarctic bacterium with high catalytic efficiency. Applications in Microbiological Biotechnology, 103: 5231–5241. https://doi.org/10.1007/s00253-019-09803-1

Thiébot, J. B., Cherel, Y., Trathan, P. N., Bost, C. A. 2012. Coexistence of oceanic predators on wintering areas explained by population-scale foraging segregation in space or time. Ecology, 93: 12–130.

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., Minh, B. Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research, 44: W232–W235. https://doi.org/10.1093/nar/gkw256.

Unknown author. 2005. Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Springer, Berlin, Heidelberg . https://doi.org/10.1007/3-540-29623-9_8147.

UniProt. 2002. P04458.KRFB_CHICK. URL https://www.uniprot.org/uniprotkb/P04458/entry (accessed 07/07/2022).

Upadhyay, V. R., Dewry, R. K., Sharma, R., Panreiphy G. S. 1977. Role of aquaporins in thermoregulation and water balance. Indian Dairyman, 86-89.

Velarde, E., Anderson, D. W. & Ezcurra, E. 2019. Seabird clues to ecosystem health. Science, 365 (6449): 116–117.

Vianna, J. A., Fernandes, F. A. N., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., Bi, K., Wang-Claypool, C. Y., Lowther, A., Parker, P., Le Bohec, C., Bonadonna, F., Wienecke, B., Pistorius, P., Steinfurth, A., Burridge, C. P., Dantas, G. P. M., Poulin, E., Simison, W. B., Henderson, J., Eizirik, E., Nery, M. F., Bowie, R. C. K. 2020. Genome-wide analyses reveal drivers of penguin diversification. Proceedings of the National Academy of Sciences, 117: 22303–22310. https://doi.org/10.1073/pnas.2006659117.

Vinogradov, A. E. 2003. DNA helix: the importance of being GC-rich. Nucleic Acids Research, 31: 1838–1844. https://doi.org/10.1093/nar/gkg296.

Wada, A. & Suyama, A. 1986. Local stability of DNA and RNA secondary structure and its relation to biological functions. Progress in Biophysics and Molecular Biology, 47: 113–157. https://doi.org/10.1016/0079-6107(86)90012-x.

Wallace, I.M., O'Sullivan, O., Higgins, D.G., Notredame, C. 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Research, 34: 1692– 1699. https://doi.org/10.1093/nar/gkl091.

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., Searle, S., White, S., Vilella, A. J., Fairley, S., Heger, A., Kong, L., Ponting, C. P., Jarvis, E. D., Mello, C. V., Minx, P., Lovell, P., Velho, T. A. F., Ferris, M., Balakrishnan, C. N., Sinha, S., Blatti, C., London, S. E., Li, Y., Lin, Y. C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backström, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A. R., Howard, J., Völker, M., Skinner, B. M., Griffin, D. K., Ye, L., McLaren, W. M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X. S., Olender, T., Lancet, D., Smit, A. F. A., Hubley, R., Konkel, M. K., Walker, J. A., Batzer, M. A., Gu, W., Pollock, D. D., Chen, L., Cheng, Z., Eichler, E. E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S. V., Yang, S. P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E. R., Wilson, R. K. 2010. The genome of a songbird. Nature, 464: 757–762. https://doi.org/10.1038/nature08819

Wiley, S. E., Murphy, A. N., Ross, S. A., van der Geer, P., Dixon, J. E. 2007. MitoNEET is an iron-containing outer mitochondrial membrane protein that regulates oxidative capacity. Proceedings of the National Academy of Sciences, 104: 5318–5323. https://doi.org/10.1073/pnas.0701078104.

Williams, T. D. 1995. The Penguins. Oxford University Press, Oxford, England. ISBN 0-19-854667-X.

Wong, K. E., Szeto, F. L., Zhang, W., Ye, H., Kong, J., Zhang, Z., Sun, X. J., Li, Y. C. 2009. Involvement of the vitamin D receptor in energy metabolism: regulation of uncoupling proteins. American Journal of Physiology-Endocrinology and Metabolism, 296: E820– E828. https://doi.org/10.1152/ajpendo.90763.2008.

World Wide Fund for Nature. 2020. Galápagos penguin facts. https://www.worldwildlife.org/species/galapagos-penguin (accessed 30/03/2020).

Wu, P., Ng, C. S., Yan, J., Lai, Y. C., Chen, C. K., Lai, Y. T., Wu, S. M., Chen, J. J., Luo, W., Widelitz, R. B., Li, W. H., Chuong, C. M. 2015. Topographical mapping of α- and β-keratins on developing chicken skin integuments: Functional interaction and evolutionary perspectives. Proceedings of the National Academy of Sciences, 112: E6770–E6779. https://doi.org/10.1073/pnas.1520566112.

Xiao, Y., Junfeng, H., Tianhong, L., Lu, W., Shulin, C., Yu, Z., Xiaohua, L., Weixia, J., Sheng, Z., Yanyun, G., Guo, L., Min, L. 2006. Cathepsin K in Adipocyte Differentiation and Its Potential Role in the Pathogenesis of Obesity. The Journal of Clinical Endocrinology & Metabolism, 91: 4520–4527. https://doi.org/10.1210/jc.2005-2486.

Xu, Y., Shi, T., Cui, X., Yan, L., Wang, Q., Xu, Xiaoyan, Zhao, Q., Xu, Xiaoxuan, Tang, Q., Tang, H., Pan, D. 2021. Asparagine reinforces mTORC1 signaling to boost thermogenesis and glycolysis in adipose tissues. EMBO Journal, 40: e108069. https://doi.org/10.15252/embj.2021108069.

Yang, L., Tan, Z., Wang, D., Xue, L., Guan, M., Huang, T., Li, R. 2014. Species identification through mitochondrial rRNA genetic analysis. Scientific Reports, 4: 4089.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Research, doi: 10.1093/nar/gkl031.

Yellow-eyed Penguin Trust. 2017. Adaptations for life in the sea. https://www.yellow-eyedpenguin.org.nz/penguins/about-the-yellow-eyed-penguin/adaptations-for-survival/. (accessed 04/06/2020).

Yong, E. 2015. National Geographic: Busting myths about penguin feathers. https://www.nationalgeographic.com/science/phenomena/2015/10/20/busting-myths-about-penguin-feathers/ (accessed 01/05/2020).

Younger, J., Emmerson, L., Southwell, C., Lelliott, P., Miller, K. 2015. Proliferation of East Antarctic Adélie penguins in response to historical deglaciation. BMC Evolutionary Biology, 15: 236. https://doi.org/10.1186/s12862-015-0502-2

Yusof, N. A., Hashim, N. H. F., Bharudin, I. 2021. Cold Adaptation Strategies and the Potential of Psychrophilic Enzymes from the Antarctic Yeast, Glaciozyma antarctica PI12. JoF, 7: 528. https://doi.org/10.3390/jof7070528.

Zhang, D., Hu, P., Liu, T., Wang, J., Jiang, S., Xu, Q., Chen, L. 2018. GC bias lead to increased small amino acids and random coils of proteins in cold-water fishes. BMC Genomics, 19: 315. https://doi.org/10.1186/s12864-018-4684-z.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xiong, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, Jie, Zheng, Q., Xiong, Y., Wang, G., Wang, B., Wang, Jingjing, Fan, Y., da Fonseca, R. R., Alfaro-Núñez, A., Schubert, M., Orlando, L., Mourier, T., Howard, J. T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M. V., Hara, E., Smith, Julia, Farré, M., Narayan, J., Slavov, G., Romanov, M. N., Borges, R., Machado, J. P., Khan, I., Springer, M. S., Gatesy, J., Hoffmann, F. G., Opazo, J. C., Håstad, O., Sawyer, R. H., Kim, H., Kim, K. W., Kim, H. J., Cho, S., Li, N., Huang, Y., Bruford, M. W., Zhan, X., Dixon, A., Bertelsen, M. F., Derryberry, E., Warren, W., Wilson, R. K., Li, S., Ray, D. A., Green, R. E., O'Brien, S. J., Griffin, D., Johnson, W. E., Haussler, D., Ryder, O. A., Willerslev, E., Graves, G. R., Alström, P., Fjeldså, J., Mindell, D. P., Edwards, S. V., Braun, E. L., Rahbek, C., Burt, D. W., Houde, P., Zhang, Y., Yang, H., Wang, Jian, Avian Genome Consortium, Jarvis, E. D., Gilbert, M. T. P., Wang, Jun, Ye, C., Liang, S., Yan, Z., Zepeda, M. L., Campos, P. F., Velazquez, A. M. V., Samaniego, J. A., Avila-Arcos, M., Martin, M. D., Barnett, R., Ribeiro, A. M., Mello, C. V., Lovell, P. V., Almeida, D., Maldonado, E., Pereira, J., Sunagar, K., Philip, S., Dominguez-Bello, M. G., Bunce, M., Lambert, D., Brumfield, R. T., Sheldon, F. H., Holmes, E. C., Gardner, P. P., Steeves, T. E., Stadler, P. F., Burge, S. W., Lyons, E., Smith, Jacqueline, McCarthy, F., Pitel, F., Rhoads, D., Froman, D. P. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science, 346: 1311–1320. https://doi.org/10.1126/science.1251385.

Zhang, J., Terán, G., Popa, M., Madapura, H., Ladds, M. J. G. W., Lianoudaki, D., Grünler, J., Arsenian-Henriksson, M., McCormack, E., Rottenberg, M. E., Catrina, S. B., Laín, S., Darekar, S. 2021. DHODH inhibition modulates glucose metabolism and circulating GDF15, and improves metabolic balance. iScienc, 24: 102494. https://doi.org/10.1016/j.isci.2021.102494.

Zhang, X., Cvetkovska, M., Morgan-Kiss, R., Hüner, N. P. A., Smith, D. R. 2020. Is Gene Duplication Driving Cold Adaptation in the Antarctic Green Alga Chlamydomonas Sp. UWO241? SSRN Journal. https://doi.org/10.2139/ssrn.3732378.

Zhao, P., Wong, K. I., Sun, X., Reilly, S. M., Uhm, M., Liao, Z., Skorobogatko, Y., Saltiel, A.R. 2018. TBK1 at the Crossroads of Inflammation and Energy Homeostasis in Adipose Tissue. Cell, 172: 731-743.e12. https://doi.org/10.1016/j.cell.2018.01.007.

Zhao, Z. D., Yang, W. Z., Gao, C., Fu, X., Zhang, W., Zhou, Q., Chen, W., Ni, X., Lin, J.-K., Yang, J., Xu, X. H., Shen, W. L. 2017. A hypothalamic circuit that controls body temperature. Proc. Natl. Acad. Sci. U.S.A. 114, 2042–2047. https://doi.org/10.1073/pnas.1616255114.

Zhao, Z.-D., Yang, W.Z., Gao, C., Fu, X., Zhang, W., Zhou, Q., Chen, W., Ni, X., Lin, J.-K., Yang, J., Xu, X.-H., Shen, W.L., 2017. A hypothalamic circuit that controls body temperature. Proceedings of the National Academy of Sciences, 114: 2042–2047. https://doi.org/10.1073/pnas.1616255114.

Zhu, R., Wang, Q., Ding, W., Wang, C., Hou, L., Ma, D. 2014. Penguins significantly increased phosphine formation and phosphorus contribution in maritime Antarctic soils. Scientific Reports, 4: 7055. https://doi.org/10.1038/srep07055

Zhou, H., Wang, H., Yu, M., Schugar, R.C., Qian, W., Tang, F., Liu, W., Yang, H., McDowell, R.E., Zhao, J., Gao, J., Dongre, A., Carman, J.A., Yin, M., Drazba, J.A., Dent, R., Hine, C., Chen, Y.-R., Smith, J.D., Fox, P.L., Brown, J.M., Li, X., 2020. IL-1 induces mitochondrial translocation of IRAK2 to suppress oxidative metabolism in adipocytes. Natural Immunology, 21: 1219–1231. https://doi.org/10.1038/s41590-020-0750-1.