UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Biology

Ph.D. COURSE IN: Biosciences

CURRICULUM: Evolution, Ecology and Conservation

SERIES XXXIV

Comparative genomics and transcriptomics reveal the genetic basis for adaptation to the extreme Antarctic environment in the Emperor penguin (*A. forsteri*)

Coordinator: Prof. Ildikò Szabò

Supervisor: Prof. Lorenzo Zane

Co-Supervisor: Prof. Emiliano Trucchi

Ph.D. student: Federica Pirri

Index

Abstract

The eco-evolutionary history of penguins is profoundly influenced by their shift from temperate to cold environments. Breeding only in Antarctica during the winter, the Emperor penguin appears as an extreme outcome of this process, with unique features related to insulation, heat production and energy management. However, whether this species actually diverged from a less cold-adapted ancestor, thus more similar in ecology to its sister species, the King penguin, is still an open question. As the Antarctic niche shift likely resulted in vast changes in selective pressure experienced by the Emperor penguin, the identification and relative quantification of the genomic signatures of selection, unique to each of these sister species, could answer this question.

Applying a suite of phylogeny-based methods on 7,651 orthologous gene alignments of seven penguins and 13 other birds, we identified a set of candidate genes showing significantly different selection regimes either in the Emperor or in the King penguin lineage. Among the candidate genes under selection in the Emperor penguin, four genes (TRPM8, LEPR, CRB1, and SFI1) were identified before in other cold adapted vertebrates, while, on the other hand, 161 genes can be assigned to functional pathways relevant to cold adaptation (cardiovascular system, lipid, fatty acid and glucose metabolism, insulation, among the others).

In order to detect signatures of more recent and ongoing selection in the two penguin species of the *Aptenodytes* genus, we also performed genome-wide analyses of selection, through the application of haplotype-based methods, on 48 individuals of Emperor and King penguins.

Both the selection analyses applied, over short and long evolutionary time, revealed a more pervasive selective shift in the Emperor penguin, supporting the hypothesis that its extreme cold adaptation is a derived state from a more King penguin-like ecology.

Moreover, also the haplotype-based methods detected genetic traits, associated with positively selected regions, that are potentially involved in the Emperor penguin adaptation to the extreme cold conditions (e.g., candidate genes involved in adipose tissue and lipid metabolism, cold-induced thermogenesis and cardiovascular functions).

Overall, the results of the selection analyses show that extreme cold adaptation in the Emperor penguin largely involved unique genetic options which, however, affect metabolic and physiological traits common to other cold-adapted homeotherms.

When a population colonizes a new environment, gene expression becomes crucial in guaranteeing population persistence and can mediate phenotypic plasticity and contribute to adaptive divergence.

To investigate the relevance of gene expression changes in adaptation to the extreme cold Antarctic environment, we explored transcriptomic differences at intraspecific and interspecific levels by analyzing QuantSeq 3'mRNA-Seq data from a large number of tissue samples (100 samples from 5 tissues of 10 individuals from natural populations of each species). Specifically, our analyses concerned tissues that could exhibit the major genetic differences for cold adaptation comparing the two species: skin (thermal insulation), liver (lipid and fatty-acid metabolism), brain (cold tolerance), muscle (thermogenesis) and kidney (osmoregulation).

In general, the expression profiles revealed a characterizing tissue-clustered pattern and a number of differentially expressed genes which, in the Emperor penguin, could be the candidates underlying its relevant adaptations to the Antarctic lifestyle. We found wide scale changes in gene transcripts regulating metabolic pathways for lipids, insulin resistance and sensitivity, storage of glucose, cold pain tolerance and fat cell differentiation in the Emperor penguin.

We especially focused on the transcriptional profile of the muscle given its critical role in whole body energy metabolism and temperature homeostasis in birds. Our findings on the most up-regulated differentially expressed genes in the muscle, together with the enrichment of GO terms and KEGG pathways, confirm a thermogenic role of this tissue in both the *Aptenodytes* penguins but probably involving different biological pathways. The non-shivering thermogenesis process appears to be the dominant mechanism of heat production in the Emperor penguin.

Moreover, we used total mRNA-Seq data from 5 tissues of 3 individuals per species to *de novo* assemble the first reference transcriptome of the Emperor and the King penguin. Our findings suggest both the transcriptomes cover a wide range of protein-coding sequences, encompassing more than 84% of genes in the Aves orthologs database and the backmapping rate of 82% (Emperor penguin) and 79% (King penguin) is comparable to that of other *de novo* transcriptomes in birds. These values indicate great levels of completion considering that the assemblies were generated from a small number of tissues and a single developmental stage.

1. Introduction

Extreme temperature conditions are known to drive a range of morpho-anatomical, behavioral, physiological and ecological adaptations that, combined with specific endocrine responses, make a wide variety of species able to live in the harshest regions of the planet.

Several eukaryotic organisms display changes in gene expression during exposure to the cold, particularly affecting all those proteins that aid the organism in tolerating extremely cold temperatures by controlling extracellular ice formation and minimizing osmotic stress during freezing (Thieringer et al., 1998); some relevant examples are represented by the cold shock proteins identified in yeast, the antifreeze proteins in Antarctic fish, ice-nucleating proteins, thermal hysteresis proteins and cryoprotectants (Costanzo et al., 1995; Storey and Storey, 1988; Devries and Hew, 1990; Johnston, 1990).

In this regard, glucocorticoid hormones activate the adrenoreceptors and thus the glycogenesis that produces anti-freezing compounds (Storey and Storey, 1988); another strategy to avoid internal ice formation consists in reducing the volume of circulating blood (Vanhoutte et al., 2001; Flavahan, 1991; Thompson, 1977) or lowering the freezing point of body fluids (Johnston, 1990).

At the molecular level, specific examples include changes in membrane fluidity, through the alteration of phospholipids saturation, and modifications in the protein translation machinery of the cell (Thieringer et al., 1998). In addition, whereas the microtubules of mammals depolymerize at low temperatures, Antarctic fish and polar foraminifera show cold-stable microtubules that are functional also at a lower temperature range (Bowser and DeLaca, 1985; Williams et al., 1985; Clarke et al., 1991).

From a biochemical point of view, the major challenge is represented by the enzymes which must overcome the exponential decay of chemical reaction rates as the temperature is lowered (Åqvist et al., 2017). Cold-adapted enzymes have evolved a more flexible surface and an enhanced plasticity to allow a lower activation enthalpy for the chemical reaction (Åqvist et al., 2017; D'Amico et al., 2002).

As stated by Allen's rule, animals adapted to cold climates have evolved shorter limbs and bodily appendages than animals adapted to warm climates. Therefore, they are characterized by lower surface area-to-volume ratios in order to minimize heat loss from the surface area (Allen, 1877).
Indeed, keeping a warm body temperature is the major challenge for cold adaptation in homeotherms (Scholander, 1955). This is achieved by three main strategies: *i*) reducing heat dissipation from the body surface by raising hairs or feathers, by decreasing peripheral circulation, by balling up or huddling (Scholander, 1955; Yudin et al., 2017); *ii*) increasing heat production via shivering and non-shivering thermogenesis through brown adipose tissue adipocytes (Yudin et al., 2017); *iii*) temporarily hibernating to survive low body temperature (Tattersall et al., 2012).
Many mammals and birds exhibit a particular vascular pattern, made up of arterio-venous countercurrent heat exchangers localized in the extremities and sometimes in the tail (Scholander, 1955). That is crucial for heat conservation and thus as adaptation to cold.
Common metabolic strategies to handle severe temperature conditions include: accumulating huge fat reserves and adipose deposits under the skin or around organs through a lipid-rich diet, lowering the metabolic activity and generally reducing body functions (Solomonov et al., 2009; Liu et al., 2014).
Although different organisms have developed different adaptive mechanisms to survive at low temperatures, there are common adaptations that have been found in most of the cold-adapted organisms which represent universal responses to cold. In addition, hibernating animals implement further specific biochemical alterations to regulate heart and brain membrane fluidity (Aloia et al., 1974; Goldman, 1975).

All the species well adapted to live in extreme environments are expected to have accumulated essential genetic changes that help them to cope with the cold (Yudin et al., 2017). However, genetic bases which make life of warm-blooded vertebrates possible in the extreme cold temperature conditions remain largely unknown (Li et al., 2014).
Some Arctic and Antarctic adaptation studies have analyzed genomic divergence between closely related species adapted to distinct environmental conditions in order to discover candidate genes that are crucial to cold adaptation (Liu et al., 2014; Kumar et al., 2015; Librado et al., 2015; Lynch et al., 2015). Although they revealed only a few common genes involved in cold tolerance in

different species, some general mechanisms and biochemical pathways relevant to cold adaptation emerged.
All the common adaptations to cold were related to the reorganization of the cardiovascular system, increased thickness and strength of the skin, brown-fat deposits for insulation, lipid and fatty-acid metabolism, insulin signaling and temperature sensation.
Overall, this overlap indicates that Arctic and Antarctic environments severely limit fitness landscapes, with only a small number of evolutionary strategies compatible with survival (Librado et al., 2015).

Among tetrapod vertebrates, birds are known for their fast adaptation rate, which sustained the colonization of a huge diversity of environments, including polar ones (Zhang et al., 2014).
Neoaves, the most diverse avian clade, underwent a rapid global expansion and radiation within a relatively short period of time (10 to 15 million years), followed by an ecological diversification of several lineages into a wide range of niches during the K-Pg transition (Jarvis et al., 2014). Sphenisciformes (penguins) is the only living clade of flightless birds which successfully colonized a wide range of ocean temperatures and depths, across the Southern Hemisphere. The current most accredited hypothesis on the origin of extant penguins, posit that they arose in tropical–warm temperate waters in the early Miocene (Stonehouse, 1975; Vianna et al., 2020) and secondary colonized Antarctica, after permanent ice sheets were established around 9.8 Mya (Gavryushkina et al., 2017; Vianna et al., 2020). The geographical distribution of fossils further suggests that *Aptenodytes* and *Pygoscelis*, the most polar-adapted penguin taxa, took advantage of a new environment generated from the expansion of Antarctic ice sheets in the Middle Miocene (Ksepka et al., 2006).

The Emperor penguin (*Aptenodytes forsteri*) is the only warm-blooded vertebrate that lives and breeds on stable fast ice, in the harshest Antarctic winter conditions, when temperatures drop regularly to −50 °C and wind speeds can exceed 100 km/h (Owens and Zawar-Reza, 2015) (Fig. 1).
They exhibit many unique features related to insulation, heat production and energy management in comparison with other birds (Scholander, 1955; Rowland et al., 2015). These birds have evolved

to face numerous challenges, such as severe cold and limited food availability during the winter, and profound seasonal changes in the length of daylight during the year (Blix, 2016; Goldsmith and Sladen, 1961).

To live in such extreme environment, Emperor penguins have developed densely packed waterproof and insulating rigid feathers (Watson, 1883; Taylor, 1986), a thick layer of subcutaneous fat (Le Maho et al., 1976), long-term fasting (up to 115 days in males), an efficient energy storage management (Groscolas, 1990; Groscolas and Robin, 2001; Cherel et al., 1994), an enhanced thermoregulation through vasoconstriction and arterio-venous heat exchange countercurrent systems in order to maintain a high core body temperature (Butler and Jones, 1997; Thomas and Fordyce, 2007), a social thermoregulatory behavior consisting in huddling together in compact groups (Le Maho et al., 1976) and a high resting metabolic rate (Le Maho et al., 1976). On the other end, its closest relative, the King penguin (*A. patagonicus*), breeds exclusively in year-round ice-free sub-Antarctic islands lacking the extreme cold environmental pressures, and hence most of the relevant adaptations featured by the Emperor penguin (Fig. 1).

Their clear ecological divergence, in contrast with a very recent history of phylogenetic separation (estimated around 1-2 Mya, Gavryushkina et al., 2017), makes this system an ideal model to investigate the genomic basis of extreme cold adaptations and understand the evolutionary drivers of diversification in these two penguin species.

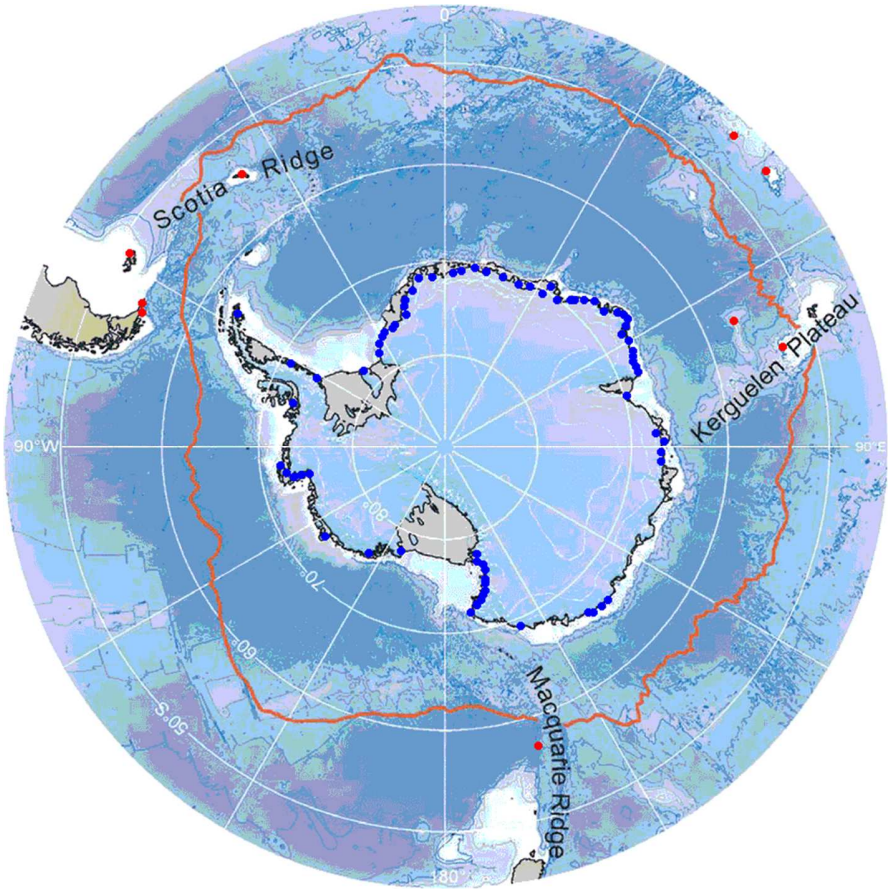

Figure 1. Distribution map of the Emperor (in blue) and the King penguin (in red) colonies. Modified from Collins and Rodhouse (2006).

Aim of the thesis

Although penguins are a relatively well-studied group, most previous studies focused on ecological (Black et al., 2018; Cherel et al., 2018; Hong et al., 2021), physiological (Fahlman et al., 2005; Groscolas and Robin, 2001; Kooyman, 2009), behavioral (Gilbert et al., 2006), or phylogenetic (Ksepka et al., 2006; Vianna et al., 2020; Pan et al., 2019) aspects of their biology, with a particular emphasis on the association between biological patterns and climate change (Younger J.L. et al., 2015; Younger J. et al., 2015; Trucchi et al., 2014; Trucchi et al., 2019; Cristofari et al., 2016; Cristofari et al., 2018) while only a few have explored genome-wide evolutionary processes among penguins (Li et al., 2014; Zhao et al., 2015; Pan et al., 2019; Vianna et al., 2020; Ramos et al.,

2018) or between penguins and other birds (Zhang et al., 2014; Jarvis et al., 2014; Borges et al., 2015).

Therefore, little is currently known about the molecular genomic basis for the unique morphological and physiological adaptations of penguins compared to other aquatic and terrestrial birds (Pan et al., 2019; Li et al., 2014; Vianna et al., 2020).

In this thesis, we explored the evolution of the Emperor penguin using a comparative genomic and transcriptomic framework in order to:

i) detect signatures of differential selection in the comparison with its less cold-adapted sibling species, the King penguin, over short and long evolutionary time;

ii) reveal genetic traits potentially associated with processes that are relevant for adaptation to cold;

iii) investigate how levels of gene expression contribute to adaptation to a more extreme environment as compared to the King penguin.

The Emperor penguin's extreme cold adaptation has been proposed to be a derived state from a less cold-adapted ancestor, possibly more ecologically comparable to the King penguin (Vianna et al., 2020). Such a dramatic ecological shift should have left a distinct signal of selection change across the Emperor penguin's genome, with some genes becoming the new targets of positive selection and others being released from the selective pressures characterizing the former habitat.

Identifying the genetic basis and exploring the phenotypic plasticity that underlie relevant traits for adaptive changes at the physiological, morphological and molecular level in the Emperor penguin, may be critical as it is predicted that future climate change will lead to ecosystem-wide changes.

Emperor penguins are sensitive indicators of local environmental changes as they breed exclusively on sea ice (Cristofari et al., 2016). This species is especially vulnerable to global warming as illustrated by the vanishing of several colonies (Trathan et al., 2011; Ancel et al., 2014) and the dramatic drop in breeding success and population size of others (Kooyman and Ponganis, 2014), in response to sea ice retreat, during the last few decades.

2. Selection-driven adaptation to the extreme Antarctic environment in the Emperor penguin

This section includes the manuscript:

Federica Pirri, Lino Ometto, Silvia Fuselli, Flávia A. N. Fernandes, Lorena Ancona, Céline Le Bohec, Lorenzo Zane, Emiliano Trucchi;

"Selection-driven adaptation to the extreme Antarctic environment in the Emperor penguin";

BioRxiv. doi: https://doi.org/10.1101/2021.12.14.471946.

Introduction

Adaptation to severe Arctic and Antarctic temperatures is rare among terrestrial vertebrates and restricted to warm-blooded lineages (Storey and Storey, 1992; Blix, 2016). The main challenge for adaptation to extreme cold is keeping adequately high core body temperature, which in homeotherms can be obtained by the combination of several physiological, morphological, and behavioral responses (Allen, 1877; Scholander, 1955; Cannon and Nedergaard, 2010; Tattersall et al., 2012; Blix, 2016; Roussel et al., 2020): *i)* Minimizing heat dissipation from the body surface *e.g.,* by reducing the surface area-to-volume ratio, by raising insulating hair or feathers, by decreasing peripheral circulation, and/or by balling up or huddling; *ii)* Increasing heat production via shivering and nonshivering thermogenesis through brown adipocytes; *iii)* Temporarily hibernating.

In contrast, the genetic basis of cold adaptation in homeotherms is not well understood. Only few candidate genes were found in common across cold-adapted mammals and birds (Yudin et al., 2017; Wollenberg Valero et al., 2014), suggesting that cold adaptation is likely the result of selection on different genes, which are nevertheless relevant to the same set of physiological and metabolic functions. For example, only four candidate genes, which are related to cardiovascular

function, were found in common in three out of six mammals dwelling in/near the Arctic or Antarctica (Yudin et al., 2017), even if different candidate genes are involved anyway in heart and vascular development and regulation (Liu et al., 2014; Vianna et al., 2020). Different genes acting in fatty-acid metabolism have been identified to be under positive selection in the polar bear and the Arctic fox (Liu et al., 2014; Castruita et al., 2020; Kumar et al., 2015). In mammals, the molecular mechanisms underpinning heat production via non-shivering thermogenesis have been extensively investigated and have been mainly linked to UCP-1, a mitochondrial uncoupling protein involved in the generation of heat in brown adipose tissue (Lowell and Spiegelman, 2000). Non-shivering thermogenesis is instead not well understood in birds, where it is not clear whether the avian homologue of UCP-1 (i.e., avian uncoupling protein, avUCP) has a similar thermogenic role or is mainly required against oxidative stress (Talbot et al., 2004). However, the same non-shivering thermogenic pathways could be used in birds (Tigano et al., 2018), and even in ectothermic vertebrates like reptiles (Akashi et al., 2016).

Birds are known for their fast adaptation rate, which allowed the colonization of a huge diversity of environments, including polar ones (Zhang et al., 2014). In particular, the clade of penguins, which likely originated in temperate environments, successfully diversified in the cold Antarctic and sub-Antarctic ecosystems (Pan et al., 2019; Vianna et al., 2020), featuring unique adaptations for insulation, heat production and energy management (Scholander, 1955; Rowland et al., 2015). Our understanding of the underlying genetic determinants of such adaptations is still rather scarce. Testing about one third of the total genes across all penguin genomes, Vianna et al. (2020) identified blood pressure, cardiovascular regulation, and oxygen metabolism in muscles as functions, potentially involved in thermoregulation, which have been the targets of positive selection. Analyses of *Spheniscus* and *Pygoscelis* mitochondrial genomes also revealed a correlation between the pattern of diversity of the ND4 gene and sea surface temperature, suggesting this gene is involved in climate adaptation (Ramos et al., 2018). Moreover, Emperor and Adelie penguin genomes show the highest rate of duplication of β-keratin genes as compared to non-penguin birds, suggesting important changes in feathers and skin structure during their evolution to increase core body insulation (Li et al., 2014). Alternative gene pathways related to

lipid metabolism and phototransduction appeared also to have been under selection in these two species (Li et al., 2014).

The Emperor penguin (*Aptenodytes forsteri*) is the only warm-blooded vertebrate thriving and breeding during the harshest Antarctic winter, facing profound seasonal changes in daylight length as well as severe cold and wind conditions (Blix, 2016; Goldsmith and Sladen, 1961). To withstand such hostile environment, the Emperor penguin shows multiple morphological, physiological, behavioral adaptations, like improved thermoregulation systems in the head, wings, and legs (Frost et al., 1975; Thomas and Fordyce, 2008), and efficient energy storage management system for long-term fasting (Groscolas, 1990; Cherel et al., 1994; Groscolas and Robin, 2001). Conversely, its sister species, the King penguin (*A. patagonicus*), breeds exclusively in year-round ice-free sub-Antarctic islands and in Tierra del Fuego. Extreme cold adaptation in the Emperor penguin has been suggested to be a derived feature from a less cold-adapted ancestor, likely more ecologically similar to the King penguin (Vianna et al., 2020). Such marked ecological transition should have left a clear signature of selection change across the genome of the Emperor penguin, with some genes becoming the novel targets of positive selection while others getting released from previous selective pressures associated with the ancestral habitat.

Here, we apply phylogeny-based tests to identify genes that markedly changed in their selection regime during the evolutionary history of the Emperor penguin, using its less cold-adapted sister species, the King penguin, as a control. If the common ancestor ecology was similar to the King penguin one (*i.e.*, not so cold-adapted), we expect a more intense selection shift across the Emperor penguin genome, with genes under positive selection related to adaptations to cold. By using a phylogenetic framework including seven species of penguins and 13 other birds, we compare the pattern of molecular evolution (Yang, 2007; Wertheim et al., 2015) between Emperor and King penguins across 7,651 orthologous genes and explore the gene ontology terms to identify the molecular functions that may have undergone positive selection in the Emperor penguin. To allow for a broader comparison with other cold-adapted vertebrates, we also investigated the overlap between the biological functions of the candidate genes identified in the Emperor penguin and the metabolic and physiological functions related to cold adaptation compiled from previous studies.

Materials and methods

Orthologous coding sequences identification

In order to test for selective signatures in coding sequences of Emperor and King penguins, we implemented a comparative phylogenomic analysis. We selected at least one species for each extant penguin genus based on the phylogeny of Pan et al. (2019), and other bird species representing the clade of Core Waterbirds (which includes penguins), the tropicbirds (one species only), and some more distant species from the Core Landbirds according to Jarvis et al. (2014). The resulting dataset included seven penguin species (*Eudyptula minor minor*, *Spheniscus magellanicus*, *Eudyptes chrysolophus*, *Pygoscelis papua*, *Pygoscelis adeliae*, *Aptenodytes patagonicus* and *Aptenodytes forsteri*) and 13 additional bird species (*Phaethon lepturus*, *Eurypyga helias*, *Gavia stellata*, *Fulmarus glacialis*, *Phalacrocorax carbo*, *Nipponia nippon*, *Egretta garzetta*, *Pelecanus crispus*, *Haliaeetus leucocephalus*, *Tyto alba*, *Cariama cristata*, *Corvus brachyrhynchos* and, as a more distant outgroup, *Opisthocomus hoazin*). All coding sequences (CDS) of each of these twenty bird species were downloaded from GigaDB and Genbank (*SI Appendix*, Table S1).

One-to-one orthologs were identified by applying a reciprocal best-hit approach using pairwise BLAST searches with an e-value cutoff of 1e-15, a nucleotide sequence identity of at least 70%, and a fraction of aligned CDS of at least 60% (Savini et al., 2021). Only CDS longer than 150 bp that were a reciprocal best-hit between the Emperor penguin and the other species were retained. The orthologous gene sequences were then aligned with MAFFT (Madeira et al., 2019) and the alignments trimmed to maintain the open reading frame using a custom perl script. The resulting nucleotide alignments were re-aligned using the PRANK algorithm (Löytynoja, 2013) in TranslatorX (Abascal et al., 2010), which aligns protein-coding sequences based on their corresponding amino acid translations. Since the results of *dN/dS* analysis could have been affected by poorly aligned regions or by regions that were too different to be considered truly orthologs, we removed CDS alignments that included internal stop codons and used a custom perl script to remove problematic regions as in Han et al. (2009) (see also Ramasamy et al., 2016). Furthermore, as phylogenetic-based selection tests are not able to properly deal with alignment gaps (Yang,

2007), we filtered all the alignments with a custom perl script and kept only sites that were unambiguously present in at least 16 of the 20 sequences and always present in both our species of interest (*i.e.,* King and Emperor penguins). The minimum length of an alignment for subsequent analyses after gap removal was set to 150 bp (50 codons).

All the scripts employed in our pipeline are made available in the Open Science Framework data repository at:

https://osf.io/zjbdx/?view_only=1b0935ab0e5a447ab3b68dad1aa5a3bd (Savini et al., 2021).

Identification of selection regime shifts

We first used CODEML in the PAML package (Yang, 2007) to estimate synonymous (*dS*) and nonsynonymous (*dN)* substitution rates and to identify genes which are characterized by lineage-specific ω values (*i.e.*, *dN/dS*, which is a proxy for the level of past selective pressure in the gene). In particular, we separately investigated the two scenarios of different ω in the King or in the Emperor lineage. If one branch in the phylogeny shows a value of ω (ω_i) significantly larger than in the other branches (*i.e.*, ω_b; background ω), such a foreground lineage may have been targeted by positive (Darwinian) or relaxed selection (allowing the accumulation of non synonymous substitutions), whereas when ω_i is lower than ω_b, the foreground lineage may have evolved under stronger selective constraints (*e.g.*, purifying selection).

We first ran the two-ratio branch model (one ω for the foreground branch, another ω for the background branches; set parameters "model=2, NSsites=0, fix_omega=0") and the one-ratio branch model (one ω estimate for all branches, as null model; set parameters "model=0, NSsites=0, fix_omega=0") on the unrooted phylogenetic tree of the species of interest. We determined the topology of such a tree (Fig. 1A) by manually combining the total evidence nucleotide tree of the avian family (Jarvis et al., 2014) and the phylogenomic reconstruction of penguins (Pan et al., 2019). The two models (two-ratios vs. one-ratio) were compared by likelihood ratio tests (LRTs). False discovery rates (FDR) were computed using the qvalue package (Storey at al., 2017) and the p.adjust function in R (R Core Team, 2013) using the Benjamini-Hochberg procedure to adjust for multiple testing. An FDR or *p*-adjusted significant threshold of 0.05 was used.

We then used a branch-site model to test for sites under selection in the candidate genes from the previous test. The parameters for the null model were set as "model=2, NSsites=2, fix_omega=1", while the parameters for the alternative model were set as "model=2, NSsites=2, fix_omega=0". LRT and FDR were computed as for the branch model tests. To control for misalignments that could have biased the results, we visually checked the sequence alignment of all candidate genes under selection using MEGA7 (Kumar et al., 2016).

We also applied aBSREL from the HyPhy package, a branch-site test that includes an adaptive branch-site random effects likelihood model (Smith et al., 2015; Pond et al., 2005), to our set of orthologous coding sequences. In contrast to the branch-site test in CODEML, which assumes 4 ω-rate classes for each branch and assigns each site to one of these classes, the aBSREL test uses AIC_c to infer the optimal number of ω-rate categories per branch, not making the assumption that all branches exhibit the same degree of substitution rate heterogeneity (Smith et al., 2015). Although we expect a broad overlap between CODEML and aBSREL results, a higher sensitivity should characterize the latter approach (Smith et al., 2015). Signatures of positive selection were searched by setting *a priori* the King and the Emperor lineage as *test* branches in the phylogeny. LRT was performed by comparing the full model to a null model where branches were not allowed to have rate classes of $\omega > 1$. A Benjamini-Hochberg correction was used to control the probability of making false discoveries and only tests with adjusted p-values < 0.05 were considered significant.

Beside presenting novel drivers of selection, the major ecological shift occurring in the Emperor penguin lineage should have also released some of the selective constraints characterizing the ancestral ecological niche. As a consequence, some genes could show a signature of relaxed selection in this lineage, a higher number than in the King penguin. To test for relaxation of selective constraints on a specific lineage we used RELAX from the HyPhy package, a general hypothesis testing framework that determines whether the strength of natural selection has been relaxed or intensified along a set of test branches defined *a priori* in a phylogenetic tree (Wertheim et al., 2015; Pond et al., 2005). It estimates a selection intensity parameter K, in which a significant $K > 1$ indicates intensification in the selection strength, whereas a significant $K < 1$ indicates relaxation in the strength of selection in the test branches (Wertheim et al., 2015). We tested whether selection pressure has increased or decreased in either the King or in the Emperor lineage

as compared to the rest of the phylogeny. In the null model, the selection intensity parameter *K* was set to 1 for all the branches of the phylogenetic tree, whereas in the alternative model the parameter *K* was inferred for every tested branch. The increase or relaxation of selection was validated by a LRT with 1 degree of freedom. Again, the Benjamini-Hochberg procedure was used to adjust for multiple testing with adjusted p-values < 0.05 considered significant.

Functional characterization of candidate genes under selection

To test whether the set of candidate genes for positive selection in the Emperor penguin lineage has a relevance for cold adaptation, we tested these genes for functional GO terms enrichment by using the g:GOSt function in g:Profiler (Raudvere et al., 2019). GO terms were assigned to candidate genes based on the Ensembl GO predictions for the flycatcher (*Ficedula albicollis*). Significantly enriched categories included at least two genes, and the Benjamini-Hochberg method was used for multiple testing correction to estimate significance (at $p < 0.05$). We used REVIGO (Supek et al., 2011) to summarize the resulting lists of GO terms in order to obtain a non-redundant and more easily interpretable set of GO terms. GO terms enrichment was performed on two lists of genes: *i)* candidate genes supported by both CODEML and aBSREL and *ii)* candidate genes supported by either CODEML or aBSREL.

To compare our results with the recent literature on genetics of cold adaptation, we compiled a list of biological/molecular functions characterizing the candidate genes for cold adaptation retrieved in previous studies in vertebrates (Table 1). The list included: *cardiovascular activity and regulation, skin thickness, immunity, lipid and fatty acid metabolism, glucose (including insulin) metabolism, thyroid hormones, non-shivering thermogenesis, shivering thermogenesis, response to oxidative stress, stress response, homeostasis, circadian rhythm, phototransduction, mitochondrial activity, feathers development*. We checked whether any of the GO terms enriched in candidate genes supported by either CODEML or aBSREL could be assigned to any of the 15 biological/molecular functions listed above. In addition, we assigned, whenever possible, genes from this list to the same biological/molecular functions, using the gene function description from the human gene database GeneCards (Stelzer et al., 2016).

Results

Signature of selection shift in the Emperor penguin lineage

We identified 7,651 orthologous coding sequences across seven penguin species and 13 other birds, corresponding to about 50% of the total number of genes in an avian genome (Zhang et al., 2014). Across all of the tests, by applying a significance threshold of FDR < 0.05, we consistently identified more candidate genes which underwent a shift in their selection regime (intensified or relaxed) in the Emperor penguin lineage than in the King penguin one (Fig. 1B). Even though we found a much larger number of genes putatively under selection using the aBSREL model (*SI Appendix*, Table S2, S3), the overlap between the CODEML (with $\omega_i > \omega_b$) branch model (*SI Appendix*, Table S4) or RELAX (with $K > 1$; *SI Appendix*, Table S5, S6) with aBSREL was on average 80% (Fig. 1B, inset).

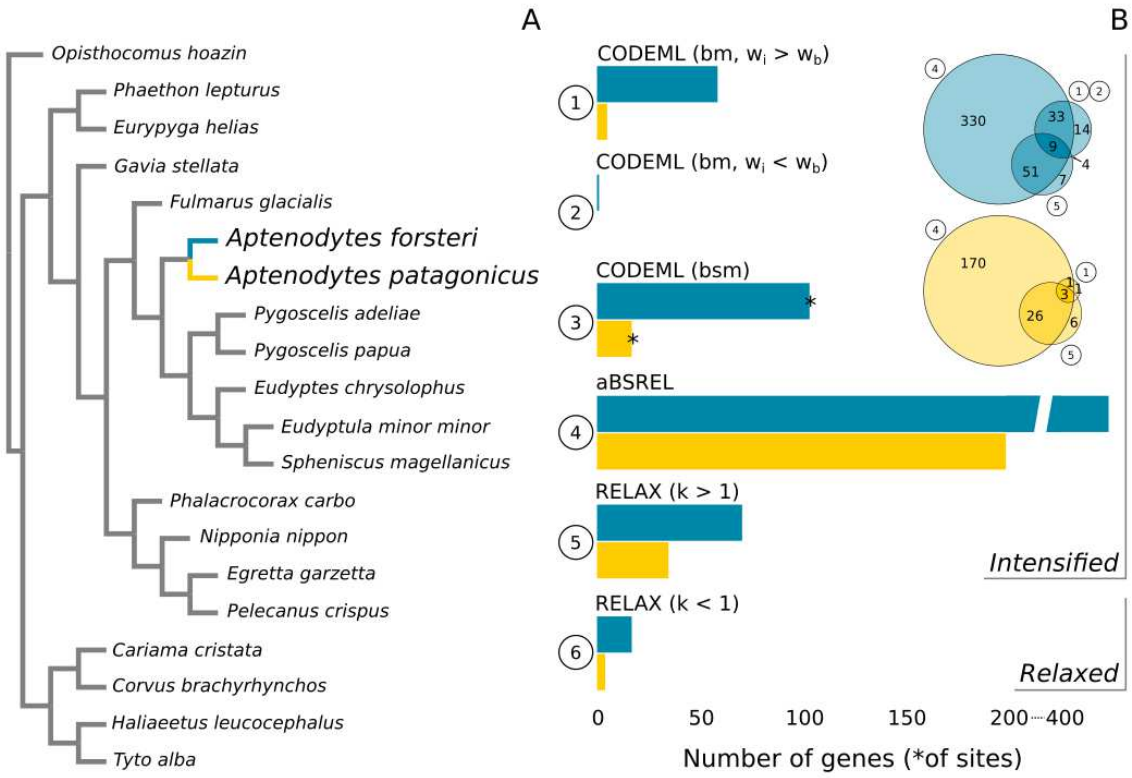

Figure 1. A. Phylogenetic tree based on Jarvis et al. 2014 and Pan et al. 2019. The Emperor and the King penguin are highlighted in blue and yellow, respectively. Note that branch length is not to scale. **B.** Comparison between Emperor and King penguins for genes with FDR > 0.05 in each of the tests performed; bm: branch model; bsm: branch-site model; w_i: ω in the target species; w_b: background ω; For sake of completeness, we also show the genes putatively under selection according to RELAX (with $K > 1$). **Inset**. Venn diagram showing the overlap among CODEML (bm), aBSREL, and RELAX ($K > 1$). Note that the total number of genes is different between the Emperor and the King penguins and the size of the circles scales to the maximum in each of the two graphs. The overlap between CODEML (bm) or/and aBSREL with RELAX ($K < 1$) is 3, 1, and 1 gene, respectively (not shown).

Using the CODEML branch test, we found 59 candidate genes with signals of positive or relaxed selection in the Emperor penguin, showing a ω significantly greater than the background, and one candidate gene under purifying selection showing a lower ω value than the background. In comparison, only five genes, with ω greater than the background, were retained as candidates of positive or relaxed selection in the King penguin lineage. Although greater than those of the background branches, the ω values of most of the candidate genes in the Emperor or the King

penguin, still remain lower than one, making it difficult to distinguish between positive selection and relaxation of purifying selection. The CODEML branch-site test indicated a total of 104 sites in the 60 candidate genes in the Emperor lineage, whereas 17 were suggested in the five genes in the King lineage (Fig. 1B; *SI Appendix*, Table S7). On the other hand, aBSREL identified a much larger number of candidate genes under positive selection than CODEML branch model (423 in the Emperor lineage and 200 in the King lineage). Between 70% and 80% of the CODEML candidate genes (42 and 4 genes considering the Emperor and the King lineage, respectively) were also significant in aBSREL results (Fig. 1B). According to RELAX, 17 genes in the Emperor lineage and in four genes in the King lineage bear a significant signature of relaxed selection ($K < 1$; Fig. 1B; *SI Appendix*, Table S8, S9). Concerning the Emperor lineage, five of the 17 genes (*i.e.*, FLVCR1, ANKRD17, ASAP1, PAK1, PHLPP1) are also candidates in either CODEML (branch model, $\omega_i > \omega_b$), aBSREL, or both, further supporting the signal of relaxed purifying selection.

Biological functions enrichment in Emperor penguin candidate genes

After correcting for multiple tests and using REVIGO's redundancy elimination algorithm, we found 16 enriched GO biological process terms in candidate genes for positive selection suggested by both CODEML and aBSREL (*SI Appendix*, Table S10). Some of these GOs are related to heart and muscle development (GO:0003306, GO:1901863), lipid (GO:0033993), glucose (GO:0071333) and sphingolipid (*i.e.*, ceramide) metabolism (GO:1905371). When considering all candidate genes supported by either CODEML or aBSREL, we retrieved 34 enriched GO biological process terms (*SI Appendix*, Table S11), 12 of which could be assigned to one of the biological/molecular functions putatively related to cold adaptation from previous studies (Table 1). In addition, qualitatively screening the biological functions of all candidate genes using the human database GeneCards, we identified 161 genes which could be assigned to the biological/molecular functions identified in previous studies (Table 1). Four genes identified as under selection in the Emperor penguin were already found in previous studies: TRPM8 (*temperature sensing*), indicated as under selection by both CODEML (branch and branch-site models) and aBSREL; LEPR (*lipid and fatty acid metabolism*) and CRB1 (*phototransduction*) were suggested as candidate genes under selection by aBSREL; SFI1 (*glucose metabolism*) which showed a significant signal of intensified selection in the RELAX test ($K > 1$). In addition, the

sphingomyelin synthase2 (SGMS2), shows a biological function that appears as markedly related to cold adaptation in the Emperor penguin, potentially regulating biological membrane fluidity at low temperatures (Wang et al., 2014). This function (named as *membrane fluidity*) was added to the list of biological/molecular functions putatively related to cold adaptation (Table 1).

Table 1. Candidate genes for positive / relaxed selection inferred by CODEML (bm, bsm), aBSREL, and RELAX ($K > 1$) which we could assign to biological functions suggested as related to cold adaptation in previous studies (see references in the table) on the basis of their GeneCards description. Candidate genes under selection in the Emperor penguins which were also found in previous studies are in bold (four genes). Note that in the *GO ID* column we associated the Gene Ontology Biological Process supported by gProfiler in candidate genes inferred by aBSREL or CODEML (bm), or in the subset of genes inferred by both (marked with an asterisk), with relevant biological functions; the GOs in this column do not only refer to the genes indicated in the previous columns.

** Weak evidence of gene expression changes in a study on feathers development (Ng et al., 2015).

Function/metabolism of	CODEML (bm)	CODEML (bsm)	aBSREL	Relax (K > 1)	Relax (K < 1)	GO ID	Previously identified genes	References
Cardiovascular			DNASE1, ADAMTS13, EIF2AK1, MYBPC3, UBA5, ENTPD1, NRP1, PKDCC, ROCK2, LMOD2, PLEKHM2, ATP11C, HGSNAT, TFRC, UFL1, SPIC, MYOCD, STXBP4, FBLN5, MYLK3, AGXT2, FLVCR1	AGXT2, MYLK3	FLVCR1, EDRF1	GO:0072359 (adj p-val = 0.01); GO:0060047 (adj p-val = 0.04); GO:0003306 (adj p-val = 0.02)*; GO:1901863 (adj p-val = 0.03)*	SPTBN5, KNG1, ICAM4, DSP, TNN, ARID5B, CUL7, EHD3, VCL, ALPK3, XIRP1, ADORA2B, SOD1, POMC, CD36, ENPEP, MME, BDKRB2	Yudin et al., 2017; Liu et al., 2014; Wollenberg Valero et al., 2014; Vianna et al. 2020
Skin thickness			PRDM5, KIAA1211, EGF, FAM83G, MATN3, DSC1, SGSM3, GAB1, RAB11FIP1, ASAH1	ASAH1			DSP, EXPH5, EVPL, DSG1	Yudin et al., 2017; Li et al., 2014
Feathers development	WNT3A, RPL31**	WNT3A	SHOC2, RPL31**, WNT3A	RPL31**			(beta-keratin gene families)	Li et al., 2014
Immunity	DHX36, ANKRD17, TEC, SLIT2, TRIM29	DHX36, ANKRD17	MILR1, TRAF3IP2, SH2D4B, SPPL2B, MAP3K5, TBK1, IFIH1, RASGRP1, IFIT5, PLD4, CD101, LGALS8, ASB9, CMAS, AOAH, C7, MRC1, JCHAIN, HERC5, ATG5, LTA4H, ANKRD17, TRIM29	HERC5, ATG5, LTA4H, DHX36, TEC	ANKRD17	GO:0002376 (adj p-val = 0.0002); GO:0006952 (adj p-val= 0.01); GO:0032607 (adj p-val=0.02)	SELPLG, ZDBF2, MAPK1, MAPK14	Yudin et al., 2017
Lipid and fatty acid	ASAP1		FAAH, VTG2, **LEPR**, PLCXD1, OSBPL2, LMAN1, PPARA, FADS1, ACSBG2, FNDC3B, GK5, ARSK, PDZRN3, DECR2, ACOT7, ACOX3, KIAA1468, PCCA, ACOT12, VPS13A	HMGCL, PCCA, ACOT12, VPS13A	ASAP1	GO:0006631 (adj p-val=0.04); GO:0006071 (adj p-val=0.05); GO:0033993 (adj p-val=0.03)*	NR1D2, LPL, CETP, SOD1, **LEPR**, APOB, ABCC6, UCP3, FASN, GLPTD1, AKT2, SLC38A4, SFI1	Kumar et al., 2015; Yudin et al., 2017; Castruita et al., 2020; Liu et al., 2014; Tigano et al., 2018; Wollenberg Valero et al., 2014; Li et al., 2014; Lin et al., 2019
Glucose (including insulin)	ADCY5, IDE, SLC2A1, PAK1	IDE, PAK1	AAED1, ALG6, SLC17A9, ABCA10, TTYH3, ABCC8, GRB14, GAB1, PHKB, STXBP4, NRG4, SLC7A6, AGL, PGM2L1, IDE, ADCY5, SLC2A1	SLC7A6, AGL, PGM2L1, **SFI1**	PAK1	GO:0071333 (adj p-val=0.05)*	**SFI1**, SLC38A4, AKT2, GLPTD1	Kumar et al., 2015; Yudin et al., 2017
Thyroid hormones			NCOA3, MARCH6			GO:0009725 (adj p-val = 0.03)	THRB, RARB	Tigano et al., 2018
Non-shivering thermogenesis			PRDM16, L2HGDH, CACNA1D, ATP1A1				UCPs, avUCP, THRB, RARB, NR1D2	Lowell and Spiegelman, 2000; Talbot et al., 2004; Tigano et al., 2018
Shivering thermogenesis							NEB, ADORA2B, SOD1, HSPB2	Yudin et al., 2017; Wollenberg Valero et al., 2014
Oxidative stress	GMEB1, ATOX1	GMEB1	MAP3K5, PTGR2, SELENOP, COQ6, GMEB1, PEX16	COQ6	VRK2		SOD1, UCP3	Wollenberg Valero et al., 2014
Stress response			EIF2AK1, CUL4B, MAP3K3, ULK2, SLC9A8, MAP4K3, TH	HSP40		GO:0006950 (adj p-val=4.0E-05)	HSP47, HSPA8, HSPB2, STUB1, MAPK1, UCP3, HSF1, MAPK14, UNG, SOD1	Wollenberg Valero et al., 2014
Homeostasis	AQP2		CHERP, AQP11, SLC9A8, ATP1A1, AQP2			GO:0006874 (adj p-val=0.03); GO:0072503 (adj p-val=0.04)	LPL, CETP, SCL6A19, ACE2, AGT	Wollenberg Valero et al., 2014; Vianna et al., 2020
Circadian rhythm	CREB1, PHLPP1		PROX1, ROCK2, GPR176, ASMT, PTCHD1		PHLPP1		ADCY2, ADCY8, CALML4, CAMK2	Lin et al., 2019
Phototransduction			RGS9, CRB2, MREG, ARR3, TLE4, ABCB5, FOXN4, GRIFIN, SPATA7, ABCA4, **CRB1**	**CRB1**		GO:0042462 (adj p-val=0.04)	CNGB1, MYO3A, UACA, **CRB1**, CRY2, MYO3B	Li et al., 2014
Mitochondrial activity	GLS	GLS	CHCHD6, PTCD3, SLC25A4, VPS13D, SPG20, SERAC1, OPA1, SLC25A3, PNPT1, LRPPRC, GLS	SLC25A3, PNPT1, LRPPRC, GLS			COX1, ATP8, ND4L, ND4	Ramos et al., 2018
Temperature sensing	**TRPM8**	**TRPM8**	**TRPM8**				**TRPM8**, TRPV3, TRPV4, TRPA1, TRPM4	Lynch et al 2015
Membrane fluidity	SGMS2	SGMS2	SGMS2			GO:1905371 (adj p-val=0.04)*	-	This study

Discussion

Emperor penguin shift towards a novel ecological niche

Among the orthologous genes tested in our analyses, a larger fraction shows signatures of novel selection regimes in the Emperor penguin lineage than in the King penguin, either as intensification or relaxation of the selection pressure (Fig. 1B). One possible explanation of this pattern is that the ancestor of both species had ecological preferences which were more similar to the King penguin, while the Antarctic ecology of the Emperor penguin is a derived, though rather recent (ca. 1-2 Mya; Gavryushkina et al., 2017), adaptation. A general shift from warmer to colder habitats has been suggested for penguins in general (Vianna et al., 2020). King and Emperor penguins critically differ in their breeding range, as they reproduce in temperate-cold sub-Antarctic islands with average winter temperature of ca. 3 °C, or on the Antarctic sea-ice featuring average winter temperature around -25 °C, sometimes at -60 °C, respectively. Upon colonization of Antarctica, novel selective pressures are expected to have appeared while others, characterizing the former ecology, should have relaxed. Relaxed purifying selection can in fact be an additional source of evolutionary novelties (Hunt et al., 2011), as it can have non-linear consequences on a trait, including stabilizing or balancing selection, pseudogenization or, on the contrary, recruitment for a different function (Lahti et al., 2009).

Rather fast adaptation to polar lifestyle is not new in homeothermic vertebrates, as suggested for the recent divergence (less than 0.5 Mya) of the polar bear from the brown bear (Liu et al., 2014; Castruita et al., 2020), or of the Arctic fox from its common ancestor with red fox (ca. 2.9 Mya, Kumar et al., 2015). One question, which could be highly relevant in the ongoing climate change scenario, is whether extreme cold adaptation is an evolutionary *cul de sac*, *i.e.*, a derived suite of traits from which reverting to a less extreme cold ecology is hampered. Interestingly, none of the extant rhino species, which are all adapted to warm climate, descend from any of the three cold-adapted species we know about, and which are all extinct (Liu et al., 2021). A similar evolutionary endpoint could have characterized the diversification of elephants, with the extinction of the cold-adapted mammoths (Lynch et al., 2015).

An alternative explanation for the higher number of genes showing evidence of selection shifts in the Emperor penguin is that selection was more efficient along the Emperor lineage due to its larger and more constant population size through time, as revealed in previous studies (Trucchi et al., 2014; Cristofari et al., 2016), when compared to the King penguin markedly oscillating demographic trajectory (Cristofari et al., 2018; Trucchi et al., 2019). In fact, at low population size genetic drift overwhelms selection, leading to the fixation of both synonymous and nonsynonymous variants, and potentially blurring the *dN/dS* ratios. However, the higher signature of relaxed purifying selection in the Emperor penguin contrasts with this alternative explanation.

Looking for a common genetic underlying of adaptation to cold in homeotherms

A large fraction of the candidate genes under selection in the Emperor penguin is involved in most of the functional pathways relevant to cold adaptation identified in previous studies (Table 1). Nonetheless, we discovered only four genes in common with those previously identified in other cold-adapted vertebrates (notably, two of which were found in the mammoth). This result is not surprising given the suggested polygenic basis of most phenotypic traits (Barghi et al., 2020) which could be shaped by the contribution of a quite large set of genes (Boyle et al., 2017). As also emerged in other vertebrates (see refs in Table 1), traits related to cardiovascular function, lipid and fatty acid metabolism, glucose metabolism, oxidative stress and stress response, insulation (including skin thickness and feathers development), phototransduction and mitochondrial activity show the largest number of candidate genes under selection in the Emperor penguin (Table 1).

Genes involved in fatty-acid metabolism have been identified to be under positive selection both in polar bear and Arctic fox, indicating similar evolutionary constraints on fat metabolism in these two cold-adapted mammal species (Liu et al 2014; Kumar et al., 2015). The storage of subcutaneous fat is also crucial in the Emperor penguin, both because it represents the main source of energy during the long fasting periods (Blem, 1990; Cherel et al., 1994; Groscolas et al., 1990) and because it provides thermal insulation (Kooyman et al., 1976). Moreover, fatty acids can stimulate muscle thermogenic processes in birds and therefore may be a very important component of the adaptive response to cold temperatures in penguins (Duchamp et al., 1999; Toyomizu et al.,

2002; Talbot et al., 2004; Rey et al., 2010). Accordingly, our results revealed candidate genes under selection in the Emperor penguin such as PPARa, regulating eating behavior (Fu et al., 2003), controlling lipid absorption in the intestine (Poirier et al., 2001) and fatty-acid oxidation (Lemberger et al., 1996), ADCY5, associated with body weight (Li and Li, 2019), and LEPR, the leptin receptor. LEPR is involved in fat and glucose metabolism, in appetite-regulating through its effects on food intake and energy consumption (Zhang et al., 1994; Halaas et al., 1995; Pelleymounter et al., 1995), and in adaptive thermogenesis (Yang et al., 2011). Yet, the adipostat activity of leptin has not been demonstrated in Aves, where the expression pattern differs greatly from that of mammals, as it seems to be missing from the adipose tissue (Friedman-Einat and Eyal Seroussi, 2019).

While shivering thermogenesis might be the main thermogenic mechanism in birds following short-term cold exposure (Teulier et al., 2010), cold acclimated birds show non-shivering thermogenesis mediated by avUCP expression within the skeletal muscle (Talbot et al., 2004). According to our analyses, some of the candidate genes could be assigned to the non-shivering thermogenesis category (Table 1), but none of them can be unambiguously associated with shivering thermogenesis. Among the former, Na,K-ATPase (ATP1A1) is a membrane enzyme that utilizes energy derived from the hydrolysis of ATP to pump Na+, wasting energy as heat, thus playing a significant role in thermal tolerance and energy balance (Geering et al., 1987; Iannello et al., 2007). It was demonstrated that its expression is affected by heat stress (Sonna et al., 2002) and consistently increases during mammal hibernation (Vermillion et al., 2015). L2HGDH was found to be associated with the TCA cycle, electron transport and glycolysis (Oldham et al., 2006) and it was identified as one of the candidate genes under positive selection in three high-altitude passerine birds (Hao et al., 2019). PRDM16 is a zinc-finger protein that activates brown fat-selective genes responsible for mitochondrial biogenesis and oxidative metabolism, while repressing the expression of a wide range of genes selective for white fat cells (Seale et al., 2007; Kajimura et al., 2015). This protein appears to play a role also in the development and function of beige cells (Ohno et al., 2012; Seale et al., 2011). Retinoic acid and thyroid hormones, whose candidate binding sites have been found in a mammal UCP-1 enhancer (Lowell and Spiegelman, 2000), together with another nuclear receptor (NR1D2), have been suggested to be involved in

thermal adaptation in birds (Tigano et al., 2018). NR1D2, as well as MARCH6 and NCOA3, also involved in thyroid hormone regulation and action (Zelcer et al., 2014; Ishii et al., 2021), influencing baseline temperature (Elliott et al., 2013) and thermoregulation in response to cold stimuli in birds (Vézina et al., 2015), were present among our candidate genes.

One of the most promising candidate genes under selection is TRPM8, which encodes the sensor for noxious cold temperature (Yin et al., 2018). Setting the physiological range of temperature tolerance and, ultimately, the width of the geographical habitat (Matos-Cruz et al., 2017), any biological thermosensory apparatus should be under strong evolutionary pressures to trigger specific responses to noxious high or low temperatures (Myers et al., 2009). Indeed, evolutionary tuning of five temperature-sensitive transient receptor potential channels, including TRPM8, has been likely key in the adaptation of the Woolly mammoth to the Arctic (Lynch et al., 2015). A previous study demonstrated the crucial role of a single-point mutation located at site 906 (as per *Gallus gallus* coordinates in Yin et al., 2018; 919 as per coordinates in Yang et al., 2020) for the activation of TRPM8 pore domain channel in the Emperor penguin (Yang et al., 2020). Interestingly our comparative selection scan found instead evidence of positive selection at two other sites (*i.e.*, I1058T, Y1069M). After aligning 541 vertebrates TRPM8 ortholog sequences available in GenBank (accessed on 22/11/2021), we found that the substitution at site 906 is not unique to Emperor penguins but it is instead widespread in birds with different ecology and habitat preference, including warm tropical regions. Also, our candidate substitution Y1069M is common in penguins and other birds. Conversely, the substitution I1058T (as per *Gallus gallus* coordinates in Yin et al., 2018) is extremely rare in birds, being present in one other species only (*Sitta europea*). This substitution is also rare in mammals and reptiles where it appears in seven and one species only, respectively. We located the single-point mutation I1058T just after the last ultra-conserved residue in one of the three helices (CTDH2) of the C-terminal cytoplasmic domain (Yin et al., 2018). Further analyses should be conducted to determine if and how this substitution affects noxious cold sensing in Emperor penguins.

Limitations of this study

There is no overlap between our set of candidate genes under selection and those discovered in previous studies on penguins (Li et al., 2014; Vianna et al., 2020). On one hand, selection tests based on phylogenies are, of course, influenced by size and composition of the set of species included in the tree. In contrast with Li et al. (2014), where 48 bird species of which only two penguins were analyzed, we selected non-target species both in close (seven penguins) and in more distant (13 other birds) clades. On the other hand, the bioinformatic pipeline applied for identifying ortholog coding sequences across the species in the phylogeny may determine which genes are included or excluded. Our pipeline, successfully tested in *Glossina* (Savini et al., 2021), aligned ca. 30% more genes than in Vianna et al. (2020) and slightly less than in Li et al. (2014). We also note that in Vianna et al. (2020), the selection scan was performed for 18 penguin species, not only on the King and Emperor ones, and searched for candidate genes in all of the penguin lineages.

Although we found an appreciable overlap between the results of CODEML and aBSREL, the latter suggested a lot more candidate genes (Fig. 1B). This could be due either to higher sensitivity of aBSREL, as due to the branch-site model applied, or to a higher false negative rate in CODEML. One big difference between these two methods is that aBSREL makes no assumptions about the selective regime on background branches while CODEML assumes negative or no selection on background branches. This could lead to a higher false negative rate when the evolutionary process along background branches deviates significantly from modeling assumptions (Kosakovsky Pond et al., 2009). In general, phylogenetic tests of adaptive evolution are not capable of avoiding false positives because they do not consider multi nucleotide mutations (simultaneous mutations at two or three codon positions) which can instead be more common than expected (Venkat et al., 2018).

Concerning our characterization of the candidate genes to associate them to the functional categories we identified from previous studies on cold adaptation, we reckon that this is by far a simple (*i.e.,* gene characterization was based on GeneCards description only) and qualitative (*i.e.*, not based on a statistical test) comparison, aiming at contextualizing our results within the published literature. As sequence orthology does not necessarily equal functional orthology, we

acknowledge that a functional characterization largely based on a human model (and sometimes on a chicken model) is a strong assumption for a distant non-model organism.

Acknowledgements

ET was supported by the PNRA16_00164 ("Programma Nazionale di Ricerca in Antartide. Bando PNRA 5 aprile 2016, n. 651. – Linea B "Genomica degli adattamenti estremi alla vita in Antartide").

References

Abascal, F., Zardoya, R. & Telford, M.J. 2010 TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**, 7-13.

Akashi, H.D., Cádiz Díaz, A., Shigenobu, S., Makino, T. & Kawata, M. 2016 Differentially expressed genes associated with adaptation to different thermal environments in three sympatric Cuban Anolis lizards. *Molecular Ecology* **25**, 2273-2285.

Allen, J.A. 1877 The influence of Physical conditions in the genesis of species. *Radical Review* **1**, 108–140.

Barghi, N., Hermisson, J. & Schlötterer, C. 2020 Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* **21**, 769–781.

Blem, C.R. 1990 Avian energy storage. *Current Ornithology* **7**, 59-113.

Blix, A.S. 2016 Adaptations to polar life in mammals and birds. *Journal of Experimental Biology* **219**,1093-1105.

Boyle, E.A., Li, Y.I. & Pritchard J.K. 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186.

Cannon, B. & Nedergaard, J. 2010 Thyroid hormones: igniting brown fat via the brain. *Nature Medicine* **16**, 965–967.

Castruita, J.A.S., Westbury, M.V. & Lorenzen, E.D. 2020 Analyses of key genes involved in Arctic adaptation in polar bears suggest selection on both standing variation and de novo mutations played an important role. *BMC Genomics* **21**, 1-8.

Cherel, Y., Gilles, J., Handrich, Y. & Le Maho, Y. 1994 Nutrient reserve dynamics and energetics during long-term fasting in the king penguin (Aptenodytes patagonicus). *Journal of Zoology* **234**, 1-12.

Cristofari, R., Bertorelle, G., Ancel, A., *et al*. 2016 Full circumpolar migration ensures evolutionary utility in the Emperor penguin. *Nature Communications* **7**, 1-9.

Cristofari, R., Liu, X., Bonadonna, F., *et al*. 2018 Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change* **8**, 245-251.

Duchamp, C., Marmonier, F., Denjean, F., Lachuer, J., *et al.* 1999 Regulatory, cellular and molecular aspects of avian muscle non-shivering thermogenesis. *Ornis Fennica* **76**, 151–165.

Elliott, K.H., Welcker, J., Gaston, A.J., *et al* 2013 Thyroid hormones correlate with resting metabolic rate, not daily energy expenditure, in two charadriiform seabirds. *Biology Open* **2**, 580–586.

Friedman-Einat, M. & Seroussi, E. 2019 Avian Leptin: Bird's-Eye View of the Evolution of Vertebrate Energy-Balance Control. *Trends in Endocrinology and Metabolism* **30**, 819-832.

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**, 231–241.

Fu, J., Gaetani, S., Oveisi, F., Lo Verme, J., *et al.* 2003 Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR-alpha. *Nature* **425**, 90 – 93.

Gavryushkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., *et al.* 2017 Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology* **66**, 57-73.

Geering, K., Kraehenbuhl, J.P. & Rossier, B.C. 1987 Maturation of the catalytic alpha unit of Na, K-ATPase during intracellular transport. *Journal of Cell Biology* **105**, 2613-2619.

Goldsmith, R. & Sladen, W.J. 1961 Temperature regulation of some Antarctic penguins. *The Journal of Physiology* **157**, 251-262.

Groscolas, R. 1990 Metabolic adaptations to fasting in emperor and king penguins. In *Penguin Biology* (ed. LS Davis & JT Darby), 269–296. San Diego: Academic Press.

Groscolas, R. & Robin, J.P., 2001 Long-term fasting and re-feeding in penguins. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **128**, 645–655.

Halaas, J.L., Gajiwala, K.S., Maffei, M., Cohen, S.L., *et al.* 1995 Weight-reducing effects of the plasma protein encoded by the obese gene. *Science* **269**, 543–546.

Han, M.V., Demuth, J.P., McGrath, C.L., Casola, C. & Hahn, M.W. 2009 Adaptive evolution of young gene duplicates in mammals. *Genome Research* **19**, 859-867.

Hao, Y., Xiong, Y., Cheng, Y., Song, G., *et al.* 2019 Comparative transcriptomics of 3 high-altitude passerine birds and their low-altitude relatives. *PNAS* **116**, 11851-11856.

Hunt, B.G., Ometto, L., Wurmb, Y., *et al.* 2011 Relaxed selection is a precursor to the evolution of phenotypic plasticity. *PNAS* **108**, 15936-15941.

Iannello, S., Milazzo, P. & Belfiore, F. 2007 Animal and human tissue Na,K-ATPase in normal and insulin-resistant states: regulation, behaviour and interpretative hypothesis on NEFA effects. *Obesity Reviews* **8,** 231-51.

Ishii, S., Amano, I., & Koibuchi, N. 2021 The Role of Thyroid Hormone in the Regulation of Cerebellar Development. *Endocrinology and Metabolism* **36**, 703–716.

Jarvis, E.D., Mirarab, S., Aberer, A.J., *et al.* 2014 Whole genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331.

Kajimura, S., Spiegelman, B.M. & Seale, P. 2015 Brown and Beige Fat: Physiological Roles beyond Heat Generation. *Cell Metabolism* **22,** 546 –559.

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 Part A* **54**, 75-80.

Kosakovsky Pond, S.L., Poon, A.F.Y. & Frost, S.D.W. 2009 Estimating selection pressures on alignments of coding sequences. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge Univ. Press, Cambridge, UK, 419-490.

Kumar, V., Kutschera, V.E. & Nilsson, M.A. 2015 Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. *BMC Genomics* **16**, 1-13.

Kumar, S., Stecher, G. & Tamura, K. 2016 MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870-1874.

Lahti, D.C., Johnson, N.A., Ajie, B.C., Otto, S.P., Hendry, A.P., *et al.* 2009 Relaxed selection in the wild. *Trends in Ecology & Evolution* **24**, 487-496.

Lemberger, T., Saladin, R., Vazquez, M., Assimacopoulos, F., *et al.* 1996 Expression of the peroxisome proliferator-activated receptor alpha gene is stimulated by stress and follows a diurnal rhythm. *Journal of Biological Chemistry* **271**, 1764-1769.

Li, F.G. & Li, H. 2019 A time-dependent genome-wide SNP-SNP interaction analysis of chicken body weight. *BMC Genomic*s **20**, 1-9.

Li, C., Zhang, Y., Li, J. *et al.* 2014 Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience* **3**, 2047-2217X.

Lin, Z., Chen, L., Chen, X., Zhong, Y., Yang, Y., Xia, W., Liu, C., Zhu, W., Wang, H., Yan, B. & Yang, Y. 2019 Biological adaptations in the Arctic cervid, the reindeer (Rangifer tarandus). *Science* **364**(6446).

Liu, S., Lorenzen, E.D., Fumagalli, M., *et al.* 2014 Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785-794.

Liu, S., Westbury, M.V., Dussex, N., Mitchell, K.J., Sinding, M.H., Heintzman, P.D., Duchêne, D.A., Kapp, J.D., von Seth, J., Heiniger, H., Sánchez-Barreiro, F. 2021 Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell* **184**, 4874-85.

Lowell, B.B. & Spiegelman, B.M. 2000 Towards a molecular understanding of adaptive thermogenesis. *Nature* **404**, 652–660.

Löytynoja, A. 2013 Phylogeny-aware alignment with PRANK. *Multiple Sequence Alignment Methods. Methods in Molecular Biology (Methods and Protocols)* **1079**, 155-170.

Lynch, V.J, Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., *et al.* 2015 Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell Reports* **12**, 217-228.

Madeira, F., Park, Y.M., Lee J., *et al.* 2019 The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, 636-641.

Matos-Cruz, V., Schneider, E.R., Mastrotto, M., *et al.* 2017 Molecular Prerequisites for Diminished Cold Sensitivity in Ground Squirrels and Hamsters. *Cell Reports* **21**, 3329-3337.

Myers, B.R., Sigal, Y.M. & Julius D. 2009 Evolution of Thermal Response Properties in a Cold-Activated TRP Channel. *PloS one* **4**, e5741.

Ng, C.S., Chen, C.K., Fan, W.L., Wu, P., Wu, S.M., Chen, J.J., Lai, Y.T., Mao, C.T., Lu, M.Y.J., Chen, D.R. & Lin, Z.S. 2015 Transcriptomic analyses of regenerating adult feathers in chicken. *BMC Genomics* **16**, 1-16.

Ohno, H., Shinoda, K., Spiegelman, B.M. & Kajimura, S. 2012 PPARg agonists induce a white-to-brown fat conversion through stabilization of PRDM16 protein. *Cell Metabolism* **15**, 395–404.

Oldham, M.C., Horvath, S. & Geschwind, D. H. 2006 Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *PNAS* **103**, 17973–17978.

Pan, H., Cole, T.L., Bi, X., *et al.* 2019 High-coverage genomes to elucidate the evolution of penguins. *GigaScience* **8**, 1-17.

Pelleymounter, M.A., Cullen, M.J., Baker, M.B., Hecht, R., *et al.* 1995 Effects of the obese gene product on body weight regulation in ob/ob mice. *Science* **269**, 540–543.

Poirier, H., Niot, I., Monnot, M.C., Braissant, O., *et al.* 2001 Differential involvement of peroxi-some-proliferator-activated receptors alpha and delta in fibrate and fatty-acid-mediated inductions of the gene encoding liver fatty-acid-binding protein in the liver and the small intestine. *Biochemical Journal* **355**, 481–488.

Pond, S.K., Frost, S. & Muse, S.V. 2005 HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679.

R Core Team. 2013 R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramasamy, S., Ometto, L., Crava, C.M., Revadi, S., Kaur, R., Horner, D.S., *et al.* 2016 The evolution of olfactory gene families in Drosophila and the genomic basis of chemical-ecological adaptation in Drosophila suzukii. *Genome Biology and Evolution* **8**, 2297–2311.

Ramos, B., González-Acuña, D., Loyola, D.E., *et al.* 2018 Landscape genomics: natural selection drives the evolution of mitogenome in penguins. *BMC Genomics* **19**, 1-17.

Raudvere, U., Kolberg, L., Kuzmin, I. *et al.* 2019 g:Profiler: a web server for functional enrichment analysis and conversions of gene lists. *Nucleic Acids Research* **47**, W191-W198.

Rey, B., Roussel, D., Romestaing, C., Belouze, M., *et al.* 2010 Up-regulation of avian uncoupling protein in cold-acclimated and hyperthyroid ducklings prevents reactive oxygen species production by skeletal muscle mitochondria. *BMC Physiology* **10**, 1-12.

Roussel, D., Le Coadic, M., Rouanet, J.L. & Duchamp, C. 2020 Skeletal muscle metabolism in sea-acclimatized king penguins. I. Thermogenic mechanisms. *Journal of Experimental Biology* **223**, p.jeb233668.

Rowland, L.A., Bal, N.C., & Periasamy, M. 2015 The role of skeletal-muscle-based thermogenic mechanisms in vertebrate endothermy. *Biological Reviews* **90**, 1279-1297.

Savini, G., Scolari, F., Ometto, L., Rota-Stabelli, O., Carraretto, D., Gomulski, L.M., Gasperi, G., Abd-Alla, A.M., Aksoy, S., Attardo, G.M. and Malacrida, A.R. 2021 Viviparity and habitat restrictions may influence the evolution of male reproductive genes in tsetse fly (Glossina) species. *BMC biology* **19**, 1-13.

Scholander, P.F. 1955 Evolution of climatic adaptation in homeotherms. *Evolution* **9**, 15-26.

Seale, P., Kajimura, S., Yang, W., Chin, S., *et al.* 2007 Transcriptional control of brown fat determination by PRDM16. *Cell Metabolism* **6**, 38–54.

Seale, P., Conroe, H.M., Estall, J., Kajimura, S., *et al.* 2011 Prdm16 determines the thermogenic program of subcutaneous white adipose tissue in mice. *Journal of Clinical Investigation* **121**, 96–105.

Smith, M.D., Wertheim, J.O., Weaver, S. 2015 Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution* **32**, 1342–1353.

Sonna, L.A., Fujita, J., Gaffin, S.L. & Lilly C.M. 2002 Effects of heat and cold stress on mammalian gene expression. *Journal of Applied Physiology* **92**, 1725-1742.

Stelzer, G., Rosen, R., Plaschkes, I., Zimmerman, S., *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Current Protocols in Bioinformatics* **54**, 1-30.

Storey, K.B. & Storey J.M. 1992 Natural freeze tolerance in ectothermic vertebrates. *Annual review of physiology* **54**, 619-637.

Storey, J.D., Bass, A.J., Dabney, A. & Robinson D. 2017 qvalue: Q-value estimation for false discovery rate control. R package version 2.15.0. http://github.com/StoreyLab/qvalue

Supek, F., Bošnjak, M., Škunca, N. & Šmuc T. 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**, e21800.

Talbot, D., Duchamp, C., Rey, B. *et al.* 2004 Uncoupling protein and ATP/ADP carrier increase mitochondrial proton conductance after cold adaptation of king penguins. *Journal of physiology* **558**, 123–135.

Tattersall, G.J., Sinclair, B.J., Withers, P.C., *et al.* 2012 Coping with thermal challenges: physiological adaptations to environmental temperatures. *Comprehensive Physiology* **2**, 2151–202.

Teulier, L., Rouanet, J.L., Letexier, D. & Romestaing, C. 2010 Cold-acclimation-induced non-shivering thermogenesis in birds is associated with upregulation of avian UCP but not with innate uncoupling or altered ATP efficiency. *The Journal of Experimental Biology* **213**, 2476-2482.

Thomas, D.B. & Fordyce R.E. 2008 The heterothermic loophole exploited by penguins. *Australian Journal of Zoology* 55, 317–321.

Tigano, A., Reiertsen, T.K., Walters, J.R., Friesen, V.L. 2018 A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird. *BioRxiv*, 507384.

Toyomizu, M., Ueda, M., Sato, S., Seki, Y., *et al.* 2002 Cold-induced mitochondrial uncoupling and expression of chicken UCP and ANT mRNA in chicken skeletal muscle. *FEBS Letters* **529**, 313-318.

Trucchi, E., Gratton, P., Whittington, J.D., *et al.* 2014 King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140528.

Trucchi, E., Cristofari, R. & Le Bohec, C. 2019 Reply to: The role of ocean dynamics in king penguin range estimation. *Nature Climate Change* **9**, 122-122.

Venkat, A., Hahn, M.W. & Thornton, J.W. 2018 Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology and Evolution* **2**, 1280–1288.

Vermillion, K.L., Anderson, K.J., Hampton, M. & Andrews, M.T. 2015 Gene expression changes controlling distinct adaptations in the heart and skeletal muscle of a hibernating mammal. *Physiological Genomics* **47**, 58-74.

Vézina, F., Gustowska, A., Jalvingh, K.M., Chastel, O. & Piersma T., 2015 Hormonal correlates and thermoregulatory consequences of molting on metabolic rate in a northerly wintering shorebird. *Physiological and biochemical zoology* **82**, 129–142.

Vianna, J.A., Fernandes, F., Frugone M.J., *et al.* 2020 Genome-wide analyses reveal drivers of penguin diversification. *PNAS* **117**, 22303-22310.

Wang, Q., Tan, X. & Jiao S. 2014 Analyzing Cold Tolerance Mechanism in Transgenic Zebrafish (Danio rerio). *PloS one* **9**, e102492.

Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L. & Scheffler K. 2015 RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular biology and evolution* **32**, 820-832.

Wollenberg Valero, K.C., Pathak, R., Prajapati, I., Bankston, S., *et al.* 2014 A candidate multimodal functional genetic network for thermal adaptation. *PeerJ* **2**, e578.

Yang, Z. 2007 PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586-1591.

Yang, J., Bromage, T.G., Zhao, Q., Xu, B.H., Gao, W.L., Tian, H.F., Tang, H.J., Liu, D.W. & Zhao, X.Q. 2011 Functional evolution of leptin of Ochotona curzoniae in adaptive thermogenesis driven by cold environmental stress. *PloS one* **6**, e19833.

Yang, S., Lu, X., Wang, Y., *et al.* 2020 A paradigm of thermal adaptation in penguins and elephants by tuning cold activation in TRPM8. *PNAS* **117**, 8633-8638.

Yin, Y., Wu, M., Zubcevic, L., *et al.* 2018 Structure of the cold- and menthol-sensing ion channel TRPM8. *Science* **359**, 237-241.

Yudin, N.S., Larkin, D.M. & Ignatieva E.V. 2017 A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genetics* **18,** 33-43.

Zelcer, N., Sharpe, L.J., Loregger, A., Kristiana, I., Cook, E.C., Phan, L., Stevenson J. & Brown A.J. 2014 The E3 ubiquitin ligase MARCH6 degrades squalene monooxygenase and affects 3-hydroxy-3-methyl-glutaryl coenzyme A reductase and the cholesterol synthesis pathway. *Molecular and cellular biology* **34,** 1262–1270.

Zhang, Y., Proenca, R., Maffei, M., Barone, M., *et al.* 1994 Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432.

Zhang, G., Li, C., Li, Q., *et al.* 2014 Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-1320.

3. Genome-wide signatures of recent selective sweeps in the Emperor penguin (*A. forsteri*) shed light on adaptation to Antarctica

This section includes the manuscript:

Federica Pirri, Céline Le Bohec, Lorenzo Zane, Emiliano Trucchi;

"Genome-wide signatures of recent selective sweeps in the Emperor penguin (*A. forsteri*) shed light on adaptation to Antarctica";

in preparation.

Introduction

Penguins (order Sphenisciformes) occupy a wide range of different habitats from polar to tropical latitudes in the Southern Hemisphere. Recent analyses in Vianna et al. (2020) supported the hypothesis of a temperate or sub-Antarctic origin of penguins followed by the colonization of Southern Ocean islands and Antarctica (Bertelli and Giannini, 2005; Ksepka and Giannini, 2006). According to this model, the common ancestor of the two largest penguin species, the Emperor and King penguins (*Aptenodytes forsteri* and *A. patagonicus*), might have lived in temperate environments and then radiated to colder thermal niches. In particular, the Emperor penguin has evolved several physiological, behavioral and morphological adaptations, such as insulating feathers (Taylor, 1986), long-term fasting (Groscolas, 1990; Cherel et al., 1994; Groscolas and Robin, 2001) and an enhanced thermoregulation system (Frost et al., 1975; Thomas and Fordyce, 2008) to deal with the extremely harsh Antarctic conditions. Such an extreme adaptation evolved in a rather short time, given the recent divergence (1-2 Mya; Gavryushkina et al., 2017) with its sister species.

The temporal dynamics of the Emperor penguin adaptation for an Antarctic lifestyle are only partially understood. Investigation of signals of long-term selection in the Emperor penguin genome through phylogeny-based methods on coding sequences (Pirri et al 2021; Vianna et al., 2020; Li et al., 2014) revealed a more pervasive selection shift in the Emperor penguin, supporting the hypothesis that its extreme cold adaptation represents a derived state from a more King penguin-like ecology. Moreover, the majority of candidate genes detected as being under selection in the Emperor penguin was related to metabolic and physiological pathways relevant to cold adaptation also in other homotherms (e.g., cardiovascular system, lipid, fatty acid and glucose metabolism, thermogenesis, etc.). On the other hand, the signature of recent, even ongoing, selection has not been uncovered yet.

Selective sweeps, combined with the hitch-hiking effect, cause an increase in the frequency of beneficial mutations and the nearby genomic variants, reducing diversity and generating extensive linkage disequilibrium (Alvarez et al., 2020).
Therefore, the haplotype that carries the advantageous allele tends to be significantly longer compared to other haplotypes at similar frequency in the population, making selection signatures detectable (Liu et al., 2013).
Several haplotype-based methods, such as integrated haplotype score (iHS) (Voight et al., 2006), extended haplotype homozygosity (EHH) (Sabeti et al., 2002) and cross population EHH (XP-EHH) (Sabeti et al., 2007) have been implemented over the years in order to identify recent selective sweeps.
In terms of scope, these methods are complementary: whereas the first two tests look for incomplete positive selection footprints, XP-EHH detects positively selected alleles that are nearly fixed in one but not all populations.
A combination of iHS and XP-EHH analyses should yield a thorough list of candidate loci underlying recent local adaptation.

Here, we performed genome-wide analyses of selection, at species level using iHS and at the between-species level using XP-EHH on 48 individuals of Emperor and King penguins.

Specifically, we are interested in mapping the footprints of recent and ongoing selection in order to verify whether a more intense selection shift affected the Emperor penguin genome after colonization of Antarctica and whether the candidate genes, associated to positively selected regions, were potentially involved in adaptation to these extreme cold conditions. Signals of selective sweeps in progress might in fact be indicative of the presence of genetic variants that have some effect on the Emperor penguin's phenotypic variation.

These results could be beneficial for understanding the mechanisms of selection between two phylogenetically close, but ecologically divergent, species and can also provide insights into the genetic features of cold adaptation in general.

Material and Methods

DNA sample collection, sequencing alignment and variant calling

Samples from a total of 48 individuals of Emperor and King penguins (24 individuals of each species) were collected from colonies at the Mertz glacier (close to the French Dumont d'Urville Station), and at the German Neumayer Station in Antarctica, and from Crozet, Heard and South Georgia Islands, respectively. Additionally, a total of 6 individuals of Adélie and Gentoo penguins (3 individuals of each species) were sampled from colonies at Dumont d'Urville in Antarctica and in Crozet Islands, to be used as outgroups.

Multiple independent short-read sequencing (on two or three lanes) was performed on PCR-free libraries of double-barcoded genomic DNA extractions on HiSeq2500/4000 (Illumina, Inc.) at the Norwegian Sequencing Centre, University of Oslo, Norway. Paired-end reads were quality filtered using Trimmomatic (Bolger et al., 2014) and URQT (Modolo and Lerat, 2015), and mapped to the reference Emperor penguin genome (RefSeq assembly accession: GCF_000699145.1) using BWA mem (Li and Durbin, 2009). Aligned reads were flagged for duplicates using the MarkDuplicates module in Picard tools (http://broadinstitute.github.io/picard/), sorted and indexed with SAMtools (Li et al., 2009). We employed Freebayes (Garrison and Marth, 2012) for haplotype-based population-aware joint variant calling using aligned reads with mapping quality higher than 20. Multiple nucleotide polymorphisms (MNPs) were then converted to SNPs using the script *vcfallelicprimitives* in the vcflib package (Garrison, 2016). SNPs were selected with the script *vcffilter* in the vcflib package according to the following criteria: snp calling quality higher than 30 (QUAL > 30), at least one read per strand (SAF > 0 & SAR > 0), and at least one read per read side (RPL > 0 & RPR > 0); indels and complex variants were discarded. SNPs were further filtered using vcftools (Danecek et al., 2011) discarding *i)* individual genotypes with coverage lower than 3 (--minDP 3), *ii)* loci with average individual coverage higher than 50 (--max-meanDP 50), and *iii)* not biallelic (--min-alleles 2; --max-alleles 2). SNPs were phased with Beagle (version 03Jul18.40b.jar; Browning and Browning, 2007) and polarized for ancestral/derived alleles leveraging the information from the two sister species and the two outgroup species by running a custom *python* script.

Identification of selective sweeps

Two complementary EHH-based statistics, the integrated Haplotype Score (iHS) and the Cross Population Extended Haplotype Homozygosity (XP-EHH), were used to assess genome-wide signatures of recent and ongoing selection in the Emperor and the King penguin.

These approaches were developed to address the main issue affecting the original EHH test, i.e., the high number of false positives caused by the considerable influence of the demographic history of the tested population (Alvarez et al., 2020).

The iHS test compares the differential levels of linkage disequilibrium (LD) surrounding a positively selected allele and the background allele at the same position within a population (Voight et al., 2006). As a result, this haplotype-based test is less sensitive to demographic history (e.g., population bottlenecks) and it is best suited to identify incomplete sweeps where the allele under selection is not yet fixed in the population.

While iHS is a robust test for the identification of incomplete sweeps, the XP-EHH has an increased power than the EHH statistics to detect (almost) complete selective sweeps (Alvarez et al., 2020).

The XP-EHH test compares the frequencies of the selected haplotypes between two populations to detect sweeps that are close to fixation (Sabeti et al., 2007).

Both large positive and negative iHS and XP-EHH scores were treated as potentially informative. In the case of iHS, negative scores could suggest that a derived allele would have swept up in frequency whereas positive values could be indicative of a selective sweep in favor of an ancestral allele.

Regarding the XP-EHH, large positive and negative values would identify selection events in the studied population (the Emperor penguin) and in the reference population (the King penguin), respectively.

All estimates were performed using the program selscan v.1.0.4 (Szpiech and Hernandez, 2014) on 437 biallelic polarized scaffolds longer than 100 kb, fitting the parameters recommended by the authors: maximum EHH extension in bp ("–max-extend" option) 1000000, maximum gap allowed between two SNPs in bp ("–max-gap" option) 200,000, EHH decay cutoff ("–cutoff" option) 0.05 and, for the iHS test, minor allele frequency (MAF) threshold ("–maf" option) 0.05.

As the recombination rate was not available for these two species, we converted physical distance into genetic distance, assuming that 1 bp corresponded to 0.000001 cM. In this way, we generated a flat recombination map for each species.

Whatever the statistics computed, the output results for each SNP were frequency-normalized over all scaffolds using the program norm, included in the selscan package. This normalization was carried out using default parameters as well: number of frequency bins ("–bins" option) 100.

Assuming a normal distribution, the most extreme scores in the tails that diverged from the null expectation (values above 5.5 and below -5.5 in the iHS test and above 4.5 and below -4.5 in the XP-EHH test), were used to identify SNPs under selection in each species.

For both iHS and XP-EHH analyses, false discovery rates (FDR) were computed using the qvalue R package (Storey and Tibshirani, 2003) and SNPs corresponding to an FDR lower than 0.05 were considered significant.

Candidate SNPs, identified as being under selection by any of the test statistics, were annotated on the Emperor penguin genome (RefSeq assembly accession: GCF_000699145.1) to identify the genes where they are located in. All the candidate SNPs that fell within the coordinates of the entire gene sequences were kept.

Functional characterization of the candidate regions

To further analyze the functions of identified genes in each species, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (Kanehisa and Goto, 2000) and Gene Ontology (GO) (Ashburner et al., 2000) enrichment analyses were performed using the functional annotation tool implemented in g:Profiler (Raudvere et al., 2019). GO and KEGG terms were assigned to candidate genes based on the Ensembl predictions for the annotated genes in flycatcher (*Ficedula albicollis*).

GO terms and the KEGG pathway with FDR < 0.05 and including at least three genes were retained.

Results

Candidate regions under selection

The iHS test detected 592 and 2,582 candidate SNPs where either the ancestral or the derived allele have been under selection in the Emperor penguin genome; in the King penguin, 417 and 2,256 candidate SNPs associated to either the ancestral or the derived allele, exhibited signals of positive selection.

The available annotation of the Emperor penguin genome allowed to identify a total of 261 potential candidate genes for recent positive selection in the Emperor penguin and 198 in the King penguin. Overall, 65 of the total candidate genes exhibited signals of positive selection in both the species.

In the Emperor penguin, 41% (244) and 45% (1,174) of the total positively selected candidate SNPs, associated to either the ancestral or the derived allele, fell into genes; whereas, in the King penguin, the findings reported 59% (246) and 68% (1,541) of the total positively selected candidate SNPs, related to either the ancestral or the derived allele, falling into genes.

A total of 26,151 and 5,571 comparison-specific outliers, representing the top 0.001% of the empirical distribution of all the XP-EHH normalized values, showed signals of positive selection in the Emperor and in the King penguin, respectively. The significant positive outliers (FDR < 0.05) overlapped with 194 genes (corresponding to 6% of the total positively selected SNPs) in the Emperor penguin and 221 genes (corresponding to 28% of the total positively selected SNPs) in the King penguin. Moreover, four of the candidate genes under selection were in common between the two species.

After crossing the output of the two tests applied in each species, a total of 23 candidate genes were identified as being under recent positive selection in the Emperor penguin and 5 in the King penguin.

Identification of functional candidate genes linked to selection sweeps

We analyzed whether candidate genes under selection were enriched for specific functions by performing GO and KEGG enrichment analyses.

In the Emperor penguin iHS output, we found 15 GO terms that were significantly enriched (FDR < 0.05), all related to nervous system development and phosphorus metabolic process. While we found no significantly enriched GO terms in the empirical tails of iHS for the King penguin.

In contrast, in the XP-EHH output, we found enrichment of several terms including numerous biological processes (Fig. 1). For example, the Emperor penguin showed an enrichment of genes involved in heart contraction (GO:0060047), blood circulation (GO:0008015) and circulatory system process (GO:0003013). Whereas, in the King penguin, GO terms related to regulation of phospholipid metabolic process (GO:1903725), homeostatic process (GO:0042592) and response to lipid (GO:0033993) were enriched.

Enriched KEGG pathways overlapped between candidate gene sets identified by iHS and XP-EHH analyses for the Emperor penguin (Fig. 2). In the iHS test, we found enrichment for pathways related to vascular smooth muscle contraction, circadian entrainment, sodium conservation, blood pressure regulation, cold-induced thermogenesis, lipolysis, stress response and circulatory system activity. Interestingly, we also found strong enrichment of the thyroid hormone signaling pathway, cortisol synthesis, insulin secretion and ATP production in candidate genes under selection identified by XP-EHH.

Although no evidence for KEGG enrichment emerged for genes that showed signals of ongoing selection in the iHS test for the King penguin, we found some relevant KEGG pathways enriched in the candidate genes detected by XP-EHH (Fig. 2). Several pathways were found to be in common with the Emperor penguin, such as: thyroid hormone signaling pathway, vascular smooth muscle contraction, stress response, circadian entrainment and sodium conservation.

Among the candidate genes which showed clear evidence of selection by both iHS and XP-EHH in the Emperor penguin, some could contribute to the adaptation to harsh cold conditions. For instance, CACNA1C and PRKG1 are involved in shivering and heat loss minimization by regulating blood vessel constriction and NPR1 plays a key role in cardiovascular homeostasis.

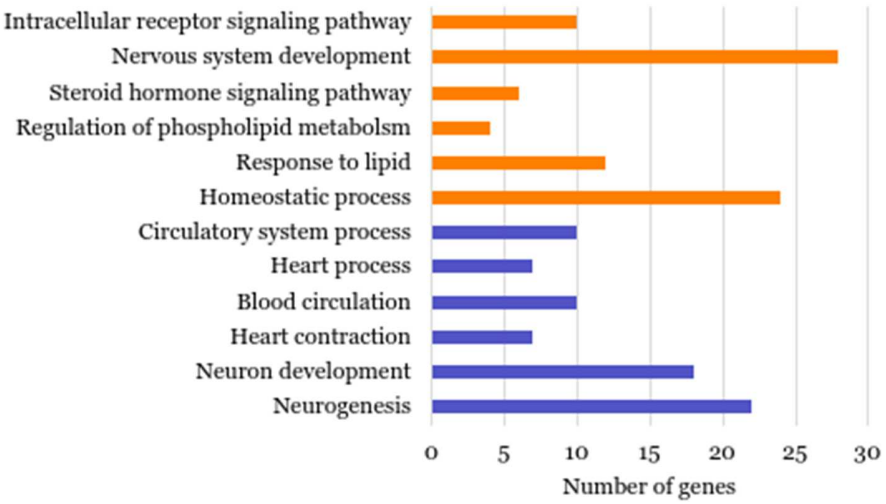

Figure 1. GO enrichment for the candidate genes found by the XP-EHH test in the Emperor (blue) and in the King penguin (orange).

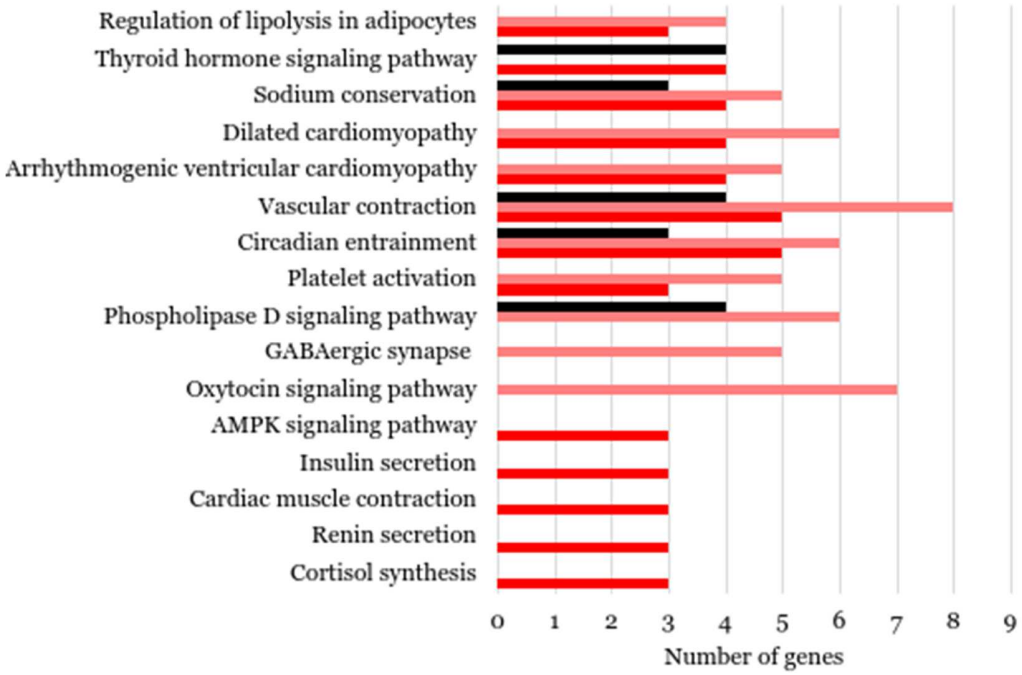

Figure 2. KEGG pathways for the candidate genes found by the iHS test in the Emperor penguin (pink) and by the XP-EHH test in the Emperor (red) and in the King penguin (black).

Discussion

Emperor penguin colonization of a more extreme habitat

We used whole-genome data to identify signals of recent positive selection in the Emperor penguin and in its closest relative, the King penguin. Both the haplotype-based selection scans applied found the strongest signatures of selection in the Emperor penguin, showing a higher number of candidate SNPs under selection. As Antarctica colonization is a relatively novel event, it is reasonable to observe a higher signature of incomplete selective sweeps in the Emperor penguin; indeed, the adaptation process to this new, more extreme, environment may have been, and continues to be, more intense and affect a wide range of phenotypic traits (and therefore genetic loci).

Moreover, this pattern further supports the hypothesis of a derived, recent, adaptation of the Emperor penguin to the Antarctic environment from a less cold-adapted ancestor, more ecologically similar to the King penguin.

The strongest selection signal was observed for the derived alleles, and we found that a large number of the putative positively selected genes (25% in the Emperor and 31% in the King penguin) were the same in the two species, suggesting some common adaptive biological pathways although we did not find the same genes in any common enriched pathways. An alternative explanation is that these genes might be located in regions of low recombination in which large portions of DNA are swept to fixation by selection events (Hudson, 1995).

Interestingly, both the iHS and XP-EHH analyses revealed a much larger number of SNPs, among those detected as being under selection, falling into annotated genes in the King than in the Emperor penguin. This could indicate that, in the latter, selective pressure has acted most on noncoding DNA regions and therefore on elements that regulate gene expression. A growing body of evidence supports the role of recurrent positive selection on noncoding DNA in the divergence observed between species (Zhen et al., 2012).

Genes under recent positive selection in the Emperor penguin are related to adaptations to the extreme Antarctic cold

Haplotype-based methods detect candidate alleles for recent and ongoing selection that are likely to underpin local environmental adaptation.

GO and KEGG enrichment analyses allowed to identify various functional terms associated directly or indirectly with environmental adaptation, such as circadian entrainment, blood pressure regulation, cold-induced thermogenesis and stress response. Specifically, some of the KEGG pathways depicted particular genetic aspects of adaptation to harsh cold in the Emperor penguin.

The same set of genes was found to be responsible for the enrichment of both the cardiovascular and lipid metabolism pathways, probably because of their pleiotropic action. In addition, other genes contributed to the enrichment of all the previously mentioned pathways.

Genes associated with cardiovascular function

In Antarctica, Emperor penguins have to cope with temperatures as low as −40 °C and forage in waters of −1.8 °C (Williams et al., 2015).

The increased internal insulation of the Emperor penguins is allowed by a powerful peripheral vasoconstriction and an enhanced countercurrent arteriovenous heat exchange system (Johansen and Bech, 1983).

Our analyses revealed 19 positively selected candidate genes associated with cardiovascular function. Some of these genes (ADCY2, ADCY5, GUCY1A2, PLA2G4A, PRKG1, PIK3CB) are involved in erythrocytes and platelets generation.

Others (ADCY5, CACNA1D, GUCY1A2, NPR1, PDE1C, CACNA1C) are linked to the control of blood pressure through renin secretion, while six candidate genes (ADCY2, ADCY5, CACNA1D, DMD, ITGB5, SGCG, ATP2A3, CACNA1C, CTNNA3, TPM4) are associated with cardiomyopathy in humans and other mammalian model organisms.

However, the most enriched pathway resulted to be the one involved in vascular smooth muscle contraction (ADCY2, ADCY5, CACNA1D, GUCY1A2, NPR1, PLA2G4A, PRKG1, RAMP3, ARHGEF11, CACNA1C, ATP2A3).

Homeotherms preserve a high core body temperature in response to cold by regulating blood pressure through the constriction of blood vessels. In penguins, peripheral vasoconstriction takes place in the appendages, preventing heat loss by reducing the temperature gradient between the body and the environment (Lewden et al., 2020). This is reasonable since flippers and feet are poorly insulated and relatively fat free structures.
Among the candidate genes related to this pathway, five (CACNA1C, PRKG1, GUCY1A2, NPR1) showed evidence of selection by both the iHS and XP-EHH tests. Moreover, PRKG1 was found to be positively selected also in previous studies on human populations living in Siberia (Cardona et al., 2014) and on Yakutian horses (Librado et al., 2015).

Genes associated with adipose tissue and lipid metabolism

Throughout the breeding season, Emperor penguins experience prolonged periods of fasting, during which they survive mainly thanks to lipid reserves previously accumulated in the subcutaneous adipose tissue (Robin et al., 1988).
Our analyses revealed signatures of recent selection in four candidate genes (ADCY2, ADCY5, NPR1, PRKG1, PIK3CB) involved in the regulation of lipolysis in adipocytes.
During times of energy deprivation, adipose tissue's lipid deposits undergo an extensive hydrolysis process to generate fatty acids and glycerol that are released into the vasculature and used as energy substrates by other organs (Duncan et al., 2007).
The ability to quickly mobilize fat stores to meet energy demands is an example of a highly adapted metabolic response.

Genes associated with cold-induced thermogenesis

In a cold environment, an efficient endothermy necessitates both an integrated system of monitoring and regulating body temperature and a powerful thermogenesis (Raccurt et al., 2008). Brown adipose tissue not only represents an isolative layer and an energy reserve, but it is where the non-shivering thermogenesis process occurs.
The oxytocin signaling pathway included seven candidate genes (ADCY2, ADCY5, CACNA1D, GUCY1A2, NPR1, PLA2G4A, RYR3). Several studies supported the role of endogenous oxytocin in stimulating brown adipose tissue thermogenesis by allowing for adaptation for cold. The

expression of this neuropeptide hormone, produced by hypothalamus, was reported to be upregulated following cold exposure and to participate in brown adipose tissue activation (Talash et al., 2021).

Environmental cold stimulates brown adipose tissue thermogenesis through complex neural networks in the central nervous system which transmit afferent signals from cutaneous thermal receptors to brain thermosensitive neurons (Morrison et al., 2012). The preoptic area is the main control centre for body temperature and contains GABAergic neurons (Tappaz et al., 1977) that exert control over the thermoefferent mechanisms (Osaka, 2004).

The iHS test found five candidate genes (ADCY2, ADCY5, CACNA1D, GABBR2, GABRG3) enriched in the GABAergic synapse pathway, indicating an adaptive response to maintain body temperature in response to extremely low temperatures.

Potential limitation of the linkage disequilibrium approach

Compared to the gene-based methods that detect selective events occurring within the deep past at a macroevolutionary level (Yang, 2007), the linkage disequilibrium-based approaches identify putative regions of partial or incomplete selective sweeps at a microevolutionary level (Sabeti et al., 2007). Variants discovered by these latter methods are likely to be newly arisen as long-haplotype tests have limited sensitivity for detecting selection on standing (pre-existing) variation (Teshima et al., 2006).

Therefore, iHS and XP-EHH are suitable tests to identify putative new selected alleles that arose after the Emperor penguin's colonization of Antarctica, as unusually long haplotypes are unlikely to survive recombination effects for more than 1,000 generations.

However, iHS performance decreases in power if the populations experienced strong bottlenecks (Huff et al., 2010). This means that the test may have been more efficient in detecting signatures of selective sweeps in the Emperor penguin, as this species has been characterized by a much more stable population size over time (Cristofari et al., 2016) compared to the King penguin, which experienced two severe bottlenecks during its demographic history (Trucchi et al., 2014; Cristofari et al., 2018)

References

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N.A. & Goyache, F. 2020 Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Scientific reports* **10**, 1-13.

Ashburner, M., Ball, C.A., Blake, J.A. *et al.* 2000 Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29.

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

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

Browning, S.R. & Browning B.L. 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics* **81**, 1084–1097.

Cardona, A., Pagani, L., Antao, T., Lawson, D.J., Eichstaedt, C.A., Yngvadottir, B., Shwe, M.T.T., Wee, J., Romero, I.G., Raj, S. & Metspalu, M. 2014 Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PloS one* **9**, p.e98076.

Cherel, Y., Gilles, J., Handrich, Y. & Le Maho, Y. 1994 Nutrient reserve dynamics and energetics during long-term fasting in the king penguin (*Aptenodytes patagonicus*). *Journal of Zoology* **234**, 1-12.

Cristofari, R., Bertorelle, G., Ancel, A., *et al.* 2016 Full circumpolar migration ensures evolutionary utility in the Emperor penguin. *Nature Communications* **7**, 1-9.

Cristofari, R., Liu, X., Bonadonna, F., *et al.* 2018 Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change* **8**, 245-251.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. & McVean, G. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Duncan, R.E., Ahmadian, M., Jaworski, K., Sarkadi-Nagy, E., & Sul, H.S. 2007 Regulation of lipolysis in adipocytes. *Annual review of nutrition* **27**, 79–101.

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**, 231–241.

Garrison, E. 2016 Vcflib, a simple C++ library for parsing and manipulating VCF files. https://github.com/vcflib/vcflib

Garrison, E. & Marth, G. 2012 Haplotype-based variant detection from short-read sequencing. *arXiv preprint* arXiv:1207.3907.

Gavryushkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., *et al.* 2017 Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology* **66**, 57-73.

Groscolas, R. 1990 Metabolic adaptations to fasting in emperor and king penguins. In *Penguin Biology* (ed. LS Davis & JT Darby), 269–296. San Diego: Academic Press.

Groscolas, R. & Robin, J.P. 2001 Long-term fasting and re-feeding in penguins. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **128**, 645–655.

Hudson, R.R. 1995 Explaining low levels of DNA sequence variation in regions of the Drosophila genome with low recombination rates. *Tempo and Mode in Evolution: Genetics and Paleontology 50 Years After Simpson* **91**, 1275.

Huff, C.D., Harpending, H.C. & Rogers, A.R. 2010 Detecting positive selection from genome scans of linkage disequilibrium. *BMC genomics* **11**, 1-9.

Johansen, K. & Bech, C. 1983 Heat conservation during cold exposure in birds (vasomotor and respiratory implications). *Polar Research* **1**, 259-268.

Kanehisa, M. & Goto, S. 2000 KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30.

Ksepka, D.T.B.S. & Giannini, N. 2006 The phylogeny of the living and fossil Sphenisciformes (penguins). *Cladistics* **22**, 412–441.

Lewden, A., Nord, A., Bonnet, B. *et al.* 2020 Body surface rewarming in fully and partially hypothermic king penguins. *Journal of Comparative Physiology B* **190**, 597–609.

Li, C., Zhang, Y., Li, J. *et al.* 2014 Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience* **3**, 2047-2217X.

Li, H. & Durbin, R. 2009 Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker A., *et al.* 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M.A., Gamba, C., Seguin-Orlando, A. & Mortensen, C.D. 2015 Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences* **112**, E6889-E6897.

Liu, X., Ong, R.T.H., Pillai, E.N., Elzein, A.M., Small, K.S., Clark, T.G., Kwiatkowski, D.P. & Teo, Y.Y. 2013 Detecting and characterizing genomic signatures of positive selection in global populations. *The American Journal of Human Genetics* **92**, 866-881.

Modolo, L. & Lerat, E. 2015 UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC bioinformatics* **16**, 1-8.

Morrison, S.F., Madden, C.J. and Tupone, D. 2012 Central control of brown adipose tissue thermogenesis. *Frontiers in endocrinology* **3**, 5.

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.

Pirri, F., Ometto, L., Fuselli, S., Fernandes, F.A., Ancona, L., Le Bohec, C., Zane, L. & Trucchi, E. 2021 Selection-driven adaptation to the extreme Antarctic environment in the Emperor penguin. *bioRxiv*.

Raccurt, M., Baudimont, F., Tirard, J., Rey, B., Moureaux, E., Géloën, A. and Duchamp, C. 2008 Growing in Antarctica, a challenge for white adipose tissue development in Adelie penguin chicks (*Pygoscelis adeliae*). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **295**, R1671-R1679.

Raudvere, U., Kolberg, L., Kuzmin, I. *et al.* 2019 g:Profiler: a web server for functional enrichment analysis and conversions of gene lists. *Nucleic Acids Research* **47**, W191-W198.

Robin, J.-P., Frain, M., Sardet, C., Groscolas, R., and Le Maho, Y. 1988 Protein and Lipid Utilization During Long-Term Fasting in Emperor Penguins. *American Journal of Physiology* **254**, R61–R68.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. & Ackerman, H.C. 2002 Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-837.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R. & Schaffner, S.F. 2007 Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918.

Storey, J.D. & Tibshirani, R. 2003 Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440-9445.

Szpiech, Z.A. & Hernandez, R.D. 2014 selscan: an efficient multithreaded program to perform EHH based scans for positive selection. *Molecular Biology and Evolution* **31**, 2824–2827.

Talash, K., Eevuri, M.R. and Diep, P.T. 2021 A potential role for endogenous oxytocin in adaptation to cold: implications for health?. *Morecambe Bay Medical Journal* **8**, 267-270.

Tappaz ML, Brownstein MJ, & Kopin IJ. 1977 Glutamate decarboxylase (GAD) and γ-aminobutyric acid (GABA) in discrete nuclei of hypothalamus and substantia nigra. *Brain Research* **125**, 109–121.

Taylor, J.R.E. 1986 Thermal insulation of the down and feathers of pygoscelid penguin chicks and the unique properties of penguin feathers. *Auk* **103**, 160–168.

Teshima K.M., Coop G., Przeworski, M., *et al.* 2006 How reliable are empirical genomic scans for selective sweeps?. *Genome Research* **16**, 702–712.

Thomas, D.B. & Fordyce R.E. 2008 The heterothermic loophole exploited by penguins. *Australian Journal of Zoology* **55**, 317–321.

Trucchi, E., Gratton, P., Whittington, J.D., *et al.* 2014 King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140528.

Vianna, J.A., Fernandes, F., Frugone M.J., *et al.* 2020 Genome-wide analyses reveal drivers of penguin diversification. *PNAS* **117**, 22303-22310.

Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. 2006 A map of recent positive selection in the human genome. *PLoS biology* **4**, p.e72.

Williams, C.L., Hagelin, J.C. and Kooyman, G.L. 2015 Hidden keys to survival: the type, density, pattern and functional role of emperor penguin body feathers. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20152033.

Yang, Z. 2007 PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24, 1586-1591.

Zhen, Y. & Andolfatto, P. 2012 Methods to detect selection on noncoding DNA. *Methods in molecular biology* **856**, 141–159.

4. Comparative transcriptomics reveal candidate genes for adaptation to Antarctic environment in the Emperor penguin

This section includes the draft manuscript:

Federica Pirri, Samuele Greco, Marco Gerdol, Alberto Pallavicini, Marine Benoiste, Clément Cornec, Lorenzo Zane, Céline Le Bohec, Emiliano Trucchi;

"Comparative transcriptomics reveal candidate genes for adaptation to Antarctic environment in the Emperor penguin";

in preparation.

Introduction

The adaptation of populations to deeply different environments is one of the major speciation mechanisms (Darwin, 1859; Schluter, 2000; Coyne and Orr, 2004; Nosil, 2012). In this process, populations adapting to alternative ecological niches will accumulate genetic differences due to divergent selection (McGirr and Martin, 2020).

The contribution of gene expression in generating morphological novelty and its role in adaptive divergence remain poorly understood; however, several studies have identified relevant candidate genes in specific tissues or life stages of ecologically divergent species, demonstrating that variation in gene expression was important in the process of ecological speciation (Tautz, 2000; Wittkopp et al., 2008; Chan et al., 2010; Jones et al., 2012; Hanikenne et al., 2013).

When a population colonizes a new environment, gene expression becomes crucial in guaranteeing population persistence and can mediate phenotypic plasticity and contribute to adaptive divergence (Pavey et al., 2010). The genetic divergence in adaptive traits over time, eventually leads to reproductive isolation between populations (Pavey et al., 2010) with a significant impact on species diversification.

Aside from changes in coding genes, changes in gene expression may speed up adaptive evolution especially under strong selection pressures over short timescales (Brauer et al., 2017; Uusi-Heikkilä et al., 2017; Jones et al., 2012).
Here we used the two penguin species of the *Aptenodytes* genus, *A. forsteri* (the Emperor penguin) and *A. patagonicus* (the King penguin), as a model system to study adaptive divergence between two phylogenetically close species and investigate how changes in gene expression contributed to Emperor penguins' adaptation to an extreme cold environment.

The Emperor penguin is the only warm-blooded vertebrate that can survive and reproduce in the harshest Antarctic winter, dealing with extreme low temperatures and high winds as well as drastic seasonal fluctuations in daylight length (Blix, 2016; Goldsmith and Sladen, 1961). Whereas, its closest relative, the King penguin, breeds only on year-round ice-free sub-Antarctic islands and in Tierra del Fuego.
The common ancestor of the two Aptenodytes species was most likely adapted to a less cold climate, as the King penguin, thus the extreme cold adaptation of the Emperor penguin would be a derived feature which evolved secondarily (Pirri et al., 2021; Vianna et al., 2020).
Emperor penguins possess a wide range of unique morphological, physiological and behavioral adaptations to withstand the extreme Antarctic environment, such as a powerful peripheral vasoconstriction to reduce heat losses (Thomas and Fordyce, 2008), densely packed feathers useful for thermal insulation (Watson, 1883; Taylor, 1986) and an efficient energy storage management system for long-term fasting (Groscolas, 1990; Cherel et al., 1994; Groscolas and Robin, 2001).
The Emperor penguin's unique tolerance to these extreme environmental conditions may have been acquired within a relatively short period of time (ca. 1-2 Mya; Gavryushkina et al., 2017).

To explore the relevance of gene expression changes in adaptation to the extreme cold Antarctic environment, we compared transcriptomic data of the Emperor penguin with those of its less cold-adapted relative, the King penguin, by performing differential gene expression analysis across multiple tissues.
Most comparative studies on cold adaptation have been undertaken using genomic approaches, whereas the gene expression patterns based on transcriptomic data have been less investigated and

strictly focused on plants (Yang et al., 2015; Miao et al., 2021; Grønvold et al., 2017; Li et al., 2019), bacteria (Raymond-Bouchard et al., 2018; Teoh et al., 2021), fishes (Ansaloni et al., 2021; Song and McDowell, 2021; Kavembe et al., 2015; Chen et al., 2008; Shin et al., 2012), insects (Schoville et al., 2021) and molluscs (Liu et al., 2020).

The recent advancement of high-throughput sequencing technologies has enabled the simultaneous quantification of thousands of genes' expression across diverse organs and tissues, enhancing our ability to understand the molecular mechanisms that drive evolution (Guo et al., 2015).

RNA-Seq generates a vast amount of data about transcription levels allowing for a thorough examination of global gene expression patterns as well as boosting the efficiency of identifying the genes of interest (Costa et al., 2010).

Gene expression patterns may reveal genes involved in adaptations that are difficult or impossible to detect in other ways and that underlie possibly relevant phenotypes for ecological divergence, also in non-model organisms (Feder and Mitchell-Olds, 2003; Costa et al., 2010). The last aspect is particularly significant because our understanding is currently confined to a small number of taxa (Ghiselli et al., 2018).

Here, we first used total mRNA-Seq data from 5 tissues of 3 individuals per species to *de novo* assemble the first reference transcriptome of the Emperor and the King penguin. Then, we explored transcriptomic differences at intraspecific and interspecific levels by analyzing QuantSeq 3'mRNA-Seq data from a large number of tissue samples (100 samples from 5 tissues of 10 individuals from natural populations of each species), in order to elucidate the molecular mechanisms behind the adaptation of these two penguin species to contrasting habitats.

Our analyses concerned tissues that could exhibit the major genetic differences for cold adaptation comparing the two species: skin (thermal insulation), liver (lipid and fatty-acid metabolism), brain (cold tolerance), muscle (thermogenesis) and kidney (osmoregulation).

In general, the expression profiles revealed a characterizing tissue-clustered pattern and a number of differentially expressed genes which, in the Emperor penguin, could be the candidates underlying its relevant adaptations to the Antarctic lifestyle.

Moreover, our *de novo* transcriptomes for the two Aptenodytes penguin species, that we describe and characterize here for the first time, will significantly contribute to the genomic resources already available for penguins, allowing future analysis and downstream applications of genome annotation, gene expression, and sequence evolution.

Material and Methods

Samples collection and preparation

From June to September 2016, samples of five different tissues (brain, liver, kidney, skin and muscle) were collected from 10 freshly-predated 3-7 months old chicks from natural populations of Emperor and the King penguin, respectively, for a total of 100 samples. Samples were collected at colonies near the Dumont d'Urville Station in Antarctica and the Alfred Faure station in Crozet Islands, respectively. All tissue samples used in this study were collected immediately after death, directly fixed in RNAlater (Applied Biosystems, Warrington, UK) and transferred to a −80°C laboratory freezer.

RNA extraction, RNA-seq library construction and sequencing

Total RNA was isolated from 40 mg of each tissue sample by a standard laboratory-based chloroform extraction after homogenization in 500 ul of TRIzol® reagent (Invitrogen, ThermoFisher Scientific).

100 µl of chloroform was added and samples were vortexed vigorously. Samples were subsequently centrifuged 12,000×g for 15 min at 8°C. The upper aqueous phase was collected and transferred to a new tube for alcohol precipitation with isopropanol. The subsequent RNA/isopropanol solution was centrifuged at 12,000×g for 10 min at 8 °C. The RNA pellet was washed with 75% ethanol and centrifuged at 7,500×g for 5 min at 8 °C. Ethanol was removed and the RNA pellet was resuspended in RNase free water and stored at -80C°.

RNA extraction from skin and muscle is challenging, as the large amount of contractile proteins, connective tissue, and collagen of these tissues are extremely difficult to homogenize.

Attempts to isolate total RNA from skin and muscle using TRIzol resulted in poor yields and low purity; therefore, RNA from these two tissues was extracted using the RNeasy Fibrous Tissue Mini Kit (Qiagen) according to the manufacturer's instructions. Also in this case, isolated RNA was dissolved in RNase free water and stored in -80C°.

An aliquot of the extracts was used to assess the concentration and purity (i.e., the A260/A280 ratio) of each RNA sample by Nanodrop 2000 (Thermo Fisher Scientific) and Qubit 4.0

fluorometer (ThermoFisher Scientific). RNA integrity was evaluated by UV transilluminator and Agilent 2100 Bioanalyzer (Agilent technologies, Santa Clara, CA).

Along with the 100 tissue samples (10 biological replicates of 5 tissues for each species), we pooled 5 different tissues, from 3 specimens of each species, based on their RNA concentrations, for a total of 6 RNA pools (3 RNA pools for each species) required for assembling the reference transcriptome for each species.

RNA-seq library preparation and sequencing was carried out by BMR Genomics Service (Padova, Italy).

Libraries were synthesized for each of the 3 pooled samples of each species using the TruSeq Stranded mRNA Sample Prep kit (Illumina, San Diego, CA), according to the manufacturer's instructions.

Poly-A mRNA was fragmented for 3 minutes at 94°C, and each purification step was carried out with 1 × Agencourt AMPure XP beads. Following the manufacturer's instructions, libraries were processed with the Illumina cBot for cluster generation on the flow cell.

Paired-end sequencing (100 bp from each end) was then performed on the Novaseq 6000 (Illumina, San Diego, CA) at a sequencing depth of 20 million reads per library.

Secondly, libraries were also prepared from the extracted RNA from 5 different tissues of 10 individuals of each species using the QuantSeq 3'mRNA-Seq Library Prep Kit (Lexogen, Vienna, Austria), per the manufacturers' instructions using 1 μg of RNA per library.

The generation of these libraries was initiated by oligo-dT priming, without the need for prior poly(A) enrichment or ribosomal RNA depletion (Moll et al., 2014).

The pooled libraries were sequenced using a NextSeq500 instrument (Illumina, San Diego, CA) for single-end 75 bp read lengths at a sequencing depth of 5 million reads per library.

QuantSeq provides an accurate method for gene expression measurement, also at low read depths. It generates only one fragment per transcript, making the number of reads mapping to a gene proportional to its expression (Moll et al., 2014).

De novo assembly of the Emperor and the King penguin transcriptome

Separate assemblies were made for *A. forsteri* and *A. patagonicus* paired-end (PE) Illumina sequencing reads using the same bioinformatic pipeline and the total mRNA Seq data from the three pools for each species.

First, raw reads quality was examined using FastQC v0.11.9 (Andrews, 2010) and MultiQC (Ewels et al., 2016). Trimming was performed on fastp v0.20.1 (Chen et al., 2018) by removing Illumina sequencing adapters and low-quality bases on a sliding window approach (size = 4, minimum average quality = 15) and enabling polyX and polyG tails elimination.

The resulting trimmed sequences shorter than 71 bp were discarded and quality of the remaining reads was evaluated again with FastQC and MultiQC.

We deemed a *de novo* assembly to be more appropriate in order to prevent any bias that a genome-guided approach might have introduced. We performed *de novo* transcriptome assembly, separately for each species, using the Oyster River Protocol (ORP) version 2.2.5 (MacManes, 2018).

By taking advantage of the strengths of different assembly tools, the ORP pipeline improves assembly quality and mapping rate, recovering expressed transcripts that may be missed by individual assemblers (MacManes, 2018).

The ORP started by error-correcting trimmed reads using Rcorrector version 1.0.4 (Song and Florea, 2015). The remaining reads were then assembled using three different *de novo* assemblers and dissimilar kmer lengths: Trinity release 2.11.0, with kmer length = 25 and without read normalization (Haas et al., 2013), rnaSPAdes version 3.14.1 with kmer length = 55 and 75 (Bushmanova et al., 2019), and Shannon version 0.0.2 with kmer length = 75 (Kannan et al., 2016). Contigs that were expressed at less than 1 transcript per million were removed with the "TPM_FILT = 1" flag.

The process above resulted in four distinct assemblies that were then merged and clustered into isoform groups using OrthoFuse (MacManes, 2018).

After Orthofuse has finished, the merged assembly was run through a modified version of TransRate v.1.0.3 (Smith-Unna et al., 2016), which is packaged with the ORP, and the best (= highest contig score) transcript from each group was placed in a new assembly file to represent the entire group. The resultant file, containing the top scoring contig for each orthogroup, was used for all downstream analyses.

Contigs derived from ribosomal and mitochondrial DNA were detected with BLASTn (Altschul et al., 1990) on the commercially available CLC Genomic Workbench 4.5.1 (CLC Bio, Katrinebjerg, Denmark), based on significant similarity (e-value threshold < 1e-37), against a dataset containing all the Sphenisciformes mitochondrial and ribosomal sequences deposited in Genbank, and consequently removed from both the assemblies. This step was then repeated using an e-value threshold < 1.66e-30 against a dataset represented by all the Neognathae complete mitochondrial and ribosomal sequences deposited in Genbank. Again, all contigs that found a match were removed from both the assemblies.

Lastly, to ensure the creation of mitochondrial and ribosomal sequences-free assemblies, the complete set of trimmed reads of each species were mapped to the filtered reference transcriptome using Salmon v1.5.2 (Patro et al., 2017) and the identity of the most overrepresented transcripts was determined with BLASTn (Altschul et al., 1990).

As a final step, only transcripts longer than 250 bp were kept.

Assembly quality assessment

The overall quality of the reference transcriptome assembly for each species was evaluated with BUSCO v.5.2.2 (Simão et al., 2015). The analysis of the Benchmarking Universal Single-Copy Orthologs allowed the estimation of completeness of our two assembled transcript sets by comparing them to the predefined set of 8338 Aves highly conserved single-copy orthologs from the OrthoDB v9.1 database (Zdobnov et al., 2017). We calculated the number of complete (length is within two standard deviations of the mean length of the given BUSCO), duplicated (complete BUSCOs represented by more than one transcript), fragmented (partially recovered BUSCOs) and missing (not recovered) in each of the two *de novo* assemblies.

We also used TransRate version 1.0.3 (Smith-Unna et al., 2016) in order to further assess the confidence and completeness of each assembly.

Additionally, we calculated the ExN50 statistics as a better measure of the assembly contiguity. According to standard Nx length statistics, at least x% of assembled transcript nucleotides are found in contigs of at least Nx length. ExN50 statistics seems more appropriate for transcriptome data (Haas et al., 2013) as in this case the N50 value is computed based on the top highly expressed transcripts, such that the subset accounts for x% of gene expression.

When the ExN50 has a maximum value of "x" greater than 90%, the assembly is deemed to have good coverage and deeper sequencing is unlikely to result in a higher-quality assembly (Galachyants et al., 2019).

The expression levels were estimated by mapping back the reads to the assembled transcripts of each species using Salmon and the statistics were generated by Trinity accessory scripts (Haas et al., 2013).

Transcript functional annotation

Functional annotation was conducted using the annot.aM pipeline [https://gitlab.com/54mu/annotaM].

First, Transdecoder v.5.5.0 (Haas et al., 2013) was used to identify open reading frames (ORFs) and translate all contigs to putative protein sequences. Then they were annotated by searching against the UniProtKB/Swiss-Prot database using BLASTp and BLASTx (Altschul et al., 1990), allowing the assignation of the homology IDs via the best hit (e-value threshold < 1e-5).

These homologies were used to associate each transcript to cell component, molecular function and biological process Gene Ontology (GO) terms (Ashburner et al., 2000) and Reactome pathways annotations (Fabregat et al., 2017).

Protein sequences were also analyzed with Hmmer v.3.3.1 (Finn et al., 2011), searching for conserved domains included in the Pfam 31.0 database (Punta et al., 2012).

lncRNA detection and annotation

Long non-coding RNAs (lncRNAs), which have lengths of 200 base pairs or longer, are unlikely to produce proteins but they still may play a key role in eukaryotic gene regulation (Han et al., 2019).

To identify putative lncRNAs in the transcriptome assembly of each species, we performed a series of filtering steps to exclude potential protein-coding RNAs.

First, we selected transcripts with a size greater than 500 bp from the set of sequences for which an ORF has not been predicted by TransDecoder.

The size cutoff was totally arbitrary, but it allowed to exclude the majority of known but still poorly understood classes of small infrastructural and regulatory RNAs, such as tRNAs, small nuclear RNAs, small nucleolar RNAs and their derivatives, microRNAs, short interfering RNAs, Piwi-interacting RNAs, transcription-initiation RNAs and small RNAs that regulate splicing (Morris and Mattick, 2014).

We then excluded from the list of putative lncRNAs, all transcripts that have been annotated by searching against the UniProtKB/Swiss-Prot database using BLASTx.

The remaining transcripts were intersected through a BLASTx search with a custom database containing all the 99.009 complete and partial protein sequences of 7 species (*Calonectris borealis*, *Fregetta grallaria*, *Fulmarus glacialis*, *Hydrobates tethys*, *Oceanites oceanicus*, *Pelecanoides urinatrix*, *Thalassarche chlororhynchos*) belonging to the Procellariiformes order, the Sphenisciformes sister taxon, deposited in NCBI. All the hits with an e-value greater than 1e-5 were kept.

A BLASTn search was then carried out on CLC Genomic Workbench against a custom database encompassing all the 104,235 mRNA sequences of the seven Procellariiformes species previously selected. Again, all the hits with an e-value greater than 1e-5 were kept.

As the final step, the putative lncRNAs were identified by searching against the reference genome of each Aptenodytes species using BLASTn on CLC Genomic Workbench based on an e-value threshold of 1e-5 and a greatest identity score of 98%.

In order to assess how many putative lncRNAs were shared by the two Aptenodytes species, we used cd-hit-est version 4.8.1 (Li and Godzik, 2006) and a 90% sequence similarity threshold.

The expression levels of the final putative lncRNAs were calculated using the transcripts per million (TPM) derived from the mapping of the trimmed reads to each species reference transcriptome using Salmon with default parameters.

Differential expression analyses

Comparing the transcriptional profiles across tissues in each species

The 3'end reads from each of the 10 biological samples of all five tissues (brain, liver, kidney, muscle, skin) of each species were trimmed using fastp by removing Illumina sequencing adapters,

low quality bases and polyX and polyG tails. Trimmed reads quality was then evaluated with FastQC and MultiQC.

Filtered reads of each species were mapped to the reference transcriptome using Salmon, in order to have a transcript-level abundance estimation across tissues in each species separately. We set the --validateMappings option to limit multi-mapping by avoiding re-computing alignment scores against redundant transcript sequences, --seqBias to correct for sequence-specific biases such as random hexamer priming bias and --gcBias to correct for fragment GC content as FASTQC and MultiQC revealed a high GC percentage in all samples.

The expected read count of each transcript across all samples of each tissue was combined into a matrix for each species and normalized using the trimmed mean of the M- values (TMM) protocol, a robust normalization method that equalizes the overall expression levels of genes across samples assuming that most of them are not differentially expressed (Robinson and Oshlack, 2010).

Only genes with at least 10 reads in at least seven samples of the same tissue were kept. This aimed at removing genes too lowly expressed for establishing any worthwhile significant differences and thus limit false positives.

In addition to a difference in sequencing depth between samples, RNA-Seq data may suffer from another problem, a mean-variance relationship, i.e. heteroskedasticity (variance of fold change depending on mean count). This phenomenon implies that weakly expressed genes seem to show much stronger differences than highly expressed genes (Love et al., 2014).

Data heteroskedasticity was evaluated by plotting the mean CPM (Counts per Million) for each transcript across all biological replicates of each tissue against the standard deviation.

The edgeR estimateDisp function (Robinson et al., 2010) was used on the design matrix to estimate the common dispersion and thus the variability across the dataset. The dispersion estimate over all genes was represented in a BCV (biological coefficient of variation) plot showing the square root of the common dispersion (i.e., BCV) for each gene. BCV is the coefficient of variation with which the (unknown) true abundance of the gene varies between biological replicates of RNA samples (Robinson et al., 2010).

Prior to differential expression (DE) analysis, we performed a multidimensional scaling (MDS) in the RStudio package (RStudio Team, 2020), based on the adjusted gene expression matrix, in order to visualize the similarity of expression patterns across samples of the same tissue in each species.

DE analyses were conducted at a transcript level in pairwise comparisons between the five different tissues (each tissue against the other four), in each species separately, using edgeR.
The glmQLFit function and the glmQLFTest function were employed to discover which genes were differentially expressed in each pairwise comparison based on a negative binomial Generalized Linear Model (GLM) (McCarthy et al., 2012).
Genes with significant results (FDR < 0.05) and fold change (FC) > 5 were identified as differentially expressed (DEGs) in each comparison.
DEGs from all the pairwise comparisons based on the same reference tissue, were merged into a single list and each DEG was kept just once so that only the most peculiar genes expressed in that specific tissue in each species were retained.
Moreover, to filter out noise from the sets of DEGs of each tissue, we set an expression threshold based on TPM, specific for each tissue of each species. Only DEGs which contributed to 90% of the total expression of that particular tissue were retained.
TPM gene expression values were added 1 unit and log 10 -transformed before being plotted in a heatmap which groups transcripts with similar expression trends based on Euclidean distance and average linkage criteria.
To determine the biological function of the DEGs for each tissue for each species, we performed functional enrichment analysis with a hypergeometric test using all annotated reference genes as the background list (Falcon and Gentleman, 2008).
Significant GO and Pfam enrichment was determined for FDR lower than 0.05, paired with a difference between observed and expected values higher than 1.

Comparing the transcriptional profiles between the two Aptenodytes species

We then performed a DE analysis to compare gene expression of the same tissue between the Emperor and the King penguin. For this purpose, we chose to align both species to the same reference genome (the Emperor penguin one) rather than to the *de novo* assembled transcriptome of each species followed by reciprocal alignment. Indeed, when we mapped the orthologous genes of the two species to the reference transcriptome assembly of each species, a strong unlikely species-specific expression signal emerged: samples of different tissues from the same species clustered together rather than with samples of the same tissue from the other species. A possible

explanation is that the reads of the ortholog genes mapped unevenly in the two transcriptomes, due to the absence of the 3'UTRs of several orthologous genes in the transcriptome of one species compared to the other.

On the other hand, mapping to the genome reduces the proportion of poorly aligned reads while increasing alignment rates (Delhomme et al., 2014).

We did not expect any particular bias from using a single reference for both species as the genetic divergence between them was estimated to be less than 1% by computing the average number of differences between pairs of sequences of the two species (Dxy) on a 2kb region located at the 3'UTR.

Trimmed 3'end reads of each sample of both species were mapped to the *A. forsteri* reference genome (RefSeq assembly accession: GCF_000699145.1) using STAR v.2.7.9a (Dobin et al., 2013).

After indexing the resulting bam files with SAMtools v.1.12 (Li et al., 2009), the count of overlapping reads for individual genes (that is, all transcript isoforms for that gene) was carried out with HTseq (Anders et al., 2015). Multi-mapped and overlapping multiple expression features reads were discarded. This did not affect the ratio of expression strength (i.e., the FC) between samples as the same fraction of reads were removed in all samples. Applying this filter, the likelihood of generating false positives has been reduced.

All counts for each sample were then merged in one matrix for each species and then normalized through the TMM method incorporated into edgeR.

Based on this normalized gene expression matrix, we performed a MDS analysis to assess the similarity of expression patterns across the same tissues in the two different species.

Only genes with an expression value greater than 0.5 CPM in at least 3 biological replicates of each tissue in each species were retained for downstream analyses; as a result, genes that had consistently very low counts across all the samples have been removed. From a statistical point of view, low counts do not provide enough statistical support to classify those genes as significantly differentially expressed (Chen et al., 2016).

edgeR was used to perform comparisons between expression levels of genes of the same tissue between the two sibling species based on a negative binomial GLM (brain Emperor vs brain King, liver Emperor vs liver King, kidney Emperor vs kidney King, skin Emperor vs skin King, muscle Emperor vs muscle King).

From each comparison, significant genes (FDR < 0.05) showing a positive -$\log_2$FC value > 2 were considered as up-regulated in the Emperor penguin while those with a negative -$\log_2$FC value < -2 were considered as up-regulated in the King penguin.

We produced a volcano plot for each pairwise tissue comparison, using the RStudio package, in order to visualize those candidate DEGs which displayed the greatest significant expression changes.

DEGs obtained from the comparative analyses in edgeR were used to perform functional annotation and pathway enrichment analyses.

We tested candidate genes for functional GO terms enrichment and KEGG biological pathways by using the g:GOSt function in g:Profiler (Raudvere et al., 2019) and flycatcher (*Ficedula albicollis*) as reference species. Significantly enriched categories included at least two genes and the Benjamini-Hochberg method was used for multiple testing correction to estimate significance ($p < 0.05$).

Interproscan v.5.34-73.0 (Jones et al., 2014) was then used to perform conserved domain search in the Pfam database.

In order to visually explore physical, biochemical and regulatory interactions between up-regulated gene products and gain insights into underlying biological processes, we used NetworkAnalyst 3.0 (Xia et al., 2015) to build tissue specific protein-protein interaction (PPI) networks for each species. Specifically, we created minimum interaction networks to keep only those nodes (proteins) that were necessary to connect the seed nodes (up-regulated gene products); the underlying protein interaction data were obtained from DifferentialNet database (Basha et al., 2018) using *H. sapiens* as a model organism.

The 20 most prominent up-regulated genes in each tissue of each penguin species were then characterized based on the gene function description from the human gene database GeneCards (Stelzer et al., 2016).

Results

The Emperor and King penguin reference transcriptome assembly and annotation

All RNA was high quality; A260/280 ratios were ~ 2 in all three RNA pools of each species and all samples resulted to have a RNA integrity number (RIN) greater than 6.

RNA-seq libraries yielded between 32 and 42 million paired-end reads per pool. Quality filtering removed approximately 6% of the raw reads. This resulted in high quality RNA-seq datasets, which contained a total of 208.5 million and 183.7 million paired-end reads for the Emperor and the King penguin respectively.

All filtered reads were combined into a single dataset for each species and used to *de novo* assemble a reference transcriptome.

In total, 106,060 contigs with an average length of 1,047 bp and an N50 length of 2,429 bp were obtained in the Emperor penguin transcriptome, accounting for a total assembly size slightly higher than 111 Mb; whereas in the King penguin transcriptome, 806,05 contigs with an average length of 1,291 bp and an N50 length of 2,511 bp were obtained, accounting for a total assembly size slightly higher than 104 Mb (Table 1).

The reference transcriptome assembly of each species was then used in the following differential expression analysis at intraspecific level.

By mapping the paired-end reads for each species back to their transcriptome, we detected good sequencing read coverage especially for the Emperor transcriptome. The transcript expression analysis showed that the assembled transcripts were covered by 82.3% of the Emperor penguin reads and 78.6% of the King penguin reads.

Assembly E90N50 (the contig N50 value computed on the set of transcripts representing the top most 90% of expression data) was 3,377 bp in the Emperor and 3,367 bp in the King hence higher values than those based on the entire data set, suggesting good quality transcriptome assembly (Table 1).

While size-based metrics can be employed to evaluate assembly continuity, they cannot be utilized to determine assembly completeness.

The quality of both the reference transcriptomes was evaluated by searching against single-copy orthologs (8338 genes shared within the Aves lineage) using BUSCO. The results showed that

7,196 (86.3%) and 7,022 (84.2%) complete BUSCOs were included in the current reference transcriptome for the Emperor and the King penguin, respectively (Table 1).

Only 9.9% (Emperor penguin) and 11.6% (King penguin) of the 8338 single-copy orthologs were classified as missing from our assemblies, suggesting good coverage and high level of completeness of the protein-coding transcriptomes for these species (Table 1). At the same time, they both encompassed a very high proportion of contigs containing complete open reading frames, as evidenced by the detection of just 3.8% (Emperor penguin) and 4.2% (King penguin) fragmented BUSCOs (Table 1).

These BUSCO values indicated great levels of completion considering that the assemblies were generated from a small number of tissues and a single developmental stage.

Table 1. General metrics and BUSCO values for *de novo* transcriptome assemblies of *A. forsteri* (Emperor penguin) and *A. patagonicus* (King penguin) based on paired-end total mRNA sequencing.

General assembly metrics	***A. forsteri***	***A. patagonicus***
NUMBER OF CONTIGS	106,06	80,605
AVERAGE CONTIG LENGTH	1,047	1,291
CONTIG N50	2,429	2,511
E90N50	3,377	3,367
TOTAL ASSEMBLED BASES	111085247	104068068
GC%	47.6%	48.2%
FUNCTIONALLY ANNOTATED CONTIGS	56,299	55,297

BUSCO results		
COMPLETENESS	86.3%	84.2%
FRAGMENTED	3.8%	4.2%
MISSING	9.9%	11.6%
COMPLETE BUSCO	7,196	7,022
COMPLETE AND SINGLE-COPY BUSCO	5,475	5,355
COMPLETE AND DUPLICATED BUSCO	1,721	1,667
FRAGMENTED BUSCO	319	352
MISSING BUSCO	823	964
TOTAL BUSCO SEARCHED	8,338	8,338

annot.aM provided functional annotation for 56,299 (53%) transcripts for the Emperor penguin and for 55,297 (69%) transcripts for the King penguin (Table 1), which is within the typical rates of annotation for non-model organisms (Baeza and MacManes, 2020). Overall, in the Emperor assembly, 55,673 contigs obtained a positive BLASTx hit and 37,849 contigs obtained a positive BLASTp hit, 37,756 were annotated with Hmmer and 46,209 were annotated with TransDecoder. In the King assembly, 54,929 contigs obtained a positive BLASTx hit and 39,036 contigs obtained a positive BLASTp hit, 38,916 were annotated with Hmmer and 47,930 with TransDecoder. Note that many transcripts were annotated by multiple tools.

Lastly, we conducted gene enrichment analysis using two databases: GO and Reactome. In the Emperor assembly, 50,315 transcripts (47%) were assigned to 15,001 GO terms from the three main gene ontologies (molecular functions, biological processes and cellular components) while in the King assembly, 49,175 transcripts (61%) were assigned to 14,854 GO terms.

A wide range of GO terms emerged in each assembly, indicating that molecular functions, biological processes and cellular components were well represented. We observed a high uniformity in GO profiles across the two transcriptomes suggesting the global similarity of the phylogenetically close-related species (Fig. 1); GO terms consistency across multiple species assemblies was already reported by previous studies (Riesgo et al., 2012; Birol et al., 2015; Kobayashi et al., 2009).

Regarding the molecular function, the most represented GO terms included metal ion binding (GO:0046872), ATP binding (GO:0005524) and oxidoreductase activity (GO:0016491).

While for the biological pro-cesses category, the most frequent GO terms were related to signal transduction (GO:0007165), lipid metabolic process (GO:0006629) and cellular response to DNA damage stimulus (GO:0006974) (Fig. 1).

The Reactome pathways enrichment resulted in 24,341 transcripts for the Emperor penguin and 23,994 transcripts for the King penguin associated with pathways of cellular processes, immune system, environmental information sensing, platelet degranulation and mitochondrial activity (Fig. 2).

We identified 3,000 putative lncRNAs with an average length of 1,000 bp from the Emperor transcriptome assembly and 2,491 putative lncRNAs with an average length of 1,058 bp in the King transcriptome assembly, covering the 2.83% and the 3.09% of the assemblies respectively.

Comparing the putative lncRNAs sequences between the two species, 295 of these elements showed a sequence similarity greater than 90% between them.

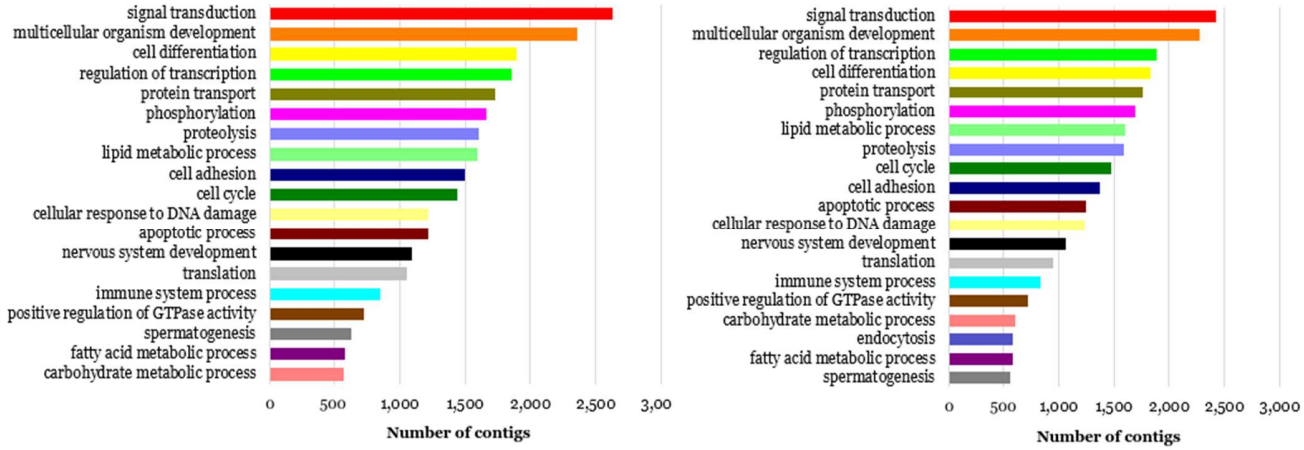

Figure 1. Functional annotations of the Emperor (left) and the King penguin (right) transcriptome using gene ontology (GO) terms.

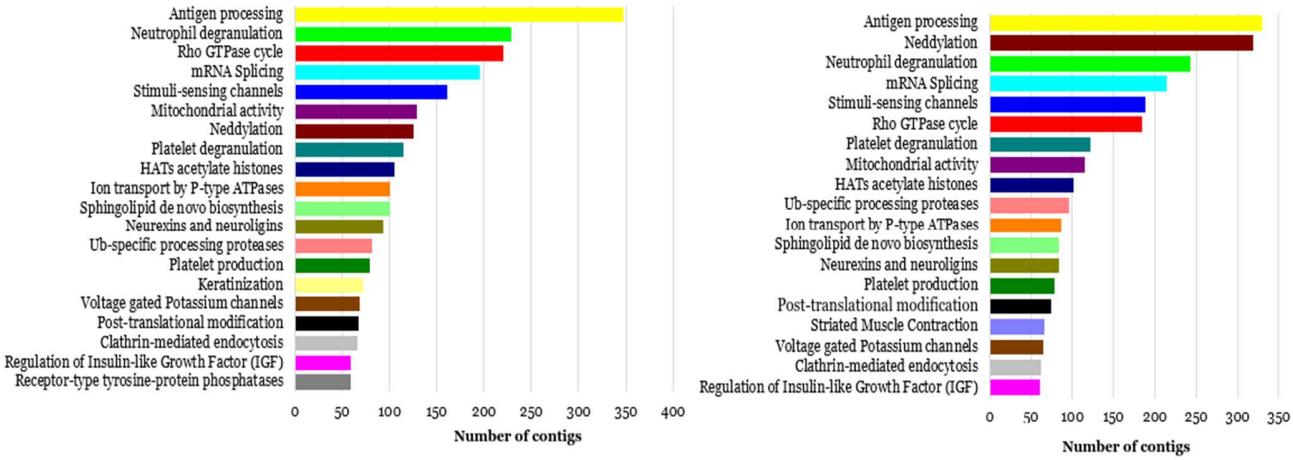

Figure 2. Functional annotations of the Emperor (left) and the King penguin (right) transcriptome using Reactome pathways.

Differential expression analysis at intraspecific level

All the 100 RNA samples from the five tissues resulted to have A260/280 ratios between 1.9 and 2.1 and a RIN greater than 6.

RNA-seq 3'end libraries yielded between 5 and 10 million single-end reads per library. Almost 95% of the raw reads passed the quality filtering. This resulted in high quality RNA-seq datasets, which contained a total of 70 million single-end reads for each of the five tissues of both the Emperor and the King penguin.

After transcript-level abundance estimation across tissues was generated by mapping the reads to the reference transcriptome, samples from each penguin species were clustered separately in MDS plots based on the adjusted gene expression matrix. The MDS plots showed that the biological replicates of the same tissue were well aggregated in each species, revealing a good quality dataset for each species (Fig. 3).

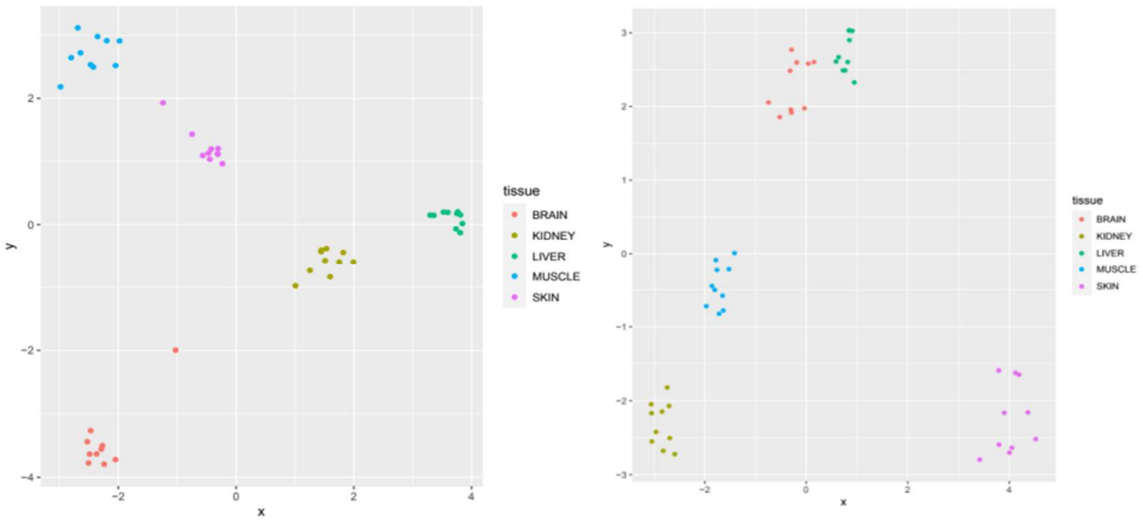

Figure 3. Multidimensional scaling (MDS) plot of the biological replicates of all tissues for the Emperor (left) and the King penguin (right).

The scatterplot of mean CPM versus standard deviation for each transcript across all biological replicates of each tissue revealed no significant variance in expression estimates for genes with low read counts, suggesting data homoscedasticity and allowing comparison of different libraries (Fig. 4).

Before testing for differential expression at intraspecific level, the BCV was employed to estimate the variability across the dataset for each species. The overall (common) BCV between the biological replicates of all tissues was estimated as 0.7, showing a high degree of heterogeneity between them, either in the Emperor and in the King penguin (Fig. 5-6).

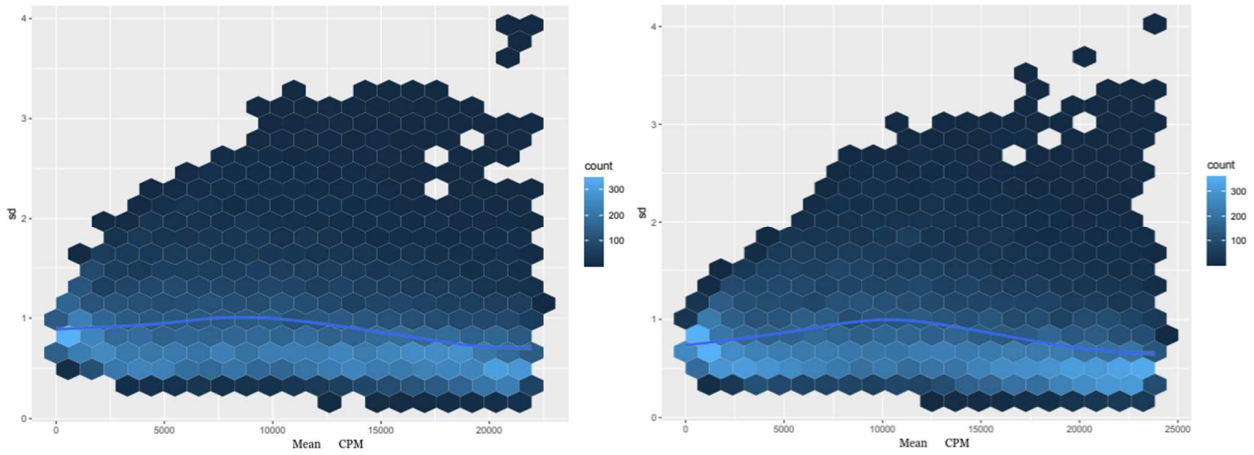

Figure 4. Scatterplot of mean CPM versus standard deviation for each contig across all biological replicates of each tissue for the Emperor (left) and the King penguin (right).

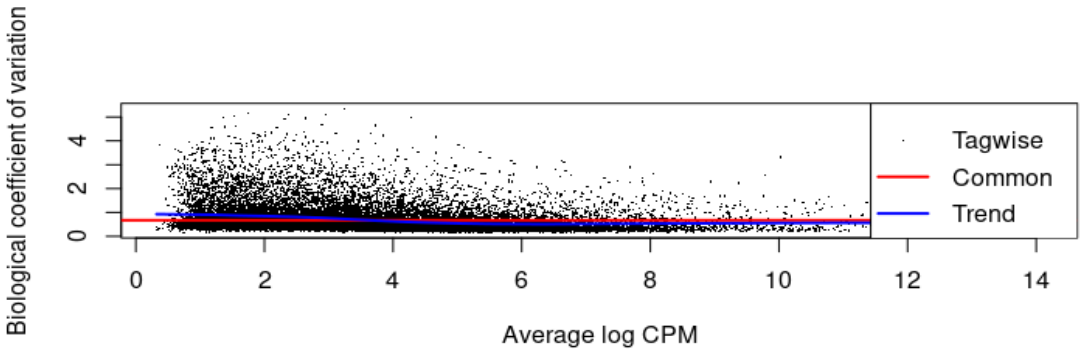

Figure 5. Biological coefficient of variation (BCV) plot for the Emperor penguin.

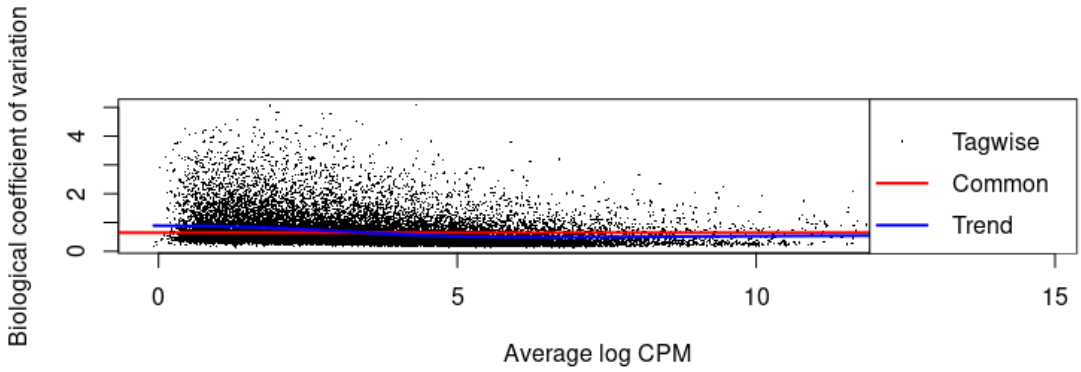

Figure 6. Biological coefficient of variation (BCV) plot for the King penguin.

A total of 378 contigs (0.36% of 106,060 contigs) showed evidence of a statistically significant difference in tissue-specific expression (FDR < 0.05) in the Emperor penguin. Specifically, 89 transcripts resulted differentially expressed in liver, 46 in kidney, 62 in brain, 122 in muscle and 59 in skin.

While in the King penguin, the pairwise tissue comparison at intraspecific level reported a total of 359 differentially expressed contigs (0.45% of 806,05 contigs). Precisely, 51 transcripts showed significant changes in expression in liver, 86 in kidney, 52 in brain, 29 in muscle and 141 in skin. From the heatmaps of the total DEGs in each species (Fig. 7-8) emerged clear differences among different tissues and a homogeneous expression pattern among the biological replicates of the same tissue.

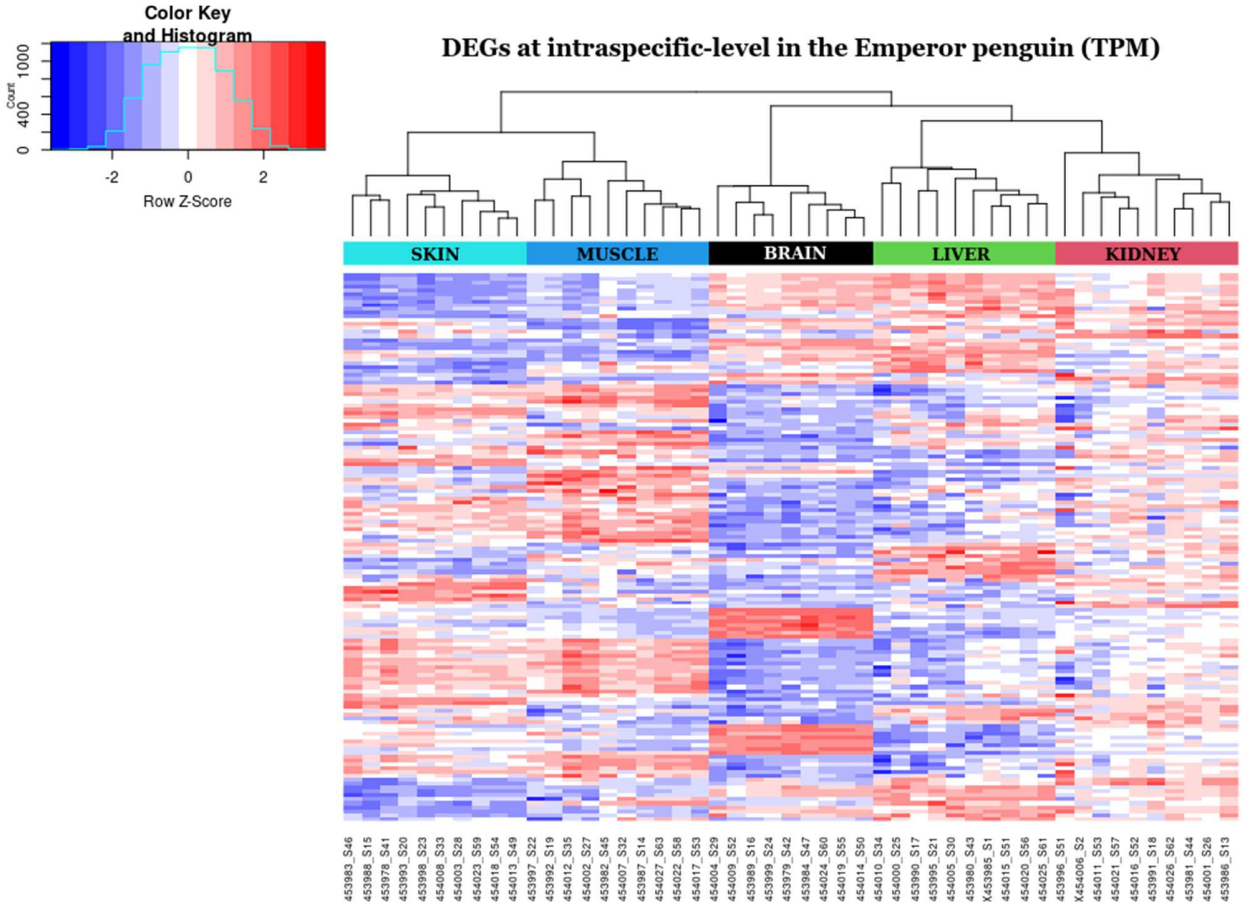

Figure 7. Heat map showing expression patterns of 378 DEGs in all 5 tissues of 10 individuals of the Emperor penguin. Expression was defined using TMM-normalized values with blue and red representing comparatively lower and higher expression levels, respectively.

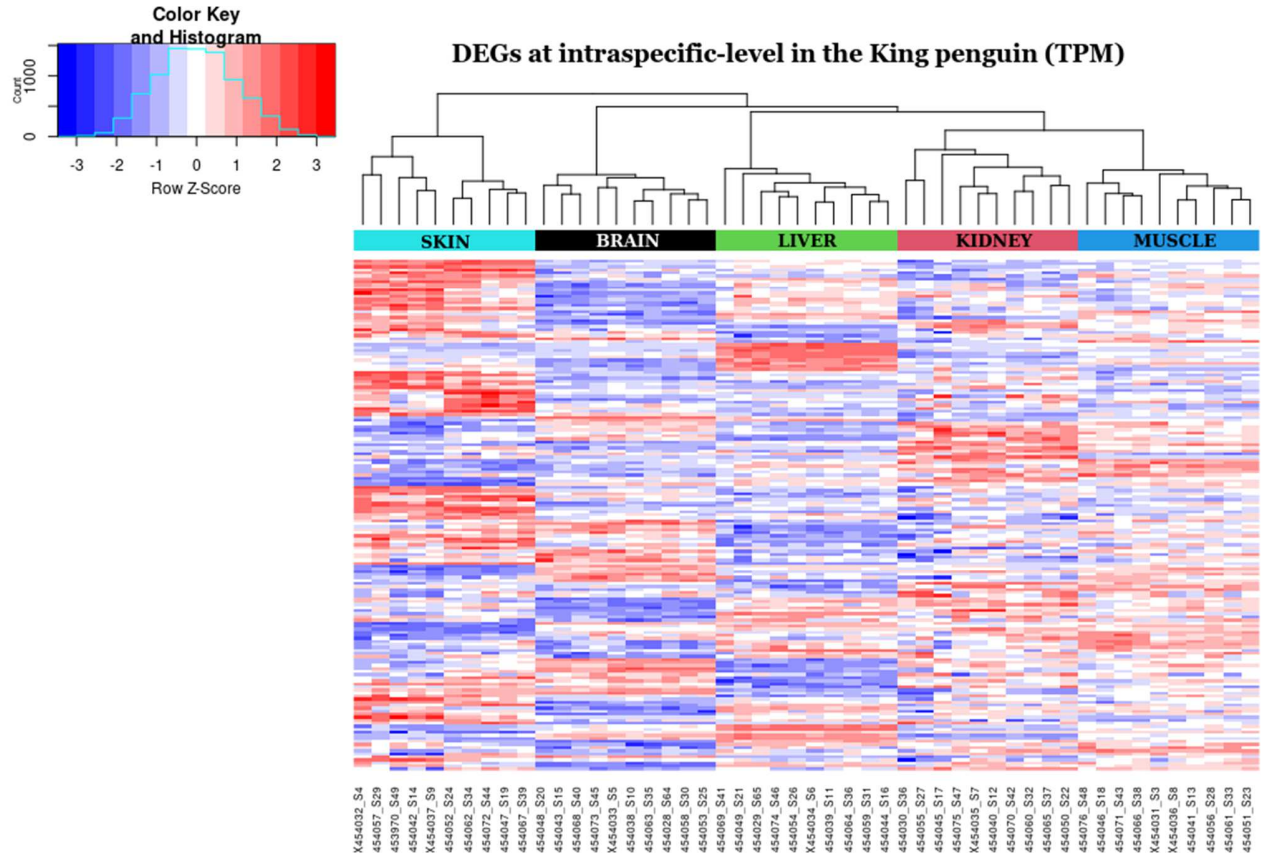

Figure 8. Heat map showing expression patterns of 359 DEGs in all 5 tissues of 10 individuals of the King penguin. Expression was defined using TMM-normalized values with blue and red representing comparatively lower and higher expression levels, respectively.

We functionally examined each tissue by testing each set of DEGs for enrichment of gene ontology terms relative to their frequency in the reference transcriptome.

Among the GO terms that were significantly enriched in the liver of the Emperor penguin, contigs involved in fatty acid beta-oxidation (GO:0006635), fatty acid (GO:0006631) and lipid metabolism (GO:0006629) and antioxidant activity (GO:0016209) were the most overrepresented (Fig. 9). Either in the skin and in the brain of the Emperor penguin, the most overrepresented GO categories were related to basic cellular functions as translation (GO:0006412), rRNA processing (GO:0006364) and GTPase activity (GO:0003924) (Fig. 9).

In contrast, the kidney DEGs showed a quite different profile, with a high representation of categories related to hemoglobin complex (GO:0005833), oxygen transport (GO:0015671),

oxygen carrier activity (GO:0005344), heme binding (GO:0020037) and oxidative phosphorylation (GO:0006119) (Fig. 9). These findings are congruent with the kidney being a hematopoietic tissue in birds.

The enriched functions of the Emperor penguin muscle showed: intracellular sequestration of iron (GO:0006880), ferric iron binding (GO:0008199), ferroxidase activity (GO:0004322), cellular iron ion homeostasis (GO:0006879) and regulation of fibroblast proliferation (GO:0048145) as the most significantly enriched GO terms (Fig. 9).

We detected just a slight overlap between the GO terms in the Emperor and the King penguin, mostly related to basic cellular functions such as translation and RNA processing (Fig. 10).

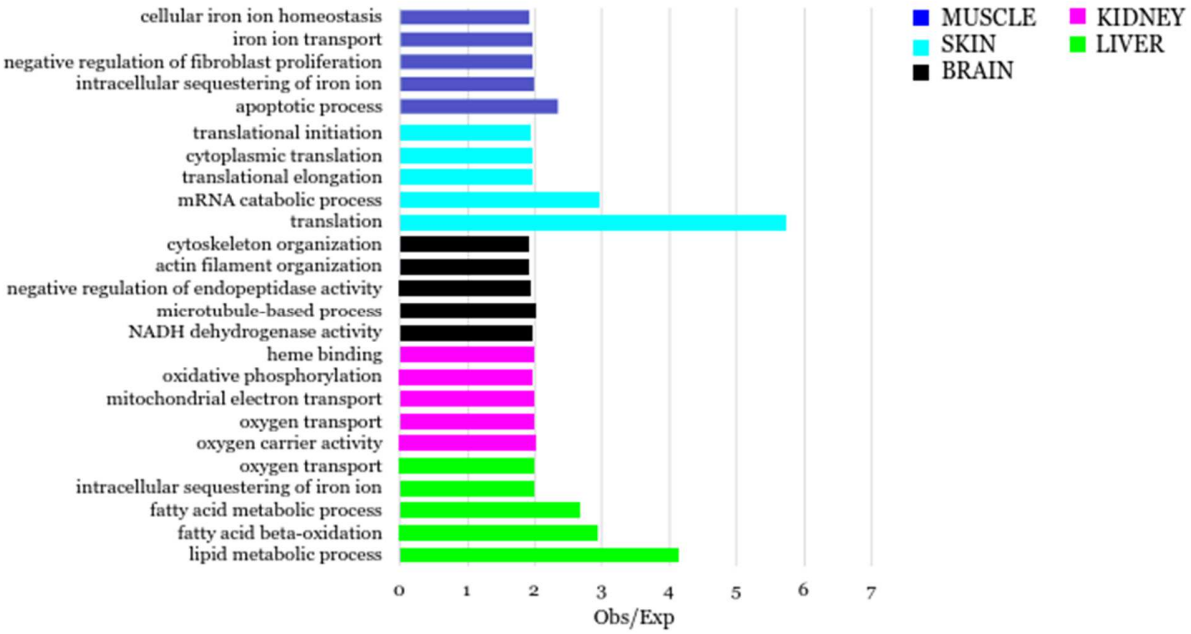

Figure 9. Classification of the DEGs by tissue after a GO enrichment analysis. Only the most overrepresented GO terms of the Emperor penguin are shown.

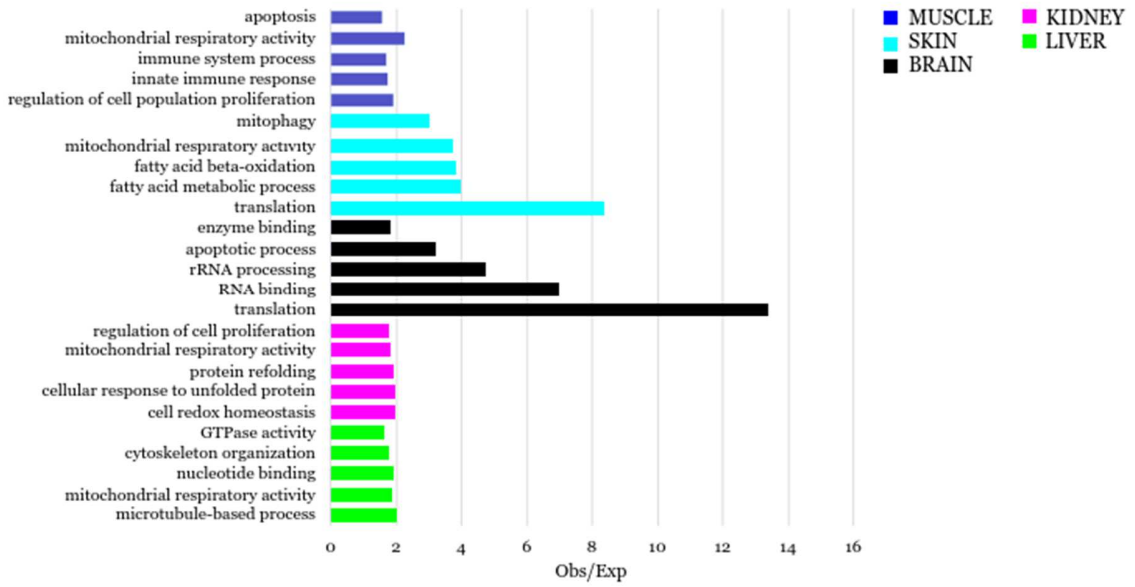

Figure 10. Classification of the DEGs by tissue after a GO enrichment analysis. Only the most overrepresented GO terms of the King penguin are shown.

The five sets of DEGs in each species were also analyzed to identify significantly enriched Pfam domain annotations. The most significantly abundant Pfam domains in the DEGs of both the Emperor penguin liver and muscle included: enzymes involved in fatty acid synthesis and steroid biogenesis (thiolase and beta-ketoacyl synthase), metalloprotein involved in oxidative stress tolerance (rubrerythrin) and iron storage proteins (ferritin like domains) (Fig. 11). In the Emperor penguin brain, the only relevant Pfam domain was the carboxymuconolactone decarboxylase, reported to increase proliferation and inhibit apoptosis of neural cells; while the most enriched Pfam domains among the DEGs in the Emperor penguin kidney resulted to be globulin and immunoglobulin (Fig. 11).

YflT was reported as one of the most abundant Pfam domains in the DEGs of the Emperor penguin skin (Fig. 11). This domain is found in general stress proteins, induced by heat shock, salt stress, oxidative stress, glucose limitation and oxygen limitation.

From the analysis of the Pfam domain enrichment in the tissue-specific DEGs of the King penguin, we found similar terms but in different tissues (Fig. 12). For example, the most abundant Pfam domains in the kidney were proteins involved in antioxidant defence (redoxin, AhpC/TSA) and enzymes related to fatty acid and steroid metabolism (thiolase, beta-ketoacyl synthase). The two latter categories resulted to be the most enriched also in the skin (Fig. 12).

The brain presented only actin, the major component of dendritic spines, as significant Pfam domain (Fig. 12). While no significant Pfam domains were reported for both the liver and the muscle.

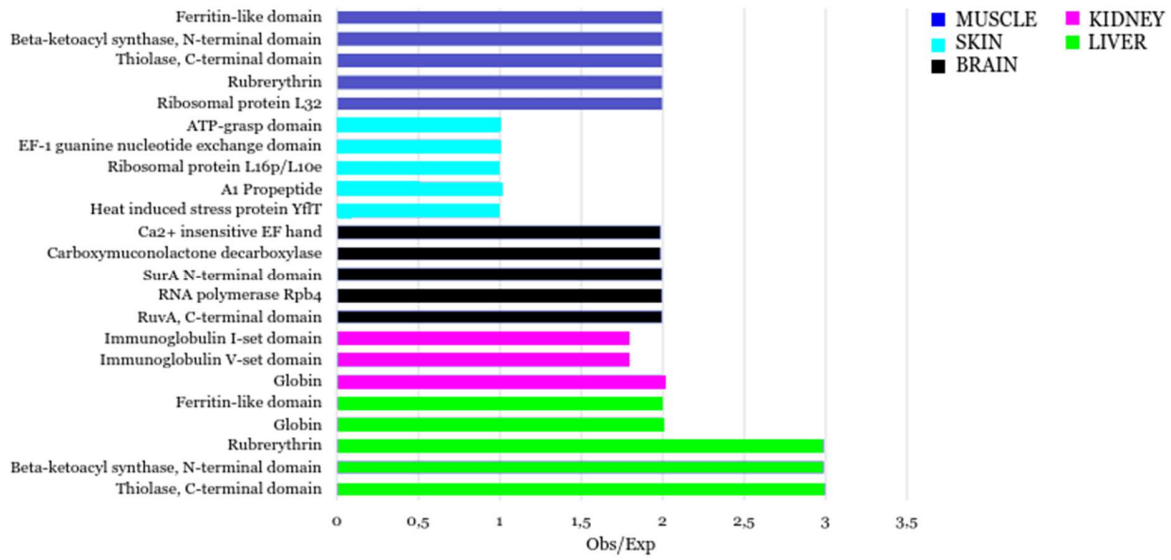

Figure 11. Classification of the DEGs by tissue after a Pfam enrichment analysis. Only the most overrepresented Pfam domains of the Emperor penguin are shown.

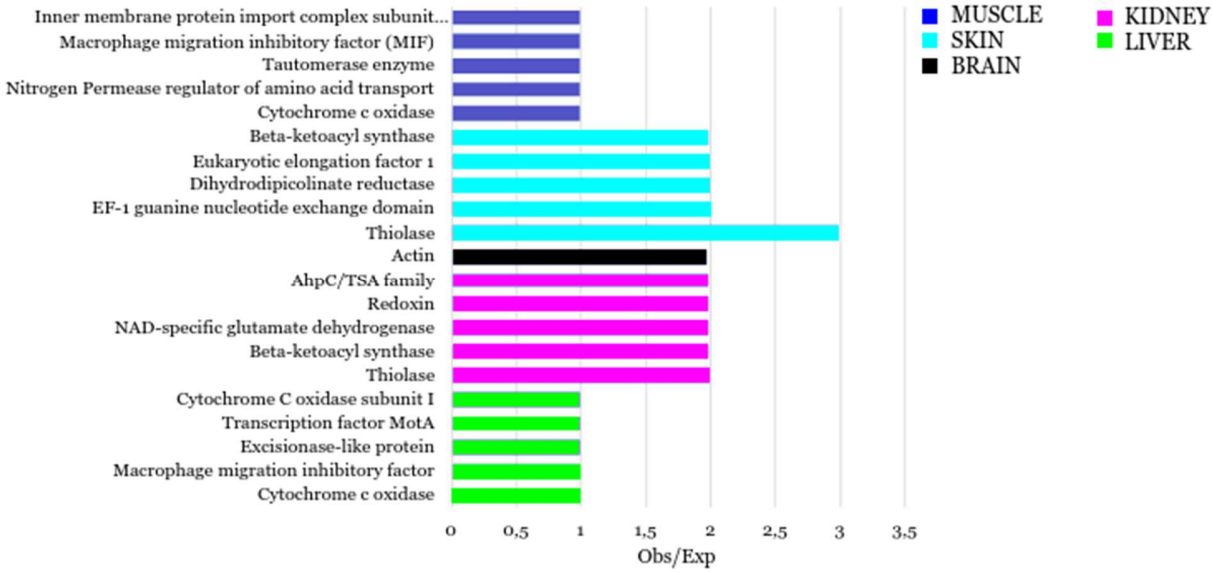

Figure 12. Classification of the DEGs by tissue after a Pfam enrichment analysis. Only the most overrepresented Pfam domains of the King penguin are shown.

Differential expression analysis at interspecific level

As comparing gene expression levels between the two penguin species, using the reference transcriptome of each of them, proved to be particularly challenging and inaccurate, we decided to use as a common reference, the Emperor penguin genome.

A total of 15,333 genes had enough reads mapped to at least one sample to pass the initial threshold for analysis of DEGs.

When applying the MDS analysis, we observed that samples of the same tissue from different species clustered together, implying a global tissue-specific expression pattern (Fig. 13).

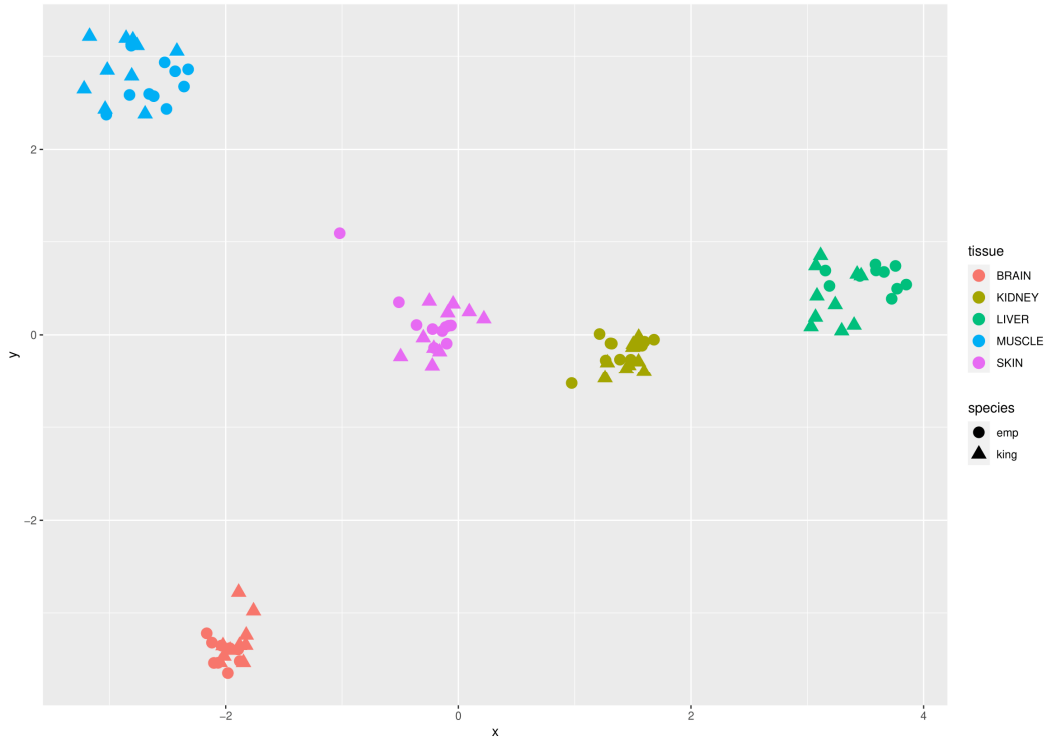

Figure 13. Multidimensional scaling (MDS) plot for the 3'end Seq libraries of the Emperor and the King penguin after mapping to the Emperor penguin reference genome.

Overall up-regulation and down-regulation was measured for the genes that were shared among the two Aptenodytes penguin species. Considering a 2-fold expression difference, a total of 2,106 genes (15% of all 15,333 expressed genes) exhibited significant differential expression between the two species considering all tissues. In particular, 1,266 and 840 genes were up-regulated in the Emperor and in the King penguin, respectively. Breaking down these figures by tissue, we identified 506 differentially expressed genes in the skin (297 up-regulated in the Emperor and 209 up-regulated in the King), 1,059 in the muscle (766 up-regulated in the Emperor and 293 up-regulated in the King), 151 in the brain (46 up-regulated in the Emperor and 105 up-regulated in the King), 466 in the liver (213 up-regulated in the Emperor and 253 up-regulated in the King) and 342 in the kidney (228 up-regulated in the Emperor and 114 up-regulated in the King).

Cluster analysis of differentially expressed genes (Fig. 14) showed that the biological replicates of the same tissue clustered per species, indicating that the samples used in this study had excellent biological repeatability. The majority of the differentially expressed genes (80% in the Emperor and 87% in the King) were tissue-specific, with skin and muscle showing greater numbers in the Emperor penguin while liver and muscle in the King penguin. Notably, the muscle exhibited the highest number of private DEGs (605), representing 24% of all up-regulated genes in the Emperor penguin.

Approximately 15% of the differentially up-regulated genes were detected in two or more tissues in each species, suggesting the presence of a shared set of enhanced cellular responses.

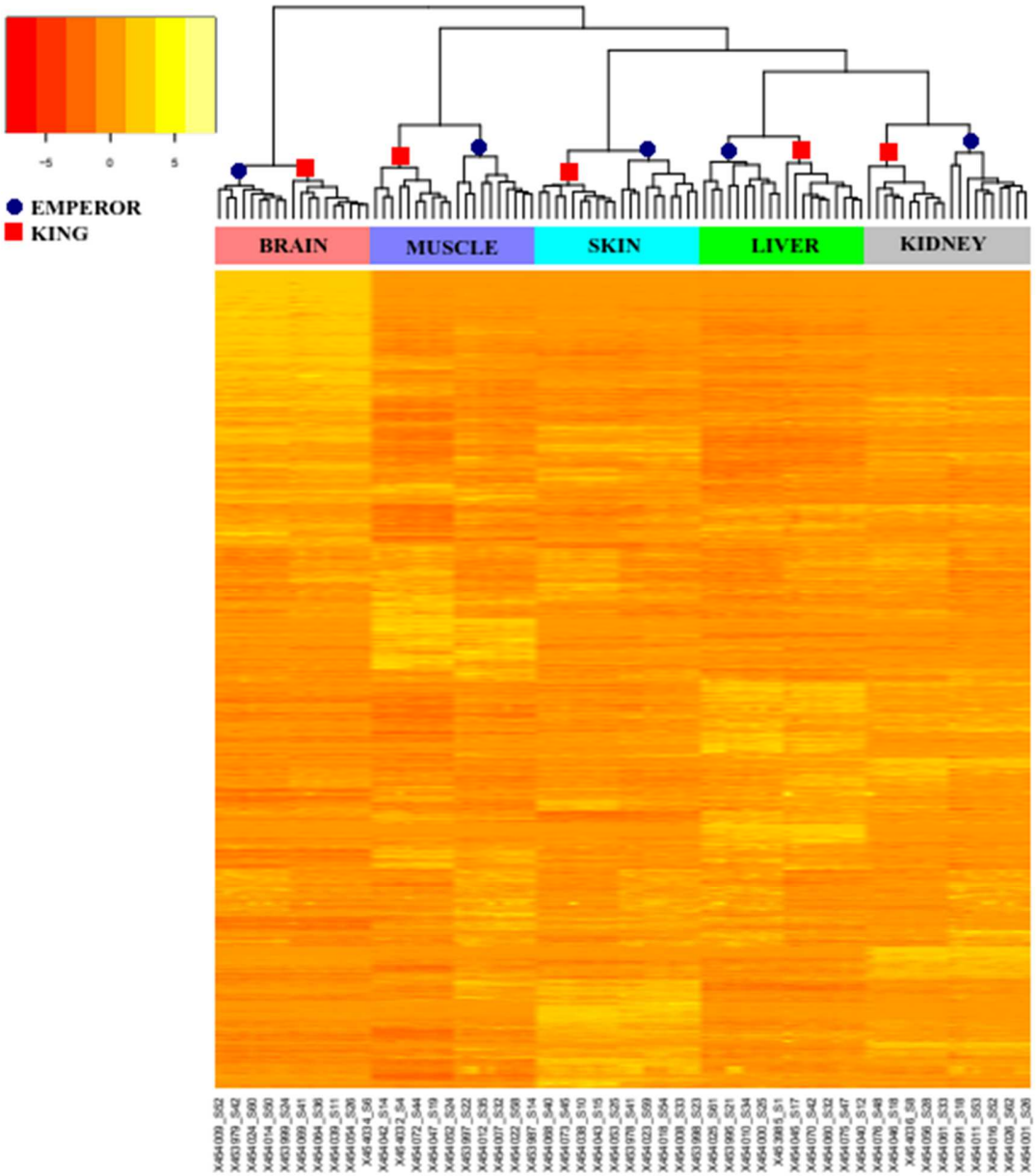

Figure 14. Heat map showing expression patterns of 2,106 DEGs in the interspecific comparison between the Emperor and the King penguin. Expression was defined using TMM-normalized values with dark orange and yellow representing comparatively lower and higher expression levels, respectively.

GO enrichment tests were used to assist in functional interpretation of the DEGs and to evaluate potential cold-adaptive response in the comparison of the same tissue between the two species. In the Emperor penguin, most of the up-regulated DEGs were associated with GO terms potentially linked to the substantial changes driven by the adaptation to the extreme Antarctic environment, such as lipid biosynthetic process (GO:0008610) and response to radiation (GO:0071478) in the liver, positive regulation of vascular permeability (GO:0043117) and triglyceride-rich lipoprotein particle remodeling (GO:0034370) in the kidney and circulatory system development (GO:0072359), blood circulation (GO:0008015), bone mineralization (GO:0030282), visual system development (GO:0150063) and fat cell differentiation (GO:0045444) in the muscle (Fig. 15). Functional enrichment analyses revealed no significant GO terms for the brain, in both the species, probably due to the small number of DEGs in this tissue. Whereas in the King penguin, we found more general GO terms enriched, such as immune response (GO:0002250) and circulatory system process (GO:0003013) in the kidney, regulation of secretion (GO:0051046) and response to stress (GO:0006950) in the liver, cell junction organization (GO:0034330) in the skin, and fatty-acid metabolism (GO:0006631) and heart contraction (GO:0002026) in the muscle (Fig. 16).

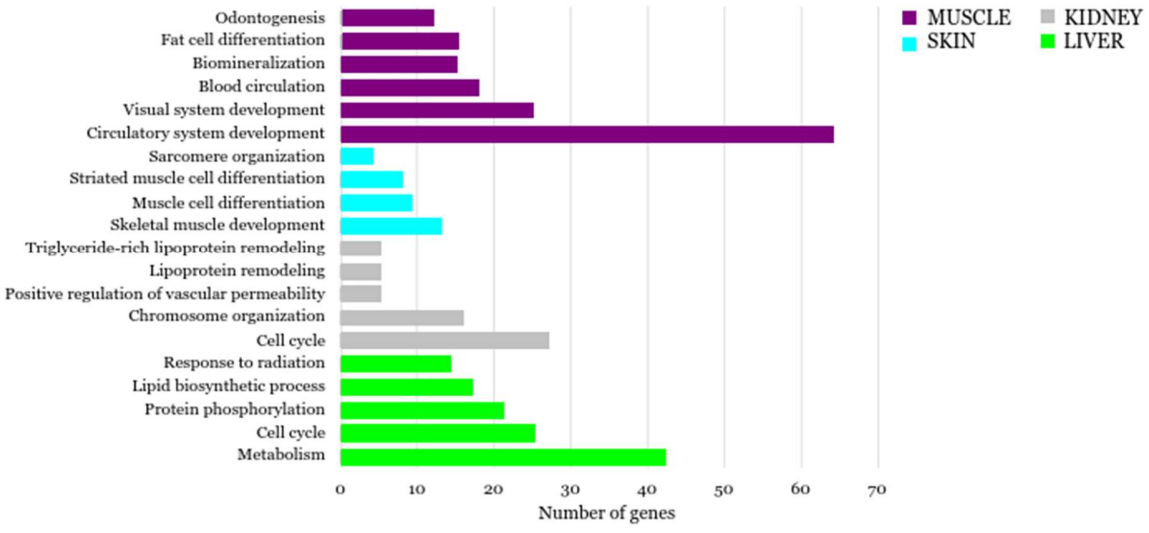

Figure 15. Classification of the up-regulated DEGs in the interspecific tissue comparison after a GO enrichment analysis. Only the most overrepresented GO terms of the Emperor penguin are shown.

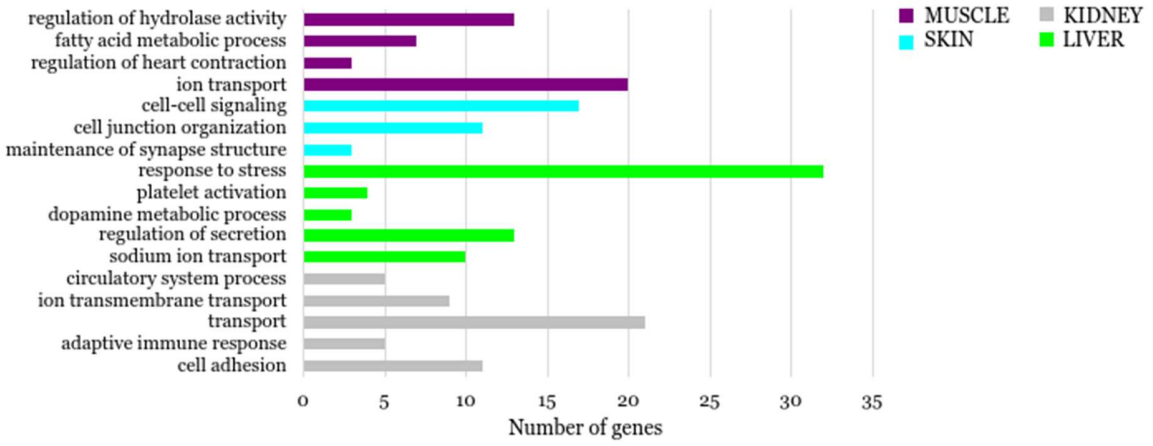

Figure 16. Classification of the up-regulated DEGs in the interspecific tissue comparison after a GO enrichment analysis. Only the most overrepresented GO terms of the King penguin are shown.

We further analysed the sequences of the up-regulated DEGs in the Emperor penguin in order to identify relevant protein domains and classify them into families. The essential terms included proteins involved in: lipid and fatty acid metabolism (perilipin, glycerol-3-phosphate dehydrogenase, lipocalin, apolipoprotein, lipase, thyroid hormone-inducible hepatic protein), angiogenesis (Von Willebrand factor, vasohibin, thrombospondin), light absorption (7 transmembrane receptor), stress response (chaperone DnaJ, AMPK, Chaperonin Cpn60, beta-thymosin), glucose/insulin metabolism (phosphoenolpyruvate carboxykinase, glucose dehydrogenase, enolase, Maf transcription factor, IGFBP, glucagon), cholesterol transport (TspO/MBR, caveolin), immune response (beta defensin, immunoglobulin), thyroid hormone biosynthesis (haem peroxidase, GPCR, thyroglobulin, iodothyronine deiodinase), mineral metabolism (stanniocalcin, calcitonin), protection (keratin), muscle contraction (calponin), hypothalamic regulation of body weight (tubby) (Fig. 17).

Many of these protein domain families were also recognized among the up-regulated DEGs in the King penguin; in particular those involved in thyroid hormone biosynthesis (iodothyronine deiodinase, haem peroxidase), regulation of body weight (tubby), lipid metabolism (thyroid hormone-inducible hepatic protein, lipocalin, lipase), glucose metabolism (glucose

dehydrogenase, phosphoenolpyruvate carboxykinase), defense response (immunoglobulin), protection (keratin), stress response (Chaperone DnaJ) and angiogenesis (Von Willebrand factor). Whereas some were different but related to the same of the main biological categories found in the Emperor penguin DEGs: avidin, 3-hydroxyacyl-CoA dehydrogenase, phospholipase A2, fatty acid hydroxylase, LBP (fatty acid and lipid metabolism), somatostatin, CBM21 (glucose/insulin metabolism) and Hsp70 protein (stress response) (Fig. 18).

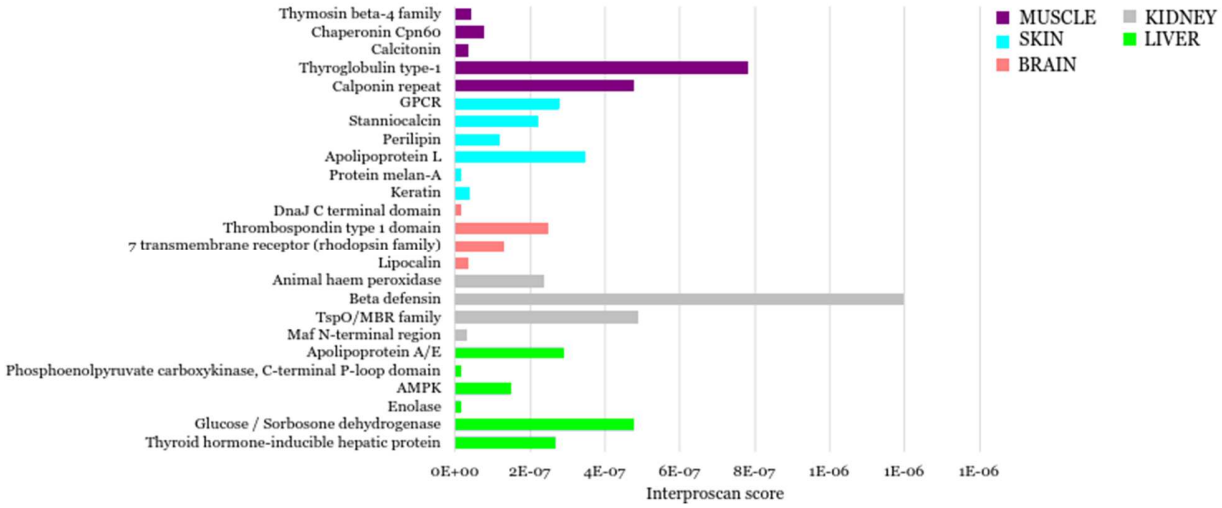

Figure 17. Most relevant protein domains in the up-regulated DEGs by tissue in the Emperor penguin.

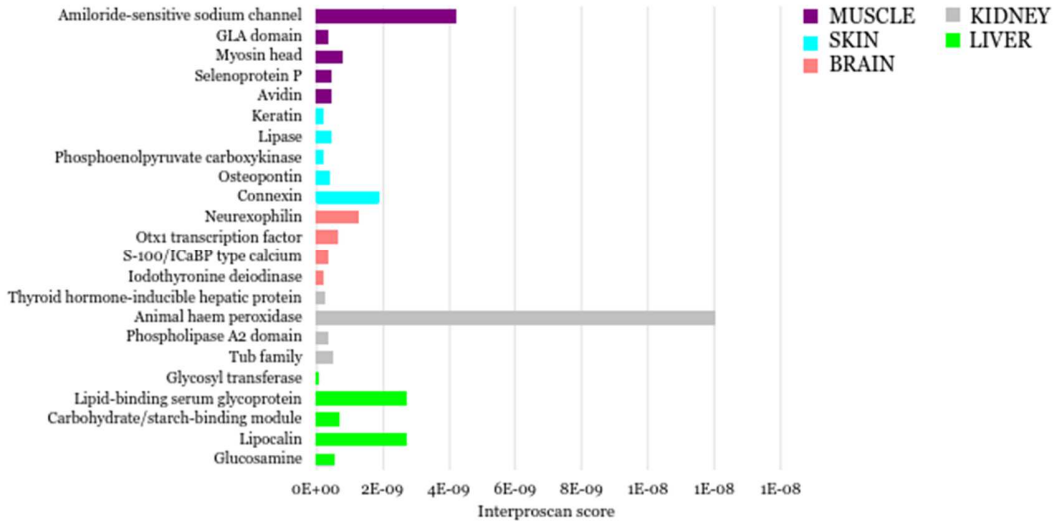

Figure 18. Most relevant protein domains in the up-regulated DEGs by tissue in the King penguin.

We also looked for significantly enriched KEGG pathways among the up-regulated DEGs of the Emperor penguin (Fig. 19). Only three tissues (liver, muscle, and skin) produced statistically significant results. Interestingly, all the relevant pathways could be grouped into two major categories: energy homeostasis and thyroid hormone metabolism. Four of the five KEGG pathways in common between the three tissues showed the highest number of candidate genes.

A total of 37 up-regulated genes were involved in the FoxO signaling pathway that promotes the transition from carbohydrate oxidation to lipid oxidation under stress conditions such as fasting and exercise. Nineteen up-regulated DEGs took part in the insulin signaling pathway playing a role in the uptake and storage of glucose, fatty acids and amino acids into liver, adipose tissue and muscle. Related to insulin resistance and sensitivity, the adipocytokine signaling pathway, encompassing 12 up-regulated DEGs, regulates energy balance in the body, but also participates in the process of inflammation, coagulation, and fibrinolysis.

Moreover, 26 up-regulated genes were included in the thyroid hormone signaling pathway which exerts a wide range of functions in terms of maintenance of homeostasis, cell proliferation and differentiation, glucose metabolism, heart rate and blood volume regulation.

In the King penguin, significantly enriched KEGG pathways emerged only from the up-regulated DEGs of the muscle (Fig. 20). Three of the enriched KEGG terms (fatty acid metabolism, PPAR signaling pathway and primary bile acid metabolism) were related to energy metabolism, suggesting again a key role of this tissue in lipid metabolism and maintenance of energy homeostasis.

A total of 11 up-regulated genes were included in pathways involved in vasoconstriction and dilation, blood pressure regulation, strengthening of cardiac contractility and angiogenesis.

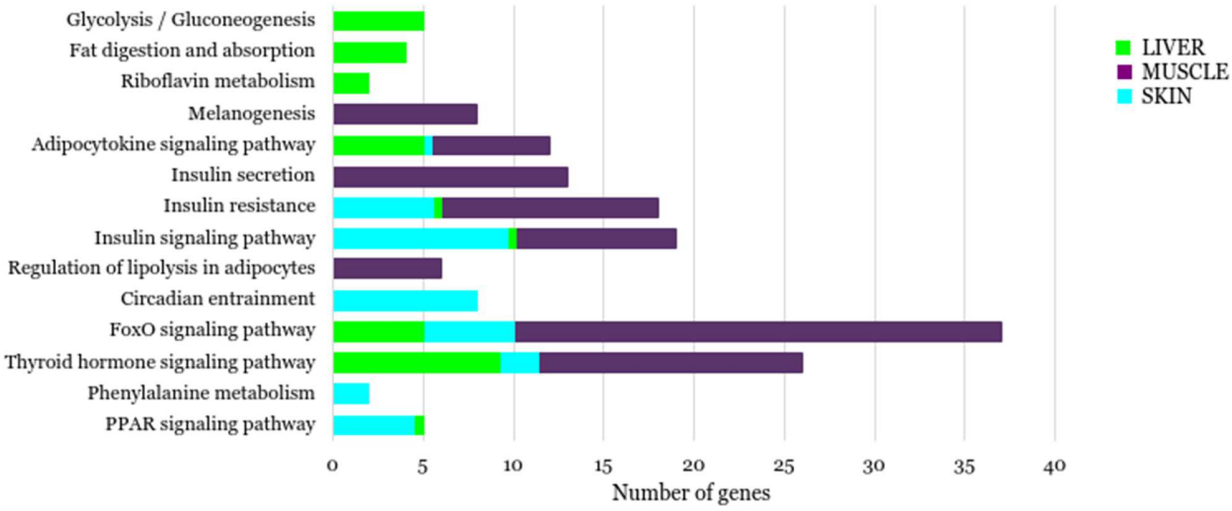

Figure 19. Classification of the up-regulated DEGs in the interspecific tissue comparison after a KEGG pathways enrichment analysis in the Emperor penguin. Only liver, muscle, and skin produced statistically significant results.

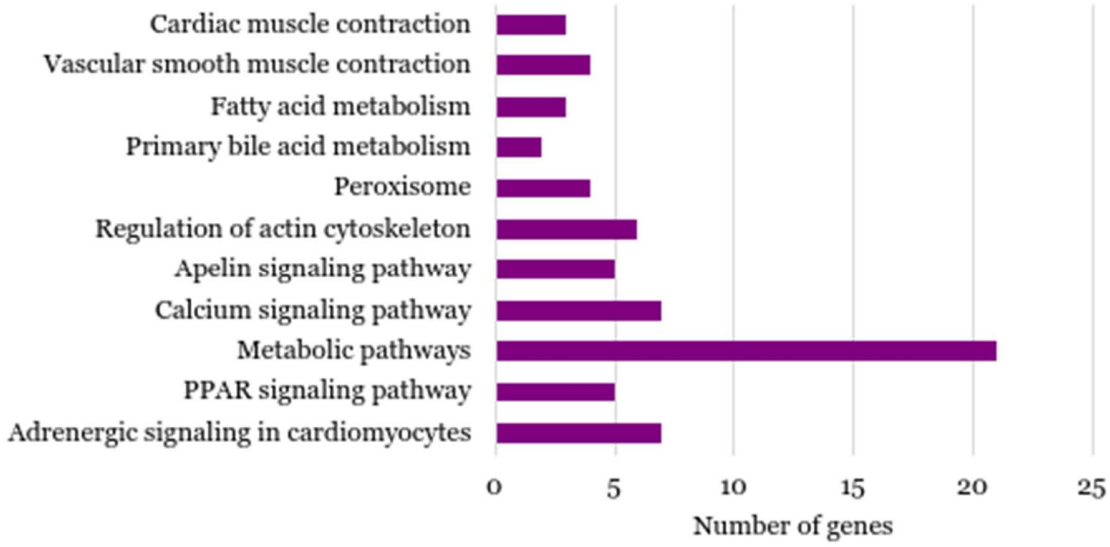

Figure 20. Classification of the up-regulated DEGs in the interspecific tissue comparison after a KEGG pathways enrichment analysis in the King penguin. Only muscle produced statistically significant results.

Figure 21 shows the volcano plot of the candidate DEGs which displayed the greatest significant expression changes in each interspecific pairwise tissue comparison. In the Emperor penguin, 3 of the 20 most up-regulated genes in the brain were associated with stress response to adverse environmental conditions (SLC9A2, DNAJB5, CGGBP1). Among the top up-regulated genes in the brain, they are worth mentioning: LARP1, involved in the mTORC pathway that regulates mass accumulation in function of growth signals and nutrient availability, and MTNR1A that mediates melatonin effects on circadian rhythm and reproductive alterations affected by day length (Fig. 21C).

LARP1 was among the ten most up-regulated DEGs also in the liver and the kidney, together with MLST8, another element of the mTOR complex, strongly indicating a constitutive system-wide need in the Emperor penguin for coping with chronic cold (Fig. 21A-B).

In the Emperor penguin kidney, RBP4 is up-regulated two times more than in the King penguin (Fig. 21B). This gene encodes for a specific retinol carrier in the blood, transporting retinoids (vitamin A and its derivatives) to target tissues, such as adipose tissue. Previous reports in humans and mice revealed that retinol transport protein Rbp are potent regulators for cold-induced thermogenic responses and cold adaptation by promoting expression of BATmarker genes, such as UCP1 in the adipocytes (Fenzi et al., 2020; Puigserver et al., 1996; Mercader et al., 2010; Ribot et al., 2004).

Also in the skin, another retinol transport protein, RBP7, is up-regulated three times more than in the King penguin (Fig. 21E).

Looking at the 20 most significantly up-regulated DEGs in the Emperor's skin, we observed that 4 of them were related to energy metabolism: LPL, key enzyme in lipid clearance from the bloodstream, lipid utilization and storage; PLIN4 and PLIN1, both involved in triacylglycerol packaging into adipocytes and lipolysis modulators; PNPLA2, enzyme that may play a role in the response of the organism to starvation, catalyzing hydrolysis of triglycerides into free fatty acids to be oxidized in situations of energy depletion (Fig. 21E).

As a sensitive neural system response to temperature fluctuations is essential for the maintenance of a constant body temperature in the homeotherms, it was not unexpected to observe TMEM100 among the 20 top up-regulated genes in the Emperor’s skin (Fig. 21E). This gene is a modulator of TRPA1 and TRPV1, two nociceptors channels that respond to extreme temperatures and cold cutaneous stimulation (Nozadze et al., 2016).

The Emperor penguin's muscle exhibited the highest number of up-regulated genes and most of the top 20 DEGs were involved in: anticoagulation and angiogenesis (GNS, FN1, EPHB2, ANGPTL2), the development of the barrier function of the epidermis (TCF7L1, LEF1), cold pain tolerance (ABLIM3) and cold-induced thermogenic responses (METRNL) (Fig. 21D).

In particular, METRNL is up-regulated three times more than in the King penguin and plays a key role in cold adaptation by stimulating energy expenditure associated with the browning of the white fat depots and improving glucose tolerance.

In the King penguin, five (OTX2, ATP8B1, SNCG, PCP4, SLITRK5) among the 20 most up-regulated DEGs in the brain play a role in brain and sensory organ development, neuronal differentiation and axon generation. Three (ARL1, TCF7L2, SHPK) are involved in insulin/glucose metabolism and two (TSHR, CRYM) are receptors for the thyroid hormone (Fig. 21B). Also among the top 20 DEGs in the kidney and the skin emerged a gene (DIO1) that catalyzes the activation/inactivation of the thyroid hormone. Among the most up-regulated DEGs of the skin, they are also worth mentioning: HSD11B1L, an enzyme that catalyzes the interconversion of inactive to active glucocorticoids required for stress responses (Liu et al., 2021) and RBP4, a specific retinol carrier, found overexpressed in the Emperor penguin's kidney as well. Looking at the 20 most significantly up-regulated DEGs in the King's liver, we observed that 3 of them (LIPA, SLC13A5, PLCE1) were related to lipid metabolism and fatty acids and cholesterol synthesis. Interestingly, we also found HSPA2, a molecular chaperone that is expressed in response to stress (Rosenzweig et al., 2019).

Noteworthy, the two most up-regulated DEGs in the King penguin's muscle (HOXC9, HOXC10) are homeoproteins involved in white adipose tissue development (Ng et al., 2017; Daneshyar et al., 2021). Several of the top 20 DEGs in this tissue were involved in energy metabolism: DGAT2 and GPAM are required for lipid accumulation (Levin et al., 2007; Yu et al., 2017), OSBPL6 regulates cellular transport and efflux of cholesterol (Hennessy et al., 2013) and CEP19 may be related to obesity, energy expenditure, whole-body fat oxidation and, glucose and insulin tolerance (Shalata et al., 2013) (Fig. 21D).

The concurrent differential up-regulation of several genes within these functional categories in the Emperor penguin compared to its less cold-adapted sibling species, the King penguin, indicates that these functions are likely to be in higher demand in the cold Antarctic environment.

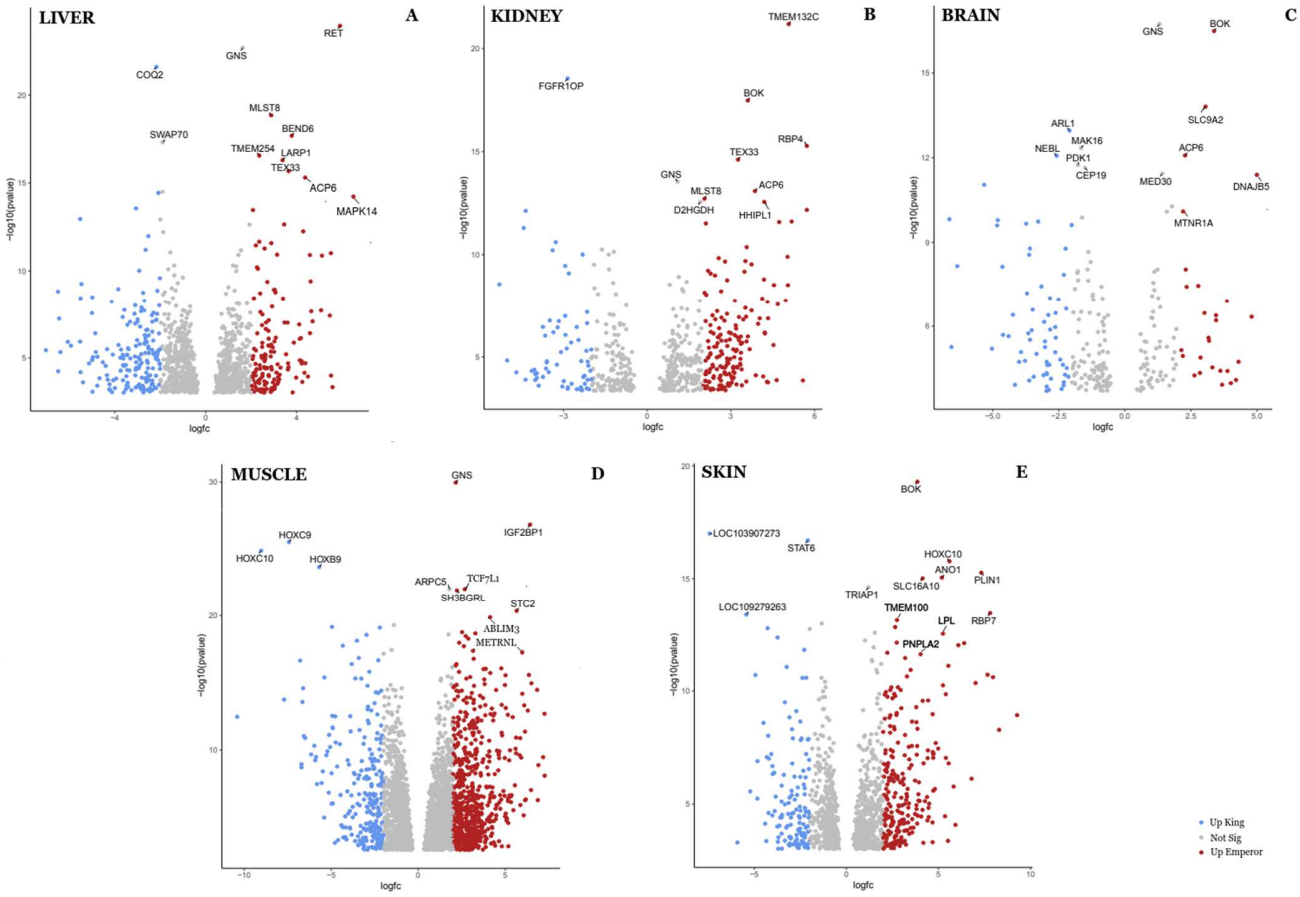

Figure 21. Volcano plot showing the up-regulated DEGs which display the greatest significant expression changes in each interspecific pairwise tissue comparison. Red points represent the up-regulated DEGs in the Emperor penguin while the blue points represent the up-regulated DEGs of the King penguin.

Discussion

Our study is the first to produce transcriptome assemblies for Emperor and King penguins from such a diverse range of tissues from multiple individuals and will therefore complement the already published resources generated from the pectoralis muscle of juvenile specimens of King penguin (Degletagne et al., 2010; Rey et al., 2016a; Rey et al., 2016b).

During the development of this work, we faced several technical challenges typical of comparative analyses of gene expression profiles on non-model species. The computation of gene expression in different species is hampered by inherent difficulties associated with varying levels of genome/transcriptome assembly and annotation, the extent of lineage-specific gene gains, losses and duplications, and transposable elements activity (Musser and Wagner, 2015).

In order to overcome the most common pitfalls of NGS assemblies and increase the robustness of our analyses, we chose to use the ORP pipeline (MacManes, 2018), able to efficiently combine multiple transcriptomes generated by different assembly algorithms to build a high-quality *de novo* transcriptome for each species.

Our findings revealed a tissue-specific expression pattern, indicating that differences among tissues were more significant than those among species.

Overall, our study recovered a high number of differentially expressed genes associated with various physiological adaptations in the Emperor penguin compared to the King penguin. The up- and down-regulation of genes might be a result of short-term adaptations through plasticity and/or long-term genetic adaptations. Both environment and species-specific differences may have influenced changes in gene expression levels between the two species (Verne et al., 2011; Romero et al., 2012).

Collectively, the results of differential gene expression and comparative transcriptome analyses provided an overview of the transcriptomic changes potentially relevant in the Emperor penguin's adaptation to an extreme cold environment.

Although further research is needed to validate the key genes identified in this study, the enormous range of up-regulated gene groups will allow more in-depth research into mechanisms of evolutionary adaptation to cold also in a broader context.

The first de novo transcriptome assembly of the two penguin species of the Aptenodytes genus

We produced the first transcriptome assembly for each of the penguin species of the *Aptenodytes* genus, by using the principal tissues involved in fattening (liver), thermogenesis (muscle), behavior (brain), thermal insulation (skin) and osmoregulation (kidney), as the initial step for the understanding of how those processes are regulated at the level of gene expression and their role in cold adaptation.

Although previous studies on *A. patagonicus* (Degletagne et al., 2010; Rey et al., 2016a; Rey et al., 2016b) investigated gene expression changes in the pectoralis muscle of juvenile King penguins during the transition from terrestrial to marine life, they did not generate a transcriptome assembly for this species.

Therefore, a high quality *de novo* assembly of blood transcriptome was produced for the chinstrap (*Pygoscelis antarcticus*) and gentoo (*P. papua*) penguins (Kim et al., 2019). In this case, blood samples were collected from only one specimen of each species and the Illumina paired-end libraries were assembled *de novo* using a single assembler. We obtained a total number of contigs four times larger than that of each *Pygoscelis* species, as expected from employing such a diverse range of tissues from multiple individuals.

While most of the significant pathways in the blood transcriptome were involved in the major innate immune systems, we observed a slight overlap in the overall composition of the biological and molecular GO terms with our transcriptome assemblies. Indeed, a considerable number of transcripts were assigned to signal transduction, developmental process and immune system activity. The enrichment of these GO terms was revealed also from the comparison with the blood transcriptome of the southern giant petrel (*Macronectes giganteus*) of the *Procellariiformes* order, the *Sphenisciformes* sister taxon (Kim et al., 2018). Moreover, the percentage of transcripts assigned to GO terms in the Emperor assembly (47%) was similar to that reported for mallards (48%) (Yin et al., 2019a). Six of the most represented GO categories in both the *Aptenodytes* penguin assemblies (i.e., signal transduction, apoptosis process, phosphorylation, cell differentiation, cell cycle and transcription regulation) were also found among the ten most common GO terms in the blue-winged teal (*Spatula discors*) transcriptome (Dolinski et al., 2020). Our findings suggest both the transcriptomes cover a wide range of protein-coding sequences, including 86.3% (Emperor penguin) and 84.2% (King penguin) of BUSCOs thought to be found

in the Aves lineage. This compares quite favorably to previous *de novo* transcriptomes in passerine species evaluated against the vertebrate gene set using BUSCO (30–62% complete) (Richardson et al., 2017).

Additionally, our backmapping rate of 82% (Emperor penguin) and 79% (King penguin) is also comparable to other *de novo* transcriptomes in birds, including European starlings (*Sturnus vulgaris*, 82%) (Richardson et al., 2017), rock doves (*Columba livia*, 70–80%) (MacManes et al., 2017), tree swallows (*Tachycineta bicolor*, 82%) (Bentz et al., 2019) and mallards (*Anas platyrhynchos*) (Yin et al., 2019a).

Subsequent studies would ideally include samples from adults and the missing tissues; however, data thus far suggest that the transcriptome from 10 chicks (5 tissues each) has good coverage in each *Aptenodytes* species.

Moreover, the percentage of functionally annotated transcripts (53% and 69% in the Emperor and in the King penguin, respectively) reflects the typical rates of annotation for non-model organisms (Baeza and MacManes, 2020).

It is worth noting that comparing *de novo* transcriptome assemblies from studies on different tissues and different species is especially challenging. The metrics that best assess the transcriptome quality (e.g., the proportion of reads mapping to an assembly, the N50 length) are affected by several factors, such as the sequencing technology and the assembly strategy used (Frias-Soler et al., 2018). However, the N50 length estimates for both the Emperor (2,429 bp) and the King penguin (2,511 bp) were much higher than that estimated for 5 different bird species (chicken, *Gallus gallus*; duck, *Anas platyrhynchos*; pigeon, *Columba livia*; goose,*Anser cygnoides*; zebra finch, *Taeniopygia guttata*), where N50 ranged from 595 bp (Chicken) to 1,533 bp (Pigeon) (Yin et al., 2019b).

An insight into the transcriptional profile of the King and Emperor penguin's muscle

The skeletal muscle of birds is commonly larger than that of reptiles and mammals of comparable size (Butler, 2000; Newman, 2011). As a result, it has been hypothesized that the enlargement of avian skeletal muscle, especially the breast and thigh muscles, provided a distinct survival advantage, allowing for more effective heat generation (Rowland et al., 2015).

The expression profile of the Emperor penguin's muscle resulted to be unexpectedly rich, encompassing the largest fraction of up-regulated genes among all the analyzed tissues at both intraspecific and interspecific level.

Although muscle is a highly specialized tissue, gene expression was not found to be limited to a narrow set of genes involved in the structural organisation of muscle fibres. This is consistent with the crucial role played by this tissue in whole body energy metabolism and temperature homeostasis in birds and mammals, producing heat through shivering and nonshivering thermogenesis (Periasamy et al., 2017). Muscular shivering is considered to be the dominant mechanism of heat production in birds (Dawson, 1975). Although this process occurs in all birds, the metabolic pathways of energy mobilization may differ quantitatively as a consequence of growth mode (Choi et al., 1993).

Emperor penguins have developed thermal and metabolic adaptations to survive the Antarctic winter conditions. These metabolic adaptations include: *i*) an enhanced ability to oxidize lipids as a fuel for energy metabolism (Teulier et al., 2012), *ii*) a large muscle mitochondrial abundance and oxidative capacities (Rey et al., 2016a) and *iii*) a high thermogenic capacity.

A total of 37 up-regulated genes were involved in the FoxO signaling pathway that promotes the transition from carbohydrate oxidation to lipid oxidation under stress conditions such as fasting and exercise (Bastie et al., 2005). This pathway has been shown to be particularly relevant also for the cold stress faced by spotted seatrout (*Cynoscion nebulosus*) populations adapted to water temperatures below 5°C (Song and McDowell, 2021).

Among the most enriched pfam domains in this tissue, we found metalloproteins involved in oxidative stress tolerance such as rubrerythrin.

Penguins have evolved powerful antioxidant capacities which include a marked reduction of ROS generation at mitochondrial level (Rey et al., 2016a) and the improvement of ROS detoxification through a higher activity of muscle antioxidant enzymes compared to other birds (Zenteno-Savin et al., 2010).

These adaptations, combined with locomotor muscles hypoperfusion and extreme hypoxemia, are crucial for their survival, as prolonged diving in cold waters is an important component of penguins' feeding behaviour (Rey et al., 2016a). Indeed, Emperor penguins can achieve depths greater than 300 m and dive durations of 22 min (Wienecke et al., 2007).

A significant gene enrichment was also observed in the thyroid hormone signaling pathway. There is evidence that an increased expression of muscle nuclear T3 receptors may help to stimulate the avian uncoupling protein (avUCP) expression in the skeletal muscle (Rey et al., 2010; Collin et al., 2005). This protein mediates the non-shivering thermogenesis process through the generation of heat in brown adipose tissue (Lowell and Spiegelman, 2000).

Fifteen up-regulated genes were assigned to the fat cell differentiation GO term. This category included: METRNL, one of the 10 most highly differentially expressed genes in muscle, which plays a key role in cold-induced thermogenic responses by stimulating several immune cell subtypes to enter the adipose tissue and activate their pro-thermogenic actions (Zheng et al., 2016); FNDC5, BMP7 and EBF2 considered to be essential for brown adipogenesis (Cao et al., 2019; Saini et al., 2015; Rajakumari et al., 2013); BMP2, PDGFRA and BBS12 that drive the differentiation of adipocyte precursors in white adipocytes (Schulz and Tseng, 2009; Shin et al., 2020); WIF1 which stimulates adipogenic gene expression (Alsaedi, 2016); SFRP2 and SOX8 associated with increased adiposity (Crowley et al., 2016); LAMA4 involved in the regulation of UCP1 expression in both white and brown adipocytes (Porras et al., 2021); TRIB2, an obesity-related gene, reported to induce thermogenesis in cold environments (Nakayama and Iwamoto, 2017); TCF7L2 and CCND1 involved in adipose tissues development and proliferation (Chen et al., 2018; Marquez et al., 2017), and ZNF516, brown adipose tissue-specific gene, which binds the promoter of the UCP1 gene to activate its transcription and thereby regulate non-shivering thermogenesis (Afonso et al., 2021).

Interestingly, looking at fig. 21D, we observe that also in the King penguin, two of the most up-regulated genes among those differentially expressed in the muscle (HOXC9, HOXC10) are involved in adipogenesis (Ng et al., 2017; Daneshyar et al., 2021). In particular, HOXC10 is selectively expressed in white adipose tissue (WAT) and plays a key role in suppressing browning of subcutaneous WAT (Ng et al., 2017). Previous studies on mice demonstrated that genetic deletion of HOXC10 caused the upregulation of the expression of UCP1 and other genes crucial for beige adipocytes biogenesis in subcutaneous WAT, impairing the ability of these mice to maintain body temperature during cold exposure (Ng et al., 2017; Angeline Tan et al., 2021). Our findings on the most up-regulated DEGs, together with the enrichment of GO terms and KEGG pathways associated with fatty-acid metabolism and energy homeostasis, suggest a thermogenic

role of the muscle also in the King penguin but conceivably favoring different biological pathways compared to the Emperor penguin.

As King penguins live and breed in temperate-cold sub-Antarctic islands with average winter temperature of ca. 3 °C, it might be reasonable that this species has evolved different mechanisms for cold adaptation other than promoting the expression of brown-fat thermogenic genes, as in the Emperor penguin. They might increase body fat by promoting lipid accumulation and storage in WAT, for example through the action of the genes DGAT2 and GPAM (Levin et al., 2007; Yu et al., 2017).

The need to accumulate huge lipid depots may arise from the highly intense exercise occurring during King penguins foraging up to 300–1600 km away from their colony, based on the seasonal changes in local prey availability (Charrassin and Bost, 2001; Bost et al., 1997).

In birds, lipid oxidation in the skeletal muscle is crucial to maintain the prolonged and energetically demanding exercise that occurs during flight migration (Vaillancourt et al., 2005; McFarlan et al., 2009; Guglielmo, 2010; Liknes and Swanson, 2011). This is made possible by the increased lipid transport and improved metabolic machinery for lipolysis and lipid oxidation in their skeletal muscles (Weber, 2009).

References

Afonso, M.S., Verma, N., van Solingen, C., Cyr, Y., Sharma, M., Perie, L., Corr, E.M., Schlegel, M., Shanley, L.C., Peled, D. & Yoo, J.Y. 2021 MicroRNA-33 inhibits adaptive thermogenesis and adipose tissue beiging. *Arteriosclerosis, Thrombosis, and Vascular Biology* **41**, 1360-1373.

Alsaedi, M. 2016 The role of WNT inhibitory factor I in adipose tissue development (Doctoral dissertation, Tennessee State University).

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.

Anders, S., Pyl, P.T. & Huber, W. 2015 HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169.

Andrews, S. 2010 FastQC: a quality control tool for high throughput sequence data.

Angeline, T.H., Michelle, S.M., Tan, S.X., Ng, Y., Gan, S.Y., Li, H., Neo, S.P., Gunaratne, J., Xu, F. & Han, W. 2021 HOXC10 suppresses browning to maintain white adipocyte identity. *Diabetes*.

Ansaloni, F., Gerdol, M., Torboli, V., Fornaini, N.R., Greco, S., Giulianini, P.G., *et al*. 2021 Cold Adaptation in Antarctic Notothenioids: Comparative Transcriptomics Reveals Novel Insights in the Peculiar Role of Gills and Highlights Signatures of Cobalamin Deficiency. *International journal of molecular sciences* **22**, 1812.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. & Harris, M.A. 2000 Gene ontology: tool for the unification of biology. *Nature genetics* **25**, 25-29.

Baeza, J.A. & MacManes, M. 2020 De novo assembly and functional annotation of the heart + hemolymph transcriptome in the Caribbean spiny lobster Panulirus argus. *Marine genomics* **54**, 100783.

Basha, O., Shpringer, R., Argov, C.M. & Yeger-Lotem, E. 2018 The DifferentialNet database of differential protein-protein interactions in human tissues. *Nucleic Acids Resorces* **46**, D522–D526.

Bastie, C.C., Nahlé, Z., McLoughlin, T., Esser, K., Zhang, W., Unterman, T. & Abumrad, N.A. 2005 FoxO1 stimulates fatty acid uptake and oxidation in muscle cells through CD36-dependent and-independent mechanisms. *Journal of Biochemical Chemistry* **280**, 14222-14229.

Bentz, A.B., Thomas, G.W., Rusch, D.B. & Rosvall, K.A. 2019 Tissue-specific expression profiles and positive selection analysis in the tree swallow (Tachycineta bicolor) using a de novo transcriptome assembly. *Scientific Reports* **9**, 1-12.

Birol, I., Behsaz, B., Hammond, S. A., Kucuk, E., Veldhoen, N. & Helbing, C. C. 2015 De novo transcriptome assemblies of Rana (Lithobates) catesbeiana and Xenopus laevis tadpole livers for comparative genomics without reference genomes. *PloS one* **10**, e0130720.

Blix, A.S. 2016 Adaptations to polar life in mammals and birds. *Journal of Experimental Biology* **219**, 1093-1105.

Bost, C.A., Georges, J.Y., Guinet, C., Cherel, Y., Pütz, K., Charrassin, J.B., Handrich, Y., Zorn, T., Lage, J. & Le Maho, Y. 1997 Foraging habitat and food intake of satellite-tracked king penguins during the austral summer at Crozet Archipelago. *Marine Ecology Progress Series* **150**, 21-33.

Brauer, C.J., Unmack, P.J. & Beheregaray, L.B. 2017 Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened fish. *Molecular Ecology* **26**, 6841–6856.

Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D. 2019 rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8**, giz100.

Butler, P.J. 2000 Energetic costs of surface swimming and diving of birds. *Physiological and Biochemical Zoology* **73**, 699–705.

Cao, R.Y., Zheng, H., Redfearn, D. and Yang, J. 2019 FNDC5: a novel player in metabolism and metabolic syndrome. *Biochimie* **158**, 111-116.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G.Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., *et al.* 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305.

Charrassin, J.B. & Bost, C.A. 2001 Utilisation of the oceanic habitat by king penguins over the annual cycle. *Marine Ecology Progress Series* **221**, 285-298.

Chen, Y., Lun, A. T., & Smyth, G. K. 2016 From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* **5**, 1438.

Chen, S., Zhou, Y., Chen, Y., Gu, J. 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 1884–1890.

Chen, X., Ayala, I., Shannon, C., Fourcaudot, M., Acharya, N.K., Jenkinson, C.P., Heikkinen, S. and Norton, L. 2018 The diabetes gene and Wnt pathway effector TCF7L2 regulates adipocyte development and function. *Diabetes* **67**, 554-568.

Chen, Z., Cheng, C.H.C., Zhang, J., Cao, L., Chen, L., Zhou, L., *et al.* 2008 Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences* **105**, 12944-12949.

Cherel, Y., Gilles, J., Handrich, Y. & Le Maho, Y. 1994 Nutrient reserve dynamics and energetics during long-term fasting in the king penguin (*Aptenodytes patagonicus*). *Journal of Zoology* **234**, 1–12.

Choi, I.H., Ricklefs, R.E. & Shea, R.E. 1993 Skeletal muscle growth, enzyme activities, and the development of thermogenesis: a comparison between altricial and precocial birds. *Physiological Zoology* **66**, 455-473.

Collin, A., Cassy, S., Buyse, J., Decuypere, E. and Damon, M. 2005 Potential involvement of mammalian and avian uncoupling proteins in the thermogenic effect of thyroid hormones. *Domestic animal endocrinology* **29**, 78-87.

Costa, V., Angelini, C., De Feis, I. & Ciccodicola, A. 2010 Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology* **2010**, 853916.

Coyne, J.A. & Orr, H.A. 2004 *Speciation* (Vol. 37). Sunderland, MA: Sinauer associates.

Crowley, R.K., O'Reilly, M.W., Bujalska, I.J., Hassan-Smith, Z.K., Hazlehurst, J.M., Foucault, D.R., Stewart, P.M. and Tomlinson, J.W. 2016 SFRP2 is associated with increased adiposity and VEGF expression. *PLoS One* **11**, p.e0163777.

Daneshyar, S., Kordi, M.R., Afshari, S. & Kadivar, M. 2021 The Effect of Endurance Training on Gene Expression of Homeobox C8 (HOXC8) and Homeobox C9 (HOXC9) in Adipose Tissue of Male Wistar Rats. *Sport Physiology* **13**, 17-38.

Darwin, C. 1859 *The origin of species by means of natural selectio*n. Pub One Info.

Dawson, W.R. 1975 Avian physiology. *Annual Review of Physiology* **37**, 441-4

Degletagne, C., Keime, C., Rey, B., de Dinechin, M., Forcheron, F., Chuchana, P., Jouventin, P., Gautier, C. & Duchamp, C. 2010 Transcriptome analysis in non-model species: a new method for the analysis of heterologous hybridization on microarrays. *BMC genomics* **11**, 1-9.

Delhomme, N., Mähler, N., Schiffthaler, B., Sundell, D., Mannapperuma, C., Hvidsten, T. R., & Street, N. R. 2014 Guidelines for RNA-Seq data analysis. *Epigenesys protocol* **67**, 1-24.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., *et al.* 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.

Dolinski, A.C., Homola, J.J., Jankowski, M.D. & Owen, J.C. 2020 De novo transcriptome assembly and data for the blue-winged teal (Spatula discors). *Data in Brief* **30**, 105380.

Ewels, P., Magnusson, M., Lundin, S., Käller, M. 2016 MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048.

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L. & Hermjakob, H. 2017 Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142.

Falcon, S. and Gentleman, R. 2008 Hypergeometric testing used for gene set enrichment analysis. In *Bioconductor case studies* (pp. 207-220). Springer, New York, NY.

Feder, M.E. & Mitchell-Olds T. 2003 Evolutionary and ecological functional genomics. *Nature Reviews Genetics* **4**, 651–657.

Fenzi, A., Kulterer, O.C., Spirk, K., Mitulović, G., Marculescu, R., Bilban, M., *et al.* 2020 Intact vitamin A transport is critical for cold-mediated adipose tissue browning and thermogenesis. *Molecular metabolism* **42**, 101088.

Finn, R. D., Clements, J. & Eddy, S. R. 2011 HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**, 29–37.

Frias-Soler, R.C., Pildaín, L.V., Hotz-Wagenblatt, A., Kolibius, J., Bairlein, F. & Wink, M. 2018 De novo annotation of the transcriptome of the Northern Wheatear (*Oenanthe oenanthe*). *PeerJ* **6**, e5860.

Galachyants, Y. P., Zakharova, Y. R., Volokitina, N. A., Morozov, A. A., Likhoshway, Y. V., & Grachev, M. A. 2019 De novo transcriptome assembly and analysis of the freshwater araphid diatom Fragilaria radians, Lake Baikal. *Scientific data* **6**, 1-11.

Gavryushkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., *et al.* 2017 Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology* **66**, 57-73.

Ghiselli, F., Iannello, M., Puccio, G., Chang, P. L., Plazzi, F., Nuzhdin, S.V., & Passamonti, M. 2018 Comparative transcriptomics in two bivalve species offers different perspectives on the evolution of sex-biased genes. *Genome biology and evolution* **10**, 1389-1402.

Goldsmith, R. & Sladen, W.J. 1961 Temperature regulation of some Antarctic penguins. *The Journal of Physiology* **157**, 251-262.

Grønvold, L., Schubert, M., Sandve, S.R., Fjellheim, S., & Hvidsten, T.R. 2017 Comparative transcriptomics provides insight into the evolution of cold response in Pooideae. *bioRxiv*, 151431.

Groscolas, R. 1990 Metabolic adaptations to fasting in emperor and king penguins. In *Penguin biology* (pp. 269-296). Academic Press.

Groscolas, R. & Robin, J.P. 2001 Long-term fasting and re-feeding in penguins. *Comp Biochem Physiol A Mol Integr Physiol* **128**, 645–655.

Guglielmo, C.G. 2010 Move that fatty acid: fuel selection and transport in migratory birds and bats. *Integrative and Comparative Biology* **50**, 336–345.

Guo, J., Liu, R., Huang, L., Zheng, X.M., Liu, P.L., Du, Y.S., *et al.* 2016 Widespread and adaptive alterations in genome-wide gene expression associated with ecological divergence of two Oryza species. *Molecular Biology and Evolution* **33**, 62-78.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. & MacManes, M.D. 2013 De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference gen-eration and analysis. *Nature Protocols* **8**, 1494–1512.

Han, L., Mu, Z., Luo, Z., Pan, Q., & Li, L. 2019 New lncRNA annotation reveals extensive functional divergence of the transcriptome in maize. *Journal of integrative plant biology* **61**, 394-405.

Hanikenne, M., Kroymann, J., Trampczynska, A., Bernal, M., Motte, P., Clemens, S., Kramer, U. 2013 Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genetics* **9**, e1003707.

Hennessy, E.J., Rayner, K.J., Sokolovska, A., Ouimet, M., Esau, C. & Moore, K.J. 2013 OSBPL6 is a Novel miR-33 Target Gene Involved in Cellular Cholesterol Efflux.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. & Hunter, S. 2014 InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., *et al.* 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Kannan, S., Hui, J., Mazooji, K., Pachter, L., Tse, D. 2016 Shannon: an information-optimal de novo RNA-seq assembler. *bioRxiv* preprint.

Kavembe, G.D., Franchini, P., Irisarri, I., Machado-Schiaffino, G., & Meyer, A. 2015 Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of earth's most extreme environments, the hypersaline soda Lake Magadi in Kenya, east Africa. *Journal of molecular evolution* **81**, 90-109.

Kim, B.M., Ahn, D.H., Kim, J.H., Jung, J.W., Rhee, J.S. & Park, H. 2018 De novo assembly and annotation of the blood transcriptome of the southern giant petrel Macronectes giganteus from the South Shetland Islands, Antarctica. *Marine Genomics* **42**, 63-66.

Kim, B.M., Jeong, J., Jo, E., Ahn, D.H., Kim, J.H., Rhee, J.S. & Park, H. 2019 Blood transcriptome resources of chinstrap (*Pygoscelis antarcticus*) and gentoo (*Pygoscelis papua*) penguins from the South Shetland Islands, Antarctica. *Genomics and Informatics* **17**.

Kobayashi, N., Watanabe, M., Horiike, T., Kohara, Y., Okada, N. 2009 Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. *Gene* **441**, 187–91.

Levin, M.C., Monetti, M., Watt, M.J., Sajan, M.P., Stevens, R.D., Bain, J.R., Newgard, C.B., Farese Sr, R.V. & Farese Jr, R.V. 2007 Increased lipid accumulation and insulin resistance in transgenic mice expressing DGAT2 in glycolytic (type II) muscle. *American Journal of Physiology-Endocrinology and Metabolism*, **293**, E1772-E1781.

Li, Y., Wang, X., Ban, Q., Zhu, X., Jiang, C., Wei, C., & Bennetzen, J.L. 2019 Comparative transcriptomic analysis reveals gene expression associated with cold adaptation in the tea plant Camellia sinensis. *BMC genomics* **20**, 1-17.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Li, W. & Godzik, A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.

Liknes, E.T. & Swanson, D.L. 2011 Phenotypic flexibility in passerine birds: seasonal variation of aerobic enzyme activities in skeletal muscle. *Journal of Thermal Biology* **36**, 430-436.

Liu, J., Sun, Z., Wang, Z., & Peng, Y. 2020 A Comparative Transcriptomics Approach to Analyzing the Differences in Cold Resistance in Pomacea canaliculata between Guangdong and Hunan. *Journal of Immunology Research* **2020**.

Liu, R., Tearle, R., Low, W.Y., Chen, T., Thomsen, D., Smith, T.P., Hiendleder, S. and Williams, J.L. 2021 Distinctive gene expression patterns and imprinting signatures revealed in reciprocal crosses between cattle sub-species. *BMC Genomics* **22**, 1-10.

Love, M.I., Huber, W. & Anders, S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Lowell, B.B. & Spiegelman, B.M. 2000 Towards a molecular understanding of adaptive thermogenesis. *Nature* **404**, 652–660.

MacManes, M.D. 2018 The Oyster River protocol: a multi-assembler and kmer approach for de novo transcriptome assembly. *PeerJ* **6**, e5428.

MacManes, M.D., Austin, S.H., Lang, A.S., Booth, A., Farrar, V. & Calisi, R.M. 2017 Widespread patterns of sexually dimorphic gene expression in an avian hypothalamic–pituitary–gonadal (HPG) axis. *Scientific Reports* **7**, 1-13.

Marquez, M.P., Alencastro, F., Madrigal, A., Jimenez, J.L., Blanco, G., Gureghian, A., Keagy, L., Lee, C., Liu, R., Tan, L. & Deignan, K. 2017 The role of cellular proliferation in adipogenic differentiation of human adipose tissue-derived mesenchymal stem cells. *Stem cells and development* **26**, 1578-1595.

McCarthy, D.J., Chen, Y. & Smyth, G.K. 2012 Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297.

McFarlan, J.T., Bonen, A. & Guglielmo, C.G. 2009 Seasonal upregulation of fatty acid transporters in flight muscles of migratory white-throated sparrows (Zonotrichia albicollis). *Journal of Experimental Biology* **212**, 2934-2940.

McGirr, J.A. & Martin, C.H. 2020 Ecological divergence in sympatry causes gene misexpression in hybrids. *Molecular ecology* **29**, 2707-2721.

Mercader, J., Palou, A. & Bonet, M.L. 2010 Induction of uncoupling protein-1 in mouse embryonic fibroblast-derived adipocytes by retinoic acid. *Obesity* **18**, 655e662.

Miao, W., Song, J., Huang, Y., Liu, R., Zou, G., Ou, L., & Liu, Z. 2021 Comparative Transcriptomics for Pepper (Capsicum annuum L.) under Cold Stress and after Rewarming. *Applied Sciences* **11**, 10204.

Moll, P., Ante, M., Seitz, A. *et al.* 2014 QuantSeq 3′ mRNA sequencing for RNA quantification. *Nature Methods* **11**, i–iii.

Morris, K.V. & Mattick, J.S. 2014 The rise of regulatory RNA. *Nature Reviews Genetics* **15**, 423–437.

Musser, J.M. & Wagner, G.P. 2015 Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called "species signal". *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**, 588-604.

Nakayama, K. & Iwamoto, S. 2017 An adaptive variant of TRIB2, rs1057001, is associated with higher expression levels of thermogenic genes in human subcutaneous and visceral adipose tissues. *Journal of physiological anthropology* **36**, 1-5.

Newman, S.A. 2011 Thermogenesis, muscle hyperplasia, and the origin of birds. *BioEssays* **33**, 653–656.

Ng, Y., Tan, S.X., Chia, S.Y., Tan, H.Y.A., Gun, S.Y., Sun, L., Hong, W. & Han, W. 2017 HOXC10 suppresses browning of white adipose tissues. *Experimental and molecular medicine* **49**, e292-e292.

Nosil, P. 2012 Ecological *Speciation* (Oxford Univ. Press).

Nozadze, I., Tsiklauri, N., Gurtskaia, G., & Tsagareli, M. G. 2016 Role of thermo TRPA1 and TRPV1 channels in heat, cold, and mechanical nociception of rats. *Behavioural pharmacology* **27**, 29–36.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. 2017 Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417–419.

Pavey, S.A., Collin, H., Nosil, P. & Rogers, S.M. 2010 The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences* **1206**, 110–129.

Periasamy, M., Herrera, J.L. and Reis, F.C. 2017 Skeletal muscle thermogenesis and its role in whole body energy metabolism. *Diabetes & metabolism journal* **41**, 327-336.

Pirri, F., Ometto, L., Fuselli, S., Fernandes, F.A., Ancona, L., Le Bohec, C., Zane, L. & Trucchi, E. 2021 Selection-driven adaptation to the extreme Antarctic environment in the Emperor penguin. *bioRxiv*.

Porras, M.A.G., Stojkova, K., Vaicik, M.K., Pelowe, A., Goddi, A., Carmona, A., Long, B., Qutub, A.A., Gonzalez, A., Cohen, R.N. and Brey, E.M. 2021 Integrins and extracellular matrix proteins modulate adipocyte thermogenic capacity. *Scientific reports* **11**, 1-14.

Puigserver, P., Vazquez, F., Bonet, M.L., Pico, C. & Palou, A. 1996 In vitro and in vivo induction of brown adipocyte uncoupling protein (thermogenin) by retinoic acid. *Biochemical Journal* **317**, 827e833.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. and Heger, A. 2012 The Pfam protein families database. *Nucleic Acids Research* **40**, D290–D301.

Rajakumari, S., Wu, J., Ishibashi, J., Lim, H.W., Giang, A.H., Won, K.J., Reed, R.R. and Seale, P. 2013 EBF2 determines and maintains brown adipocyte identity. *Cell Metabolism* **17**, 562-574.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. & Vilo, J. 2019 g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research* **47**, W191-W198.

Raymond-Bouchard, I., Tremblay, J., Altshuler, I., Greer, C.W., & Whyte, L.G. 2018 Comparative transcriptomics of cold growth and adaptive features of a eury-and steno-psychrophile. *Frontiers in microbiology* **9**, 1565.

Rey, B., Spée, M., Belouze, M., Girard, A., Prost, J., Roussel, D. and Duchamp, C. 2010 Oxygen recovery up-regulates avian UCP and ANT in newly hatched ducklings. *Journal of Comparative Physiology B* **180**, 239-246.

Rey, B., Degletagne, C., Bodennec, J., Monternier, P.A., Mortz, M., Roussel, D., Romestaing, C., Rouanet, J.L., Tornos, J. & Duchamp, C. 2016a Hormetic response triggers multifaceted anti-oxidant strategies in immature king penguins (Aptenodytes patagonicus). *Free Radical Biology and Medicine* **97**, 577-587.

Rey, B., Degletagne, C. & Duchamp, C. 2016b Transcriptomic data analysis and differential gene expression of antioxidant pathways in king penguin juveniles (*Aptenodytes patagonicus*) before and after acclimatization to marine life. *Data in brief* **9**, 549-555.

Ribot, J., Felipe, F., Bonet, M.L. & Palou, A. 2004 Retinoic acid administration and vitamin A status modulate retinoid X receptor alpha and retinoic acid receptor alpha levels in mouse brown adipose tissue. *Molecular and Cellular Biochemistry* **266**, 25e30.

Richardson, M.F., Sherwin, W.B. & Rollins, L.A. 2017 De novo assembly of the liver transcriptome of the European starling, Sturnus vulgaris. *Journal of Genomics* **5**, 54–57.

Riesgo, A., Andrade, S.C., Sharma, P.P., Novo, M., Pérez-Porro, A.R., Vahtera, V., *et al.* 2012 Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Frontiers in zoology* **9**, 1-24.

Robinson, M.D. & Oshlack, A. 2010 A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Romero, I.G., Ruvinsky, I., Gilad, Y. 2012 Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* **13**, 505–516.

Rosenzweig, R., Nillegoda, N.B., Mayer, M.P. and Bukau, B. 2019 The Hsp70 chaperone network. *Nature Reviews Molecular Cell Biology* **20**, 665-680.

Rowland, L.A., Bal, N.C. & Periasamy, M. 2015 The role of skeletal-muscle-based thermogenic mechanisms in vertebrate endothermy. *Biological Reviews* **90**, 1279-1297.

RStudio Team 2020 RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Saini, S., Duraisamy, A.J., Bayen, S., Vats, P. and Singh, S.B. 2015 Role of BMP7 in appetite regulation, adipogenesis, and energy expenditure. *Endocrine* **48**, 405-409.

Schluter, D. 2000 *The Ecology of Adaptive Radiation* (Oxford Univ. Press).

Schoville, S.D., Simon, S., Bai, M., Beethem, Z., Dudko, R.Y., Eberhard, M.J., *et al*. 2021 Comparative transcriptomics of ice-crawlers demonstrates cold specialization constrains niche evolution in a relict lineage. *Evolutionary applications* **14**, 360.

Schulz, T.J. & Tseng, Y.H. 2009 Emerging role of bone morphogenetic proteins in adipogenesis and energy metabolism. *Cytokine & growth factor reviews* **20**, 523-531.

Shalata, A., Ramirez, M.C., Desnick, R.J., Priedigkeit, N., Buettner, C., Lindtner, C., Mahroum, M., Abdul-Ghani, M., Dong, F., Arar, N. & Camacho-Vanegas, O. 2013 Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *The American Journal of Human Genetics* **93**, 1061-1071.

Shin, S.C., Kim, S.J., Lee, J.K., Ahn, D.H., Kim, M.G., Lee, H., *et al.* 2012 Transcriptomics and comparative analysis of three Antarctic notothenioid fishes. *PloS one* **7**, e43762.

Shin, S., Pang, Y., Park, J., Liu, L., Lukas, B.E., Kim, S.H., Kim, K.W., Xu, P., Berry, D.C. & Jiang, Y. 2020 Dynamic control of adipose tissue development and adult tissue homeostasis by platelet-derived growth factor receptor alpha. *Elife* **9**, e56189.

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.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd., J.M., Kelly, S. 2016 TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research* **26**, 1134-1144.

Song, L., Florea, L. 2015 Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, 48.

Song, J., & McDowell, J.R. 2021 Comparative transcriptomics of spotted seatrout (Cynoscion nebulosus) populations to cold and heat stress. *Ecology and evolution* **11**, 1352-1367.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y. & Kaplan, S. 2016 The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1-30.

Tautz, D. 2000 Evolution of transcriptional regulation. *Current Opinion in Genetics and Development* **10**, 575–579.

Taylor, J.R.E. 1986 Thermal insulation of the down and feathers of pygoscelid penguin chicks and the unique properties of penguin feathers. *The Auk* **103**, 160–168.

Teoh, C.P., Lavin, P., Lee, D.J.H., González-Aravena, M., Najimudin, N., Lee, P.C., *et al.* 2021 Genomics and transcriptomics analyses provide insights into the cold adaptation strategies of an Antarctic bacterium, Cryobacterium sp. SO1. *Polar Biology*, 1-15.

Teulier, L., Dégletagne, C., Rey, B., Tornos, J., Keime, C., de Dinechin, M., Raccurt, M., Rouanet, J.L., Roussel, D. & Duchamp, C. 2012 Selective upregulation of lipid metabolism in skeletal muscle of foraging juvenile king penguins: an integrative study. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2464-2472.

Thomas, D.B. & Fordyce, R.E. 2008 The heterothermic loophole exploited by penguins. *Australian Journal of Zoology* **55**, 317–321.

Uusi-Heikkilä, S., Sävilammi, T., Leder, E., Arlinghaus, R. & Primmer, C.R. 2017 Rapid, broad-scale gene expression evolution in experimentally harvested fish populations. *Molecular Ecology* **26**, 3954–3967.

Vaillancourt, E., Prud'Homme, S., Haman, F., Guglielmo, C.G. & Weber, J.M. 2005 Energetics of a long-distance migrant shorebird (Philomachus pugnax) during cold exposure and running. *Journal of Experimental Biology* **208**, 317-325.

Verne, S., Jaquish, B., White, R. *et al.* 2011 Global transcriptome analysis of constitutive resistance to the white pine weevil in spruce. *Genome Biology and Evolution* **3**, 851–867.

Vianna, J.A., Fernandes, F., Frugone M.J., *et al.* 2020 Genome-wide analyses reveal drivers of penguin diversification. *PNAS* **117**, 22303-22310.

Watson, M. 1883 Report on the Anatomy of the Spheniscidae Collected by HMS Challenger, During the Years 1873–1876. Edinburgh: Neill and Company.

Weber, J.M. 2009vThe physiology of long-distance migration: extending the limits of endurance metabolism.*Journal of Experimental Biology* **212**, 593-597.

Wienecke, B., Robertson, G., Kirkwood, R. & Lawton, K. 2007 Extreme dives by free-ranging emperor penguins. *Polar Biology* **30**, 133-142.

Wittkopp, P.J., Haerum, B.K. & Clark, A.G. 2008 Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics* **40**, 346–350.

Xia, J., Gill, E. E., & Hancock, R. E. 2015 NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature protocols* **10**, 823-844.

Yang, Q.S., Gao, J., He, W.D., Dou, T.X., Ding, L.J., Wu, J.H., *et al*. 2015 Comparative transcriptomics analysis reveals difference of key gene expression between banana and plantain in response to cold stress. *BMC genomics* **16**, 1-18.

Yin, Z., Zhang, F., Smith, J., Kuo, R. & Hou, Z.C. 2019a Full-length transcriptome sequencing from multiple tissues of duck, Anas platyrhynchos. *Scientific Data* **6**, 1-9.

Yin, Z.T., Zhu, F., Lin, F.B., Jia, T., Wang, Z., Sun, D.T., Li, G.S., Zhang, C.L., Smith, J., Yang, N. & Hou, Z.C. 2019b Revisiting avian 'missing' genes from de novo assembled transcripts. *BMC Genomics* **20**, 1-10.

Yu, H., Zhao, Z., Yu, X., Li, J., Lu, C. & Yang, R. 2017 Bovine lipid metabolism related gene GPAM: Molecular characterization, function identification, and association analysis with fat deposition traits. *Gene* **609**, 9-18.

Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simao, F.A., Ioannidis, P., Seppey, M., Loetscher, A. & Kriventseva, E.V. 2017 OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research* **45**, D744–D749.

Zheng, S.L., Li, Z.Y., Song, J., Liu, J.M. and Miao, C.Y. 2016 Metrnl: a secreted protein with new emerging functions. *Acta pharmacologica sinica* **37**, 571-579.

Zenteno-Savin, T., Leger, J.S. & Ponganis, P.J. 2010 Hypoxemic and ischemic tolerance in emperor penguins. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **152**, 18-23.

5. Final remarks

One of the primary goals in evolutionary biology is to identify and analyze those genetic traits that contribute to key phenotypes involved in adaptation and survival (Ellegren and Sheldon, 2008; Stinchcombe and Hoekstra, 2008). Not long ago, large-scale genomic analyses were limited to model organisms such as *Drosophila*, *Arabidopsis*, *Saccharomyces*, and domestic animals and plants. However, in the last few years, the 'next-generation' or massive parallel sequencing technologies (Rothberg and Leamon, 2008) have become integral parts in molecular ecology studies (Vera et al., 2008).

Importantly, the generation of large amounts of DNA and RNA sequence data from related species has enabled comparative genomic and transcriptomic analyses for the identification of specific loci of interest (Wang et al., 2009).

Transcriptome sequencing identifies the fraction of genes in the genome that are functionally active in a particular tissue or species. Comparative genomics, along with expression profiling, genetic mapping, and candidate gene approaches, is one of the primary methods for determining the genetic basis of phenotypic variation (Ellegren and Sheldon, 2008).

One typical application in comparative genomics is to examine orthologous gene sequences from two or more species for divergence rates and test whether this divergence deviates from rates expected under a neutral scenario (Ellegren, 2008).

Accelerated divergence beyond neutral expectations would suggest evidence of adaptive evolution where positive selection has increased the rate of fixation of advantageous alleles (Nielsen, 2005; Wright and Andolfatto, 2008). Sequence conservation beyond neutral expectations, on the other hand, would be indicative of purifying selection due to functional constraints (Ponting, 2008).

We assessed the divergence rates by analysing the rates of accumulation of non-synonymous substitution (*dN*) and synonymous substitution (*dS*) in protein-coding sequences, in the Emperor and in the King lineage separately.

As a complementary selection analysis, we also looked for extended regions of linkage-disequilibrium (or, equivalently, long haplotypes) that have swept to high prevalence quickly within each species. Indeed, a population-wide reduction of genetic diversity generated by the

fixation of a beneficial allele and the surrounding variants on the same haplotype is a hallmark of a selective sweep (Smith and Haigh, 1974).

Differences in gene expression can directly underpin ecologically significant phenotypic variation in populations (Oleksiak et al., 2002; Whitehead and Crawford, 2006; Hoekstra and Coyne, 2007), providing selection with the raw material on which to act (Ellegren and Sheldon, 2008).

Gene expression and function play a key role in understanding adaptation, phenotypic diversification, and ecological speciation because they relate the genotype to the phenotype (Sanetra et al., 2005; Manceau et al., 2010; Pavey et al., 2010).

Several studies have successfully isolated and described specific candidate genes in different tissues or life stages of ecologically divergent species, demonstrating that variation in gene expression do play a role in ecological speciation (Chan et al., 2010; Nosil, 2012; Wittkopp et al., 2008; Jones et al., 2012; Hanikenne et al., 2013). In Darwin's finches and African cichlids, for example, changes in bmp4 expression are linked to the phenotypic variation of beak and jaw morphology, ecologically relevant traits required for each group's adaptive radiation (Abzhanov et al., 2004; Albertson et al., 2005); pelvic reduction evolved in three-spined sticklebacks by a decrease in Pitx1 expression, with fixation in populations reliant on factors such as the presence of predators and calcium limitation (Chan et al., 2010).

Moreover, previous comparative gene expression analyses between recently diverged ecotypes revealed ecologically relevant traits that are not immediately evident from morphological inspections, such as the activation of different immune pathways in response to parasite exposure or the overexpression of specific pathways related to energy metabolism (Ghalambor et al., 2015; Jeukens et al., 2010; Lenz et al., 2013).

Comparative transcriptomic approaches lead to a more inclusive understanding of the role of gene expression in adaptive evolution by detecting differential expression in novel genes as well as the interaction of many genes of small effects (Mackay et al., 2009; Stapley et al., 2010), both of which are likely to contribute to the emergence of novel phenotypes.

Although our comparative genomics and transcriptomics study is mostly descriptive, it has provided insights regarding candidate genes that are potentially linked to adaptation to an extreme cold environment, namely Antarctica.

We cross-referenced the results of multiple selection analyses over long and short periods of time performed through, respectively, phylogeny-based methods and linkage-disequilibrium tests, with the expression pattern of five different tissues.

No statistically significant overlap emerged between the sets of upregulated DEGs and the candidate genes under positive selection. This could suggest that the evolution of gene expression and sequence divergence between species may involve different genes with related functions (Kavembe et al., 2015). Therefore, the Emperor penguin adaptation responses to cold may involve different sets of genes, some showing divergence in gene expression and others in coding sequences.

Kozak et al. (2014) have proposed that gene expression and coding sequences may evolve independently but converge in function to guarantee the best phenotype for each specific environment.

However, the link between gene expression and protein sequence evolution is still a topic of discussion. Some research suggests that gene expression and sequence evolution are unrelated, as proposed by our study, while others have found a positive correlation between the two evolutionary processes in a variety of species, including yeast (Kim and Yi, 2007), Drosophila (Lemos et al., 2005), ants (Hunt et al., 2013), and mammals (Khaitovich et al., 2005; Warnefors and Kaessmann, 2013).

Several factors have been proposed to influence the evolution of protein sequences and gene expression, including protein–protein interactions, mutation rates, and selection strength, which determine whether the two processes are linked or decoupled (Kavembe et al., 2015).

Nevertheless, the high number of differentially expressed genes suggests that the phenotypic changes observed in the Emperor penguin may be the result not only of adaptation to cold, but also of gene expression plasticity to environmental conditions.

In addition, expressional changes can also be driven by non-coding regulatory mutations, which we could not investigate using the transcriptomic data collected in this work. Noncoding regulatory

regions, such as cis-regulatory elements, may be a major contributor causing expressional divergence, according to several studies (Osada et al., 2017; Signor et al., 2016).

Noncoding regions, which have weaker pleiotropic effects than coding genes, can have significantly greater evolutionary flexibility and have been implicated in several episodes of adaptive evolution (King and Wilson, 1975; Sackton et al., 2019).

Future research utilizing well-annotated genomes and resequencing data from several individuals could more efficiently investigate adaptive evolution in noncoding regions.

6. References for the introduction and the final remarks

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R. & Tabin, C.J. 2004 Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462-1465.

Albertson, R.C., Streelman, J.T., Kocher, T.D. & Yelick, P.C. 2005 Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proceedings of the National Academy of Sciences* **102**, 16287-16292.

Aloia, R.C., Pengelley, E.T., Bolen, J.L. & Rouser, G. 1974 Changes in phospholipid composition in hibernating ground squirrel, Citellus lateralis, and their relationship to membrane function at reduced temperature. *Lipid* **9**, 993–999.

Allen, J.A. 1877 The influence of Physical conditions in the genesis of species. *Radical Review* **1**, 108–140.

Ancel, A., Cristofari, R., Fretwell, P.T., Trathan, P.N., Wienecke, B., Boureau, M., Morinay, J., Blanc, S., Le Maho, Y. & Le Bohec, C. 2014 Emperors in hiding: when ice-breakers and satellites complement each other in Antarctic exploration. *PLoS One* **9**, e100404.

Åqvist, J., Isaksen, G.V. & Brandsdal, B.O. 2017 Computation of enzyme cold adaptation. *Nature Reviews Chemistry* **1**, 1-14.

Black, C., Southwell, C., Emmerson, L., Lunn, D. & Hart, T. 2018 Time-lapse imagery of Adélie penguins reveals differential winter strategies and breeding site occupation. *PloS one* **13**, e0193532.

Blix, A.S. 2016 Adaptations to polar life in mammals and birds. *Journal of Experimental Biology* **219**, 1093-1105.

Borges, R., Khan, I., Johnson, W.E., Gilbert, M.T.P., Zhang, G., Jarvis, E.D., O'Brien, S.J. & Antunes, A. 2015 Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC genomics* **16**, 1-14.

Bowser, S.S. & DeLaca, T.E. 1985 Rapid intracellular motility and dynamic membrane events in an Antarctic foraminifera. *Cell Biology International Reports* **9**, 901-910.

Butler, P.J. & Jones, D.R. 1997 Physiology of diving of birds and mammals. *Physiological reviews* **77**, 837-899.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J. & Myers, R.M. 2010 Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302-305.

Cherel, Y., Gilles, J., Handrich, Y. & Le Maho, Y. 1994 Nutrient reserve dynamics and energetics during long-term fasting in the king penguin (*Aptenodytes patagonicus*). *Journal of Zoology* **234**, 1-12.

Cherel, Y., Parenteau, C., Bustamante, P. & Bost, C.A. 2018 Stable isotopes document the winter foraging ecology of king penguins and highlight connectivity between subantarctic and Antarctic ecosystems. *Ecology and evolution* **8**, 2752-2765.

Clarke, A. 1980 A reappraisal of the concept of metabolic cold adaptation in polar marine invertebrates. *Biological Journal of the Linnean Society* **14**, 7-92.

Collins, M.A. & Rodhouse, P.G. 2006 Southern ocean cephalopods. *Advances in marine biology* 50, 191-265.

Costanzo, J.P., Lee, R.E., DeVries, A.L., Wang, T. & Layne J.R. 1995 Survival mechanisms of vertebrate ectotherms at subfreezing temperatures: Applications in cryomedicine. *FASEB Journal* **9**, 351–358.

Cristofari, R., Liu, X., Bonadonna, F., *et al.* 2018 Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change* **8**, 245-251.

Cristofari, R., Bertorelle, G., Ancel, A., Benazzo, A., Le Maho, Y., Ponganis, P.J., Stenseth, N.C., Trathan, P.N., Whittington, J.D., Zanetti, E. & Zitterbart, D.P. 2016 Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature communications* **7**, 1-9.

D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M.A., Feller, G. & Gerday, C. 2002 Molecular basis of cold adaptation. Philosophical Transactions of the Royal Society of London. *Series B: Biological Sciences* **357**, 917-925.

Devries, P.L. & Hew, C.L. 1990 Biochemistry of fish antifreeze proteins. *FASEB Journal* **4**, 2460–2468.

Ellegren, H. 2008 Comparative genomics and the study of evolution by natural selection. *Molecular ecology* **17**, 4586-4596.

Ellegren, H. & Sheldon, B.C. 2008 Genetic basis of fitness differences in natural populations. *Nature* **452**, 169–175.

Fahlman, A., Schmidt, A., Handrich, Y., Woakes, A.J. & Butler, P.J. 2005 Metabolism and thermoregulation during fasting in king penguins, Aptenodytes patagonicus, in air and water. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **289**, R670-R679.

Flavahan, N.A. 1991 The role of vascular alpha-2-adrenoceptors as cutaneous thermosensors. *Physiology (Bethesda)* **6**, 251–255.

Gavryushkina, A., Heath, T.A., Ksepka, D.T., Stadler, T., Welch, D. & Drummond, A.J. 2017 Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology* **66**, 57-73.

Ghalambor, C.K., McKay, J.K., Carroll, S.P. & Reznick, D.N. 2007 Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional ecology* **21**, 394-407.

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

Goldman, S.S. 1975 Cold resistance of the brain during hibernation. III. Evidence of a lipid adaptation. *American Journal of Physiology* **288**, 834–838.

Goldsmith, R. & Sladen, W.J.L. 1961 Temperature regulation of some Antarctic penguins. *The Journal of physiology* **157**, 251-262.

Groscolas, R. 1990 Metabolic adaptations to fasting in emperor and king penguins. In *Penguin biology* (pp. 269-296). Academic Press.

Groscolas, R. & Robin, J.P. 2001 Long-term fasting and re-feeding in penguins. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **128**, 643-653.

Hanikenne, M., Kroymann, J., Trampczynska, A., Bernal, M., Motte, P., Clemens, S. & Krämer, U. 2013 Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS genetics* **9**, e1003707.

Hoekstra, H.E. & Coyne, J.A. 2007 The locus of evolution: evo devo and the genetics of adaptation. *Evolution: International Journal of Organic Evolution* **61**, 995-1016.

Hong, S.Y., Gal, J.K., Lee, B.Y., Son, W.J., Jung, J.W., La, H.S., Shin, K.H., Kim, J.H. & Ha, S.Y. 2021 Regional Differences in the Diets of Adélie and Emperor Penguins in the Ross Sea, Antarctica. *Animals* **11**, 2681.

Hunt, B.G., Ometto, L., Keller, L. & Goodisman, M.A. 2012 Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Molecular biology and evolution* **30**, 263-271.

Jarvis, E.D., Mirarab, S., Aberer, A.J., *et al.* 2014 Whole genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331.

Jeukens, J., Renaut, S., ST-CYR, J.É.R.Ô.M.E., Nolte, A.W. & Bernatchez, L. 2010 The transcriptomics of sympatric dwarf and normal lake whitefish (Coregonus clupeaformis spp., Salmonidae) divergence as revealed by next-generation sequencing. *Molecular ecology* **19**, 5389-5403.

Johnston, I.A. 1990 Cold adaptation in marine organisms. *Philosophical Transactions of the Royal Society B* **326**, 655-667.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S. & Birney, E. 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.

Kavembe, G.D., Franchini, P., Irisarri, I., Machado-Schiaffino, G. & Meyer, A. 2015 Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of earth's most extreme environments, the hypersaline soda Lake Magadi in Kenya, east Africa. *Journal of molecular evolution* **81**, 90-109.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M. & Pääbo, S. 2005 Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850-1854.

Kim, S.H. & Yi S.V. 2007 Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica* **131**, 151–156.

King, M.C. & Wilson, A.C. 1975 Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-116.

Kooyman, G.L. & Ponganis, P.J. 2014 Chick production at the largest emperor penguin colony decreases by 50% from 2008–10. *Antarctic Science* **26**, 33-37.

Kooyman, G. 2009 Milestones in the study of diving physiology: Antarctic emperor penguins and Weddell seals. *Smithsonian at the Poles: Contributions to International Polar Year Science*.

Kozak, G.M., Brennan, R.S., Berdan, E.L., Fuller, R.C. & Whitehead, A. 2014 Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution* **68**, 63-80.

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

Kumar, V., Kutschera, V.E. & Nilsson, M.A. 2015 Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. *BMC Genomics* **16**, 1-13.

Le Maho, Y., Delclitte, P. & Chatonnet, J. 1976 Thermoregulation in fasting emperor penguins under natural conditions. *American Journal of Physiology-Legacy Content* **231**, 913-922.

Lemos, B., Bettencourt, B.R., Meiklejohn, C.D. & Hartl, D.L. 2005.Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Molecular biology and evolution* **22**, 1345-1354.

Lenz, T.L., Eizaguirre, C., Rotter, B., Kalbe, M. & Milinski, M. 2013 Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Molecular ecology* **22**, 774-786.

Li, C., Zhang, Y., Li, J., *et al.* 2014 Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience* **3**, 2047-217X.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M.A., Gamba, C., Seguin-Orlando, A. & Mortensen, C.D. 2015 Tracking the origins of Yakutian horses and the genetic basis

for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences* **112**, E6889-E6897.

Liu, S., Lorenzen, E.D., Fumagalli, M., *et al.* 2014 Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell **157**, 785-794.

Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W. & Schuster, S.C. 2015 Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell reports* **12**, 217-228.

Mackay, T.F., Stone, E.A. & Ayroles, J.F. 2009 The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**, 565-577.

Manceau, M., Domingues, V.S., Linnen, C.R., Rosenblum, E.B. & Hoekstra, H.E. 2010 Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 2439-2450.

Nielsen, R. 2005 Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197–218.

Nosil, P. 2012 Ecological speciation. *Oxford University Press*.

Oleksiak, M.F., Churchill, G.A. & Crawford, D.L. 2002 Variation in gene expression within and among natural populations. *Nature genetics* **32**, 261-266.

Osada, N., Miyagi, R. & Takahashi, A. 2017 Cis-and trans-regulatory effects on gene expression in a natural population of Drosophila melanogaster. *Genetics* **206**, 2139-2148.

Owens, I. & Zawar-Reza, P. 2015 Antarctica's Role in the Global Atmospheric System. *Exploring the Last Continent*, p.91.

Pan, H., Cole, T.L., Bi, X., *et al.* 2019 High-coverage genomes to elucidate the evolution of penguins. *GigaScience* **8,** 1-17.

Pavey, S.A., Collin, H., Nosil, P. & Rogers, S.M. 2010 The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences* **1206**, 110.

Ponting, C.P. 2008 The functional repertoires of metazoan genomes. *Nature Reviews Genetics* **9**, 689–698.

Ramos, B., González-Acuña, D., Loyola, D.E. *et al.* 2018 Landscape genomics: natural selection drives the evolution of mitogenome in penguins. *BMC Genomics* **19**, 1-17.

Rothberg, J.M. & Leamon, J.H. 2008 The development and impact of 454 sequencing. *Nature Biotechnology* **26**, 1117–1124.

Rowland, L.A., Bal, N.C. & Periasamy, M. 2015 The role of skeletal-muscle-based thermogenic mechanisms in vertebrate endothermy. *Biological Reviews* **90**, 1279-1297.

Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M. & Edwards, S.V. 2019 Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74-78.

Sanetra, M., Begemann, G., Becker, M.B. & Meyer, A. 2005 Conservation and co-option in developmental programmes: the importance of homology relationships. *Frontiers in zoology* **2**, 1-17.

Scholander, P.F 1955 Evolution of climatic adaptation in homeotherms. *Evolution* **9**, 15-26.

Signor, S.A., Liu, Y., Rebeiz, M. & Kopp, A. 2016 Genetic convergence in the evolution of male-specific color patterns in Drosophila. *Current Biology* **26**, 2423-2433.

Smith, J.M. & Haigh, J. 1974 The hitch-hiking effect of a favourable gene. *Genetics Research* **23**, 23-35.

Solomonov, N.G., Anufriev, A.I., Yadrikhinskii, V.F. & Isaev, A.P. 2009 Body temperature changes in purebred and hybrid Yakut horses under the conditions of Yakutia. *Doklady Biological Sciences* **427**, 358–361.

Stapley, J., Reger, J., Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P. & Slate, J. 2010 *Adaptation genomics: the next generation. Trends in ecology & evolution* **25**, 705-712.

Stinchcombe, J.R. & Hoekstra H.E. 2008 Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170.

Stonehouse, B. 1975 Adaptation in polar and subpolar penguins (*Spheniscidae*). In *Antarctic ecology* (ed. M. W. Holdgate), pp. 526–541. San Diego, CA: Academic Press.

Storey, K.B. & Storey, J.M. 1988 Freeze tolerance in animals. *Physiological Reviews* **68**, 27–84.

Tattersall, G.J., Sinclair, B.J., Withers, P.C., *et al.* 2012 Coping with thermal challenges: physiological adaptations to environmental temperatures. *Comprehensive Physiology* **2**, 2151–202.

Taylor, J.R. 1986 Thermal insulation of the down and feathers of pygoscelid penguin chicks and the unique properties of penguin feathers. *The Auk* **103**, 160-168.

Thieringer, H.A., Jones, P.G. & Inouye, M. 1998 Cold shock and adaptation. *BioEssays* **20**, 49–57.

Thomas, D.B. & Fordyce, R.E. 2007 The heterothermic loophole exploited by penguins. *Australian Journal of Zoology* **55**, 317-321.

Thompson, G.E. 1977 Physiological effects of cold exposure. *International Review of Physiology* **15**, 29–69.

Trathan, P.N., Fretwell, P.T. & Stonehouse, B. 2011 First recorded loss of an emperor penguin colony in the recent period of Antarctic regional warming: implications for other colonies. *PLoS one* **6**, e14738.

Trucchi, E., Gratton, P., Whittington, J.D., *et al.* 2014 King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140528.

Trucchi, E., Cristofari, R. & Le Bohec, C. 2019 Reply to: The role of ocean dynamics in king penguin range estimation. *Nature Climate Change* **9**, 122-122.

Vanhoutte, P.M. 2011 Physical factors of regulation. *Comprehensive Physiology,* (Wiley, New York).

Vera, J.C., Wheat, C.W., Fescemyer, H.W., *et al.* 2008 Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**, 1636–1647.

Vianna, J.A., Fernandes, F., Frugone, M.J., *et al.* 2020 Genome-wide analyses reveal drivers of penguin diversification. *PNAS* **117**, 22303-22310.

Wang, Z., Gerstein, M. & Snyder, M. 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63.

Warnefors, M. & Kaessmann, H. 2013 Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome biology and evolution* **5**, 1324-1335.

Watson, M. 1883 Report on the anatomy of the *Spheniscidae* collected by HMS Challenger, during the years 1873–1876. Edinburgh, UK: Neill and Co.

Whitehead, A. & Crawford, D.L. 2006 Variation within and among species in gene expression: raw material for evolution. *Molecular ecology* **15**, 1197-1211.

Williams, R.C., Correia, J.J. & DeVries, A.L. 1985 Formation of microtubules at low temperature by tubulin from Antarctic fish. *Biochemestry* **24**, 2790-2798.

Wittkopp, P.J., Haerum, B.K. & Clark, A.G. 2008 Regulatory changes underlying expression differences within and between Drosophila species. *Nature genetics* **40**, 346-350.

Wright, S.I. & Andolfatto, P. 2008 The impact of natural selection on the genome: emerging patterns in Drosophila and Arabidopsis. *Annual Review of Ecology Evolution and Systematics* **39**, 193–213.

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**, 1-11.

Younger, J.L., Clucas, G.V., Kooyman, G., Wienecke, B., Rogers, A.D., Trathan, P.N., Hart, T. & Miller, K.J. 2015 Too much of a good thing: sea ice extent may have forced emperor penguins into refugia during the last glacial maximum. *Global Change Biology* **21**, 2215-2226.

Yudin, N.S., Larkin, D.M. & Ignatieva E.V. 2017 A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genetics* **18**, 111.

Zhang, G., Li, C., Li, Q., *et al.* 2014 Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-20.

Zhao, H., Li, J. & Zhang, J. 2015 Molecular evidence for the loss of three basic tastes in penguins. *Current Biology* **25**, R141-2.

7. SI Appendix

Supplementary tables for the section 2

Table S1. GigaDB and Genbank accession ID of the coding sequences (CDS) of the twenty bird species selected for the phylogeny

Species	Accession ID	Database
Phaethon lepturus	GCA_000687285.1	GenBank
Eurypyga helias	GCA_000690775.1	GenBank
Gavia stellata	GCA_000690875.1	GenBank
Fulmarus glacialis	GCA_000690835.1	GenBank
Phalacrocorax carbo	GCA_000708925.1	GenBank
Nipponia nippon	GCA_000708225.1	GenBank
Egretta garzetta	GCA_000687185.1	GenBank
Pelecanus crispus	GCA_000687375.1	GenBank
Haliaeetus leucocephalus	GCA_000737465.1	GenBank
Tyto alba	GCA_000687205.1	GenBank
Cariama cristata	GCA_000690535.1	GenBank
Corvus brachyrhynchos	GCA_000691975.1	GenBank
Opisthocomus hoazin	GCA_000692075.1	GenBank
Aptenodytes forsteri	Aptenodytes_forsteri	GigaDB
Aptenodytes patagonicus	KP FORT 001	GigaDB
Pygoscelis adeliae	Pygoscelis_adeliae	GigaDB
Pygoscelis papua	Gentoo penguin DNA -4	GigaDB
Eudyptula minor minor	Gonzo	GigaDB
Spheniscus magellanicus	AH 6	GigaDB
Eudyptes chrysolophus	MP PEI 1	GigaDB

Table S2. Candidate genes predicted to be under positive selection by aBSREL models in A. forsteri (FDR < 0.05)

Gene ID	Gene name	Full adaptive model	Full adaptive model (non-synonymous subs/site)	Full adaptive model (synonymous subs/site)	LRT	Adjusted p-value
AF_67	RAB11FIP1	0,03978832569	0,03756351761	0,002224808077	15,23266442	0,004735497959
AF_126	SIX4	0,3441046301	0,343446601	0,0006580290316	14,22141505	0,007197335296
AF_224	RECK	0,1463463798	0,1459333673	0,0004130124426	24,53948108	0,0001031920533
AF_262	CTNNB1	0,1493238387	0,1491977367	0,0001261019944	22,04681764	0,0002728009514
AF_284	MATN3	18,92142651	18,91413236	0,007294147998	16,11095243	0,003275988575
AF_393	AAED1	0,01953908628	0,01932694688	0,0002121394001	10,96131878	0,02844513253
AF_407	DAPK1	0,8098414789	0,8086136977	0,001227781231	44,15418911	0,00000003670802221
AF_471	HFM1	0,00879112879	0,007957081776	0,0008340470144	15,8110221	0,003733917226
AF_475	ZNF326	0,01009790176	0,009474285168	0,0006236165973	10,90570251	0,02908619585
AF_505	VTG2	0,01056709205	0,009231676141	0,001335415905	12,3313369	0,01597452425
AF_532	FAAH	0,1486163283	0,148387298	0,0002290302404	19,72381036	0,0006899259259
AF_539	ATP1A1	0,0045410133	0,003540288914	0,001000724387	19,40315246	0,0007940719858
AF_569	BLVRA	0,009176526122	0,009169842134	0,000006683988159	13,1744513	0,01108689193
AF_617	ORAI2	0,1204426367	0,1191761458	0,001266490874	22,835274	0,0002067553426
AF_653	CHD1	0,00366681623	0,003661335911	0,000005480319194	13,31745026	0,01041231951
AF_753	SLCO2B1	1,53892817	1,537672809	0,001255361024	22,45030622	0,000233533378
AF_755	PGM2L1	2,548265627	2,546378343	0,001887283342	14,48694696	0,006529131699
AF_767	KIAA1211L	0,05639786074	0,05358806488	0,002809795854	19,32306804	0,0008225005249
AF_798	EEF2KMT	7,480391256	7,47683077	0,003560485915	32,43752956	0,000003532612265
AF_812	CCDC78	0,4645971468	0,4620782485	0,002518898309	40,67091371	0,0000001141742667
AF_827	DECR2	1,127009506	1,1266572	0,0003523059627	12,56959359	0,01433239235
AF_864	DNASE1	0,08281133932	0,07238054746	0,01043079186	13,78534023	0,00868524499
AF_867	CLCN7	0,02132897875	0,01988934777	0,001439630972	31,85409517	0,000004434647767
AF_964	CCNL1	0,142891058	0,1422133439	0,0006777140744	15,15261744	0,004869908448
AF_1044	ACMSD	6,074454798	6,07150117	0,002953628114	10,97731619	0,02834500441
AF_1081	UBA5	0,02760754756	0,02759661161	0,00001093594528	17,17620541	0,0020476022
AF_1154	TRAIP	0,02395098152	0,02108769185	0,002863289671	15,19354682	0,004811779067
AF_1397	WDR91	1,151496155	1,150012696	0,001483458679	15,31660935	0,004590771994
AF_1408	NUP205	0,5655074625	0,5635384856	0,001968976899	24,16836947	0,0001187902981
AF_1424	SPPL2B	0,09318646639	0,09309602034	0,00009044604672	28,79132923	0,00001635403894
AF_1497	ATG5	0,8201047647	0,8197896349	0,0003151298028	29,29253855	0,00001334030508
AF_1557	RASGRP1	0,007786515805	0,006253846927	0,001532668879	17,53637812	0,001753367932
AF_1588	AQP2	0,01382013194	0,01262341776	0,001196714184	16,50612816	0,002761966709
AF_1620	NADSYN1	0,02463090053	0,02292391302	0,001706987508	41,37933785	0,00000009315064475
AF_1642	FADS1	3,628145165	3,625037933	0,003107232386	10,20621951	0,03929380701
AF_1775	PCCA	0,7648488704	0,7639763486	0,000872521782	17,93037528	0,001491480849
AF_1783	DOCK9	0,01792926443	0,01723753967	0,0006917247652	13,15480668	0,01116276863
AF_1787	IPO5	0,1220430759	0,1210062034	0,001036872438	23,41509412	0,0001612436106
AF_1799	CNOT6L	0,001534889751	0,001475056985	0,00005983276553	12,23167376	0,01670106848
AF_1807	LOC103906701	0,366533191	0,3612674233	0,005265767622	13,2091632	0,01096226353
AF_1816	JCHAIN	10,77552827	10,76832797	0,007200292626	21,97362283	0,000279273065
AF_1882	KIF20A	0,04869304532	0,04764856366	0,001044481659	42,14103884	0,00000007952623791
AF_1887	TMEM68	0,1204860162	0,1162487623	0,004237253928	24,76516954	0,00009295190657
AF_1912	USP53	0,006054971834	0,005812362898	0,000242608936	25,82288114	0,00005818049088
AF_1917	PRDM5	4,079948322	4,076727551	0,003220770996	14,76791604	0,005728264803
AF_1929	PTCD3	0,00921734749	0,007922864097	0,001294483393	14,21802588	0,007197335296
AF_1975	ATXN1L	0,02137468629	0,02136486398	0,000009822306982	22,58436207	0,0002253074507
AF_2053	ADAD1	4,033811871	4,030321859	0,003490012375	29,26333547	0,00001337397837
AF_2077	MYRF	0,2893620799	0,2886663865	0,0006956933596	22,51520215	0,0002308235853
AF_2083	LMAN1	0,00295427087	0,002950300764	0,000003970105988	12,26214011	0,01649402119
AF_2109	PLEKHM2	1,269284393	1,268165549	0,001118843695	18,95591141	0,0009647814647
AF_2179	VPS13D	0,008764636026	0,008260184461	0,0005044515649	11,03963978	0,02773965306
AF_2266	BOC	0,002312002236	0,001873143668	0,0004388585682	11,62245584	0,0216244603
AF_2282	NME7	0,06470575529	0,06461877215	0,00008698313518	12,15978793	0,01721832703
AF_2333	DACT1	0,01632246372	0,01565816008	0,00066430364	25,27772448	0,00007315760713
AF_2373	PSMC6	1,675165805	1,665237367	0,009928437524	14,66774347	0,006003237488
AF_2396	L2HGDH	0,007668108554	0,005914862833	0,001753245721	10,50968082	0,03483108728
AF_2501	CBX7	0,1420097652	0,1419000797	0,0001096855037	22,35473643	0,0002416773208
AF_2518	SGSM3	0,5203066568	0,5189417723	0,001364884486	22,67445553	0,0002208660149
AF_2675	SPG20	0,02931811527	0,02723157383	0,002086541444	9,8522667	0,04598483916
AF_2773	TBC1D9	0,004914651648	0,002685978369	0,00222867328	16,33249122	0,002977310999
AF_2789	GAB1	0,04923557324	0,04912591907	0,0001096541685	27,40643562	0,0000306161936
AF_2812	COCH	0,01731891179	0,01730826488	0,00001064690551	14,23614743	0,007197335296
AF_2852	BRMS1L	0,01132246109	0,01017352006	0,001148941035	12,51335502	0,01465849019
AF_2853	MBIP	0,04193802384	0,03967564561	0,002262378234	15,61848747	0,004050264847
AF_2889	AGL	0,1418661565	0,1390376657	0,002828490814	66,87906083	2,40E-12
AF_2912	ABCA4	0,3394585731	0,3384368122	0,001021760962	30,7158949	0,000007145043674
AF_2934	BRDT	1,171623807	1,170830269	0,0007935385179	36,82858007	0,00000064207998
AF_2954	JMJD1C	0,1866937015	0,1856161309	0,001077570657	23,46840075	0,0001586972209
AF_3056	SLC6A17	0,2825163628	0,2819921229	0,0005242398926	13,34553852	0,01033031961
AF_3175	COQ9	11,28423583	11,28178076	0,002455069428	42,046746	0,00000007974715523
AF_3429	ZCCHC4	0,02145941441	0,01973715793	0,001722256484	21,72239629	0,0003069250126
AF_3445	TMA16	0,4944993733	0,4944324818	0,00006689145004	41,66894926	0,00000009188883773
AF_3475	GALNT7	0,05854879404	0,03891157761	0,01963721643	22,57746357	0,0002253074507
AF_3490	SPCS3	0,06083411427	0,05559048664	0,00524362763	13,6583595	0,009124118166
AF_3520	SUSD5	0,2260853005	0,2255291055	0,0005561950056	14,18489382	0,007288902697
AF_3559	STARD3NL	1,781957483	1,780455617	0,0015018653	17,60083887	0,001712473949
AF_3752	HNF4G	0,07383319862	0,07376917012	0,00006402850067	14,92828196	0,005340961046
AF_3784	NEDD4	0,007867344004	0,00744343847	0,0004239055339	29,64929325	0,00001143422481
AF_3839	MAGI3	0,01492588857	0,0119220488	0,003003839769	26,5290658	0,00004373746908
AF_3995	ITPR3	0,01359517896	0,01153521953	0,002059959425	32,66550056	0,000003252830472
AF_4022	ZNF76	2,075821736	2,072063911	0,003757825333	19,81570631	0,0006654562324
AF_4157	FANCA	0,04163997639	0,03987345235	0,00176652404	12,81989673	0,01289845235
AF_4405	COPB2	0,003382067176	0,003077928579	0,0003041385963	11,80625502	0,02000600088
AF_4464	RPAC2	0,0635537822	0,05574498825	0,007808793955	13,70892113	0,008979953594
AF_4473	LOC103894128	0,01944187994	0,01886969321	0,0005721867356	29,91099906	0,00001041890587
AF_4476	EXOC1	0,1065868812	0,1057529498	0,0008339313523	43,77873236	0,00000004054858036
AF_4505	KIAA2022	0,03285119131	0,03164966737	0,001201523936	25,68993935	0,00006057186228
AF_4510	NME8	0,02258568456	0,0219900496	0,0005956349615	33,36194355	0,000002452653903
AF_4525	DPY19L1	0,8182403043	0,8165752796	0,001665024678	18,42291996	0,001219095109
AF_4576	HERC5	10,15353877	10,15012397	0,003414797855	15,18142184	0,004823372845
AF_4632	NPNT	0,915919142	0,9144191119	0,001500030088	13,66937384	0,009117862086
AF_4640	SGMS2	0,469739201	0,4695225555	0,0002166455254	26,67318644	0,0000419456323
AF_4656	EGF	0,003541120473	0,00214877122	0,001392349253	10,25543171	0,03870671253
AF_4666	HGSNAT	0,7933310732	0,7927564759	0,0005745972484	33,50895364	0,000002402727129
AF_4780	KIF13A	0,00945767291	0,008894608783	0,000563064127	15,44083728	0,004412036259
AF_4800	JAKMIP1	0,006780313096	0,005263738593	0,001516574503	12,62705578	0,01396409478

AF_4804	EVC2	0,009139362228	0,008791149845	0,0003482123829	41,18885575	0,00000009459493518
AF_4871	TP53BP2	0,008828493449	0,007884811957	0,0009436814923	14,35638971	0,006899963935
AF_4919	ZNF512B	0,0494718192	0,04486307598	0,004608743225	14,94127993	0,005334210714
AF_4948	SLC17A9	9,592798121	9,587249289	0,005548831908	11,60112729	0,02176278925
AF_4967	OSBPL2	0,1569758082	0,1560461371	0,0009296711032	26,49456859	0,0000440602787
AF_5022	HERC1	0,005245301286	0,004417134627	0,0008281666587	27,9951148	0,00002303997516
AF_5025	TBC1D2B	0,004162930193	0,00339313016	0,0007698000328	9,688079163	0,04937725887
AF_5042	RBM19	0,01000283162	0,00937909807	0,0006237335461	33,92347968	0,000002025680885
AF_5053	IRAK1BP1	0,06337511533	0,06334588215	0,00002923317857	14,2532382	0,007197335296
AF_5071	TFAP2C	0,7188867334	0,7185770624	0,0003096710258	34,15064323	0,000001842483609
AF_5128	ACOT7	29,60381824	29,59718598	0,0066322606	21,13433397	0,0003926042443
AF_5155	PRDM16	0,009940803745	0,009488432163	0,0004523715827	37,61013056	0,0000004575096084
AF_5226	ACAP3	1,003614538	1,001131029	0,002483509415	12,99795666	0,01200732941
AF_5343	EP400	0,1150294599	0,1103427452	0,004686714648	18,10481487	0,001403732804
AF_5352	ZDHHC8	0,03830667807	0,03705876978	0,001247908287	33,46172869	0,000002416338058
AF_5369	SEPTIN2	1,909258359	1,907081141	0,002177218284	25,88565162	0,0000569794185
AF_5373	UFD1L	0,02801107365	0,02662521742	0,001385856228	12,0102065	0,01825336473
AF_5390	SCN8A	0,007425941976	0,007375021108	0,0000509208685	33,20317029	0,0000026107793229
AF_5464	CD101	0,01898172726	0,0167812132	0,002200514066	13,46780486	0,009805205741
AF_5508	C7	0,01332810461	0,01144798466	0,001880119954	22,40326497	0,0002374761789
AF_5516	SELENOP	0,2466970662	0,242742043	0,00395502321	13,83298998	0,008507205667
AF_5623	IDUA	0,2397789785	0,2378589348	0,001920043743	57,05977363	0,0000000001663416904
AF_5652	NEK4	0,7581716173	0,7564679075	0,001703709806	17,47838987	0,001790203317
AF_5751	MAP4K3	14,66609571	14,66040348	0,005692227596	41,17534904	0,00000009459493518
AF_5762	HNRNPLL	0,007797190854	0,007232229832	0,0005649610 22	19,21583022	0,0008593548994
AF_5846	GRB14	0,01006703456	0,009351716978	0,0007153175826	19,30402232	0,0008262364449
AF_5849	IFIH1	1,112654948	1,111564765	0,001090183169	26,36421253	0,00004612420485
AF_5866	MYO10	0,02321004545	0,02320028941	0,000009756041587	11,71040062	0,02082314261
AF_5911	STK36	0,03436092576	0,03374225233	0,0006186734243	19,93302898	0,0006412721739
AF_5943	SMARCAL1	0,5482235134	0,5473190464	0,0009044670416	39,60681089	0,0000001821791806
AF_5979	DIRC2	0,01305891456	0,01084909637	0,002209818196	9,799296511	0,04689719347
AF_5984	ADCY5	1,103921479	1,103588253	0,0003332260808	74,26338727	2,12E-13
AF_5987	CCDC14	0,0305091642	0,02925655797	0,001252606235	26,99928615	0,00003694608098
AF_6080	ARAP2	0,05158722253	0,05116430874	0,000422913793	13,86523737	0,008425725666
AF_6230	CUL4B	0,007016997298	0,004677539947	0,00233945735	11,44550269	0,02328629733
AF_6305	UFL1	0,5520319313	0,5509267844	0,001105146894	20,37581776	0,0005483757567
AF_6321	COQ3	0,009966581185	0,008790617018	0,001175964167	11,62806416	0,0216244603
AF_6396	MATN2	76,93773028	76,92720736	0,01052292833	60,01790401	4,54E-11
AF_6428	PLCG1	3,766835923	3,763865664	0,002970258406	28,17008723	0,00002182395621
AF_6444	ATRIP	0,826562458	0,825510247	0,001052211022	31,0343879	0,0000062575413 96
AF_6455	NBAS	0,001412499983	0,001080740158	0,0003317598248	10,96005133	0,02844513253
AF_6460	GREB1	0,03241624331	0,03175392508	0,0006623182301	64,38662242	6,26E-12
AF_6466	ROCK2	0,008449948599	0,007868163878	0,0005817847213	25,78843661	0,00005818049088
AF_6467	PQLC3	40,56335263	40,55436553	0,008987104862	11,18368032	0,02608280139
AF_6523	AKAP11	0,00825123253	0,007359456234	0,0008917762966	32,3046771	0,0000037177599 27
AF_6540	DIAPH3	0,03012633944	0,02931124288	0,0008150965607	18,24482378	0,001320680642
AF_6604	MRC1	0,00239331999	0,001176886811	0,001216433179	10,32843756	0,03775007209
AF_6608	SLC39A12	0,008581028928	0,008577155282	0,000003873645773	16,69277537	0,002556544635
AF_6700	EDEM3	0,004357704614	0,003198556722	0,001159147892	12,22295038	0,01672638451
AF_6764	STX6	0,124444855	0,1216400366	0,002804818449	21,23120164	0,0003762162058
AF_6792	MAP3K3	1,58664796	1,584931662	0,001716297786	22,65415592	0,0002215112826
AF_6794	DLEC1	2,110929476	2,110066959	0,0008625175625	45,75506885	0,00000002140975404
AF_6831	WNT3A	3,94639955	3,945296909	0,001102640964	33,38096408	0,0000024526539 03
AF_6861	HRAS	2,322802122	2,321018907	0,001783214186	25,33336037	0,0000717847392
AF_6879	AP2A2	0,004726739816	0,003995685639	0,0007310541775	18,04358598	0,001440876688
AF_6922	ATP11A	0,2737805531	0,2726085633	0,001171989812	14,25117298	0,007197335296
AF_6996	RPL31	0,2160326998	0,216014005	0,00001869487392	13,19615707	0,01100020325
AF_7017	GMEB1	2,931421898	2,929578889	0,001843009364	15,39300465	0,0045022272 11
AF_7112	SRSF10	1,34082902	1,340098115	0,0007309049577	20,61811772	0,0004934399512
AF_7253	GTF2IRD1	0,02196983565	0,01865322187	0,003316613786	24,15078056	0,0001187902981
AF_7276	UNC13C	0,02223902557	0,02183811405	0,0004009115199	49,26832903	0,000000005478789589
AF_7370	TMC3	0,03235078533	0,03197082473	0,0003799606008	17,45212477	0,001805484889
AF_7436	ELP2	0,03568769623	0,03566942553	0,00001827070224	21,55770364	0,0003252035147
AF_7489	CUNH18orf63	0,01907021904	0,01764266773	0,00142755131	18,01064022	0,001458156507
AF_7503	MYLK3	0,05456947156	0,05277131118	0,00179816038	15,93248378	0,003554357963
AF_7512	PHKB	0,01341722578	0,01100354306	0,002413682718	19,41628267	0,000792875358
AF_7564	SLC7A6	0,2615421031	0,2613760966	0,0001660065391	29,83809698	0,00001066821507
AF_7642	IARS2	2,015634505	2,014124206	0,001510298873	19,93224513	0,0006412721739
AF_7649	ASB9	1,937031878	1,936164617	0,0008672608589	28,00820986	0,00002303997516
AF_7678	CDKL5	0,005686387081	0,002764532136	0,002921854945	14,12189587	0,007453344956
AF_7709	SMS	0,01585477328	0,01276152485	0,003093248426	21,31755911	0,0003624335635
AF_7711	PTCHD1	0,274252728	0,2734066306	0,000846097412	10,32456479	0,03775007209
AF_7751	EXPH5	0,006877410336	0,004919859205	0,001957551131	10,98527823	0,02830494523
AF_7762	MARCH6	1,196042983	1,194843402	0,001199580813	12,94650879	0,0122108129
AF_7895	GLS	0,0153220458	0,01531663302	0,000005412780182	29,30605089	0,00001334030508
AF_7908	ATP6V1C1	0,0673397153	0,06350905571	0,003830659589	32,54548799	0,000003399662175
AF_8088	GAN	0,08461075488	0,08273966745	0,001871087426	15,37909205	0,004516559759
AF_8090	PKD1L2	0,4965526433	0,4950327396	0,001519903617	22,01723792	0,0002750727968
AF_8127	SLC25A4	0,07242602985	0,06855132334	0,003874706503	30,76283202	0,00000707319326
AF_8156	ASAH1	0,180978473	0,1790564627	0,001922010334	56,10039996	0,000000000230379858
AF_8173	SLC7A2	0,2158252023	0,2057370341	0,01008816826	35,56437263	0,0000010038002
AF_8191	KIAA1211	1,347269666	1,345599928	0,00166973789	18,15845725	0,001372802496
AF_8264	MYBPC3	0,002465328147	0,001323866208	0,001141461939	10,81497555	0,03037211338
AF_8271	ARFGAP2	1,215035571	1,212752196	0,002283375027	13,64285207	0,009151103905
AF_8290	PEX16	0,02533374716	0,0228588797	0,002474867463	13,70609651	0,008979953594
AF_8317	ITPKA	0,07948197886	0,07942590878	0,00005607008548	19,80603098	0,0006654562324
AF_8333	PLA2G4E	1,045559116	1,044903361	0,0006557548897	26,80361516	0,00003969861473
AF_8336	AOAH	1,004717416	1,003816178	0,0009012381161	17,93449351	0,001491480849
AF_8341	CAPN3	1,393148687	1,391217407	0,001931279959	26,55568843	0,00004359321896
AF_8360	PTGR2	6,114255848	6,111060549	0,003195298252	12,39452189	0,01551853998
AF_8401	SEC24C	0,08123627963	0,08067922173	0,000557057892	53,03899402	0,0000000009332595946
AF_8439	NOLC1	0,2877836602	0,2870802633	0,0007033969589	15,31851306	0,004590771994
AF_8507	IFIT5	0,8703109592	0,8688983555	0,001412603683	17,81251205	0,001553340243
AF_8543	IDE	0,06954494136	0,0652883135	0,004256627859	40,78200642	0,0000001114783137
AF_8564	ENTPD1	0,9232658192	0,9226229107	0,0006429084951	44,69119138	0,00000003252280495
AF_8615	SLC2A1	1,46749876	1,466357737	0,001141023196	12,12370595	0,01748439292
AF_8676	SMC3	0,001555045025	0,001225269526	0,0003297754991	11,3339364	0,02450283608
AF_8679	SHOC2	0,006466755992	0,006047314469	0,0004194415235	13,45599039	0,009832643163
AF_8703	TDRD1	0,02383866312	0,02297535507	0,0008633080505	16,1104096	0,003275988575
AF_8721	CCDC172	0,02139381718	0,02138911033	0,000004706850261	20,22886338	0,000577192499

AF_8749	MCMBP	0,01768397501	0,01629924446	0,001384730553	24,41705036	0,0001087843763
AF_8859	KIAA0100	0,5151053434	0,512962425	0,002142918406	15,74542065	0,0038143613
AF_8931	ULK2	0,4685110285	0,466283928	0,002227100427	11,27028372	0,02516972731
AF_8948	RAD51C	0,2599204861	0,2591794804	0,0007410057144	17,04494169	0,002159639103
AF_8972	INTS2	0,02673584555	0,02648060925	0,0002552363015	44,42907002	0,000000003476696855
AF_8983	APPBP2	0,009512621027	0,008953854224	0,0005587668028	28,01091184	0,00002303997516
AF_9000	SYNRG	0,001817541614	0,001204579517	0,0006129620973	11,12583421	0,0267139225
AF_9101	LRPPRC	0,01882777214	0,01808437273	0,0007433994089	32,83424921	0,000003038031195
AF_9132	HDHD2	0,04337220122	0,04336310223	0,000009098987899	23,64911572	0,0001484958218
AF_9208	FRMPD4	0,003030208313	0,002473209253	0,0005569990596	12,1091944	0,01756255054
AF_9256	TLK1	0,002431916282	0,001967820711	0,0004640955707	10,36047204	0,03735043747
AF_9261	DYNC1I2	1,176785899	1,175581336	0,001204562944	31,45697981	0,000005205579133
AF_9269	PPP1R9A	0,01063592033	0,009171691092	0,001464229241	12,685698	0,01367824218
AF_9324	NRP1	0,1333378255	0,1329186701	0,0004191554272	15,36390424	0,004527384238
AF_9395	LEPR	0,01815337243	0,01753063618	0,0006227362465	17,30625269	0,001934517579
AF_9464	NRG3	0,1428630036	0,1375500413	0,005312962331	21,96118644	0,000279273065
AF_9465	SH2D4B	0,9731184149	0,9724262155	0,0006921994229	34,19633193	0,000001836086873
AF_9481	NRG4	0,1593165742	0,159262557	0,00005401721511	10,07382642	0,04170853662
AF_9570	GLIPR1	5,004217405	5,002901418	0,001315987432	14,55531131	0,006330469463
AF_9601	CACNA1C	0,06601467318	0,06266040964	0,003354263541	28,91877129	0,00001570693992
AF_9671	CMAS	0,8330908392	0,8321350861	0,0009557530784	24,1412006	0,0001187902981
AF_9733	RAB28	0,02591849541	0,02391896919	0,001999526222	16,38930276	0,002917141322
AF_9766	NCAPG	0,003322747161	0,002163973817	0,001158773344	10,38238793	0,03703197152
AF_9868	KIAA1468	0,3420213356	0,3409977429	0,001023592702	10,52841969	0,03465890023
AF_9957	RASSF6	4,081119081	4,079204661	0,00191442074	18,80027671	0,001033089575
AF_9961	ANKRD17	0,3210492809	0,3201732193	0,000876061543	22,69096804	0,0002206534446
AF_9984	XPO4	0,006674037121	0,005944214685	0,0007298224363	21,91195547	0,0002816939081
AF_10151	CTNNBL1	0,5995043483	0,5982384568	0,001265891573	17,88148044	0,001515023888
AF_10182	PROX1	0,3310229367	0,3305490473	0,0004738893888	34,93819995	0,000001291340669
AF_10189	FLVCR1	1,160623638	1,159496641	0,001126996925	20,18185253	0,0005877013744
AF_10200	INTS7	0,005519745352	0,003355153893	0,002164591459	13,53318841	0,009547827175
AF_10205	TRAF5	0,9876421331	0,9852680321	0,002374100997	21,63483425	0,0003147808264
AF_10208	HHAT	0,07034864676	0,06917693409	0,001171712673	25,80326173	0,00005818049088
AF_10241	HN1	0,2806232341	0,2805158291	0,0001074049677	11,55207272	0,02224746507
AF_10274	MYCBPAP	0,4444195016	0,4434113147	0,001008186909	17,97118255	0,001480514134
AF_10344	HOMER3	1,606971453	1,605087682	0,001883771142	20,01990655	0,0006236505749
AF_10462	NCLN	1,50075604	1,49811764	0,00263839976	9,826268444	0,04648243294
AF_10487	FCHO2	0,03962909106	0,03745220037	0,002176890685	28,82779723	0,00001624661897
AF_10514	PCDH11X	0,005605937955	0,005599075267	0,000006862688305	40,51142824	0,0000001199216584
AF_10519	SMARCB1	0,1795297329	0,1775822607	0,001947472196	22,22507981	0,0002544694618
AF_10646	RALGPS1	0,1207896812	0,1171060447	0,003683636471	11,25503775	0,02529419107
AF_10666	GUF1	0,02711422799	0,02040706234	0,006707165651	15,23696085	0,004735497959
AF_10728	MYOCD	1,392528108	1,390967346	0,001560762587	9,799544073	0,04689719347
AF_10832	POC5	4,685859869	4,684948015	0,000911853748	23,64804072	0,0001484958218
AF_10875	FAM189A1	0,00678089685	0,006010508014	0,0007703888352	15,9491091	0,003538791404
AF_10879	TARS3	1,685980369	1,684493039	0,00148733002	15,88845053	0,003605374903
AF_10888	MYO1E	0,3490732525	0,3480539258	0,001019326722	19,9444929	0,0006412721739
AF_10980	IQCH	0,06409245137	0,06260364698	0,001488804388	35,62799999	0,0000009934392667
AF_11037	OVOSTATIN	0,01068600381	0,00978594134	0,0009000624739	24,98791252	0,00008385862313
AF_11098	CRB2	0,05078646755	0,04743731574	0,003349151809	44,08538814	0,000000036708022221
AF_11099	STRBP	0,005906829765	0,004926662138	0,0009801676271	13,33619199	0,01034666184
AF_11108	ZBTB26	0,005145463897	0,005137646074	0,000007817822596	15,14047698	0,004869908448
AF_11206	COQ6	0,03072372939	0,02753137104	0,003192358352	17,29328301	0,001938818244
AF_11306	SPATA7	0,154469859	0,1514558425	0,003014016512	22,92118804	0,0001995063419
AF_11340	FBLN5	0,01399570927	0,01214473967	0,001850969608	10,2330742	0,03893079788
AF_11365	SPI	0,01256683467	0,01076210572	0,001804728947	13,0311361	0,01184437918
AF_11394	PKDCC	0,06646315002	0,06639261544	0,00007053458193	10,23745471	0,03893079788
AF_11417	DYNC1H1	0,04078894646	0,04020711951	0,0005818269522	20,22921994	0,000577192499
AF_11476	CEP170B	0,01953696314	0,01848662137	0,001050341773	39,5591675	0,0000001821791806
AF_11477	PLD4	1,975296584	1,973675324	0,001621260392	29,66419432	0,00001143422481
AF_11502	ACOT12	0,9144216886	0,9136783643	0,0007433242721	20,79618341	0,00045706111158
AF_11533	ANKRD27	0,3692949929	0,368360055	0,0009349379122	10,00309342	0,04291507833
AF_11577	TBK1	1,779249672	1,777257898	0,001991774174	23,40579309	0,0001612436106
AF_11606	ACSBG2	1,029856358	1,02816679	0,001689567121	17,83250184	0,001544704043
AF_11735	UBASH3B	0,001752235249	0,001648236651	0,0001039985983	10,2294175	0,03893079788
AF_11796	CYFIP1	0,272226405	0,2718186587	0,0004077463288	32,91868003	0,000002960691945
AF_11804	PLCXD1	0,2166149969	0,2111653288	0,005449668119	14,47459566	0,006546817455
AF_11820	ASMT	9,637590343	9,633554807	0,004035535762	31,52550751	0,000005101655625
AF_11822	ZBED1	0,006095513195	0,005113829529	0,0009816836655	14,22564464	0,007197335296
AF_11851	PLEKHG1	0,01280637439	0,01147481866	0,001331555722	17,88003358	0,001515023888
AF_11892	SERAC1	0,01682154245	0,01633024826	0,0004912941936	31,85036482	0,0000044346477 67
AF_11938	MAP3K5	0,3513339573	0,3502085801	0,001125377227	19,61666904	0,0007243276299
AF_11986	ENPP3	0,6782808561	0,6778047243	0,00047613181	47,00232291	0,00000001394050434
AF_12020	OTOA	0,1541432555	0,1535220303	0,0006212251944	15,3592274	0,004527384238
AF_12034	CUNH16orf52	0,1305708194	0,1305001146	0,00007070471932	19,05284497	0,0009234621018
AF_12116	CLEC16A	0,00499552975	0,004265390717	0,0007301390327	16,35552869	0,002955017286
AF_12130	DIS3	0,002423790662	0,002422649933	0,000001140729739	11,42668662	0,02344502645
AF_12166	KIF25	0,021343334	0,02072893779	0,0006143962098	24,22001669	0,0001170622884
AF_12181	ERMARD	0,08453182397	0,08330574262	0,001226081349	21,02048312	0,0004132307583
AF_12214	SIPA1L2	0,001706956146	0,001093460181	0,0006134959649	11,11927941	0,0267317854
AF_12236	LGALS8	5,318558341	5,317586561	0,0009717796977	38,27099521	0,0000003376870397
AF_12263	DENND2D	3,05952287	3,05572448	0,003798389525	13,34794691	0,01033031961
AF_12309	IPO9	0,8042759515	0,8031402014	0,00113575007	30,07490732	0,000009723234942
AF_12310	KDM5B	0,3279725277	0,3273010628	0,000671464863	42,6317459	0,00000006841784304
AF_12349	RNGTT	0,006748648503	0,006740489868	0,000008158634841	13,56940592	0,009405106785
AF_12459	GPNMB	0,2668340778	0,2575612884	0,009272789413	16,57799037	0,002674918073
AF_12465	RAPGEF5	0,01889986361	0,01839942191	0,0005004416999	26,98885371	0,00003694608098
AF_12469	ABCB5	0,003787852021	0,003657991636	0,0001298603848	20,07487745	0,0006133313711
AF_12473	TWISTNB	0,008063425918	0,008059197212	0,000004228706331	10,91468463	0,02902933432
AF_12524	ATP11C	0,0118689251	0,01061775949	0,001251165609	10,31173321	0,03790123428
AF_12614	NCOA3	1,044490885	1,043192859	0,001298026163	48,44038076	0,000000007463786289
AF_12625	SLC25A3	364,6342247	364,6013479	0,0328767763	15,28595127	0,004644702801
AF_12725	OPA1	0,3024994997	0,3016890638	0,0008104358947	9,769219148	0,04750345054
AF_12757	TFRC	0,005932000962	0,003732076954	0,002199924008	12,66167202	0,01376304684
AF_12764	LRCH3	0,7834875496	0,7814672727	0,002020276954	22,1414702	0,0002618633167
AF_12765	IQCG	0,5251012186	0,5242630146	0,0008382040036	23,16964402	0,0001801480674
AF_12795	ABCC5	0,001988698841	0,001508580674	0,0004801181673	11,8363059	0,01981558631
AF_12855	FNDC1	0,1200977544	0,1194923257	0,0006054286713	23,88316416	0,0001341136431
AF_12857	FANCL	0,01104571877	0,01103928792	0,000006430845254	21,69188788	0,0003097148811
AF_12861	PUS10	0,1353225192	0,1340656052	0,001256914046	22,19162487	0,0002570543295

AF_12863	KIAA1841	0,01704920592	0,01388639616	0,003162809757	18,32559323	0,001274147333
AF_13012	SLC9A8	0,05279039255	0,05057658751	0,002213805035	20,6539129	0,000488043873
AF_13038	MEST	0,01860537543	0,01498521545	0,003620159975	10,77392452	0,03092911775
AF_13198	NAE1	0,03636405251	0,0312353947	0,005128657805	17,10002466	0,002109537088
AF_13199	CCDC79	0,01323731664	0,009782123075	0,003455193565	15,14098517	0,004869908448
AF_13214	TRAT1	2,851200333	2,849313836	0,001886497255	18,59581788	0,001128376368
AF_13261	MAP3K7CL	3,318271913	3,316266302	0,002005610787	20,60886772	0,0004934399512
AF_13270	TIAM1	0,01886357836	0,01845503992	0,0004085384402	16,61127788	0,002652423538
AF_13275	URB1	0,01896577496	0,01822561061	0,0007401643585	26,39976405	0,00004575316475
AF_13392	EFHC2	1,91030806	1,908301679	0,002006380286	20,03440441	0,0006225083081
AF_13473	CRB1	0,1598558885	0,1572670449	0,002588843597	18,63820068	0,001109931475
AF_13632	DCHS2	0,01157540067	0,01157276165	0,000002639015996	11,51891867	0,02249999658
AF_13648	GLRB	0,01202984642	0,01056643496	0,001463411457	12,88656039	0,01254748328
AF_13668	PDZRN3	0,008537851061	0,008534366134	0,000003484926135	28,76720785	0,00001636278065
AF_13769	RGS9	0,006021303633	0,004998317552	0,00102298608	16,15572942	0,00322796075
AF_13777	ABCA10	1,13778764	1,136149473	0,001638166458	35,36509908	0,00000106393119
AF_13824	PPARA	0,01131840318	0,01078165592	0,0005367472563	14,17951193	0,007288902697
AF_13926	GPR176	0,1215294414	0,1214882902	0,00004115122313	24,33714381	0,0001122670847
AF_14001	SLC4A1AP	6,570386418	6,567875788	0,002510630009	20,40355977	0,0005438954367
AF_14010	KCNK16	1,538622147	1,537485662	0,001136485816	17,64933502	0,001678651887
AF_14025	SLC35B2	0,0061324904	0,005356154334	0,0007763360664	10,08545863	0,04156478997
AF_14042	EPHX1	0,0521990067	0,04907549785	0,003123508855	11,7272589	0,02070406814
AF_14101	UBE3B	2,473793829	2,472426632	0,001367197479	26,89042217	0,00003840729621
AF_14105	FOXN4	2,251175117	2,249105152	0,002069964979	26,59146545	0,00004325480657
AF_14110	SVOP	1,260238329	1,258970696	0,001267633098	35,84345646	0,0000009115943046
AF_14164	KNTC1	0,005241578998	0,004435929073	0,0008056499247	14,93795178	0,005334210714
AF_14165	RSRC2	0,0115592052	0,01072946168	0,0008297435187	15,7489292	0,0038143613
AF_14199	P2RX4	6,142745091	6,140162724	0,00258236709	26,24005587	0,00004861612489
AF_14240	DNAH10	1,213024613	1,211230778	0,001793835501	14,8797613	0,00545342123
AF_14293	SIN3B	0,02822956777	0,02698157918	0,001247988595	36,62686662	0,0000006853780293
AF_14300	CHERP	0,01688180952	0,01645649813	0,0004253113872	12,67719549	0,01369625122
AF_14331	CHAF1A	0,004575334716	0,003729171772	0,0008461629439	12,96404653	0,01214623461
AF_14343	CEP192	6,154238662	6,149546456	0,004692206798	18,74885738	0,001054987138
AF_14621	LOC103903316	0,003998493062	0,003371692487	0,0006268005754	10,75989083	0,03106953792
AF_14622	valA	0,275089235	0,2747380807	0,0003511543192	20,10501314	0,0006074542916
AF_14728	CHCHD6	14,11627194	14,10496955	0,01130238587	35,99578438	0,0000008638019293
AF_14772	ARF4	0,02031024575	0,02029828522	0,000011960533032	10,12302513	0,04088209977
AF_14783	FAM208A	0,09765536521	0,09743287823	0,0002224869763	12,70514717	0,01358518138
AF_14795	CACNA1D	2,014551314	2,012220591	0,002330722913	41,34381426	0,00000009315064475
AF_14803	RHOA	0,0905631409	0,09048798405	0,00007515684962	10,63463008	0,03293118247
AF_14805	CBFA2T2	0,02738147276	0,02314399054	0,004237482218	14,16621036	0,007313277971
AF_14808	EDEM2	0,9952138131	0,993207423	0,002006390082	11,23011994	0,0255461708
AF_14877	SUMO1	0,4851387184	0,4850287723	0,0001099460721	12,05013368	0,01793999278
AF_14884	TRAK2	0,002509145438	0,001810790373	0,0006983550655	11,03558264	0,02773965306
AF_14894	FAM126B	0,008003581223	0,007482301797	0,0005212794258	13,87806384	0,008399359311
AF_14895	ORC2	0,0268451121	0,02683086393	0,00001424816832	41,51900488	0,00000009188883773
AF_14923	GTF3C3	0,2500293402	0,2486308942	0,001398445982	17,13629781	0,002080219925
AF_14936	INPP5B	0,4827641573	0,4816089587	0,001155198563	22,60934945	0,000224920156
AF_14953	STK40	0,005088646176	0,00434126912	0,0007473770558	10,72968295	0,03146691585
AF_14968	AGO3	0,01866136335	0,01754160411	0,001119759231	11,16372981	0,02627703608
AF_15036	YARS	0,009833433495	0,007559311133	0,002274122362	11,52724447	0,02246667711
AF_15280	TRPM8	0,2532695124	0,2523353915	0,0009341209363	24,31741812	0,0001124229825
AF_15307	MREG	0,1070936174	0,1070283721	0,0000652453506	9,931593102	0,04428207066
AF_15354	AGXT2	0,01022199828	0,009563731156	0,0006582671236	14,21997271	0,007197335296
AF_15375	PDZD2	0,0779070751	0,0775053774	0,0004016977047	22,95031747	0,0001984538852
AF_15382	APH1A	0,0982923112	0,09658492472	0,001707386476	17,54142625	0,001753367932
AF_15421	TLE4	0,2576819117	0,2572341749	0,0004477367506	10,33330952	0,03775007209
AF_15426	VPS13A	0,03807496994	0,03713535852	0,0009396114258	45,66387865	0,00000002140975404
AF_15504	PACS2	0,1367160786	0,1363119338	0,0004041448222	26,16736629	0,00004994132816
AF_15518	ACOX3	0,02429256467	0,02428787386	0,000004690809109	20,31203329	0,0005630121111
AF_15608	USP6NL	0,00916706987	0,006771909296	0,002395160573	13,57588248	0,009404508786
AF_15646	HSPA14	0,05924882961	0,05635405767	0,002894771942	10,98829738	0,02830494523
AF_15710	FAM179A	0,01538644056	0,01437337311	0,001013067449	21,90581397	0,0002816939081
AF_15763	DSC1	0,005581450024	0,005180416633	0,0004010333913	12,52808445	0,01459267036
AF_15828	WWC1	0,03445973961	0,03373306233	0,0007266772826	23,59158114	0,0001510705047
AF_15869	AFAP1L1	0,9440083286	0,9427131802	0,00129514836	22,45328945	0,000233533378
AF_15924	UIMC1	0,02073651934	0,01956924046	0,001167278878	12,87378737	0,01259051098
AF_16096	SPOCK1	0,09207915351	0,08871949116	0,003359662355	14,34198351	0,006926075861
AF_16232	RAC1	0,012486128	0,01181707786	0,0006690501391	22,9466486	0,0001984538852
AF_16233	AQP11	8,697188071	8,694510574	0,002677496624	20,85154359	0,0004471587929
AF_16351	FOLH1B	0,03287269411	0,03195599949	0,0009166946186	21,63543681	0,0003147808264
AF_16447	EIF2AK1	0,03051825929	0,03045053435	0,00006772494567	13,58446374	0,009394039244
AF_16500	GRIFIN	0,0458192666	0,04581227938	0,000006987224221	19,18781848	0,0008672000429
AF_16502	TTYH3	0,09239911559	0,09231317363	0,00008594196221	42,3911723	0,00000007350331379
AF_16567	TOM1L2	0,4674341354	0,4667503793	0,0006837560419	10,004131	0,04291507833
AF_16568	DRC3	0,4150412829	0,4143344688	0,0007068141427	15,90768	0,003584771225
AF_16581	FAM83G	1,306908631	1,30438457	0,002524060496	41,57315335	0,00000009188883773
AF_16751	RASA2	0,01030588393	0,009700126009	0,0006057579245	12,10183038	0,01757770101
AF_16756	GK5	0,6507002745	0,6498188393	0,0008814352149	11,61919467	0,0216244603
AF_16776	GRIK4	2,669365072	2,667329241	0,002035830826	36,59851399	0,0000006853780293
AF_16779	TMEM136	0,02006504086	0,01877310066	0,001291940198	15,79180384	0,003755474746
AF_16783	TRIM29	0,02194445052	0,02065662214	0,001287828375	17,92918951	0,001491480849
AF_16790	FGD6	0,01902910961	0,01661895308	0,002410156534	19,54338983	0,0007476379241
AF_16799	LTA4H	0,02098161647	0,01911363397	0,001867982502	31,83374595	0,000004434647767
AF_16828	LMOD2	3,78860204	3,785384772	0,003217267648	18,54218591	0,001153658776
AF_16835	IQUB	0,3910552716	0,3903328286	0,0007224430584	19,87043829	0,0006544956179
AF_16836	ATP6AP1	0,147821687	0,1455122342	0,002309452808	35,50898966	0,000001010545493
AF_16867	MDFIC	26,60651394	26,60174685	0,004767087208	18,89987147	0,0009874728638
AF_16937	PKN3	0,00501285254	0,003824853164	0,001187999376	11,91091436	0,01913714523
AF_16944	KYAT1	0,09659550055	0,09074181815	0,005853682406	31,06970091	0,000006233130315
AF_17026	ADAMTS13	11,86514811	11,86122736	0,003920746695	37,10190558	0,0000005746430133
AF_17113	IFFO1	0,02146014097	0,02145037573	0,000009765239648	11,73360023	0,02069539674
AF_17152	ZYX	1,120612987	1,119336387	0,001276600268	21,49276918	0,0003339329224
AF_17339	STIL	0,09317566221	0,09310628826	0,00006937395353	9,952216005	0,04392789491
AF_17367	CC2D1B	0,005296772346	0,004446239018	0,0008505333279	13,85139554	0,008456593404
AF_17377	ZYG11B	0,01372164763	0,01282870213	0,0008929454923	17,47765812	0,001790203317
AF_17412	USP24	0,06424886691	0,06366709996	0,0005817669493	12,96304662	0,01214623461
AF_17441	ALG6	0,05709463351	0,05663136903	0,0004632644736	36,45413677	0,0000007192508452
AF_17459	SUDS3	70,38629047	70,37249321	0,01379725802	46,06709203	0,00000002041132013
AF_17490	TM9SF2	0,008754830559	0,006922652649	0,001832177911	16,59378138	0,002664766814

AF_17538	FAM162B	0,6810530544	0,6746913071	0,006361747253	15,07192491	0,005022567467
AF_17593	STXBP4	0,007275735737	0,005531611466	0,001744124272	12,09287541	0,01760713523
AF_17641	MILR1	0,01563041921	0,01426280171	0,001367617498	11,82657272	0,01985738016
AF_17652	PSMD12	0,01214793827	0,01024474829	0,001903189987	16,25940115	0,003076318981
AF_17653	HELZ	0,2350236519	0,2343826218	0,0006410300136	23,58216135	0,0001510705047
AF_17826	LIMD2	16,29431379	16,29079534	0,0035184548	21,92703059	0,0002816939081
AF_17836	CDC27	0,03033322529	0,02892258853	0,001410636764	16,78791126	0,002447333112
AF_17906	ARR3	0,06929399564	0,06736850641	0,001925489224	10,50859622	0,03483108728
AF_17913	FNDC3A	0,01323979093	0,01203938489	0,001200406048	14,86787641	0,005466557203
AF_17945	POF1B	0,003764213756	0,003762620378	0,000001593377149	10,25951784	0,03870671253
AF_17947	FNDC3B	0,004914860153	0,004024141611	0,0008907185417	12,71984003	0,01352527589
AF_18159	ARSK	0,008774920918	0,008168778641	0,0006061422771	13,6551254	0,009124118166
AF_18225	WEE2	0,02691658058	0,02277707934	0,004139501242	19,8839262	0,0006535336967
AF_18243	SVOPL	0,1007933191	0,09694915952	0,003844159566	19,84147914	0,0006605917169
AF_18269	TH	0,009722248323	0,009714519737	0,000007728586594	22,29008073	0,00024796371
AF_18295	CRNKL1	0,5838243385	0,5832332323	0,0005911061323	33,73969962	0,000002179976614
AF_18346	PNPT1	0,01501970879	0,0136120458	0,001407662992	14,41684037	0,006716466652
AF_18631	ABCC8	0,01044263352	0,009131848594	0,001310784928	31,96411953	0,000004342938808
AF_18668	ST5	0,004954102911	0,003450331129	0,001503771781	10,0328922	0,04247787112
AF_18676	TMEM41B	0,01280678167	0,01280085299	0,000005928681768	20,24308439	0,000577192499
AF_18686	RNF141	3,57930238	3,577819496	0,001482883981	36,20685391	0,0000007951871368
AF_18738	ELP4	2,534276948	2,533197341	0,001079607515	10,26397757	0,03870671253
AF_18857	CELSR1	0,04036886118	0,03914367914	0,001225182043	22,48491382	0,0002327211435
AF_18889	ATXN7L1	0,005309448674	0,004791370254	0,0005180784204	13,50647971	0,009646481236
AF_18986	TRAF3IP2	0,3127519785	0,3120624634	0,0006895151168	11,27010845	0,02516972731

Table S3. Candidate genes predicted to be under positive selection by aBSREL models in A. patagonicus (FDR < 0.05)

Gene ID	Gene name	Full adaptive model	Full adaptive model (non-synonymous subs/site)	Full adaptive model (synonymous subs/site)	LRT	Adjusted p-value
AF_57	DDB1	0,005937256786	0,003816996435	0,002120260351	12,65945415	0,02702669133
AF_214	NFX1	0,1215834258	0,1214902985	0,00009312724318	28,31927794	0,00003462003615
AF_218	TNS3	0,06475598911	0,06245494489	0,002301044226	41,6297189	0,0000001076302137
AF_285	WDR35	0,004366266236	0,004103820558	0,0002624456784	12,2118068	0,03224737962
AF_341	OPN4	18,64797477	18,64446331	0,003511456269	19,76381782	0,001144790352
AF_536	CASQ2	2,30682873	2,30593735	0,0008913794655	44,9319042	0,00000003088874603
AF_791	ABCA3	0,0228990592	0,02096447155	0,001934587652	20,15170574	0,0009675710921
AF_1084	DNAJC13	0,1260427786	0,1255885172	0,0004542613353	21,1763807	0,0006172599229
AF_1498	AIM1	0,4318026211	0,4311080841	0,0006945370328	31,98630951	0,000008858236032
AF_1638	NAT10	1,295176535	1,293324391	0,00185214437	12,557389	0,02801770992
AF_1698	TMEM248	0,04155954323	0,03531877362	0,006240769616	17,30733287	0,003250466448
AF_1727	CEP126	0,0106435014	0,009887081362	0,0007564200379	17,51882338	0,00303242132
AF_1784	SLC15A1	0,1155202955	0,1138052254	0,001715070054	19,08108845	0,001495208243
AF_1795	NUDT14	7,346380748	7,342040484	0,004340263939	12,72341343	0,02637023142
AF_1885	TGS1	0,9212590184	0,9190020599	0,002256958548	20,16956129	0,0009675710921
AF_2026	SCLT1	0,0362501338	0,03573915626	0,0005109775401	14,04905263	0,01435926029
AF_2078	DAGLA	0,3889938515	0,3877082609	0,001285590582	24,11500136	0,0001907040367
AF_2094	ECE1	0,1768274972	0,1732710105	0,003556486673	14,12523029	0,013905064
AF_2274	CCDC80	0,8409098272	0,8394889408	0,001420886431	18,94700597	0,001586524715
AF_2331	ZPLD1	0,01138594996	0,01137010762	0,00001584234518	17,31808283	0,003250466448
AF_2397	SOS2	0,01197343273	0,01163223884	0,0003411938901	31,07172932	0,00001211700761
AF_2403	POLE2	0,4648494109	0,4640828638	0,0007665471081	11,31862209	0,04626827607
AF_2501	CBX7	0,1115915889	0,1115746732	0,00001691570334	22,90411252	0,0003135345715
AF_2516	TNRC6B	0,09650260734	0,09176857126	0,004734036083	31,8771465	0,0000090729082096
AF_2550	MEI1	0,8980263035	0,8978212401	0,0002050634317	28,8349229	0,00002866744234
AF_2630	ATP7B	5,385564968	5,381311937	0,004253030723	15,10593516	0,008874247141
AF_2637	CKAP2	0,07329966265	0,07142948562	0,001870177034	21,20869589	0,0006132203986
AF_2702	ALOX5AP	12,85255407	12,85021811	0,002335959454	29,51480309	0,00002289533282
AF_2715	FLT3	86,9053855	86,89511285	0,01027265134	39,84005141	0,0000002408430043
AF_2734	NUP58	0,646803744	0,6462148664	0,0005888775714	27,62047299	0,00004685178587
AF_2837	BAZ1A	0,004478273113	0,00315739915	0,001320873963	19,43910981	0,001301960977
AF_2954	JMJD1C	0,004825940621	0,004214545055	0,0006113955661	11,87052306	0,03707797701
AF_3110	DYSF	0,0190449074	0,01211991786	0,006924989539	26,02817922	0,00008546331066
AF_3231	ZDHHC1	0,02957741287	0,02876729329	0,0008101195774	11,94496346	0,03609706339
AF_3359	COL3A1	0,06392492151	0,06379783217	0,0001270893435	12,65159143	0,02702669133
AF_3417	CNTN5	0,008446959317	0,007104054928	0,001342904389	11,4932885	0,04323623222
AF_3573	PSMA5	0,1272491479	0,1271773918	0,00007175613363	12,34330626	0,03057114021
AF_3806	PPP1R21	2,310173053	2,307885904	0,002287148705	16,96122623	0,003840600344
AF_3816	SYCP1	0,5968355815	0,5960697152	0,000765866358	11,87359714	0,03707797701
AF_3839	MAGI3	0,02640441398	0,02560561342	0,00079880056	17,48113501	0,003044648715
AF_3906	SLC26A9	0,8401014525	0,8379518741	0,002149578351	25,49443551	0,000104307093
AF_4207	CACNB4	0,8133276157	0,8130654314	0,0002621843164	61,37504353	1,92E-11
AF_4405	COPB2	0,04975699746	0,04812351352	0,00163348394	15,31447996	0,008095727715
AF_4494	LNX2	0,00999752087	0,00882379806	0,00117372281	27,79587967	0,00004371597081
AF_4545	GOLGA4	0,01043916717	0,009585539135	0,0008536280333	20,3811921	0,0008780463366
AF_4651	CFI	0,009802179791	0,008954659774	0,0008475200172	13,52973498	0,01820145221
AF_4689	ADGRL3	11,28889889	11,28473103	0,004167859745	35,8354935	0,000001525397869
AF_4724	C3	0,1210142544	0,1209499008	0,00006435357467	12,08231755	0,03404816296
AF_4869	CAPN8	0,02306175111	0,02212865713	0,0009330939781	25,51920655	0,000104307093
AF_4955	OGFR	0,2139703472	0,2100508983	0,003919448902	32,65089359	0,0000065534639 65
AF_4961	CABLES2	10,29838566	10,29568735	0,002698310482	38,17475937	0,0000005102870979
AF_4973	TAF4	0,07434304415	0,0720697597	0,002273284451	41,83413208	0,0000001020140432
AF_5019	CA12	0,04989817566	0,04845814259	0,001440033068	14,4427141	0,01207669995
AF_5091	DPM1	0,9940052605	0,9938876027	0,000117657811	54,48970093	0,00000000004013031574
AF_5241	PTPN1	0,3932645643	0,3926030608	0,0006615034975	22,26084629	0,0003879292634
AF_5347	CHFR	0,0466541824	0,04511326915	0,001540913249	18,22301085	0,002194349056
AF_5419	CD3D	0,3261298979	0,3102344846	0,01589541327	16,01016986	0,005940428242
AF_5565	ZNF654	0,1488430892	0,1487683151	0,00007477401938	28,94230134	0,00002774271789
AF_5570	CD200	15,73677197	15,73052902	0,006242956027	19,81071991	0,001128026379
AF_5615	GAK	0,08683904118	0,08571334788	0,001125693307	52,87304854	0,00000000008112921834
AF_5865	LY75	3,512209313	3,511070099	0,001139213737	18,55725516	0,001874189439
AF_5896	SLC23A3	0,01875259688	0,01720124022	0,001551356663	11,16096182	0,0496015022
AF_5947	TMEM169	0,5468122407	0,5465149977	0,0002972429473	34,96450476	0,000002275769571
AF_5970	TFCP2L1	0,005271244559	0,005267530807	0,000003713752527	19,98777543	0,001041240433
AF_6157	ARHGAP21	0,9851172736	0,9849471124	0,0001701612221	29,33093431	0,00002416332797
AF_6166	DNAJC13	0,8634884026	0,8619839937	0,0015044089	24,86767695	0,0001360173224
AF_6198	COL4A5	0,00708270317	0,005500055117	0,001582648053	12,33965266	0,03057114021
AF_6340	PPIF	5,946557408	5,944009458	0,002547950001	16,11612038	0,005671162177
AF_6369	CDH17	0,4458470753	0,4453726632	0,0004744120952	19,52918736	0,001265805263
AF_6433	FAM83D	0,6181919159	0,6173526063	0,0008393096024	18,68390608	0,001796307321
AF_6452	FAM49A	0,1245999598	0,1231064365	0,001493523275	11,4470776	0,04402981684
AF_6454	DDX1	0,084665123	0,0816170281	0,003048094904	49,54142743	0,0000000003909759153
AF_6591	FAM188A	3,043542224	3,041021535	0,00252068987	44,03719957	0,00000004512560925
AF_6649	SLC17A5	0,006489439712	0,006478549352	0,00001089035999	14,64739178	0,01103452145
AF_6680	UGGT2	0,05938174984	0,05746137467	0,001920375175	19,20868438	0,001425530652
AF_6686	PDC	8,822326155	8,819536929	0,002789226166	30,78681608	0,00001326053273
AF_6753	TOR1AIP2	0,747478393	0,7468068629	0,0006715300893	23,51808116	0,0002502762899
AF_6791	CRYD	0,0339350755	0,03138360271	0,002551472788	12,28429642	0,0312621719
AF_6901	TNFSF13B	0,1753293474	0,1689726398	0,006356707562	19,16958051	0,00144190114
AF_7006	EIF5B	0,06438983034	0,06411205399	0,0002777763449	13,73440845	0,01651796065
AF_7288	TMOD3	6,680221854	6,677833329	0,002388525235	17,38557355	0,003171026065
AF_7443	EXOC3	0,01992274379	0,01842913429	0,001493609501	31,68155704	0,0000097131319 36
AF_7486	CNDP1	13,77947704	13,77547716	0,003999873677	12,59792596	0,02760851487
AF_7552	DPEP2	2,477066367	2,476090856	0,0009755113167	21,6348358	0,0004999456272
AF_7615	ENDOV	0,2092887314	0,2076667221	0,001622009238	30,67394	0,00001359944136
AF_7639	SLC30A10	0,4408796802	0,4321053288	0,008774351395	30,63671497	0,00001359944136
AF_7884	SMARCD3	0,04219367715	0,04086487596	0,001328801196	24,7717462	0,0001389685219
AF_7903	TMEFF2	0,114813153	0,1134745615	0,001338591459	15,62412941	0,007067866967
AF_8067	USP10	0,008389001634	0,006990045023	0,001398956611	19,30746193	0,001379303658
AF_8091	GCSH	114,2037949	114,1987998	0,004995117907	26,6949493	0,00006938885791
AF_8125	CENPU	1,276125107	1,27499598	0,001129127848	12,50645047	0,02858347334
AF_8128	CFAP97	1,155060648	1,15337955	0,001681097249	19,49854761	0,001274455305
AF_8314	RTF1	0,01616115098	0,01612500485	0,00003614612483	11,54430061	0,04235534591
AF_8336	PLA2G4E	0,1612694654	0,1608109976	0,0004584678594	23,49716465	0,0002502762899
AF_8402	SYNPO2L	0,08197724841	0,07866437743	0,003312870972	62,95063231	1,04E-11
AF_8533	RBP4	2,363669703	2,362139384	0,001530318449	22,6432056	0,000345360127
AF_8573	GOT1	2,422247909	2,419506566	0,002741343911	13,43529678	0,01886002144
AF_8702	CCDC186	0,01002221987	0,007353311669	0,0026689082	14,16759755	0,01378376332

AF_8752	WDR11	0,1461561642	0,1438428488	0,002313315401	23,3415679	0,0002608237381
AF_8774	ADAM9	0,006535936278	0,005501729257	0,001034207022	11,75965813	0,0383904482
AF_8799	BCCIP	0,3273939273	0,3273036919	0,00009023545302	36,04471603	0,000001426414755
AF_8923	TRPV1	0,3330752988	0,3325756602	0,0004996386053	18,55226595	0,001874189439
AF_8924	TRPV3	0,02583313376	0,02204617717	0,003786956594	22,38489193	0,0003762666613
AF_8965	RPS6KB1	0,1695594819	0,1694306037	0,000128878211	26,2429858	0,00008020581505
AF_9307	OSBPL6	0,7087270231	0,7079967564	0,0007302666988	43,34042205	0,00000005330741007
AF_9437	HSDL2	9,831365043	9,829351684	0,002013358979	26,54191774	0,00007242736882
AF_9565	KCNC2	0,0385876983	0,03853338366	0,00005431463228	11,85111612	0,03707922591
AF_9572	KRR1	0,06214049059	0,06100617066	0,001134319925	12,94767748	0,02368938086
AF_9658	SLCO1C1	16,19914585	16,19393326	0,005212583656	29,30452619	0,00002416332797
AF_9667	ETNK1	1,927112457	1,925392604	0,001719852722	21,72439402	0,0004826945679
AF_9693	ASUN	0,01328404667	0,01324965872	0,00003438794802	15,88288159	0,006289965573
AF_9717	SCUBE1	0,02071437324	0,01940862847	0,001305744772	15,5836303	0,007164986827
AF_9727	PARVG	2,765459629	2,765208466	0,0002511632831	56,28562407	0,000000000209898955
AF_9977	MICU2	0,03369875291	0,02839806267	0,005300690241	14,51478196	0,01172091522
AF_10112	TBCD	1,473139253	1,47150554	0,001633712746	22,82651015	0,0003222417228
AF_10515	DIAPH2	0,01093249756	0,009600196701	0,001332300855	17,69152841	0,002822662853
AF_10527	DDX51	0,9378915608	0,9369184145	0,0009731462513	41,91745476	0,0000001020140432
AF_10579	CUNH12orf43	0,3067495143	0,2972392045	0,009510309743	13,26596401	0,02029717954
AF_10588	RNF10	0,4436387246	0,4426528457	0,0009858789244	19,24383315	0,001412246449
AF_10741	GALNT1	0,05249943537	0,05242477381	0,00007466156625	26,81436721	0,00006650121601
AF_10753	FH	9,109293291	9,1050282	0,004265090676	18,66067928	0,001802974854
AF_10875	FAM189A1	0,009046161982	0,00822764125	0,0008185207323	11,24250753	0,04783957162
AF_10955	CTDSPL2	0,2586738739	0,2586127689	0,00006110493966	23,07278526	0,000291471255
AF_11068	PHC3	0,1043667433	0,1042776234	0,00008911992996	22,03450677	0,0004204624574
AF_11070	SKIL	0,03518970408	0,03470647056	0,0004832335171	29,50270627	0,00002289533282
AF_11086	GOLGA1	0,2649574131	0,2641556125	0,0008018005803	16,60798332	0,004522251395
AF_11248	GPATCH2L	0,5272433474	0,526413675	0,0008296723463	22,01912677	0,0004204624574
AF_11322	PSMC1	0,03608322582	0,0360594605	0,00002376531881	34,3364666	0,000003010004238
AF_11503	SSBP2	2,889753203	2,886874064	0,002879139425	17,4881303	0,003044648715
AF_11754	NCAPD3	0,004277597829	0,003415605102	0,0008619927269	26,02197676	0,00008546331066
AF_12130	DIS3	0,0265844982	0,02619206382	0,000392434382	16,47673277	0,004796872524
AF_12249	LMOD1	1,707949253	1,701768289	0,006180963946	23,7982331	0,0002206959013
AF_12343	UBE2J1	0,08217479008	0,08214845041	0,00002633967597	45,11499681	0,00000003088874603
AF_12372	NUDCD1	0,6474017875	0,64694441	0,00045737748	44,95066892	0,00000003088874603
AF_12511	TNMD	0,2487106681	0,2486043251	0,0001063429585	33,31839975	0,000004846414014
AF_12856	VRK2	0,7497148463	0,7490378137	0,0006770325602	28,29904589	0,00003462003615
AF_12981	LARP7	2,197850282	2,195756042	0,002094240158	20,87719327	0,0007037982417
AF_13114	HM13	4,578041731	4,574028862	0,004012869542	13,46165636	0,01872325621
AF_13220	IKZF1	0,4960034325	0,4952452513	0,0007581811588	22,42235955	0,0003762666613
AF_13259	CCT8	8,997546764	8,994026772	0,003519992113	29,11579565	0,00002598415929
AF_13272	SCAF4	0,4716506825	0,4707024848	0,0009481976874	25,78741068	0,00009343959957
AF_13292	SON	0,2346211319	0,223208303	0,01141282895	71,12322062	4,24E-13
AF_13382	CUNH21orf33	0,3145793075	0,3144467905	0,000132517005	25,21814982	0,0001172534902
AF_13596	SRP72	0,02897017681	0,02869252851	0,0002776483007	20,8297916	0,0007140178182
AF_13611	SH3D19	3,013563446	3,011529417	0,002034029729	11,78537422	0,0380948584
AF_13616	TMEM154	0,1356337981	0,130445048	0,005188750067	26,06113563	0,00008546331066
AF_13632	DCHS2	4,894872716	4,8853898	0,009482915365	69,9531184	5,30E-13
AF_13635	FGA	0,02268006755	0,02146318058	0,001216886974	22,5340846	0,0003607525068
AF_13646	CTSO	0,1320539404	0,1313226428	0,000731297664	24,88376938	0,0001360173224
AF_13984	TALDO1	1,556107402	1,55500152	0,00110588189	22,07535572	0,0004171007351
AF_14038	MRPL14	212,5345299	212,5193111	0,01521882664	43,74964358	0,00000004598401072
AF_14073	FEZ2	0,09731074674	0,09726594276	0,00004480397809	26,24128272	0,00008020581505
AF_14104	MYO1H	0,00950198102	0,008570097574	0,0009318834465	26,57231915	0,00007242736882
AF_14160	TCTN1	1,345407591	1,344529592	0,0008779987846	13,37036521	0,01937140393
AF_14262	PLIN3	20,67591391	20,66950574	0,006408168858	23,46570251	0,0002511206374
AF_14348	PSMG2	0,06679916162	0,06539511107	0,001404050548	15,8543084	0,006337645332
AF_14366	RALBP1	5,387881296	5,38587416	0,002007135953	31,50036849	0,00001033269931
AF_14429	YTHDC2	2,561987549	2,56095862	0,001028929029	55,79094924	0,000000000235356172
AF_14480	SEPT2	0,2494469262	0,2482360804	0,001210845809	23,20354862	0,00027620468
AF_14481	FARP2	0,01651661482	0,0123462809	0,004170333927	28,76834849	0,00002903567217
AF_14563	IWS1	0,01279021242	0,007457027493	0,005333184923	13,95422893	0,01496884219
AF_14715	UROC1	2,000200281	1,999309926	0,000890354592	43,78477964	0,00000004598401072
AF_15048	PPIE	0,5395259464	0,5394727057	0,00005324072776	22,79339639	0,0003239195542
AF_15058	PLEKHJ1	28,89213649	28,88745705	0,00467943846	26,25672607	0,00008020581505
AF_15086	COG3	0,5369797312	0,5364435958	0,0005361353999	22,40473377	0,0003762666613
AF_15093	CPB2	0,1281658489	0,1280722853	0,00009356360449	14,14427404	0,01385893593
AF_15149	MAP3K13	0,04484568375	0,04092908707	0,003916596681	16,77068296	0,004196674108
AF_15160	CLDN18	0,3560735771	0,3011926911	0,05488088593	27,19663795	0,00005587893808
AF_15461	PIP5K1B	0,02356036353	0,01945238511	0,004107978418	12,08549077	0,03404816296
AF_15487	ATF6	9,895815174	9,891212037	0,004603136747	18,13050126	0,002281115606
AF_15529	HTT	4,085212757	4,082987655	0,002225102845	11,58363008	0,04173899905
AF_15558	LETM1	0,02425854547	0,02142006516	0,002838480302	38,62300544	0,0000004246402379
AF_15623	CAMK1D	0,1351989764	0,1299962754	0,005202700978	11,8493755	0,03707922591
AF_15642	PRPF18	0,3555883672	0,3554802453	0,0001081218483	22,19351729	0,0003971292174
AF_15711	WDR43	0,2422732853	0,2417170643	0,0005562209572	11,96470837	0,03593343789
AF_15753	MEP1B	40,57629195	40,56990371	0,006388235188	70,11910142	5,30E-13
AF_15815	LCP2	0,01263990306	0,01261164507	0,00002825798433	24,83930242	0,0001361280674
AF_15900	BRD8	0,02948859475	0,02947729122	0,00001130353151	17,56778632	0,002981094261
AF_15959	NIPAL4	7,924006813	7,920356133	0,003650679888	40,45011765	0,0000001854923599
AF_16001	ATOX1	1,483983821	1,483858444	0,0001253774638	23,37064844	0,0002601791948
AF_16408	CALM2	0,6508186902	0,6497704949	0,001048195239	12,36721934	0,03048965803
AF_16658	PLK1	0,05615625732	0,054427514	0,001728743319	22,33124933	0,0003824248869
AF_16805	IL17REL	13,66595343	13,66206709	0,003886340809	16,1713283	0,005554379625
AF_17266	ATP6V0B	9,827781262	9,825749531	0,002031730657	22,30875436	0,0003826968443
AF_17368	ORC1	0,6165499894	0,6105594671	0,005990522262	30,89050128	0,00001292017622
AF_17412	USP24	0,002603733902	0,002164618161	0,0004391157419	13,82968248	0,01584038194
AF_17513	FAM122B	1,96981135	1,965795706	0,004015643797	11,36589829	0,04540622131
AF_17750	TBL1X	1,940134343	1,937430497	0,002703846587	19,6254957	0,001216502581
AF_17802	SCN4A	1,401322444	1,400118561	0,001203883282	15,25039157	0,008306136691
AF_17859	NLRX1	0,4709383127	0,4704025835	0,0005357291877	31,39029	0,00001061556539
AF_17888	HEPH	2,785244099	2,784785105	0,0004589934151	25,78459945	0,00009343959957
AF_18206	KIAA1551	0,003909069604	0,003178850134	0,0007302194707	20,93846262	0,0006889718636
AF_18227	AGK	6,087107303	6,084514386	0,002592916646	28,68015833	0,00002974152019
AF_18243	SVOPL	14,90774948	14,90194904	0,005800432981	14,73519546	0,01062585061
AF_18275	HSP90B1	0,1708640743	0,1707363217	0,0001277526462	15,533971	0,007297335766
AF_18433	MCPH1	0,1376417404	0,1376232463	0,00001849417126	27,31769241	0,00005354338037
AF_18616	LDHA	0,07563674084	0,07465889987	0,0009778409774	25,47923096	0,000104307093
AF_18858	TRMU	0,01637471855	0,01636704877	0,000007669783809	11,39126431	0,04505711381
AF_18859	GTSE1	24,75972041	24,75315282	0,006567588785	20,60242586	0,0007929791936

Table S4. Candidate genes predicted to be under positive selection by CODEML branch models in A. forsteri and A. patagonicus (FDR < 0.05)

Lineage	Gene ID	Gene name	Gene description	Likelihood Null Model	Likelihood Alternative Model	dN	dS	ω background	ω foreground	Adjusted p-value
A. forsteri	AF_4464	POLR1D	DNA-directed RNA polymerases I and III subunit RPAC2	-1073650663	-106755323	0,028517	0,036234	0,05409	0,787	0,04364804417
A. forsteri	AF_4207	CACNB4	Calcium voltage-gated channel auxiliary subunit beta 4	-2433426797	-2418667129	0,011123	0,008163	0,0308	1,3626	6,05E-05
A. forsteri	AF_4640	SGMS2	Sphingomyelin synthase 2	-297537272	-29663895l7	0,007836	0,003428	0,06737	2,2862	0,0050595612 94
A. forsteri	AF_4559	ASAP1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1	-6296899247	-6290455269	0,002263	0,002701	0,05016	0,8377	0,03520841431
A. forsteri	AF_4973	TAF4	TATA-box binding protein associated factor 4	-2421578647	-2411889524	0,0101	0,001486	0,05708	6,7988	0,00264625877 4
A. forsteri	AF_5071	TFAP2C	Transcription factor AP-2 gamma	-2007770972	-1995712463	0,009457	0,009425	0,02139	1,0033	0,0004080053271
A. forsteri	AF_5352	ZDHHC8	Zinc finger DHHC-type palmitoyltransferase 8	-627925246	-62729554	0,005392	0,00724	0,0939	0,7448	0,04000852649
A. forsteri	AF_5369	SEPTIN2	Septin-2	-232271942	-2307164315	0,00471	0,008085	0,03002	0,5826	3,81E-05
A. forsteri	AF_5373	UFD1L	Ubiquitin recognition factor in ER associated degradation 1	-2206514633	-2197794535	0,007362	0,009361	0,02117	0,7865	0,00630123608 3
A. forsteri	AF_5984	ADCY5	Adenylate cyclase 5	-4773199909	-4765414796	0,00906	0,016187	0,07598	0,5597	0,01267149056
A. forsteri	AF_6996	RPL31	Ribosomal protein L31	-1191786381	-1184957119	0,019555	0,031658	0,04402	0,6177	0,027956754
A. forsteri	AF_7017	GMEB1	Glucocorticoid modulatory element binding protein 1	-2945127035	-2939114773	0,003496	0,011999	0,01184	0,2914	0,04617422471
A. forsteri	AF_6879	AP2A2	Adaptor related protein complex 2 subunit alpha 2	-6843745376	-6835199098	0,002413	0,003286	0,01918	0,7344	0,00705873412 8
A. forsteri	AF_6831	WNT3A	Wnt family member 3A	-2339459998	-2319832376	0,01204	0,011843	0,01389	1,0166	1,42E-06
A. forsteri	AF_7895	GLS	Glutaminase	-3206570464	-3196047445	0,007334	0,003744	0,04538	1,959	0,00117305132
A. forsteri	AF_7908	ATP6V1C1	ATPase H+ transporting V1 subunit C1	-3093405935	-3086738859	0,011505	0,02447	0,07101	0,4702	0,03216134871
A. forsteri	AF_8749	MCMBP	Minichromosome maintenance complex binding protein	-5495090419	-5485779913	0,007091	0,006285	0,07152	1,1281	0,00369682409 1
A. forsteri	AF_8543	IDE	Insulin degrading enzyme	-6904648351	-6896112142	0,005545	0,027849	0,03653	0,1991	0,00705873412 8
A. forsteri	AF_8972	INTS2	Integrator complex subunit 2	-10334032792	-10327810599	0,002343	0,002821	0,05648	0,8306	0,04157785835
A. forsteri	AF_8983	APPBP2	Amyloid beta precursor protein binding protein 2	-4370923681	-4354724513	0,004569	0,002776	0,00889	1,646	3,20E-05
A. forsteri	AF_9132	HDHD2	Haloacid dehalogenase like hydrolase domain containing 2	-2586355115	-2580233995	0,013557	0,005209	0,13516	2,6025	0,04359203902
A. forsteri	AF_9261	DYNC1I2	Dynein cytoplasmic 1 intermediate chain 2	-4339346636	-4331810605	0,003983	0,004721	0,03087	0,8437	0,0158345096
A. forsteri	AF_9769	SLIT2	Slit guidance ligand 2	-8311327808	-8305127747	0,001522	0,00295	0,02838	0,5159	0,04157785835
A. forsteri	AF_9862	PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	-9858850225	-9850650474	0,002592	0,001983	0,06277	1,3074	0,00943659766 7
A. forsteri	AF_9984	XPO4	Exportin 4	-7237695777	-7227132988	0,002114	0,002742	0,02107	0,771	0,00117305132
A. forsteri	AF_9961	ANKRD17	Ankyrin repeat domain 17	-15007969513	-14992613022	0,003813	0,003778	0,05029	1,0093	3,81E-05
A. forsteri	AF_10572	SFI1	SFI1 centrin binding protein	-3940653339	-3932635808	0,001847	0,060968	0,54726	0,0303	0,01081225409
A. forsteri	AF_10519	SMARCB1	Actin dependent regulator of chromatin	-2590451276	-2582187836	0,006447	0,012508	0,01855	0,5154	0,00917450537 5
A. forsteri	AF_11846	PCMT1	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	-1193624067	-1187040217	0,005729	0,003375	0,02551	1,6972	0,033011879
A. forsteri	AF_13292	SON	SON DNA binding protein	-4358166602	-43499761	0,028991	0,072681	0,10877	0,3989	0,00943659766 7
A. forsteri	AF_12993	ARSJ	Arylsulfatase family member J	-4474783053	-4467715932	0,004169	0,005722	0,02725	0,7286	0,02348588544
A. forsteri	AF_13270	TIAM1	TIAM Rac1 associated GEF 1	-13787238749	-13779297985	0,002102	0,002069	0,05724	1,016	0,01121603335
A. forsteri	AF_262	CTNNB1	Catenin beta 1	-5927657146	-5916428017	0,002978	0,002427	0,01269	1,2269	0,00066380076
A. forsteri	AF_475	ZNF326	Zinc finger protein 326	-3086139466	-3074893861	0,005437	0,00556	0,03129	0,9779	0,00066380076
A. forsteri	AF_13872	TEC	Tec protein tyrosine kinase	-483256932	-4825957223	0,003355	0,004604	0,0502	0,7287	0,033011879
A. forsteri	AF_14165	RSRC2	Arginine and serine rich coiled-coil 2	-2871993365	-2864573572	0,00697	0,00595	0,0417	1,1714	0,0175597951
A. forsteri	AF_13907	SLC2A1	Solute carrier family 2 member 1	-4273606712	-426656548	0,00519	0,005713	0,02372	0,9085	0,02348588544
A. forsteri	AF_14293	SIN3B	SIN3 transcription regulator family member B	-8947294844	-8931805361	0,003435	0,007034	0,01865	0,4884	3,81E-05
A. forsteri	AF_14863	CREB1	cAMP responsive element binding protein 1	-2182817514	-2176441134	0,0045	0,003487	0,03012	1,2903	0,03725722575
A. forsteri	AF_14970	AGO4	Protein argonaute-4	-5831073447	-5824543194	0,002199	0,002652	0,01364	0,8293	0,033011879
A. forsteri	AF_14805	CBFA2T2	CBFA2/RUNX1 partner transcriptional co-repressor 2	-486238352	-4850439733	0,008462	0,020958	0,03234	0,4038	0,00041126 13842
A. forsteri	AF_15874	UBTD2	Ubiquitin domain containing 2	-1426090087	-1411632319	0,015211	0,012841	0,02579	1,1845	7,23E-05
A. forsteri	AF_16001	ATOX1	Antioxidant 1 copper chaperone	-698565592	-692311733	0,024614	0,070414	0,0778	0,3496	0,0413429436
A. forsteri	AF_16232	PAK1	p21 (RAC1) activated kinase 1	-3902088349	-3893452186	0,005017	0,002471	0,0505	2,0305	0,00669710640 5
A. forsteri	AF_16462	TRRAP	Transformation/transcription domain associated protein	-1100710008	-1094267791	0,007178	0,010158	0,007	0,7067	0,03520841431
A. forsteri	AF_16779	TMEM136	Transmembrane protein 136	-2554230876	-2547536678	0,011799	0,005151	0,10675	2,2907	0,03175666705
A. forsteri	AF_16783	TRIM29	Tripartite motif containing 29	-4563715231	-4556368765	0,005731	0,006242	0,05422	0,9181	0,01861547154
A. forsteri	AF_17377	ZYG11B	Protein zyg-11 homolog B	-5952338226	-5941118625	0,004142	0,004208	0,03019	0,9843	0,00066380076
A. forsteri	AF_17652	PSMD12	Proteasome 26S subunit, non-ATPase 12	-2457245614	-2449415142	0,006546	0,007682	0,03985	0,8522	0,01233487709
A. forsteri	AF_17836	CDC27	Cell division cycle 27	-5874928607	-5868764267	0,003246	0,0087	0,03319	0,3731	0,04213245123
A. forsteri	AF_867	CLCN7	Chloride voltage-gated channel 7	-7135912477	-7125413896	0,004365	0,00829	0,02712	0,5265	0,00117305132
A. forsteri	AF_980	DHX36	DEAH-box helicase 36	-8387828847	-8379740228	0,002498	0,001433	0,04758	1,7426	0,0102635496
A. forsteri	AF_18295	CRNKL1	Crooked neck pre-mRNA splicing factor 1	-5307866494	-5301030989	0,004075	0,00459	0,03059	0,8878	0,027956754
A. forsteri	AF_18677	IPO7	Importin 7	-6939111933	-692536631	0,003364	0,001617	0,03016	2,081	0,00012 1914818
A. forsteri	AF_1408	NUP205	Nucleoporin 205	-6807066793	-6800013146	0,001181	0,007783	0,0716	0,1518	0,02348588544
A. forsteri	AF_1588	AQP2	Aquaporin 2	-3064184948	-3057602995	0,008733	0,004594	0,07474	1,9011	0,033011879
A. forsteri	AF_2852	BRMS1L	BRMS1 like transcriptional repressor	-1970591685	-1959043846	0,008633	0,008542	0,04364	1,0107	0,0005615834
A. forsteri	AF_3784	NEDD4	NEDD4 E3 ubiquitin protein ligase	-6386335381	-6380399578	0,004129	0,001973	0,11255	2,0925	0,04955413591
A. forsteri	AF_15280	TRPM8	Transient receptor potential cation channel subfamily M member 8	-9444417165	-9435416073	0,001314	0,01346	0,06013	0,0976	0,01875770167
A. forsteri	AF_2954	JMJD1C	Jumonji domain containing 1C	-13288222845	-132622255	0,001813	0,006211	0,12672	0,2919	4,26E-09
A. patagonicus	AF_4494	LNX2	Ligand of Numb protein X 2	-6658627707	-6649558437	0,005742	0,004195	0,08009	1,3688	0,01208858
A. patagonicus	AF_13259	CCT8	Chaperonin Containing TCP1 Subunit 8	-3462889754	-3455267095	0,009668	0,021091	0,0495	0,4584	0,038017819
A. patagonicus	AF_15160	CLDN18	Claudin 18	-3346636113	-3331828403	0,030984	0,243117	0,07568	0,1274	0,00020154 2642
A. patagonicus	AF_4689	ADGRL3	Adhesion G protein-coupled receptor L3	-310995719	-3099585622	0,008685	0,021118	0,02611	0,4113	0,00502278586 3
A. patagonicus	AF_5762	HNRNPLL	Heterogeneous nuclear ribonucleoprotein L like	-3456725962	-3446531149	0,00538	0,002824	0,02437	1,9055	0,00503879558

Table S5. Candidate genes for which RELAX found signals of intensified selection (K > 1) in A. forsteri (FDR < 0.05)

Gene ID	Gene name	Gene description	K	P-value	Adjusted p-value
AF_10572	SFI1	SFI1 Centrin Binding Protein	5,135876433	0,0000000266	0,00009690635
AF_10879	TARS3	Threonyl-TRNA Synthetase 3	24,97403162	0,0000000367	0,00009690635
AF_10888	MYO1E	Myosin IE	40,54140993	0,0000000801	0,0001256878
AF_11206	COQ6	Coenzyme Q6, Monooxygenase	50	0,0000000952	0,0001256878
AF_11502	ACOT12	Acyl-CoA Thioesterase 12	5,020793978	0,000000144	0,0001520928
AF_12020	OTOA	Otoancorin	45,73485854	0,000000192	0,000168992
AF_12479	SNX13	Sorting Nexin 13	3,675279974	0,000000299	0,0002255741429
AF_12625	SLC25A3	Solute Carrier Family 25 Member 3	4,621722815	0,000000469	0,000309598625
AF_12863	KIAA1841	SANT And BTB Domain Regulator Of CSR	2,559263135	0,000000653	0,0003831658889
AF_13392	EFHC2	EF-Hand Domain Containing 2	4,016546382	0,00000118	0,000623158
AF_13473	CRB1	Crumbs Cell Polarity Complex Component 1	10,54184525	0,00000213	0,001022593636
AF_13872	TEC	Tec Protein Tyrosine Kinase	50	0,00000307	0,001351055833
AF_14164	KNTC1	Kinetochore Associated 1	50	0,00000363	0,001474617692
AF_14805	CBFA2T2	CBFA2/RUNX1 Partner Transcriptional Co-Repressor 2	6,349842484	0,00000517	0,001946928667
AF_14936	INPP5B	Inositol Polyphosphate-5-Phosphatase B	45,74236636	0,00000553	0,001946928667
AF_1497	ATG5	Autophagy Related 5	11,97966152	0,00000666	0,00219821625
AF_15354	AGXT2	Alanine--Glyoxylate Aminotransferase 2	50	0,0000072	0,002236658824
AF_15426	VPS13A	Vacuolar Protein Sorting 13 Homolog A	4,692523499	0,00000912	0,002675706667
AF_15710	FAM179A	TOG Array Regulator Of Axonemal Microtubules 2	21,04335833	0,0000136	0,003780084211
AF_15874	UBTD2	Ubiquitin Domain Containing 2	5,526435947	0,000015	0,00396075
AF_16568	DRC3	Dynein Regulatory Complex Subunit 3	7,066727352	0,0000169	0,004249947619
AF_16779	TMEM136	TLC Domain Containing 5	10,86567826	0,0000185	0,004440840909
AF_16799	LTA4H	Leukotriene A4 Hydrolase	4,703763483	0,0000218	0,004906929167
AF_17653	HELZ	Helicase With Zinc Finger	50	0,0000223	0,004906929167
AF_1775	PCCA	Propionyl-CoA Carboxylase Subunit Alpha	50	0,0000259	0,005471116
AF_17836	CDC27	Cell Division Cycle 27	4,318131157	0,000029	0,005890346154
AF_1783	DOCK9	Dedicator Of Cytokinesis 9	4,53879479	0,0000321	0,006278522222
AF_18346	PNPT1	Polyribonucleotide Nucleotidyltransferase 1	5,359372452	0,0000343	0,006301993333
AF_18686	RNF141	Ring Finger Protein 141	50	0,0000357	0,006301993333
AF_1887	TMEM68	Transmembrane Protein 68	50	0,0000358	0,006301993333
AF_2282	NME7	NME/NM23 Family Member 7	8,933255746	0,0000415	0,006848796875
AF_2359	WDHD1	WD Repeat And HMG-Box DNA Binding Protein 1	3,463530025	0,0000595	0,00952180303
AF_2812	COCH	Cochlin	4,984052411	0,0000709	0,01101243824
AF_2889	AGL	Amylo-Alpha-1, 6-Glucosidase, 4-Alpha-Glucanotransferase	50	0,0000734	0,01107501143
AF_2954	JMJD1C	Jumonji Domain Containing 1C	50	0,0000781	0,01145683611
AF_3445	TMA16	Translation Machinery Associated 16 Homolog	5,608371484	0,0000863	0,01231757568
AF_3520	SUSD5	Sushi Domain Containing 5	6,091471237	0,000092	0,01278557895
AF_3784	NEDD4	NEDD4 E3 Ubiquitin Protein Ligase	3,408788434	0,0001034193348	0,01400403864
AF_3816	SYCP1	Synaptonemal Complex Protein 1	45,98022099	0,0001324912403	0,017492156
AF_3995	ITPR3	Inositol 1,4,5-Trisphosphate Receptor Type 3	2,343454673	0,0001435518409	0,01810492095
AF_4510	NME8	NME/NM23 Family Member 8	50	0,000143989146	0,01810492095
AF_4536	AVL9	AVL9 Cell Migration Associated	4,035448742	0,0001482325433	0,01820502468
AF_4576	HERC5	HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase 5	50	0,000162617641	0,01951781277
AF_4632	NPNT	Nephronectin	5,143337172	0,0001908419878	0,0223963675
AF_4804	EVC2	EvC Ciliary Complex Subunit 2	5,857175477	0,0002091719408	0,0227100099
AF_5022	HERC1	HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase	12,18225009	0,0002092371906	0,0227100099
AF_5029	Hsp40	DnaJ Heat Shock Protein Family	5,696852219	0,0002202738908	0,0227100099
AF_5042	RBM19	RNA Binding Motif Protein 19	5,275586623	0,0002208192394	0,0227100099
AF_5623	IDUA	Alpha-L-Iduronidase	5,857869718	0,0002248507643	0,0227100099
AF_5943	SMARCAL1	Actin Dependent Regulator Of Chromatin, Subfamily A Like 1	5,061129038	0,0002249640859	0,0227100099
AF_6396	MATN2	Matrilin 2	50	0,0002257383151	0,0227100099
AF_6608	SLC39A12	Solute Carrier Family 39 Member 12	2,990206511	0,0002316498662	0,0227100099
AF_6996	RPL31	Ribosomal Protein L31	50	0,0002322174843	0,0227100099
AF_7116	HMGCL	3-Hydroxy-3-Methylglutaryl-CoA Lyase	50	0,0002587200563	0,02454949982
AF_7276	UNC13C	Unc-13 Homolog C	4,541280096	0,0002632738049	0,02454949982
AF_7370	TMC3	Transmembrane Channel Like 3	4,870346811	0,0002649728252	0,02454949982
AF_7436	ELP2	Elongator Acetyltransferase Complex Subunit 2	50	0,0002777042813	0,02528545361
AF_7503	MYLK3	Myosin Light Chain Kinase 3	7,40606769	0,000309483481	0,027396877
AF_755	PGM2L1	Phosphoglucomutase 2 Like 1	50	0,0003112691953	0,027396877
AF_7564	SLC7A6	Solute Carrier Family 7 Member 6	3,588501998	0,0003317227172	0,02871848639
AF_7895	GLS	Glutaminase	4,076720222	0,0004077740001	0,03473313701
AF_8156	ASAH1	N-Acylsphingosine Amidohydrolase 1	48,43880498	0,0004240274354	0,03554426803
AF_8173	SLC7A2	Solute Carrier Family 7 Member 2	38,12050412	0,0004489436309	0,0370448643
AF_8703	TDRD1	Tudor Domain Containing 1	5,16834605	0,0004968792884	0,04002047086
AF_8782	CPXM2	Carboxypeptidase X, M14 Family Member 2	3,865603805	0,0005001611583	0,04002047086
AF_8972	INTS2	Integrator Complex Subunit 2	17,57559971	0,0005828530974	0,0451634167
AF_8983	APPBP2	Amyloid Beta Precursor Protein Binding Protein 2	2,937027461	0,0005979239328	0,0451634167
AF_9101	LRPPRC	Leucine Rich Pentatricopeptide Repeat Containing	3,107315367	0,0006002563028	0,0451634167
AF_9269	PPP1R9A	Protein Phosphatase 1 Regulatory Subunit 9A	9,516051849	0,0006039642174	0,0451634167
AF_9464	NRG3	Neuregulin 3	6,285501843	0,0006071960965	0,0451634167
AF_980	DHX36	DEAH-Box Helicase 36	4,556533591	0,0000412	0,006848796875

Table S6. Candidate genes for which RELAX found signals of intensified selection (K > 1) in A. patagonicus (FDR < 0.05)

Gene ID	Gene name	Gene description	K	P-value	Adjusted p-value
AF_15753	A2ML1	Alpha-2-Macroglobulin Like 1	28,96233276	1,71E-13	0,000000000891936
AF_11047	GOLGA1	Golgin A1	3,844974319	2,99E-12	0,00000000779792
AF_8752	SPTLC3	Serine Palmitoyltransferase Long Chain Base Subunit 3	14,30144609	0,000000164	0,00026732
AF_8402	CCT8	Chaperonin Containing TCP1 Subunit 8	3,84427245	0,000000205	0,00026732
AF_2550	PLEKHH1	Pleckstrin Homology, MyTH4 And FERM Domain Containing H1	11,09394818	0,000000361	0,0003765952
AF_9727	SEPT2	Septin 2	47,28415224	0,00000124	0,001077973333
AF_17888	UROC1	Urocanate Hydratase 1	10,52857883	0,00000308	0,00204076
AF_4689	AIM1	Crystallin Beta-Gamma Domain Containing 1	50	0,00000313	0,00204076
AF_2516	CPB2	Carboxypeptidase B2	5,115426218	0,00000383	0,002219697778
AF_4955	CLDN18	Claudin 18	50	0,00000702	0,003661632
AF_7552	LETM1	Leucine Zipper And EF-Hand Containing Transmembrane Protein 1	38,8346529	0,00000788	0,003736552727
AF_5241	MEP1B	Meprin A Subunit Beta	29,50697063	0,0000124	0,005389866667
AF_5889	RARS	Arginyl-TRNA Synthetase 1	3,038511142	0,0000135	0,005416615385
AF_8924	RBBP6	RB Binding Protein 6, Ubiquitin Ligase	38,87934461	0,000017	0,006333714286
AF_1498	HEPH	Hephaestin	3,810692817	0,0000211	0,007337173333
AF_15558	ECE1	Endothelin Converting Enzyme 1	3,117268144	0,0000266	0,007997866667
AF_2837	TNRC6B	Trinucleotide Repeat Containing Adaptor 6B	2,891330917	0,0000268	0,007997866667
AF_12888	MEI1	Meiotic Double-Stranded Break Formation Protein 1	14,12997858	0,0000276	0,007997866667
AF_3806	BAZ1A	Bromodomain Adjacent To Zinc Finger Domain 1A	10,40442415	0,0000382	0,01048690526
AF_13823	WDR35	WD Repeat Domain 35	4,690558028	0,000043	0,0112144
AF_9658	PPP1R21	Protein Phosphatase 1 Regulatory Subunit 21	50	0,0000609	0,0146048
AF_14715	ADGRL3	Adhesion G Protein-Coupled Receptor L3	47,00028235	0,000063	0,0146048
AF_2094	OGFR	Opioid Growth Factor Receptor	46,1	0,0000644	0,0146048
AF_13259	PTPN11	Protein Tyrosine Phosphatase Non-Receptor Type 11	10,07599425	0,0000708	0,0153872
AF_285	ANKZF1	Ankyrin Repeat And Zinc Finger Peptidyl TRNA Hydrolase 1	8,37681621	0,0000919	0,019174016
AF_15160	DDX1	DEAD-Box Helicase 1	4,647765809	0,0001058074996	0,02114278623
AF_8479	PDC	Phosducin	4,806996748	0,0001162518623	0,02114278623
AF_6454	DPEP2	Dipeptidase 2	18,77623626	0,0001207897653	0,02114278623
AF_15827	CFAP97	Cilia And Flagella Associated Protein 97	12,97242384	0,0001224644622	0,02114278623
AF_15093	SYNPO2L	Synaptopodin 2 Like	50	0,0001225926527	0,02114278623
AF_8128	ZNF488	Zinc Finger Protein 488	29,48334142	0,0001256568967	0,02114278623
AF_14480	WDR11	WD Repeat Domain 11	47,32465792	0,000184692065	0,0301048066
AF_16647	TRPV3	Transient Receptor Potential Cation Channel Subfamily V Member 3	50	0,0002285431435	0,03612366777
AF_11086	SLCO1C1	Solute Carrier Organic Anion Transporter Family Member 1C1	50	0,000247227765	0,03792764771
AF_6686	PARVG	Parvin Gamma	49,74626866	0,0003276404311	0,04882778539

Table S7. Candidate genes with positively selected sites in A.forsteri and A.patagonicus, predicted by CODEML branch-site models (FDR < 0.05)

Lineage	Gene ID	Gene name	Gene description	Likelihood Null Model	Likelihood Alternative Model	LRT	d.f	Adjusted p-value	BEB Analysis
A. forsteri	AF_4494	LNX2	Ligand of Numb protein X 2	-6578,64252	-6575,726939	5,831162	1	0,0324436654545455	12 W 0.951*
									18 T 0.827
									26 D 0.819
A. forsteri	AF_4640	SGMS2	Sphingomyelin synthase 2	-2943,85942	-2933,884229	19,950382	1	3,86036485714286e-05	281 L 0.999**
									282 L 1.000**
									283 T 0.977*
									284 I 0.976*
A. forsteri	AF_5071	TFAP2C	Transcription factor AP-2 gamma	-1987,396934	-1974,065937	26,661994	1	2,35416e-06	158 C 0.996**
									159 F 0.998**
									160 R 0.873
A. forsteri	AF_5373	UFD1L	Ubiquitin recognition factor in ER associated degradation 1	-2194,528675	-2188,770013	11,517324	1	0,00195360033333333	1 M 0.998**
									2 T 0.936
									3 V 0.933
A. forsteri	AF_7017	GMEB1	Glucocorticoid modulatory element binding protein 1	-2937,594909	-2931,005189	13,17944	1	0,000916500761904762	1 M 0.997**
									2 N 0.817
A. forsteri	AF_6831	WNT3A	Wnt family member 3A	-2311,378253	-2300,910396	20,935714	1	2,75173333333333e-05	71 R 1.000**
									72 A 0.884
									73 V 0.878
									77 P 0.948
									78 P 0.830
A. forsteri	AF_7895	GLS	Glutaminase	-3109,236662	-3094,430938	29,611448	1	7,17944e-07	23 R 0.965*
									218 A 1.000**
									220 R 1.000**
									221 H 0.963*
A. forsteri	AF_8749	MCMBP	Minichromosome maintenance complex binding protein	-5446,9791	-5439,253415	15,45137	1	0,00028782904	49 V 0.908
									50 K 0.853
									51 E 0.655
									146 G 0.942
									147 C 0.941
									331 Q 0.938
A. forsteri	AF_8543	IDE	Insulin degrading enzyme	-6780,555004	-6769,523983	22,062042	1	1,99450044444444e-05	384 Q 0.749
									548 H 0.547
									557 R 0.774
									566 L 0.529
A. forsteri	AF_8972	INTS2	Integrator complex subunit 2	-10187,45183	-10167,08127	40,741114	1	9,52e-09	484 S 0.997**
									485 V 0.974*
									486 A 0.998**
A. forsteri	AF_9132	HDHD2	Haloacid dehalogenase like hydrolase domain containing 2	-2559,627097	-2545,489555	28,275084	1	1,19272e-06	21 L 0.791
									22 R 0.950
									23 S 0.950
									107 C 0.797
A. forsteri	AF_9261	DYNC1I2	Dynein cytoplasmic 1 intermediate chain 2	-4323,901549	-4307,812753	32,177592	1	3,1892e-07	299 Q 0.999**
									300 Q 0.998**
A. forsteri	AF_9961	ANKRD17	Ankyrin repeat domain 17	-14885,71439	-14879,29238	12,844022	1	0,00100095704347826	177 T 0.599
									1081 V 0.724
									1150 A 0.563
A. forsteri	AF_11846	PCMT1	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	-1167,054052	-1158,865431	16,377242	1	0,0002144912	195 D 0.909
A. forsteri	AF_13292	SON	SON DNA binding protein	-4086,87545	-4066,972694	39,805512	1	9,52e-09	6 S 0.694
									26 R 0.629
A. forsteri	AF_13270	TIAM1	TIAM Rac1 associated GEF 1	-13709,29822	-13704,80283	8,990766	1	0,00683392948148148	963 A 0.821
									1319 E 0.538
									1485 V 0.728
A. forsteri	AF_262	CTNNB1	Catenin beta 1	-5888,046367	-5881,508793	13,075148	1	0,000924923636363636	1 Q 0.643
									703 A 0.889
									704 V 0.885
									705 S 0.777
A. forsteri	AF_475	ZNF326	Zinc finger protein 326	-3074,599753	-3070,989783	7,21994	1	0,0169057660689655	219 K 0.946
									220 K 0.998**
									221 K 0.819
									223 K 0.946
									226 T 0.947
A. forsteri	AF_13872	TEC	Tec protein tyrosine kinase	-4800,546021	-4796,991208	7,109626	1	0,0173787826666667	149 M 0.862
									172 K 0.995**
									477 G 0.868
A. forsteri	AF_14805	CBFA2T2	CBFA2/RUNX1 partner transcriptional co-repressor 2	-4820,717669	-4815,448563	10,538212	1	0,0031805368	359 E 0.675
									362 G 0.657
									363 G 0.704
									364 G 0.996**
									368 A 0.967*
A. forsteri	AF_15280	TRPM8	Transient receptor potential cation channel	-9342,079787	-9334,315259	15,529056	1	0,00028782904	1049 T 0.808
									1060 M 0.870
A. forsteri	AF_16232	PAK1	p21 (RAC1) activated kinase 1	-3875,695768	-3865,027497	21,336542	1	2,6201488e-05	471 F 0.999**
									473 K 0.998**
									474 F 0.945
A. forsteri	AF_16462	TRRAP	Transformation/transcription domain associated protein	-1093,238656	-1089,565938	7,345436	1	0,0163280677142857	41 Q 0.997**
A. forsteri	AF_16779	TMEM136	Transmembrane protein 136	-2514,418204	-2505,261884	18,31264	1	8,49814133333333e-05	61 F 0.999**
									62 S 0.979*
									65 Q 0.960*
									66 V 0.998**
A. forsteri	AF_16783	TRIM29	Tripartite motif containing 29	-4507,973276	-4500,069787	15,806978	1	0,0002649858	487 D 0.934
									489 S 0.932
									490 L 0.999**
									491 K 0.995**
									492 G 0.933
A. forsteri	AF_17377	ZYG11B	Protein zyg-11 homolog B	-5910,2849	-5907,044897	6,480006	1	0,0239304327741935	1 M 0.597
									53 V 0.795
A. forsteri	AF_17652	PSMD12	Proteasome 26S subunit, non-ATPase 12	-2445,634772	-2437,477009	16,315526	1	0,0002144912	113 T 0.997**
									114 R 0.995**
									115 H 0.856
A. forsteri	AF_17836	CDC27	Cell division cycle 27	-5856,389289	-5846,05131	20,675958	1	2,84526123076923e-05	179 G 0.835
									688 G 0.995**
									690 T 0.840
									691 F 0.627
A. forsteri	AF_867	CLCN7	Chloride voltage-gated channel 7	-7091,645003	-7075,991059	31,307888	1	3,7434e-07	683 P 0.862
									687 G 0.856
									689 C 0.999**
									692 R 0.966*
									693 Q 0.832
A. forsteri	AF_980	DHX36	DEAH-box helicase 36	-8350,885653	-8346,101466	9,568374	1	0,00517736307692308	40 G 0.952*
									261 S 1.000**
									340 V 0.950*
									798 T 0.953*
A. forsteri	AF_1408	NUP205	Nucleoporin 205	-6746,865304	-6735,552885	22,624838	1	1,673837e-05	448 S 0.984*
A. forsteri	AF_2954	JMJD1C	Jumonji domain containing 1C	-13118,79475	-13108,34812	20,893262	1	2,75173333333333e-05	255 Q 0.970*
A. patagonicus	AF_2954	JMJD1C	Jumonji domain containing 1C	-13116,26543	-13110,24293	12,044998	1	0,007062689848	818 P 0.908
									1543 K 0.914
									1544 S 0.992**
A. patagonicus	AF_4689	ADGRL3	Adhesion G protein-coupled receptor L3	-3095,661257	-3077,545649	36,231216	1	5,95e-08	230 H 0.999**
									231 S 0.809
									232 D 0.817

									236 D 1.000**
A. patagonicus	AF_13259	CCT8	Chaperonin Containing TCP1 Subunit 8	-3424,544245	-3410,031141	29,026208	1	1,61862666666667e-06	248 I 0.832
									302 T 0.998**
									303 C 1.000**
									304 P 0.840
									379 I 0.837
A. patagonicus	AF_4494	LNX2	Ligand of Numb protein X 2	-6581,073176	-6572,401156	17,34404	1	0,00053018716	191 D 0.924
									192 L 0.904
									193 P 0.904
									194 Q 0.910
									260 V 0.945

Table S8. Candidate genes for which RELAX found signals of relaxed selection (K < 1) in A. forsteri (FDR < 0.0

Gene ID	Gene name	Gene description	Gene function	K	P-value	Adjusted p-value
AF_1808	SHROOM3	Shroom Family Member 3	This gene encodes a PDZ-domain-containing protein that belongs to a family of Shroom-related proteins. This protein may be involved in regulating cell shape in certain tissues. A similar protein in mice is required for proper neurulation	0,061487	0	0,00000040874
AF_9961	ANKRD17	Ankyrin Repeat Domain 17	It has been suggested that this protein plays a role in both DNA replication and in both anti-viral and anti-bacterial innate immune pathways.	0,089538	0,0000038	0,001935467
AF_16351	FOLH1B	Putative N-acetylated-alpha-linked Acidic Dipeptidase	Metallopeptidase activity and dipeptidase activity	0,366502	0,0000341	0,008288242
AF_12856	VRK2	Vaccinia Related Kinase 2	Serine/threonine kinase that regulates several signal transduction pathways. Isoform 1 modulates the stress response to hypoxia and cytokines.	0,069616	0,0000382	0,008583765
AF_12356	RARS2	Arginyl-tRNA Synthetase 2, mitochondrial	This nuclear gene encodes a protein that localizes to the mitochondria, where it catalyzes the transfer of L-arginine to its cognate tRNA, an important step in translation of mitochondrially-encoded proteins.	0,129315	0,0000603	0,01212347
AF_475	ZNF326	Zinc Finger Protein 326	Core component of the DBIRD complex, a multiprotein complex that acts at the interface between core mRNP particles and RNA polymerase II (RNAPII) and integrates transcript elongation with the regulation of alternative splicing. May play a role in neuronal differentiation and is able to bind DNA.	0,000031	0,0000729	0,01367746
AF_10189	FLVCR1	Feline Leukemia Virus Subgroup C Cellular Receptor 1	This gene encodes a member of the major facilitator superfamily of transporter proteins. The encoded protein is a heme transporter that may play a critical role in erythropoiesis by protecting developing erythroid cells from heme toxicity.	0,335002	0,0000939	0,01594213
AF_14165	RSRC2	Arginine/Serine-rich Coiled-coil 2	Diseases associated with RSRC2 include Strabismus and Myopia	0,243106	0,0001179351	0,01917073
AF_18677	IPO7	Importin 7	Functions in nuclear protein import, either by acting as autonomous nuclear transport receptor or as an adapter-like protein in association with the importin-beta subunit KPNB1	0	0,0002248237	0,02816098
AF_9862	PHLPP1	PH Domain and Leucine Rich Repeat Protein Phosphatase 1	The encoded protein promotes apoptosis by dephosphorylating and inactivating the serine/threonine kinase Akt, and functions as a tumor suppressor in multiple types of cancer.Involved in the hippocampus-dependent long-term memory formation (By similarity). Involved in circadian control by regulating the consolidation of circadian periodicity after resetting (By similarity). Involved in development and function of regulatory T-cells	0	0,0002440689	0,02913572
AF_964	CCNL1	Cyclin L1	Involved in pre-mRNA splicing. Functions in association with cyclin-dependent kinases (CDKs)	0,374898	0,0003203334	0,03435755
AF_4559	ASAP1	ArfGAP with SH3 domain, Ankyrin repeat and PH domain 1	May coordinate membrane trafficking with cell growth or actin cytoskeleton remodeling by binding to both SRC and PIP2. May function as a signal transduction protein involved in the differentiation of fibroblasts into adipocytes and possibly other cell types	0,01036983409	0,0003301746439	0,05882572017
AF_11868	FBXO5	F-box Protein 5	Regulator of APC activity during mitotic and meiotic cell cycle	0	0,0003327825771	0,05882572017
AF_5373	UFD1L	Ubiquitin Fusion Degradation 1 like	Essential component of the ubiquitin-dependent proteolytic pathway which degrades ubiquitin fusion proteins	0,293617511	0,0003696194827	0,06067039795
AF_8796	EDRF1	Erythroid Differentiation Regulatory factor 1	Transcription factor involved in erythroid differentiation. Involved in transcriptional activation of the globin gene	0,3761251066	0,0004624720448	0,0673835077
AF_401	PTCH1	Patched 1	This gene encodes a member of the patched family of proteins and a component of the hedgehog signaling pathway. Hedgehog signaling is important in embryonic development and tumorigenesis.	0,04620246266	0,0004691628038	0,0673835077
AF_16232	PAK1	p21 (RAC1) activated kinase 1	Protein kinase involved in intracellular signaling pathways downstream of integrins and receptor-type kinases that plays an important role in cytoskeleton dynamics, in cell adhesion, migration, proliferation, apoptosis, mitosis, and in vesicle-mediated transport processes. Plays a role in the regulation of insulin secretion in response to elevated glucose levels.	0,313	0,0005412858331	0,0731691085

Table S9. Candidate genes for which RELAX found signals of relaxed selection (K < 1) in A. patagonicus (FDR < 0.05)

Gene ID	Gene name	Gene description	Gene function	K	P-value	Adjusted p-value
AF_10582	ACADS	Acyl-CoA Dehydrogenase Short Chain	This gene encodes a tetrameric mitochondrial flavoprotein, which is a member of the acyl-CoA dehydrogenase family. This enzyme catalyzes the initial step of the mitochondrial fatty acid beta-oxidation pathway.	0,270306	0,0000000331	0,000063
AF_14861	CPO	Carboxypeptidase O	This gene is a member of the metallocarboxypeptidase gene family.	0	0,00000279	0,002174
AF_6089	ZNF438	Zinc Finger Protein 438	Isoform 1 acts as a transcriptional repressor.	0	0,0000491	0,015632
AF_8002	TBC1D31	TBC1 Domain Family Member 31	Diseases associated with TBC1D31 include Branchiootorenal Syndrome 1; dominant disorder characterized by sensorineural, conductive, or mixed hearing loss, structural defects of the outer, middle, and inner ear, branchial fistulas or cysts, and renal abnormalities ranging from mild hypoplasia to complete absence	0,045369	0,0000000261	0,000063

Table S10. GO biological process terms in candidate genes for positive selection suggested by both CODEML and aBSREL

GO category	Term ID	Adjusted p-value	Genes Number	Genes names
Wnt signaling pathway involved in heart development	GO:0003306	0,02090722984	2	WNT3A,CTNNB1
biosynthetic process	GO:0009058	0,02090722984	16	WNT3A,ZDHHC8,CTNNB1,NEDD4,ANKRD17,RPL31,ZNF326,SMARCB1,TRIM29, TFAP2C,MCMBP,ADCY5,GMEB1,SGMS2,CBFA2T2,SIN3B
positive regulation of muscle tissue development	GO:1901863	0,02753468434	2	WNT3A,CTNNB1
cellular nitrogen compound biosynthetic process	GO:0044271	0,03517376825	13	WNT3A,CTNNB1,NEDD4,RPL31,ZNF326,SMARCB1,TRIM29,TFAP2C,ADCY5, GMEB1,SGMS2,CBFA2T2,SIN3B
cellular localization	GO:0051641	0,03517376825	9	WNT3A,XPO4,CTNNB1,NEDD4,NUP205,TRIM29,ADCY5,APPBP2,AP2A2
dopaminergic neuron differentiation	GO:0071542	0,03517376825	2	WNT3A,CTNNB1
regulation of nucleic acid-templated transcription	GO:1903506	0,03517376825	10	WNT3A,CTNNB1,NEDD4,ZNF326,SMARCB1,TRIM29,TFAP2C,GMEB1,CBFA2T2, SIN3B
RNA metabolic process	GO:0016070	0,03517376825	12	WNT3A,CRNKL1,CTNNB1,NEDD4,ZNF326,SMARCB1,TRIM29,TFAP2C,GMEB1, CBFA2T2,SIN3B,INTS2
response to lipid	GO:0033993	0,03517376825	4	CTNNB1,NEDD4,ADCY5,SLC2A1
developmental induction	GO:0031128	0,03517376825	2	WNT3A,CTNNB1
sister chromatid cohesion	GO:0007062	0,03582990517	2	CTNNB1,MCMBP
gene expression	GO:0010467	0,04040696313	13	WNT3A,CRNKL1,CTNNB1,NEDD4,RPL31,ZNF326,SMARCB1,TRIM29,TFAP2C, GMEB1,CBFA2T2,SIN3B,INTS2
ceramide phosphoethanolamine metabolic process	GO:1905371	0,04481644826	1	SGMS2
organic cyclic compound metabolic process	GO:1901360	0,04481644826	13	WNT3A,CRNKL1,CTNNB1,NEDD4,ZNF326,SMARCB1,TRIM29,TFAP2C,ADCY5, GMEB1,CBFA2T2,SIN3B,INTS2
nuclear transport	GO:0051169	0,04481644826	3	XPO4,NEDD4,NUP205
cellular response to glucose stimulus	GO:0071333	0,04721095806	2	SMARCB1,ADCY5

Table S11. GO biological process terms in candidate genes for positive selection suggested by either CODEML or aBSREL

GO category	Term ID	Adjusted p-value	Genes Number	Genes names
response to stress	GO:0006950	0,00004058722031	61	INTS7,UBA5,PLD4,TBK1,UBASH3B,MAP3K5,ENPP3,CLEC16A,LGALS8,OPA1, FANCL,PPARA,P2RX4,CHAF1A,EDEM2,SUMO1,ATG5,RASGRP1,UIMC1,AQP11, EIF2AK1,MDFIC,ZYX,STXBP4,JCHAIN,TH,PNPT1,CELSR1,PLEKHM2,DACT1, PSMC6,CTNNB1,COCH,MBIP,NEDD4,MAGI3,DAPK1,FANCA,NME8,AOAH,C7, NEK4,IFIH1,SMARCAL1,CUL4B,UFL1,ATRIP,EDEM3,MAP3K3,DLEC1,WNT3A, HRAS,MYLK3,CAPN3,RAD51C,ANKRD17,PHLPP1,TEC,PAK1,DHX36,IPO7
response to external stimulus	GO:0009605	0,00005231652761	45	STRBP,ANKRD27,TBK1,UBASH3B,MAP3K5,ENPP3,CLEC16A,LGALS8,CRB1, GLRB,PPARA,P2RX4,CEP192,RHOA,ATG5,PDZD2,RASGRP1,RAC1,EIF2AK1, KYAT1,ZYX,STXBP4,JCHAIN,USP53,PLEKHM2,BOC,CBX7,GAB1,COCH,DAPK1, FANCA,AOAH,NPNT,C7,IFIH1,WNT3A,HRAS,CAPN3,NRP1,NRG3,ANKRD17,SLIT2, PHLPP1,DHX36,IPO7
immune system process	GO:0002376	0,0002485224227	43	FLVCR1,UBA5,MYO1E,PLD4,TBK1,UBASH3B,ENPP3,LGALS8,GPNMB,ATP11C, SIX4,TFRC,TRAT1,SPPL2B,RHOA,ATG5,RASGRP1,RAC1,EIF2AK1,ZYX,STXBP4, JCHAIN,PLEKHM2,CTNNB1,COCH,DAPK1,FANCA,IRAK1BP1,C7,IFIH1,UFL1, PLCG1,WNT3A,HRAS,LEPR,CACNA1C,ANKRD17,SLIT2,PHLPP1,TEC,CREB1, DHX36,IPO7
regulation of catabolic process	GO:0009894	0,0007657388642	24	TRAF5,TBK1,CLEC16A,KIF25,PPARA,WDR91,SUMO1,AQP11,PNPT1,VPS13D, DACT1,PSMC6,SGSM3,NEDD4,DAPK1,EGF,HERC1,UFL1,NBAS,ASB9,SLC25A4, LRPPRC,LEPR,DHX36
reproduction	GO:0000003	0,006569724938	25	MYCBPAP,STRBP,SIX4,IQCG,FANCL,GLRB,INPP5B,VPS13A,FNDC3A,WEE2,TH, ADAD1,CTNNB1,BRDT,FANCA,NME8,HFM1,TFAP2C,SELENOP,SMC3,RAD51C, LEPR,TAF4,SLIT2,DHX36
gical process involved in interspecies interaction between organ	GO:0044419	0,007656893518	22	SMARCB1,TBK1,LGALS8,CEP192,RHOA,ATG5,RASGRP1,KYAT1,ZYX,STXBP4, JCHAIN,PLEKHM2,COCH,NEDD4,DAPK1,C7,IFIH1,HRAS,NRP1,ANKRD17,DHX36, IPO7
circulatory system development	GO:0072359	0,01032228054	24	PROX1,FLVCR1,MYO1E,CRB2,LGALS8,RHOA,AGO3,ATG5,TH,CTNNB1,GAB1, EGF,PLCG1,ROCK2,MAP3K3,WNT3A,MYBPC3,NRP1,LEPR,CACNA1C,ANKRD17, SLIT2,CREB1,DHX36
hepatocyte growth factor receptor signaling pathway	GO:0048012	0,01149124281	3	RAC1,NRP1,PAK1
defense response	GO:0006952	0,01465043969	22	PLD4,TBK1,ENPP3,LGALS8,PPARA,RASGRP1,EIF2AK1,ZYX,STXBP4,JCHAIN, PLEKHM2,COCH,DAPK1,FANCA,AOAH,C7,IFIH1,DLEC1,HRAS,MYLK3,ANKRD17, IPO7
organophosphate ester transport	GO:0015748	0,01656475837	7	ATP11C,ABCC5,SLC35B2,SLC17A9,OSBPL2,ATP11A,SLC25A4
mitotic cell cycle process	GO:1903047	0,01727762692	16	DYNC1H1,KIF25,GPNMB,NAE1,KNTC1,CEP192,RHOA,TOM1L2,STIL,KIF20A, DACT1,EGF,SMC3,RAD51C,NCAPG,ANKRD17
interferon-alpha production	GO:0032607	0,01727762692	3	TBK1,IFIH1,DHX36
ubiquinone metabolic process	GO:0006743	0,02100251913	3	COQ6,COQ9,COQ3
chemical synaptic transmission	GO:0007268	0,02127127929	14	CYFIP1,GLRB,P2RX4,TH,CTNNB1,ITPR3,PLCG1,WNT3A,HRAS,UNC13C,PTCHD1, ITPKA,PPP1R9A,NRG3
tissue remodeling	GO:0048771	0,02173939945	7	UBASH3B,GPNMB,TFRC,CRB1,ATG5,CTNNB1,LEPR
locomotion	GO:0040011	0,02252432364	31	PROX1,CRB2,LGALS8,GPNMB,SIX4,IQCG,P2RX4,ARF4,RHOA,INPP5B,PDZD2, VPS13A,WWC1,RAC1,CELSR1,BOC,CBX7,CTNNB1,GAB1,NME8,EGF,ACAP3, PLCG1,ROCK2,MAP3K3,WNT3A,HRAS,NRP1,NRG3,SLIT2,PAK1
muscle structure development	GO:0061061	0,0267082668	14	PROX1,SIX4,RHOA,ATG5,LMOD2,BOC,CTNNB1,NPNT,WNT3A,ATP11A,MYBPC3, CAPN3,ANKRD17,CREB1
cellular calcium ion homeostasis	GO:0006874	0,0281040751	10	UBASH3B,P2RX4,CHERP,ATG5,PACS2,ITPR3,ADCY5,PLCG1,CAPN3,CACNA1C
behavior	GO:0007610	0,03220405041	14	STRBP,GLRB,GPR176,ARF4,VPS13A,TH,CELSR1,ITPR3,ZDHHC8,SELENOP, ADCY5,PTCHD1,LEPR,CREB1
regulation of respiratory burst	GO:0060263	0,03840462882	2	RAC1,JCHAIN
heart contraction	GO:0060047	0,0396681173	7	P2RX4,CACNA1D,SUMO1,ATG5,TH,ATP1A1,CACNA1C
eye photoreceptor cell development	GO:0042462	0,04234538582	3	CRB2,CRB1,TH
cellular divalent inorganic cation homeostasis	GO:0072503	0,04391777767	10	UBASH3B,P2RX4,CHERP,ATG5,PACS2,ITPR3,ADCY5,PLCG1,CAPN3,CACNA1C
fatty acid metabolic process	GO:0006631	0,04495075639	9	ACOT12,ACSBG2,PPARA,EPHX1,ACOX3,AOAH,ACOT7,ASAH1,PTGR2
cellular pigmentation	GO:0033059	0,04803600723	3	ANKRD27,MREG,KIF13A
glycerol metabolic process	GO:0006071	0,04988690227	2	GK5,COQ3
morphogenesis of a branching epithelium	GO:0061138	0,03716719742	8	PROX1,SIX4,CELSR1,CTNNB1,NPNT,EGF,SLIT2,PAK1
cellular localization	GO:0051641	0,000003537184997	59	PKDCC,DYNC1H1,ANKRD27,UBASH3B,SERAC1,KIF25,SIX4,TFRC,CRB1,GLRB, WDR91,NUP205,P2RX4,CHERP,CEP192,ARF4,SUMO1,TRAK2,ATG5,MREG, PACS2,RASGRP1,RAC1,AQP11,TOM1L2,TRIM29,STIL,STXBP4,LIMD2,IPO5, CELSR1,LMAN1,PLEKHM2,CTNNB1,STARD3NL,NEDD4,ITPR3,COPB2,EGF, KIF13A,OSBPL2,ADCY5,PLCG1,ROCK2,AKAP11,STX6,WNT3A,HRAS,AP2A2, SRSF10,UNC13C,EXPH5,PEX16,CAPN3,SEC24C,APPBP2,TLK1,XPO4,IPO7
regulation of signaling	GO:0023051	0,00001120879736	63	PROX1,TRAF5,NCLN,SMARCB1,RALGPS1,UBA5,CRB2,TBK1,UBASH3B,CYFIP1, MAP3K5,CLEC16A,SIPA1L2,GPNMB,OPA1,TFRC,TRAT1,RGS9,PPARA,P2RX4, CHERP,RHOA,CBFA2T2,STK40,RASGRP1,WWC1,RASA2,MDFIC,STXBP4,DACT1, SGSM3,CTNNB1,MBIP,NEDD4,MAGI3,ITPR3,DAPK1,FANCA,NPNT,EGF,PRDM16, MAP4K3,GRB14,STK36,ADCY5,UFL1,PLCG1,MAP3K3,WNT3A,HRAS,UNC13C, ITPKA,CAPN3,SHOC2,PPP1R9A,NRP1,NRG3,ANKRD17,SLIT2,PHLPP1,CREB1, PAK1,DHX36
cellular component biogenesis	GO:0044085	0,01053640548	48	PROX1,FBLN5,DYNC1H1,ANKRD27,CYFIP1,SIX4,OPA1,TFRC,IQCG,NUP205, CHAF1A,CEP192,RHOA,SUMO1,AGO3,ATG5,PACS2,RAC1,AQP11,LMOD2,IQUB, ZYX,STIL,JCHAIN,WEE2,CRNKL1,PNPT1,TMEM41B,DACT1,CTNNB1,NME8, OSBPL2,MYO10,STK36,CUL4B,ROCK2,SLC39A12,WNT3A,HRAS,SRSF10,CDKL5, PEX16,CAPN3,SMC3,NRP1,ASAP1,SLIT2,PAK1

anatomical structure morphogenesis	GO:0009653	0,0165389987	44	PROX1,FLVCR1,MYO1E,CRB2,PKDCC,ANKRD27,CYFIP1,LGALS8,SIX4,OPA1, TFRC,CRB1,PPARA,FOXN4,RHOA,STK40,AGO3,RAC1,LMOD2,TH,PNPT1, CELSR1,BOC,CTNNB1,GAB1,COCH,NPNT,EGF,HERC1,MYO10,PLCG1,ROCK2, MAP3K3,WNT3A,MYBPC3,CAPN3,NOLC1,NRP1,LEPR,NRG3,CACNA1C,SLIT2, CREB1,PAK1
cell junction	GO:0030054	0,0006013854666	37	MYO1E,CRB2,CYFIP1,CRB1,GLRB,PDZRN3,RGS9,P2RX4,ARF4,RHOA,RAC1, GRIK4,ZYX,ARR3,TH,USP53,SGSM3,CTNNB1,GAB1,MAGI3,TP53BP2,ATP1A1, PLCG1,WNT3A,HRAS,UNC13C,CDKL5,PTCHD1,SLC7A2,MCMBP,RAD51C, PPP1R9A,NRP1,NRG3,CACNA1C,ASAP1,PAK1
autophagy	GO:0006914	0,001862677973	15	UBA5,TBK1,CLEC16A,KIF25,LGALS8,ATG5,VPS13A,PACS2,TMEM41B,VPS13D, DAPK1,HERC1,UFL1,SLC25A4,LEPR
response to hormone	GO:0009725	0,02567556464	14	UBA5,PPARA,RHOA,STXBP4,TH,NEDD4,ATP1A1,GRB14,UFL1,SLC2A1,LEPR, SLIT2,CREB1,PAK1