UNIVERSITÀ POLITECNICA DELLE MARCHE

UNIVERSITÉ DE STRASBOURG

Ph.D. Course in Life and Environmental Sciences (UNIVPM)

Ph.D. Course in Life and Health Sciences (UNISTRA)

ONGOING SELECTIVE FORCES DRIVING KING PENGUIN EVOLUTION

Ph.D. Candidate:

Flávia Akemi
NITTA FERNANDES

Supervisors:

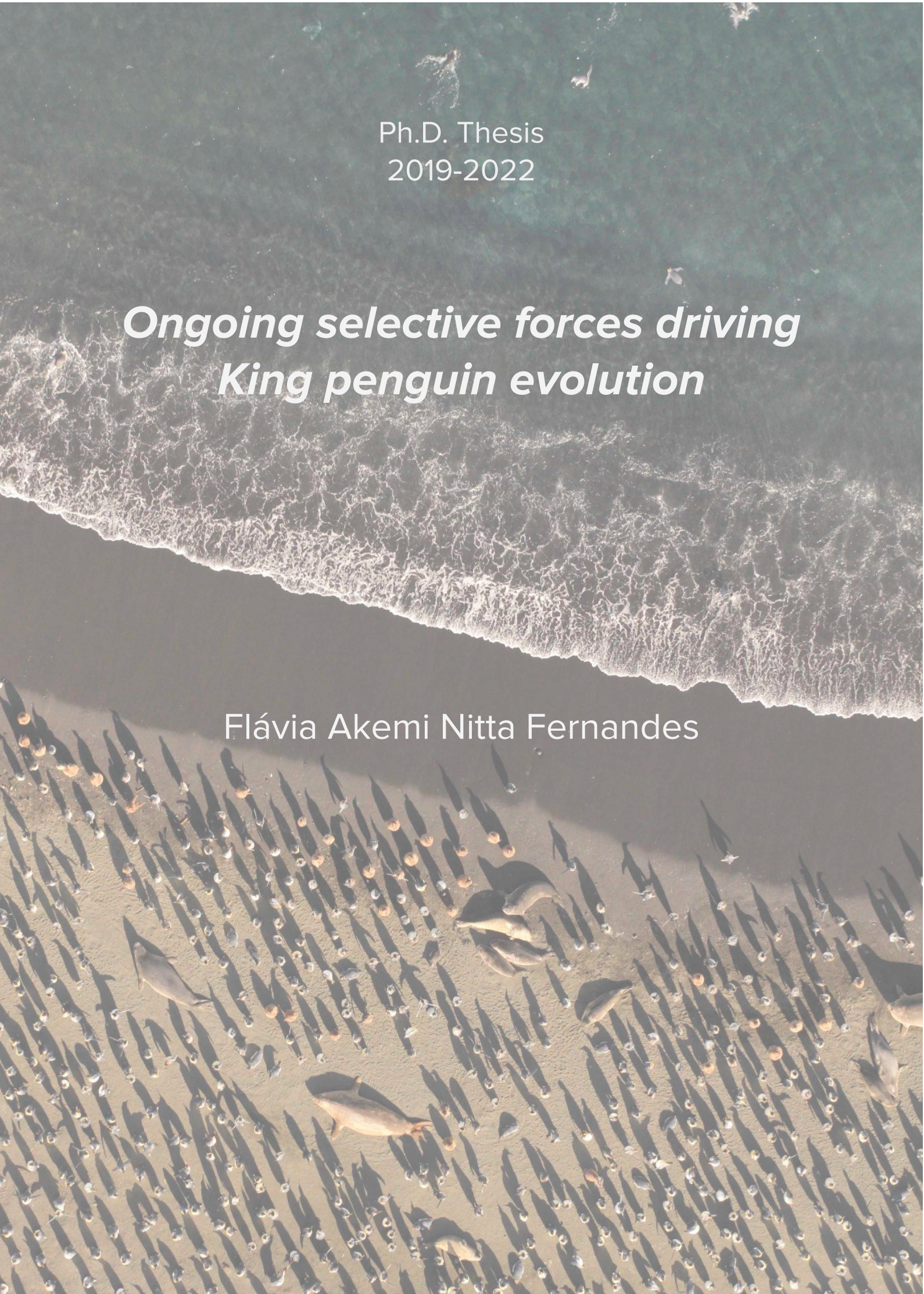
Prof. Emiliano TRUCCHI
(UNIVPM)

Dr. Céline LE BOHEC
(UNISTRA)

XXXV cycle
2019-2022

Ph.D. Thesis
2019-2022

Ongoing selective forces driving King penguin evolution



Flávia Akemi Nitta Fernandes

Acknowledgments

The amount of people I would like to thank for their academic and emotional support during these three years of Ph.D. cannot fit in a page or two, but I will try.

I cannot start this acknowledgment without thanking the two people who gave me unconditional support throughout my life and through this process, my parents, Elisa and Edson *Obrigada por sempre estarem no último vagão do trem*, independently of where we are in the world. For some mysterious reason you never questioned my dream of working with penguins, so here we are.

I would like to thank my dear supervisors, Emiliano and Céline. Three years after the moment I decided to take my two bags from Brazil and start this journey, I could not be more sure about how lucky I was to have them as tutors. Thanks to their support and guidance through all the way, and especially during the difficult moments.

A thank you to all my family who recharged my energies to finish this Ph.D. after three years apart. For Batian and her handmade sushi and origami that gave me strength. To Tim, for always being light and oasis even with the distance.

Thank you to Claudine, Gilles, Lucille, Bruno, and Marie Thérèse, for being my family and support through this period. Thanks to Josie, Joan, and Luisa who are a family in Ancona, to my housemates and former housemates, Umberto, Nicola, Federico, Alessandro, and Iva, and to my almost housemates Leo and Alessandra.

To my long standing friends from the motherland, Larissa and Bruna, thanks for always being there, even when I was in Greenland. To my dear friend Érica, who was there all the time, “na saúde e na doença”. To Rodrigo, Bruna, Lucas Canesin, Jair, Pedro, Helena, Simone, Thais, my dear friends from Brazil who even with the long distance were present in this process. A special thanks to Letícia, for following me and bringing me sanity during this whole process.

To my friends, that are much more than lab colleagues, Lorena Francesco, and Federica, you gave light and support to moments that seemed a little difficult and showed me that the Ph.D. can be very funny when you have good friends around, and a billiardino table. Thanks to the other friends and colleagues that passed through our lab, Sebastiano, Tijana, Daniela, and Masoud. Thanks to Piergiorgio for showing that late chicks are the best. Thanks to Alessandra, who is one of few other Brazilian people in the world that also decided to do a Ph.D. in Ancona, for my luck.

To my friends, a special thanks to Gaël and Robin, with whom it was a pleasure to share the field, the boat, and the work. Thanks to Téo and his cafézinhos. To Manfred and Nina for the friendship and good moments in the field and in life.

For all the aid in the field work, I would like to thank Pierre and Elodie, who overwintered for the program 137 in Crozet and, without their work, this thesis would not have been done. To Benoit, for all the help and good moments shared in the freezers of Strasbourg and outside, apart from all his contribution for the 137 Program. To all the other great people who I had the chance to meet during the two missions in the TAAF, Sandra, Julie, Léo, Ralph, Cécile, Pauline, Mathis, Rémi, Tatiana, Michel, Elodie, Natasha, Laura, Théo, Yoann, Lucie, Alex, Thibault, Mathilde.

Thanks to Paolo Gratton and JP Robin, who followed and helped the development of this project and my life as a PhD. Thank you to Alessio Iannucci for the library preparation and sequencing, to Samuele Greco and Marco Gerdol for the help with the transcriptomic analyses.

"*...gli errori sono le chances che ha la vita di cambiare*"
E.T.

Table of contents

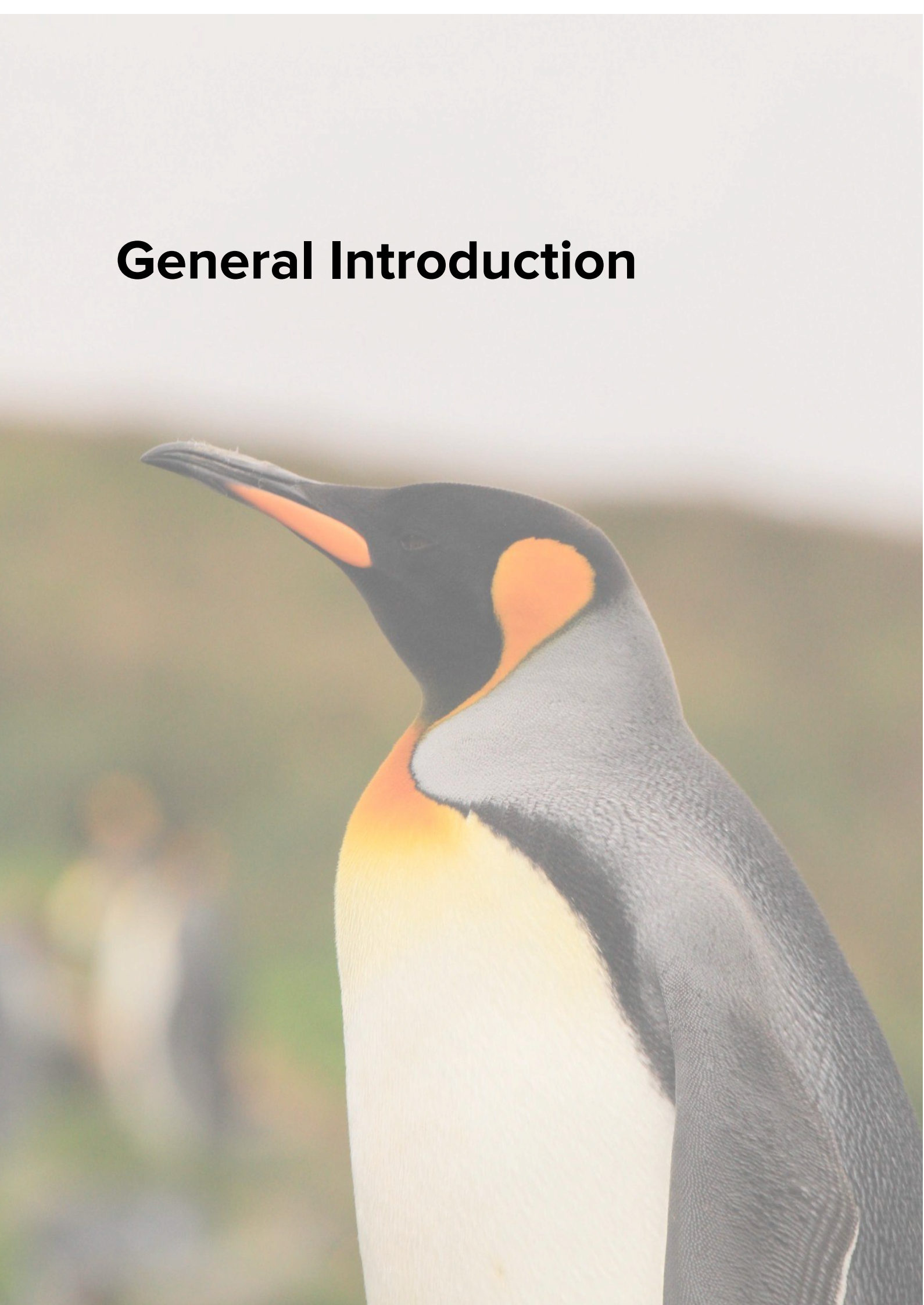
Abstract

According to life history theory, the most energetically costly activities should match the period of highest resource abundance, which otherwise will result in a “mismatch” interaction (i.e., the match-mismatch hypothesis). A possible interpretation of the match-mismatch hypothesis is that a mismatch refers to a reduction in the fitness of the individuals phenologically out of its optimum timing. However, when external environmental conditions are unpredictable, or when natural systems become more asynchronous, such as under climate change, mismatches could provide adaptive responses at the population level. Under this scenario, the main objective of this thesis is to assess whether individuals born under mismatched environmental conditions can increase the adaptive potential of the whole population. To do so, we used genomes, transcriptomes, and life history data derived from individuals born under match and mismatch conditions in a wild population of king penguins (*Aptenodytes patagonicus*). At every breeding season, two phenological peaks of hatching take place in king penguin colonies, generating two phenological groups or chicks, early and late, which are born in match and mismatch conditions, respectively. Our results indicate that late-born chicks have a lower accumulation of highly deleterious mutations. This finding is independent of the survival of the chick past the first winter selection. In contrast, we detect a higher accumulation of highly deleterious variants in early chicks, caused by a high frequency of these mutations in chicks that do not survive during the first year of life. These results suggest that the entire late-chick group starts with a lower genetic load threshold in comparison to the early-chick group. Indeed, the early-chick group is more heterogeneous at birth, likely due to a weaker selective pressure when adults breed under matched conditions when there is a peak of resources. We also find evidence of an upregulation of genes related to growth efficiency and tumour suppression in the late-chick group. Late chicks are born a month after their early conspecifics and consequently have less time to grow and accumulate enough body mass until the beginning of the winter, when all chicks pass through a period of long fasting. The gene expression patterns observed in late chicks hints at possible plastic adaptations to grow fast, and also likely to the oxidative molecules produced in this process. Finally, life history data shows patterns indicative of equal post-fledging survival in both phenological groups, even if late chicks fledge at smaller sizes and at lower body condition. Considering future climate change scenarios, in which the King penguin is also predicted to be subjected to drastic habitat losses, mismatched adaptations, such as the ones detected here, could represent an adaptive strategy that avert population from extinction under unpredictable and variable environments.

Key words:

Match-mismatch, genetic load, phenotypic plasticity, viability selection, predictive adaptive response, fitness, King penguin

General Introduction



1. Life on a constantly changing Earth

1.1. Geological and climatic drivers of biodiversity

Environments are constantly changing in a natural manner. Since the origins of life on Earth to the present day, our planet has passed through several modifications in atmosphere composition, temperature, sea-level, and many other abiotic conditions (Crowley 1983). Such modifications have historically pressured organisms that thrived under previous environments to cope with new ones. As a general consequence, individuals that are not able to survive and reproduce under the new conditions (i.e., individuals that are not already adapted or will not adapt fast enough) will die before leaving descendants. At the population level, if a significant proportion of individuals is not adapted to the novel conditions, the population's fate is to decline. In the most severe aftermath of population decrease, the whole species can disappear through an extinction process (Hallam 1987; Hallam and Wignall 1999; Stanley 2016; Bond and Grasby 2017).

A remarkable example of massive species disappearance, estimated to have been the major extinction event on Earth (i.e., >80% of species extinct), was related to changes in the Earth's atmospheric composition after the Great Oxidation Event (GOE) (Hodgskiss et al. 2019). The atmosphere composition before the GOE, which was characteristic of the Archean eon (4 to 2.5 billion years (Ga) ago, one-third of Earth's history), contained only a negligible portion of O_2 (less than 10^{-6} times the present O_2 concentration, Zahnle et al. 2006), while being mostly composed by gases such as CO_2, CH_4, and N_2 (reviewed in Catling and Zahnle 2020) (**Figure 1**). It is believed that such a weakly reducing anoxic atmosphere would have restricted life to unicellular and prokaryotic organisms, mostly assembled in aquatic microbial mats (i.e., a type of biofilm formed by archaea and bacteria) (Lepot 2020). Only after the GOE (from 2.5 to 2.0 Ga), which was responsible for the first significant rise in atmospheric O_2 concentrations (see **Figure 1**), multicellular and eukaryotic organisms diverged and colonised the planet (Lyons et al. 2014).

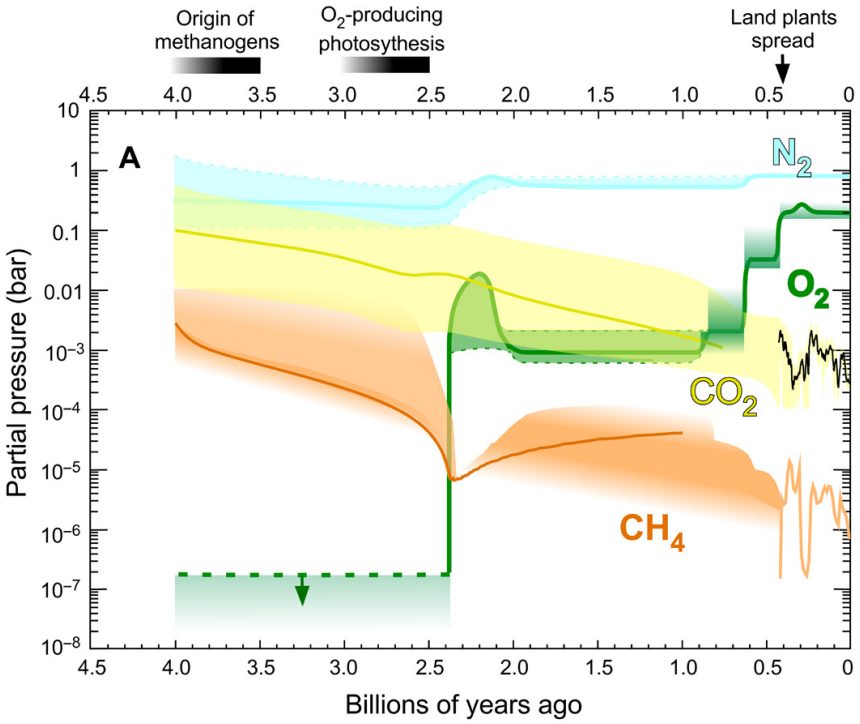


Figure 1. Overview of the post-Archean atmospheric evolution adapted from Catling and Zahnle (2020). The graph shows the partial pressure of four atmospheric gases (N_2 in blue, O_2 in green, CO_2 in yellow, and CH_4 in orange) during the last 4.5 billion years inferred by rocks, with the degrees of uncertainty shown by the coloured shadows around the solid lines.

While the higher O_2 availability allowed the radiation of the diversity of life forms known nowadays, at the same time, it also caused a mass extinction of previously existing species (Hodgskiss et al. 2019), which were challenged with strong metabolic changes (Chen et al. 2020). From what is estimated by studies on present day hydrothermal vents, the pre-GOE atmosphere was also rich in reduced arsenic (As) species (Zhu et al. 2014). Such reduced As species are highly toxic to most of the extant organisms, but were likely non-lethal to the Archean life (Oremland et al. 2009).

After the GOE, the increased amount of dissolved O_2 would have generated a higher abundance of oxidised As species, pressuring Archean organisms to adapt to the new metabolite availability (Zhu et al. 2014). By using a molecular clock analysis, a previous study showed that some Archean microbial mats' could have been able to metabolise oxidised As owing to a set of As detoxification genes (Chen et al. 2020). This study also showed that the same set of As detoxification genes were evolutionarily maintained in more recent lineages, including extant bacteria, archaea, and even eukaryotes.

This example shows that, while new conditions may extirpate populations and species that do not adapt fast enough, they also allow the appearance of new adaptations. Genes that originated under a specific environmental pressure (i.e., As-rich atmosphere) were conserved through many present day lineages, which do not necessarily face the same pressures of when the adaptation originated. However, such adaptation may have been maintained due to other similar stress sources that still exist nowadays, such as heavy metal pollution in the case of As detoxification (Chen et al. 2020).

This process, in which a current adaptation originates before the existence of the current pressure, is also known as adaptation through standing variation, a subject that we will come back to in the next sections. Moreover, as it is shown by the GOE example, the perpetuation of adaptations and the persistence of species under new conditions will happen when at least part of the population is able to adapt to the novel environment and generate offspring that will carry the adaptation (Darwin 1859).

The aforementioned atmospheric changes represent exceptional modifications that happened throughout long geological periods and had a drastic impact on the life forms and ecosystem composition of our planet (Ligrone 2019). In addition to such changes, species are more frequently subjected to other major fluctuations that affect the globe in a more cyclic manner, such as climate changes (National Research Council et al. 2008). For example, our planet has experienced at least five major long periods of surface temperature reduction (reviewed in Adamo et al. 2021). These so-called ice ages or icehouse periods intercalate with greenhouse periods, which are characterised by a rise in greenhouse gases' levels (e.g., CO_2, CH_4, water vapour) and sea surface temperatures, leading to the absence of glaciers in the whole planet.

If we only consider the temperature oscillations since the Cambrian period, which is marked by the origins of modern multicellular fauna and flora (approximately 500 million years ago) (Butterfield 2007), greenhouse periods have been more frequent than icehouse ones (**Figure 2**). The beginning of this period was likely marked by a greenhouse climate (Hearing et al. 2018), which was also characteristic of two other global warm peaks in the last 100 million years. Of the latter, the first and more extreme peak is estimated to have occurred during the Cretaceous (around 92 million years ago), while a less extreme rise in temperatures happened during the Paleocene-Eocene (around 54-56 million years ago) (Dunkley Jones et al. 2013; Scott and Lindsey 2020).

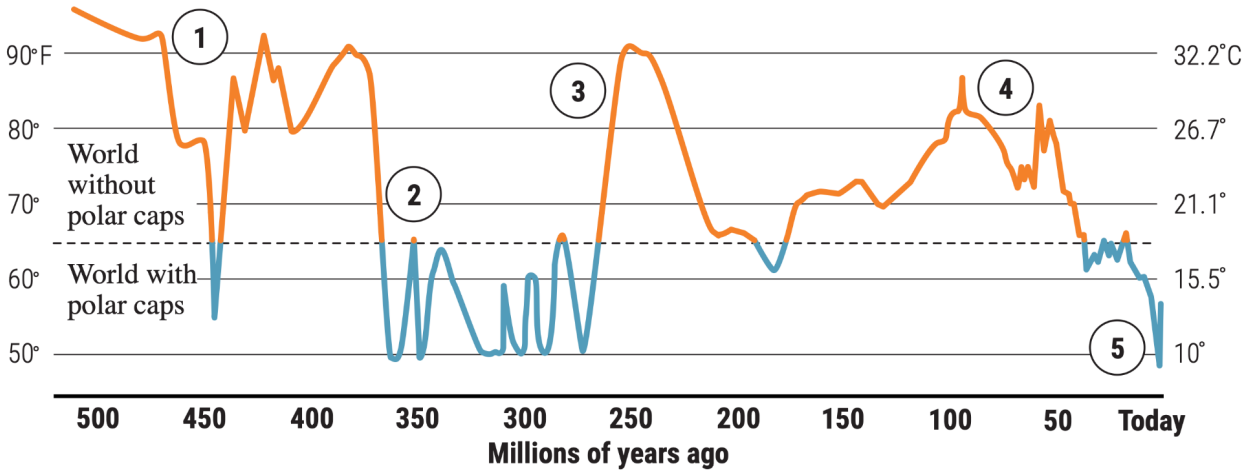


Figure 2. Earth's average surface temperature curve during the past 500 million years, adapted by N. Desai/SCIENCE from the Smithsonian Institution National Museum of Natural History (Voosen 2019). Numbers indicate 1) marine life diversification; 2) divergence of land plants, which started absorbing higher levels of carbon dioxide, and formation of the modern polar ice caps; 3) divergence of mammals; 4) divergence of humans; and 5) onset of current climate warming.

The warmer average temperatures and the absence of polar ice shelves during greenhouse periods would have allowed even cold-blooded warm-adapted species to inhabit polar regions, as in the case of turtle and plesiosaurs (extinct marine reptiles) fossils from the Cretaceous found in Siberia at 66–72° of latitude (Zverkov et al. 2023). In addition, pollen, spore, and other organic assemblages recovered from ocean sediments have revealed that the early Eocene Arctic vegetation was composed by many present day subtropical taxa, including Arecaceae (palm family) and Bombacoideae (balsa, baobab subfamily) species (Willard et al. 2019). Such examples illustrate how intensely global climate shifts can impact species distribution and ecosystem composition (i.e., on a spatial scale).

1.2. Spatial and temporal scale impacts of environmental changes

Spatial scale impacts involve expansion and contraction of species' range, as well as distributional shifts due to species' tracking of specific environmental conditions, as in the cases outlined above. For instance, when organisms cannot adapt to temperatures out of their physiological range, they suffer the pressure of shifting to a cooler (in the case of a climate warming) or warmer (in the case of a glaciation) environment (Parmesan and Yohe 2003). This is especially the case of ectotherms (i.e., cold-blooded organisms), which represent an extreme case of restrained physiological limits. However, even if ectotherms are more sensitive to temperature maxima, all species are subjected to the pressure of tracking the environmental conditions that fall within their adaptive limit (Moritz et al. 2008; Vitasse et al. 2021).

During climatic maxima events, species can concentrate in regions that are relatively buffered from the climatic changes taking place globally (i.e., climate change refugia) (Ashcroft 2010). However, the unprecedented pace of contemporary climate changes and habitat changes are especially worrying as not all individuals have the dispersal capacity to reach a climate refugium, which leads to the local or global extinction of many species (Díaz et al. 2019; Exposito-Alonso et al. 2022). A study using data from 538 animal and plant taxa has estimated that 57–70% species will be unable to disperse to a climate refugia under the current climate changes, based on their past rates of dispersal (Román-Palacios and Wiens 2020).

Local extinctions due to current warmer temperatures are already widespread in both terrestrial and marine ecosystems (Pinsky et al. 2019), always having a higher impact on ectothermic species due to their lower efficiency in regulating physiological performance with a higher range of temperature changes (Deutsch et al. 2008; Somero 2010; Pinsky et al. 2019). The main causes of current climate change-related species extinctions, however, are not due to direct factors, such as the challenge of temperature adjustment (Parmesan 2006). A recent study has shown that, although the local extinction probability of 11 large terrestrial mammalian species in China increases with temperature, this variable is not the direct cause of such extinctions (Wan et al. 2019). Instead, extinctions have been caused by

the rise in human population density, which also corresponds with a period of temperature increase (Wan et al. 2019). Their estimates of local extinction sensitivity showed that if human density reached an extreme of 400 people per square kilometre, extinction rates would range from 34% to up to 99% within a period of 50 years. This is because high density in human populations is linked with higher poaching, road kills, habitat loss, deforestation, among other land-use changes caused by local scale activities (Kramer-Schadt et al. 2004; Jiang et al. 2014).

As shown by the previous example, climate changes will affect natural systems through complex, indirect, factors that lead to disruption or breakdown in species interactions. Among the disturbances in species interactions that lead to the majority of anthropogenic local extinctions, prey-predator trophic relationships are one of the most affected (Cahill et al. 2013). Prey-predator trophic relationships may be affected in three main ways. First, a species decline can lead to the decrease of a second species, which depends upon the first one (e.g., prey or food resource) (e.g., Durance and Ormerod 2010; Schweiger et al. 2012). Second, an increase in predators (Harley 2011) or in other species that may have negative effects on the focal species (e.g., pathogens and competitors) can also cause the decline of the focal species (Benning et al. 2002; Pounds et al. 2006; Suttle et al. 2007). Finally, a temporal mismatch between two interacting species can lead to a mistiming of activities (e.g., reproduction, migration) with resources (Visser et al. 1998).

Temporal scale impacts of climate changes hence represent a more indirect but widespread disturbance in ecosystems: new climatic conditions lead to a desynchronization of relationships among species (e.g., trophic, symbiotic) and/or between species and abiotic factors (e.g., rain, snow, light rhythm). Abiotic events that usually take place at a specific time of the year, as, for example, the meltdown of winter snow, can happen in advance due to the average increase in global temperatures. Consequently, species that rely their yearly activities on the timing of such events, as, following the previous example, the start of a new breeding season in ground squirrels (*Urocitellus richardsonii*), have to adjust their phenology (i.e., the timing of cyclic biological events, and the biotic and abiotic causes of their timing) to the new conditions in order to be less affected (Sheriff et al. 2011; Kucheravy et al. 2021).

In the context of contemporary climate changes, the timing of migration events is also being affected in different terrestrial (Walther et al. 2002; Saino et al. 2011; Mayor et al. 2017) and marine species (Ramp et al. 2015; van Weelden et al. 2021). A long-term study with two sympatric whale species revealed an earlier arrival of individuals to feeding grounds over a 26 years' period, which the authors relate to earlier ice-breaks and, consequently, bloom of primary productivity (Ramp et al. 2015). Even though these whale populations were able to adjust their annual cycles following changes in external cues, showing some degree of plastic response, this is not always the case for other species.

Another study investigated changes in migratory arrival dates in relation to the vegetation green-up (i.e., a proxy of food availability after winter) for 48 songbird species in North America (Mayor et al. 2017). This study detected a phenological tracking response for

39 out of 48 species, meaning that most birds adjusted their arrival times towards the direction of the green-up over a 12 year period. Despite the phenological adjustment of most species, nine out of the 48 did not show the same plasticity, and, consequently, had an increased lag between migration arrival and food peaks.

There is plethora of evidence of unequal phenological shifts in different species as a consequence of current climate changes, resulting in the desynchronization of interspecies' interactions (e.g., Hughes 2000; Walther et al. 2002; Parmesan and Yohe 2003; Dunn 2004; Marvelde et al. 2011; Kharouba et al. 2018). As mentioned before, most of the affected species-to-species interactions consist of trophic relationships between prey availability and the life cycle of its predator (Edwards and Richardson 2004; Visser and Both 2005; Twining et al. 2022). A classic example of such trophic mismatch concerns the availability of winter moths (*Opheroptera brumata*) and its predators, the great tit (*Parus major*) (Perrins 1970; Visser et al. 1998; Buse et al. 1999) and flycatchers (*Ficedula spp.*) (Both and Visser 2001; Sanz et al. 2003), two common insectivorous birds. These birds depend on the abundance of moth caterpillars in order to feed their nestlings during the breeding season, while caterpillars are only available for a short period of time during spring (Visser and Both 2005). For this reason, the peak of caterpillars constrains the reproductive success of tits and flycatchers.

Long-term studies on both great tit and flycatcher populations have shown that warmer spring temperatures lead to an anticipated phenology of both birds and their prey (Visser et al. 1998; Charmantier et al. 2008). However, caterpillar advancement is usually higher than the advancement of the bird's breeding cycle, causing a trophic desynchronisation between predator and prey (Visser et al. 1998; Both and Visser 2001; Cresswell and Mccleery 2003; Sanz et al. 2003; Both et al. 2009). Consequently, the mismatch between the bird's breeding activity and the caterpillar's peak of abundance results in reduced clutch sizes and poor body conditions at fledging (Perrins and McCleery 1989; Van Noordwijk et al. 1995; Verboven et al. 2001).

On the other hand, previous studies have also shown that such desynchronization and the phenological shifts will not happen in a uniform manner throughout the whole population (Visser and Both 2005; Cole et al. 2021). This happens because other factors are involved in the abundance of caterpillars, and not only temperature changes. The most straightforward factor is the relationship between caterpillar abundance and oak tree (*Quercus robur*) density and health (Wint 1983). As caterpillars are highly abundant in oak foliage, birds will have higher reproductive success in sites with high oak density, as it has been shown by a recent study with great tits (Cole et al. 2021).

In their 60-year study, Cole and collaborators investigated the variation of more than 13,000 great tits' laying date in relation to environmental factors, with a special focus on oak health. The variation in the laying date was mostly related to the health of oak trees, being that individuals anticipated more their breeding and were more successful when reproducing in a healthy tree area. Oak foliage blossoming, in healthy trees, could represent a cue for the birds to start breeding or it could grant higher food availability, although this question could not be answered by the study. This example evidentiates that

climate changes may not affect all individuals from a population in a similar manner. Therefore, fine-scale individual heterogeneity must be accounted for when studying species response to environmental changes.

In summary, environmental changes affect species in spatial and temporal scales, through direct and indirect ways. Although direct impacts are usually more straightforward to quantify, especially under the complexity of natural conditions, most species will be affected in indirect ways (Cahill et al. 2013).

1.3. Climate change impacts on polar regions and seabirds: a close-up

The uncoupling of trophic activities has an even stronger detrimental impact in high latitude ecosystems (e.g., polar regions), where there is an optimum time-window for life-cycle events (e.g., reproduction, moulting) due to the strong environmental seasonality (e.g., Moline et al. 2008; Ji et al. 2013; Kroeker et al. 2020). In such regions, even slight environmental changes can move species out of their phenological optimum and, if organisms are not able to adapt to the new timing, the phenological shift will negatively affect individual fitness (i.e., survival and reproduction) and population dynamics (Visser and Gienapp 2019).

Changes in the timing of primary productivity peaks in such regions are especially concerning, as all the trophic webs are/will be disrupted (Gradinger 1995). Warming climates are causing an earlier retreat of sea ice in several regions of the Arctic Ocean, which affects the phenology of pelagic phytoplankton (Ji et al. 2013). Pelagic plankton bloom, on the other hand, affects the whole pelagic food web, ultimately disturbing the phenology of top predators such as cod and seabirds (Darnis et al. 2012; Moody et al. 2012; Ji et al. 2013).

In the case of seabirds, species with different foraging strategies may respond differently to changes in sea ice and subsequent prey availability. Northern Fulmars (*Fulmarus glacialis*), which are long-distance foragers, seem to be less affected by temporal variation in sea ice conditions, being able to forage in further locations when the local food web is disrupted by less sea ice. Alternatively, the shallow-divers Kittiwakes (*Rissa tridactyla*) show changes in prey type when ice cover is lower (Moody et al. 2012).

In the opposite polar extreme, the Antarctic is also facing pervasive losses of ice sheet extent in the Western sector of the continent (Steig et al. 2009; Schneider et al. 2012; Jun et al. 2020). Sea ice contractions in the western Antarctic Peninsula (WAP) have also coincided with drastic reductions of phytoplankton productivity during summer, likely affecting other mesopelagic fish and local penguin populations (Montes-Hugo et al. 2009). Warming temperatures will likely cause the rearrangement of whole local food webs that are dependent on ice-edge diatom algae, such as the Antarctic krill (*Euphausia superba*), the Antarctic silverfish (*Pleuragramma antarcticum*), and antarctic penguin species such as the Adélie penguin (*Pygoscelis Adeliae*) (Arrigo and Thomas 2004).

Polar regions are also especially touched by habitat losses due to the rapid pace of ice melting, which affects the foraging habits of local species such as polar bears (Boonstra et al. 2020; Robinson 2022). Habitat loss is amongst the main causes of species extinction,

and is of special concern for small-ranged taxa or endemic species (Myers et al. 2000; Pimm et al. 2014). A recent assessment has shown that 89% of seabirds that are affected by climate changes, are also affected by other threats (Dias et al. 2019). Seabirds are amongst the most threatened group of birds (Croxall et al. 2012), and apart from climate change, invasive species and bycatch (i.e., capture of unwanted species by commercial fishing nets) are the top threats for their persistence. In addition to that, other types of human activities are highly detrimental to this and other natural systems, such as overfishing, pollution, among others (Seabloom et al. 2002; Laidre et al. 2015; Trathan et al. 2015).

2. The match-mismatch hypothesis (MMH) and adaptive mismatch

Environmental disturbances can result in phenological shifts (Parmesan and Yohe 2003). Some individuals may be able to track those shifts, at least to some extent (Charmantier et al. 2008; Ramp et al. 2015; Mayor et al. 2017). In other cases, only some individuals in the population will be able to track the changes, while others will not (Visser et al. 2003; Cole et al. 2021). However, what makes a species, or a population, as a whole, able to track environmental changes? The answer to that question is not a simple one due to the complexity of natural ecosystems. We will first focus on constraints that can shape evolution and adaptation from an eco-evolutionary point of view, and in the next section we will detail the mechanisms through which species may adapt.

From an eco-evolutionary perspective, a species can adapt to new conditions if the change is within the limits of its life history (i.e., timing of reproduction and survival) (Pelletier et al. 2009). In this sense, external and internal factors will restrain the flexibility of species to adapt to new pressures. For example, when a predator's reproductive success depends on a seasonal prey availability, the predator will suffer the pressure of synchronising its reproduction with the prey.

This phenological intertwine between resource and consumer availability is known as the match-mismatch hypothesis (MMH) (Cushing 1974; Cushing and Saleem 1982; Cushing 1990), and a schematic representation can be visualised in **Figure 3**. The MMH (Cushing 1974; Cushing and Saleem 1982; Cushing 1990) postulates that individuals must synchronise the most energetically demanding activities (e.g., migration, reproduction) with the peak of environmental resources in order to thrive.

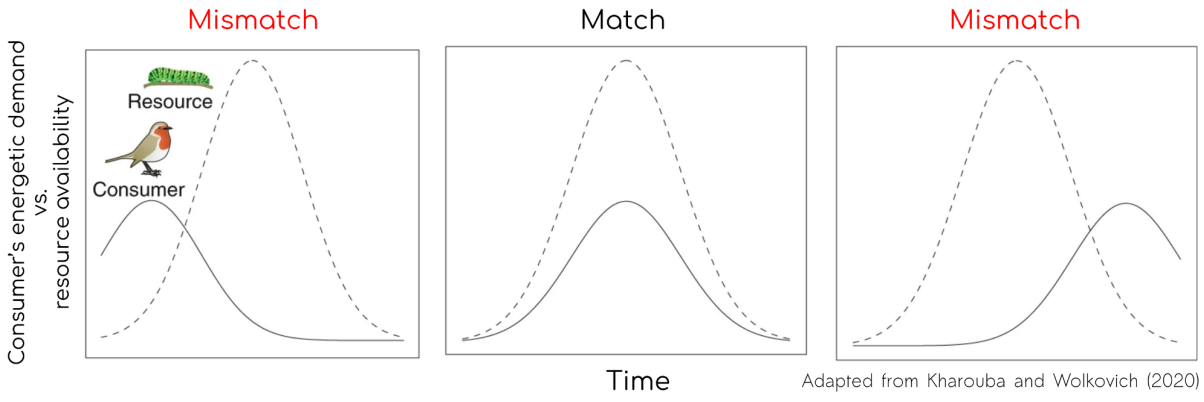


Figure 3. Representation of the match-mismatch hypothesis (MMH), adapted from (Kharouba and Wolkovich 2020). The three panels show the variation in resources (dashed line), and energetic demands of the consumer (solid line) through time.

In a previous work, using data from four high latitude fish species, Cushing observed that the variation in spawning dates were linked to the production cycles of their prey (Cushing 1969). By using long-term data, this study showed that fish species from higher

latitudes had more marked spawning seasons than fish from lower latitudes, which spawned all year round. This restriction in the reproduction of high latitude fish was then related to the fact that the marine productivity at polar regions is more constricted by the strong seasonality of light and wind (Cushing 1969).

Following studies showed that the MMH theory could be applied to other terrestrial and marine systems (Durant et al. 2005; Durant et al. 2007), if two main assumptions were met. First, the consumer's fitness must be in-part controlled by the availability of its resource. Second, the consumer and its resource must show a degree of seasonality, in the way that resource availability will restrict the ideal period of growth and reproduction of the consumer. Under this scenario, a match between consumer and resource could be translated into the maximisation of the total fitness of the consumer (Kharouba and Wolkovich 2020). Consequently, mismatches result in a reduction of consumer fitness.

However, even when the aforementioned assumptions are met, there are cases in which mismatches do not necessarily decrease lifetime fitness. Mostly, this scenario happens when the trade-off between different selective pressures generates a "middle solution" that is fitter in the long-term (Visser et al. 2012; Visser and Gienapp 2019; Petrullo et al. 2023). For example, in a literature review on different bird species' phenology, Visser et al. (2012) showed that even if breeding early in the season allows for a more abundant food supply (i.e., match) (Dunn 2004), some bird species still show higher chick survival later in the season (i.e., mismatch). According to the authors, this *adaptive mismatch* happens because colder temperatures in the beginning of the reproductive season have a stronger impact on chick mortality than lower food abundance at the end of the season (Visser et al. 2012).

In a more recent study, Petrullo et al. (2023) showed that individual red squirrels (*Tamiasciurus hudsonicus*) that always "play safe" by producing smaller clutch sizes (even in years when food availability is high) have a higher total fitness in comparison to individuals that produce big clutches only when food availability is plentiful. This type of mismatch strategy, which sacrifices short-term success in exchange of lower variation in long-term success, can be considered as a bet-hedging response (Seger and Brockmann 1987; Philippi and Seger 1989; Simons 2011). The bet-hedging strategy derives from the logic of "not putting all of your eggs in the same basket", and can be of great value when conditions are unpredictable (Seger and Brockmann 1987). Natural selection may favour bet-hedgers under unpredictable conditions, when environmental cues are not reliable and individuals may apply more conservative life history strategies (Slatkin 1974). However, empirical evidence of bet-hedging in nature is still scarce owing to the difficulty of recognising a bet-hedging strategy in a population, as it will be only adaptive after several generations (Simons 2011). Consequently, this model has been mostly empirically tested in short-lived organisms, such as bacteria (Veening et al. 2008; Beaumont et al. 2009) and annual plants (Childs et al. 2010), while evidence is less widespread for wild longer-lived organisms, such as birds (Nevoux et al. 2010; Capilla-Lasheras et al. 2021).

3. How can species adapt to a changing environment?

The idea that species can go extinct if failing to adapt to environmental conditions has been acknowledged long before the discovery of genetic inheritance and evolutionary mechanisms that regulate genetic variation (Darwin 1859). More than a century later, the relevance of this topic has only increased, as the intensification of global environmental changes poses bigger challenges for species persistence, causing unprecedented rates of extinction across the planet (Parmesan 2006; Pimm et al. 2014; Pyron and Pennell 2022). In this context, one question that can be posed is, how do species adapt to changes in their environment?

According to Gienapp et al. (2007), there are three main ways that a species will respond to environmental changes without going extinct. The first strategy involves a distributional range shift and subsequent colonisation of new environments through niche tracking. The other two strategies reflect *in situ* responses, without a change in the species' distribution range. The second strategy regards species adaptation to the new conditions through plastic adjustments to the environment. Such sort of adaptation is also known as phenotypic plasticity and can be defined as the production of different phenotypes by the same individual genotype under different environments. The third possible response to environmental changes is through genetic adaptation (i.e., involving genetic differentiation specific to each environment) (Gienapp et al. 2007).

The measurement of a species response to changes, by using the three components mentioned above, allows the assessment of the species adaptive potential (Waldvogel et al. 2020). The species adaptive potential will, in turn, help us evaluate its vulnerability and plan for conservation management actions. As the first response regarding rapid niche tracking through species dispersal may be limited for most living species (Román-Palacios and Wiens 2020), adaptations through phenotypic plasticity and inheritable genetic components can be observed in a more generalised set of cases.

3.1. Phenotypic plasticity

Phenotypic plasticity refers to when a genotype produces different phenotypes under different environments. A phenotypic trait can be any morphological (e.g., birds' beak), physiological (e.g., level of corticosteroid hormone under a stressful condition), or behavioural (e.g., personality) trait of an organism (Pigliucci 2001). Another example is the level of gene transcripts under different developmental stages (i.e., gene expression). Even the more complex life history traits (e.g., number of offspring generated in a breeding season) can also be considered as a phenotype.

The production of distinct phenotypic traits will, however, have a limitation under each environment. Such limits of a genotype's phenotypic traits are defined by its reaction norms, as illustrated in **Figure 4**. Reaction norms quantify the change of a phenotypic trait as a function of the variation in an environmental variable of interest, and can provide a

measure of the sensitivity of that genotype (Aubin-Horth and Renn 2009). We can use as an example the rate of fungi growth according to different soil temperatures (i.e., diameter of growth as a function of soil temperature) (McLean et al. 2005). In this example, the authors measured the thermal reaction norms of two fungi species isolated from urban and rural environments with different temperatures. Urban isolated genotypes showed greater tolerance under higher temperatures (i.e., higher reaction norms towards warm temperatures), while rural isolated fungi had higher resistance to lower temperatures (i.e., higher reaction norms towards cold temperatures). This exemplifies that the reaction norms of different genotypes can change in different directions under the same environment.

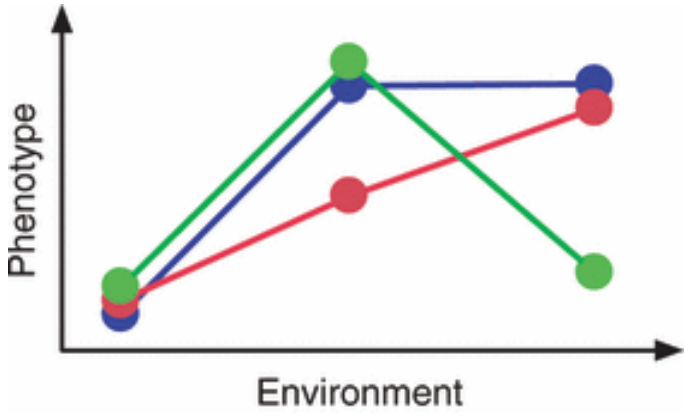


Figure 4. Reaction norm generic representation, from (Aubin-Horth and Renn 2009). The three coloured lines represent reaction norms in relation to the environment, which could also represent different time steps in the development of an organism (y-axis). The red reaction norm shows a progression of the phenotypic value from developmental stages 1 to 3. The green line represents a phenotype that reaches a maximum value during a transition developmental phase (e.g., gene expression level). The blue line shows a phenotype that changes during the transitional phase, but keeps the same value until the end of development.

Compared to genetic adaptation, phenotypic plasticity is usually considered as a faster mechanism of adjustment to novel environmental conditions (Barrett and Schluter 2008). However, adaptation via phenotypic plasticity will only be efficient if it proceeds in the same direction as the new pressure (Ghalambor et al. 2007). In other words, a trait is more likely to be adaptive under a new condition if it goes in the same direction of the pressures (Ghalambor et al. 2007). For example, resistance to colder temperatures is expected to be adaptive if the future trends of the species indicate a shift towards higher latitudes or altitudes, as refugia to climate change (Leonard and Lancaster 2020; De Lisle et al. 2022).

When considering phenotypic plasticity in the context of mismatch, one possible way of identifying adaptations to unfavourable conditions is through the assessment of changes in gene expression between individuals under match and mismatch (Ghalambor et al. 2015). Gene expression represents the process by which the information encoded in a gene is turned into a function (e.g., through the transcription of messenger RNA molecules that can code for a protein). Therefore, gene expression data can be considered as a snapshot of

the physiological status of the individual at the time of sampling (Evans and Hofmann 2012; Philipp et al. 2012).

Consequently, gene expression will also change according to developmental stages and sampled tissues (Cardoso-Moreira et al. 2019). A stage of development that can be highly informative about different responses to pressures is an individual's early-life (i.e., from birth to maturation), as it represents a period during which organisms are especially sensitive to external changes, as their organs and structures are still being developed (Lindström 1999).

In fact, early development conditions play a key role in the determination of the individual's adult phenotype (Pantalacci and Sémon 2015; Silbereis et al. 2016). Stressful early-life conditions can affect the phenotype of the adult and, consequently, its fitness, in both a negative or positive way. Negative effects of facing stressful conditions at birth and early development can be later expressed through a reduction in reproductive success and/or increased adult mortality (Taborsky 2006; Mugabo et al. 2010; Millon et al. 2011; Hamel et al. 2009; Hayward et al. 2013; Pigeon et al. 2019). The detrimental effects of a poor early development environment in the life-time of the individual are known as *silver spoon* effects (Grafen 1988, **Box 1**).

In contrast to the detrimental consequences of stressful early-life environments, an alternative effect is that early-life constraints can provide cues that will allow adult phenotypes to be more efficiently adapted to limiting conditions (Gluckman et al. 2005; Monaghan 2008; Vincenzi et al. 2013). This is known as the *predictive adaptive response* (PAR) hypothesis, and takes place when the newborn faces restrictions that will be also encountered later in adulthood (Gluckman et al. 2005a, **Box 1**).

Considering the increasing asynchrony of natural systems due to climate change (Kharouba et al. 2018), PAR could provide species with adaptive mismatched phenotypes that could reduce extinction risks under more frequently unpredictable environments. While some experiments have shown that such phenological plasticity can be realised by mismatched offspring in the laboratory (De Lisle et al. 2022), studies conducted in wild populations show more pessimistic results (Oostra et al. 2018). More specifically, the contribution of adaptive plasticity to unpredictable conditions, in which environmental cues are not reliable, may be limited by the extent of genetic diversity in the population (Oostra et al. 2018), a topic which is still a matter of debate in our race against species extinction.

Box 1. Outcomes on fitness consequences of early-life conditions

Environmental conditions experienced at birth and/or during early development can have an impact on the later survival and reproductive performance of an individual (i.e., fitness). In this context, different outcomes of early-life conditions can be fitted into two main hypotheses: the silver spoon and the predictive adaptive response hypothesis.

Silver spoon hypothesis (Grafen 1988)
The silver spoon hypothesis posits that favourable early-life conditions will lead to higher

fitness in adult-life (and unfavourable early-life conditions will lead to reduced adult fitness). For example, in an experimental setting in which several clutches of zebra finches (*Taeniopygia guttata*) were submitted to different degrees of parental foraging cost, individuals reared under lower food availability conditions showed shorter lifespans than conspecifics reared under less harsh conditions (Briga et al. 2017).

Predictive adaptive response (PAR) hypothesis (Gluckman et al. 2005a)
The predictive adaptive response, or environmental matching hypothesis, stipulates that environmental cues experimented during early development can influence the development of adaptive phenotypes later in the individual's life. However, this will only hold true if the environmental pressures encountered in adult-life match conditions experienced during early development. In this way, early-life pressures can shape individuals towards an early plastic adaptive response.
For example, in seasons of high population density, when migrating is expected to increase chances of survival, the offspring of the Migratory locust (*Locusta migratoria*), develops wing shapes and metabolism which are better suited to migration (Gluckman et al. 2005b). In this case, even though the offspring does not necessarily express phenotypes that are more beneficial to survival at the larval stage in a high population density scenario, their adult phenotypes will lead to a better fitness when they face high population density events later in life.

3.2. Adaptations at the genomic level

Local and global extinctions cause the permanent loss of biodiversity, which can also be translated into loss of genetic diversity (Exposito-Alonso et al. 2022). Genetic diversity, in turn, provides populations with evolutionary material to deal with new environmental pressures (e.g., some alleles could be adaptive under novel conditions) (Waldvogel et al. 2020). If we consider the evolution of an adaptive phenotypic trait in the population, heritable genetic variation represents a key opportunity to further increase variation throughout generations, whereas phenotypic plasticity could not, by definition, allow an adaptive trait to evolve in the population (**Figure 5**).

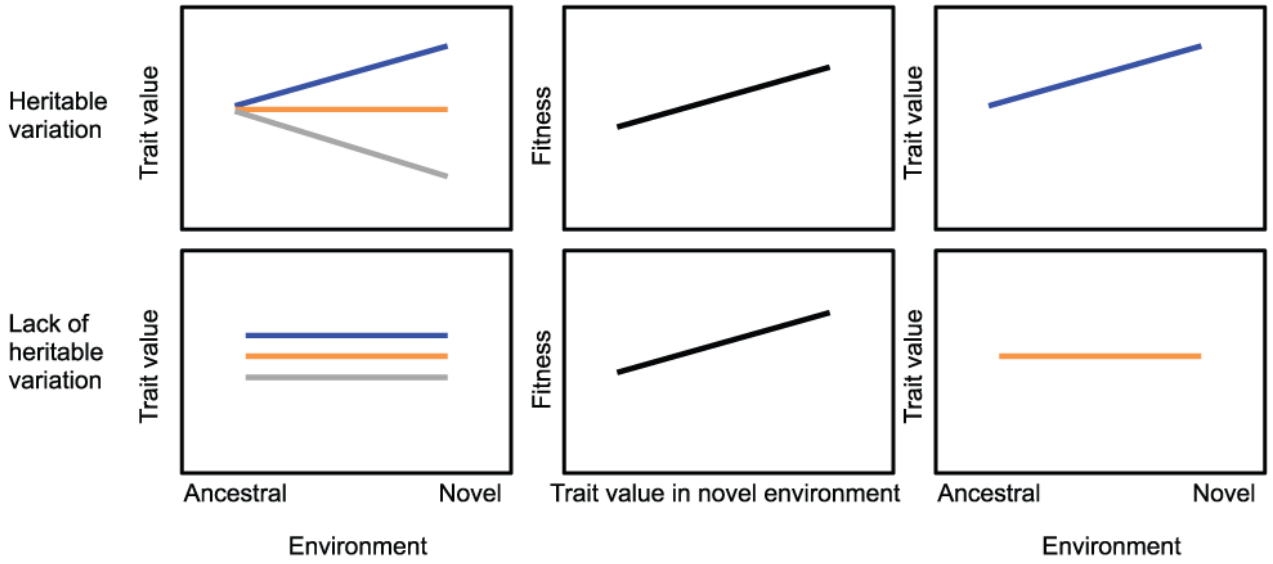


Figure 5. Adaptive trait variation under new environment conditions, from Diamond and Martin (2016). Graphs exemplify the variation in a given trait under new environments when the trait is heritable (top graphs) and when it is not heritable (bottom graphs). Middle column represents the action of selection upon the expressed trait, showing that both heritable and non-heritable variation are adaptive in the case. However, as it is shown by the last column, only the heritable adaptive trait should allow the population to evolve towards a new trait optimum to that environment.

As mentioned above, genetic adaptation is generally considered to proceed much slower than phenotypic plasticity, as new mutations can take several generations to increase in frequency in a population (Barrett and Schluter 2008). However, genetic-level adaptations can also happen in just a few generations in cases where the population has enough pre-existing genetic variation (i.e., standing variation) that allow it to adapt to new conditions (Barrett and Schluter 2008; Bitter et al. 2019). From a conservation genetics perspective (i.e., the realm of population genetics focused on the use of evolutionary and molecular genetics applied to species conservation, Frankham 2010), species holding higher levels of genetic variation have higher chances of harbouring alleles that could be adaptive under new conditions.

As genetic diversity scales with effective population size (Coop 2020), larger populations usually tend to harbour higher levels of genetic variability, therefore, are usually thought to have higher adaptive potential and, hence, lower extinction risk (Kardos et al. 2021). The flip side of the coin is that, although large populations can harbour more potentially adaptive alleles due to high genetic diversity, they also tend to hold higher levels of deleterious variation (i.e., mutations that cause fitness reduction) in heterozygosity, the so-called, masked genetic load (Bertorelle et al. 2022). This is because deleterious mutations with strong impact on fitness are usually recessive and detected at a low frequency in the population. Thus, highly deleterious mutations will be removed by natural selection only when individuals express them in homozygosity (i.e., realised genetic load) (Bertorelle et al. 2022). Masked genetic load can be especially problematic in cases when a large population suffers from a rapid demographic decline (e.g., a population bottleneck), in which the chances of combinations between recessive deleterious alleles in homozygosity

increases, likely leading to population extinction (Lansch-Justen et al. 2022). Consequently, higher levels of genetic diversity can include alleles which will be the seed of future adaptations, but it can also hide masked deleterious mutations in heterozygous genotypes as a trojan horse of future population demise.

The risk posed by masked genetic load is even more problematic when considering a common measure used in conservation genetics, known as genetic rescue. The genetic rescue of an endangered population consists in a managed reintroduction of individuals from another larger population. From one side, the insertion of new individuals increases genetic variation and reduces inbreeding in the short-term, leading to an apparent increase in fitness (Frankham 2015). However, the new genetic variation may also contain levels of masked load that can lead to the extirpation of the already endangered population, as shown by the classic example of the Isle Royale grey wolf (*Canis lupus*) population (Robinson et al. 2019). The interbreeding of this isolated population with a single migrant individual from the mainland generated the collapse of the whole population, due to the input of highly deleterious mutations caused by this individual.

In summary, both genetic variability and genetic load can be considered when evaluating a population's extinction risk, even though there is no consensus about which one represents a better approximation of fitness (Teixeira and Huber 2021; Kardos et al. 2021). In this scenario, understanding whether one factor could predict the other could provide a useful tool for conservation genetics. From a methodological point of view, genetic diversity represents a more accessible proxy, as it can be generated by less expensive sequencing methods, such as RADseq or low coverage sequencing (Peterson et al. 2012; Therkildsen and Palumbi 2017). Genetic load quantification, on the other hand, requires more resources as higher sequencing coverage of well-annotated and "phylogenetically-aligned" genomes, where fitness effects of all mutations are known, either by their predicted effect on the mRNA/protein (SnpEff, Cingolani et al. 2012) or by their degree of conservation across different lineages (GERP or PhyloP scores, Cooper et al. 2005; Siepel et al. 2005). Despite being more resource dependent, the scan for genetic load can be essential for conservation genetics, especially before performing measures of genetic rescue, as mentioned beforehand.

4. The King penguin: a system to study the evolutionary contribution of ecological mismatch under climate change scenario

In the context of the MMH, individuals born under mismatched conditions are expected to be subjected to stronger viability selection (i.e., selective pressures that affect the individual from the zygotic stage until adulthood). Such stronger selection leads to higher mortality rates and, in some cases, to poorer fledging conditions and reduced fitness for the surviving offspring (e.g., Perrins and McCleery 1989; Van Noordwijk et al. 1995; Verboven et al. 2001). However, little is known about the actual effect of selection on the genetic composition of surviving mismatched individuals, not to mention their potential adaptive contribution to the population long-term evolutionary trajectory.

Considering the unprecedented pace of current global changes, characterised by the unpredictability and intensification of natural events (Planton et al. 2008; Stott 2016), understanding the evolutionary mechanisms that maintain mismatched strategies may give a hint on the adaptations to future conditions (Lof et al. 2012). In other words, individuals that can survive even under mismatched conditions should harbour genetic and plastic components that are needed to endure non-optimum conditions, and could represent a condition for the population to adapt to rapid changes. Thus, identifying such mechanisms and preserving individuals with this life history strategy could represent a way of buffering species extinctions before it is too late. A study investigating the genetic and plastic contribution of mismatched phenological strategies to long-term fitness is still missing in literature, at least until the time of writing.

As it was stressed out in the previous sections, the complexity of natural systems and interspecies relationships makes such studies even more challenging. Even if laboratory conditions may provide a better control for confounding variables, the answer to such a complex question necessarily involves a complex system. Studying species under their natural wild conditions has the benefit of accounting for realistic natural selection pressure that cannot be mirrored in controlled laboratory conditions. In addition, for most endangered taxa (e.g., large mammals, polar species, seabirds) it is not physically nor ethically feasible to perform studies under controlled experimental conditions. Therefore, to understand the evolutionary mechanisms that can allow species to avoid extinction in their wild system is essential to find a naturally designed experiment testing the effects on fitness of match and mismatch phenologies.

In this thesis, we investigated genetic patterns and plastic adaptations linked to survival in individuals born under natural mismatched conditions. To that aim, we used genomes, transcriptomes, and life history traits of a long-lived seabird species regularly producing offspring under matched and mismatched conditions, the King penguin (*Aptenodytes patagonicus*) (**Figure 6**). In terms of conservation, king penguins are considered of least concern (LC) by the IUCN Red List of Threatened Species (IUCN 2020),

due to the large population size (*ca.* 1.1 million breeding pairs, Bost et al. 2013; Weimerskirch et al. 2018) and high interconnectivity of its breeding colonies (i.e., absence of genetic structure between breeding areas) (Cristofari et al. 2018). Although the large and highly interconnected global population classifies the species as non-vulnerable, king penguins are still subjected to threats (IUCN 2020). According to the most updated report on the species conservation status, apart from a less widespread threat posed by invasive terrestrial predators in some populations (< 50% of the population), the main threat to the species is climate change and habitat shifting (> 90% of the population) (IUCN 2020).

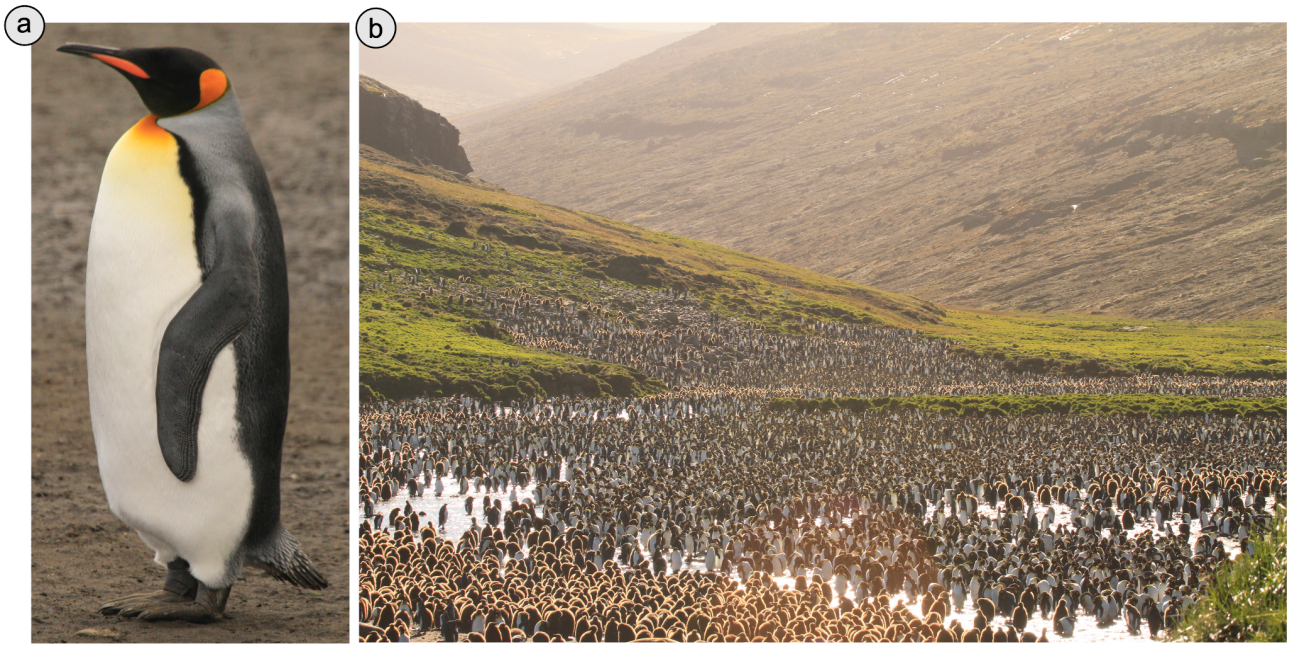


Figure 6. The studied species, the King penguin, and a breeding colony of the species. a) an adult King penguin; b) the colony of La Baie du Marin in Possession Island, Crozet archipelago. Authorial photos taken in December 2021, Possession Island, Crozet Archipelago.

Climate changes pose the biggest threats for king penguins, as projected warmer sea surface temperatures (SST) are estimated to cause a poleward shift in the main foraging grounds of the species during the summer season (the Antarctic Polar Front (APF), Le Bohec et al. 2007; Péron et al. 2012; Cristofari et al. 2018), a critical energy-intensive period for chick growth before winter fasting. The APF consists of an upwelling zone rich in the species' main prey during the Austral summer, the myctophid fish (i.e., lantern fish, family Myctophidae) (Bost et al. 1997). Consequently, the farther the APF is located from the colony, the higher the foraging distance an adult needs to cover during incubation and the chick feeding (Bost et al. 2015). King penguins are flightless birds with equal parental care, meaning that males and females take shifts between feeding at sea and staying in land during incubation and chick caring. Consequently, too long foraging trips lead to egg or chick abandoning by the partner that is on land (Olsson 1997; Gauthier-Clerc et al. 2002).

Niche modelling studies estimate that the position of the APF by 2100 will hinder the permanence of current king penguin colonies that will be too distant from the foraging area to be reached during the breeding season (Péron et al. 2012; Cristofari et al. 2018). As it can be seen in the projections from Cristofari et al. (2018) (**Figure 7**), in the worst-case predicted

scenario, several existing colonies will probably be too far north of the APF to continue being viable for breeding. The endangered colonies include the islands from the Crozet archipelago (number 8 from **Figure 7**), which used to hold the highest number of breeding pairs from the species before a drastic decline of 88% of individuals from its largest colony in the last 30 years (Weimerskirch et al. 2018). The causes of such rapid massive mortality is still unknown, but the main hypotheses involve: a strong Dipole event (i.e., drastic fluctuation in temperature and rainfall) in 1997; predation from invasive species known in the island, like feral cats (*Felis catus*) and house mice (*Mus musculus*); or a parasitic infection, as it has also previously been reported in another colony from Marion island (Cooper et al. 2009). Anyhow, this case raises the concern that currently stable populations can suddenly suffer from rapid drops that could even result in the whole species extinction.

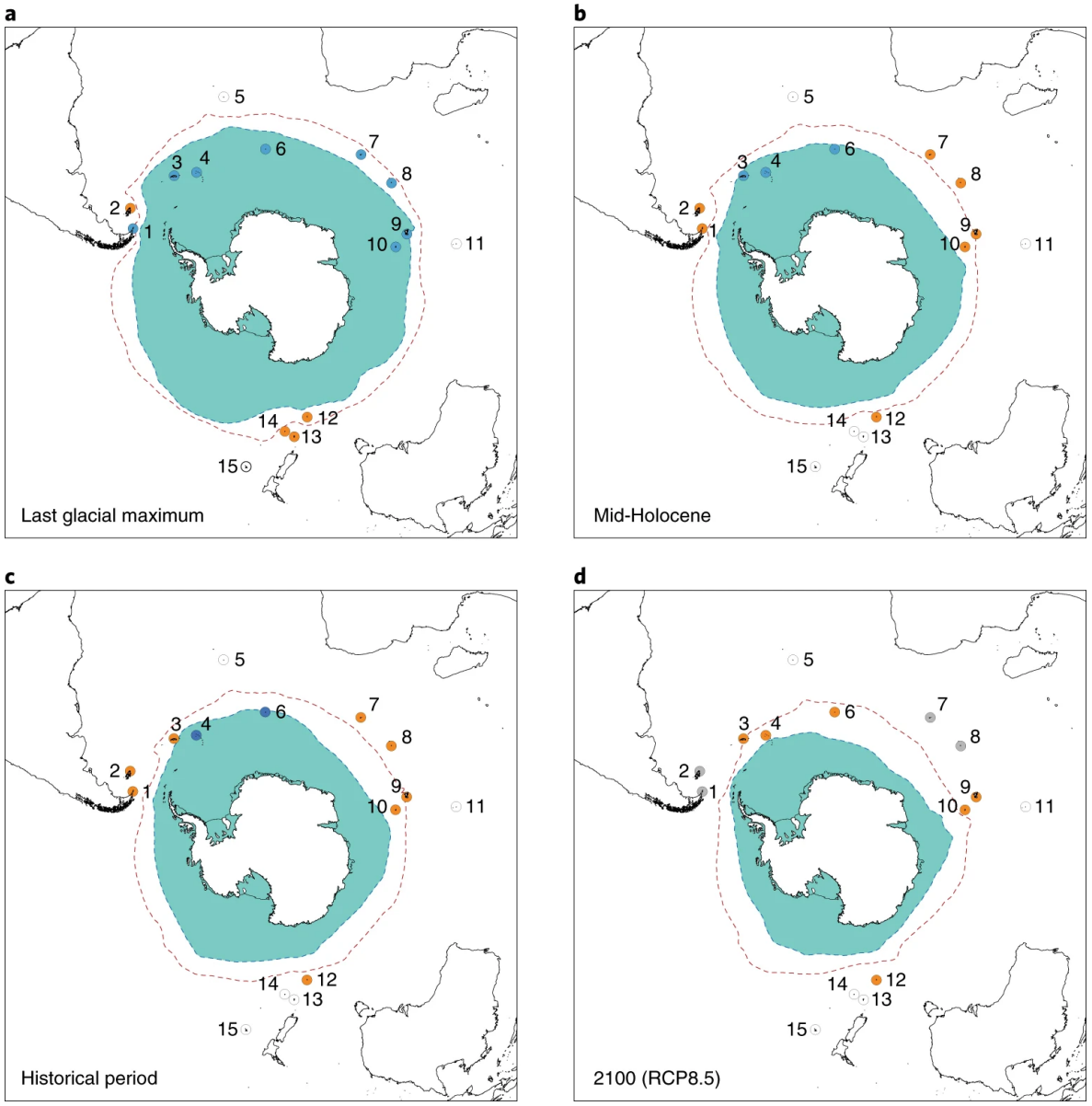


Figure 7. Predicted past and future APF position and King penguin colonies at different periods, from Cristofari et al. (2018). a) during the last glacial maximum (21-19 thousand years ago, ka); b) the mid-Holocene (6 ka); c) the historical period of the study (1981–2005); and the projection for 2100 according to the worst-case greenhouse gas concentration scenario (RCP8.5). At each box, orange dots represent areas with presence of King penguin colonies, blue dots represent areas where colony foundation is precluded by sea- or land-ice

extent, grey dots represent areas that will be too far from the APF for foraging, and white dots represent areas that were never occupied by a King penguin colony. Red dashed lines represent the position of the APF based on SST in February (5°C) and blue dashed lines and light blue background represent the APF and the extent of sea-ice in September (sea-ice concentration > 15%) respectively. Colony numbers represent: 1) Tierra del Fuego (Chile); 2) Falklands (Malvinas); 3) South Georgia; 4) South Sandwich; 5) Gough; 6) Bouvet; 7) Marion and Prince Edward; 8) Crozet; 9) Kerguelen; 10) Heard and McDonald; 11) Amsterdam; 12) Macquire; 13) Auckland; 14) Campbell; 15) Chatham islands.

The King penguin is the second largest living penguin species, after the Emperor penguin. The species breeds in high-density colonies in majorly flat or slightly sloped areas (Bauer 1967; Barrat 1976). These colonies show a widespread range along the sub-Antarctic area, with a major density occurring in the Indian Ocean sector of the Southern Ocean (e.g., Crozet and Kerguelen archipelagos, Marion and Prince Edward islands, and Heard and McDonald islands) (**Figure 8**). King penguin colonies are also also present in the Macquarie (Pacific) and South Georgia (Atlantic) islands, with growing colonies in the Falkland (Malvinas) (de Hoyo et al. 1992) and South Sandwich islands (Convey et al. 1999), and recently established colonies in Tierra del Fuego, Chile (Kusch and Marín 2012) and South Shetland islands (Gryz et al. 2018).

King penguin colonies are mostly located in islands around the Antarctic Polar Front (APF, dashed line in **Figure 8**), an upwelling zone rich in the species' main prey during the Austral summer, the myctophid fish (i.e., lantern fish, family Myctophidae) (Bost et al. 1997). As explained before, the position of the APF is a key factor for King penguin reproductive success (Le Bohec et al. 2007; Péron et al. 2012; Bost et al. 2015). The studied colony in the Crozet archipelago (indicated with an asterisk in **Figure 8**) is located 400-500 Km north of the APF during the summer season (Descamps et al. 2002). The Crozet archipelago used to hold the largest colony of the species in one of its five islands, the *Île aux Cochons* until the recent drastic population decline (Weimerskirch et al. 2018).

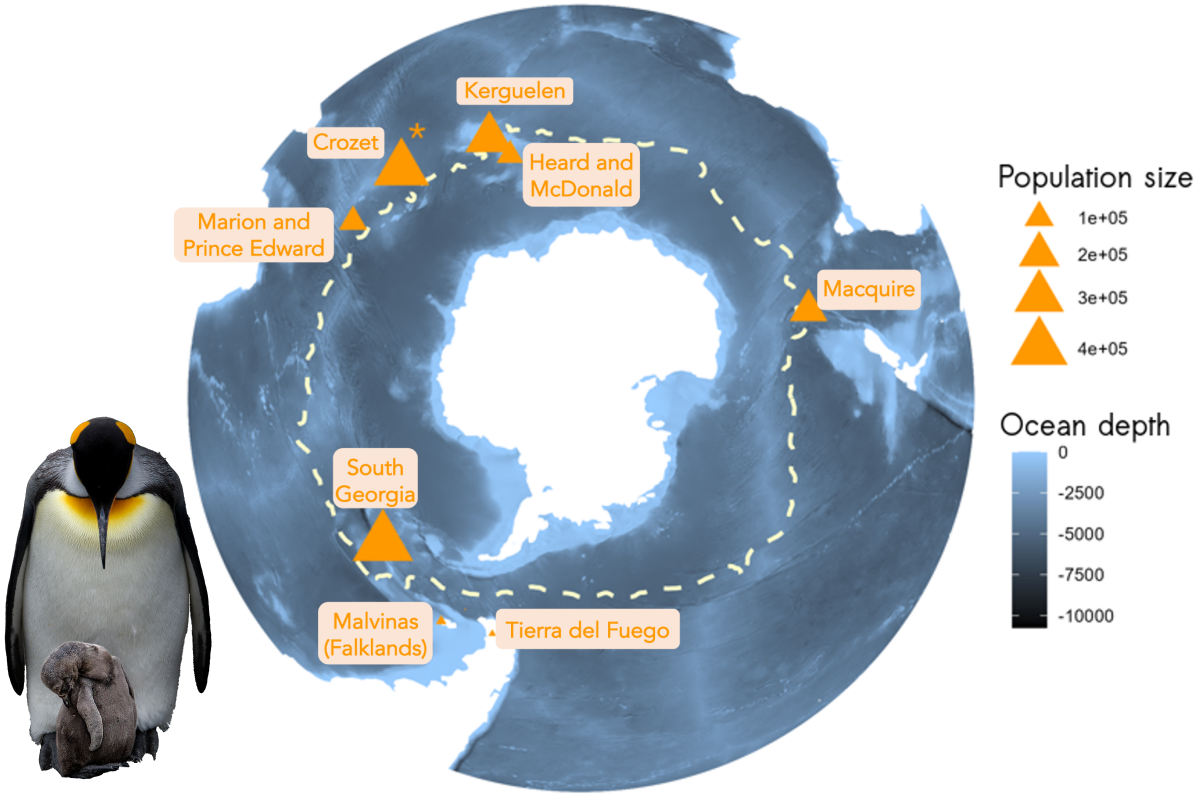


Figure 8. Current King penguin breeding distribution. Orange triangles represent the number of breeding individuals per location (each location is composed of at least one colony), with the archipelago of Crozet, which contains the studied colony, indicated with an asterisk. Dashed line represents the APF position based on (Park et al. 2019).

In the wild, the known lifespan of the species is 30 years old, and individuals usually reach sexual maturity around 3 years of age (Kriesell et al. 2021). King penguins do not have clear visual cues that differentiate between sexes (i.e., the species is monomorphic), although males are slightly bigger in size than females, in addition to an acoustic variation in the calls between sexes (Kriesell et al. 2018). The *Aptenodytes* genus, which includes the King and the Emperor penguins, is characterised by the absence of nests for the incubation and chick brooding phases. Instead, king penguins make use of their brooding pouch (i.e., a featherless region above the feet) (**Figure 9a and b**), where a single egg is incubated for approximately 54 days (Barrat 1976; Descamps et al. 2002). The brooding pouch is also used to protect the chick from cold temperatures and predators (skuas, *Catharacta antarctica*, and giant petrels, *Macronectes spp.*) during the chick's first month of life (Borboroglu and Dee Boersma 2015).

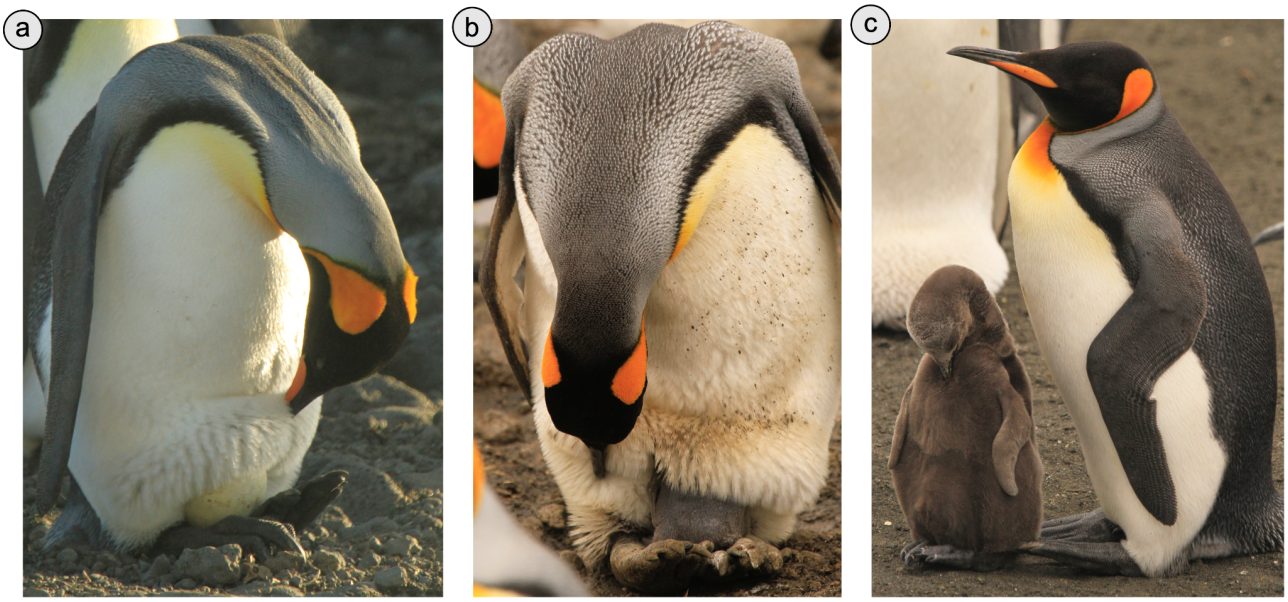


Figure 9. Incubation and chick brooding in the King penguin. a) incubating adult, showing the egg protected by the brooding pouch; b) brooding-stage adult with the newborn chick in the brooding pouch; c) parent and chick (around 1 month old) when the chick is thermally independent and can stay out of the pouch. Authorial photos taken at the study colony in Possession Island, Crozet Archipelago, 2022.

The breeding cycle of the King penguin is one of the longest among birds, taking around 14 months from couple formation to chick fledging (Stonehouse 1960; Barrat 1976; Descamps et al. 2002). Chicks are completely dependent on parental feeding from hatching to fledging, a period that takes ca. 11 months, which is interrupted by the winter period (Descamps et al. 2002). As body mass accumulation relies on the feeding frequency, chick's growth can be divided in three phases: I) an initial growth period during the first summer; II) a period of body mass decrease during winter as a consequence of less frequent feeding, which also results in high chick mortality; and III) and a second growth period after winter, due to the resume of frequent feeding that goes until the chick fledges (**Figure 10)** (Cherel et al. 1987; Descamps et al. 2002; Stier et al. 2014). Throughout this first year of life, chicks also show a pattern of heterothermy (i.e., body temperature is both regulated internally and by the environment), reducing peripheral body temperatures during winter fasting, likely as a strategy of energy saving (Eichhorn et al. 2011).

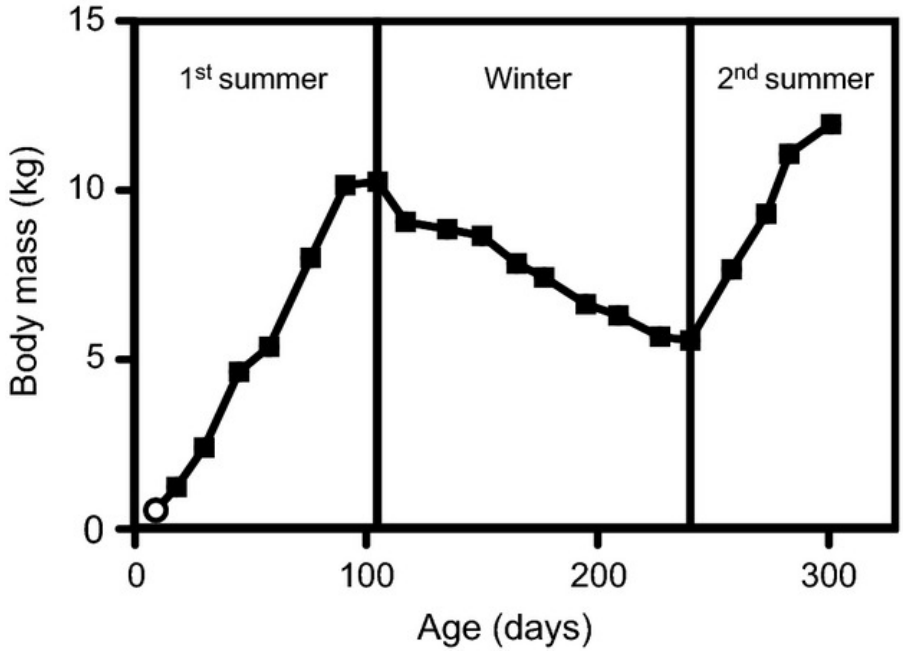


Figure 10. Body mass trajectory of a king penguin chick during the first year of life, from 28 January 2010 to 16 November 2010, from Stier et al. (2014). The highest rate of body mass accumulation takes place during the first summer, decays during winter, and resumes during the second summer until the chick stops being fed by the parents.

Due to the long duration of the reproductive cycle, individuals that successfully fledge a chick in one year and try to reproduce in the consecutive season will have a late start in the second attempt, generating reproductive asynchrony in the colonies (Descamps et al. 2002). Consequently, two phenological groups of chicks are born every year, an early (matched) and a late (mismatched) group, which are known for their marked differences in survival (Olsson 1996; Weimerskirch et al. 1992; Stier et al. 2014). Even though early and late chicks are only born one month apart, late chicks show a much higher mortality throughout the first year, as they have less time to grow and accumulate enough body mass until the beginning of the austral winter (Stier et al. 2014).

As described by Descamps (2002), early breeders in Crozet begin incubation around November or December, while late breeders start from January to February. From incubation until chick's thermal emancipation, around one month old, parents take regular shifts to forage and provide food to the chick. This is characterised as a first growth period of the chick, which lasts until the beginning of the austral winter. From May until late August, adults have to forage farther away from the colony, closer to the Antarctic pack ice (Bost et al. 2004). This is due to the seasonal drop of myctophid availability in the APF (Koslov et al. 1991). During these months, chicks are left in crèches and are rarely fed, which coincides with the period of highest mortality rates, which is especially higher in the late-chick group (Weimerskirch et al. 1992; Stier et al. 2014). Chicks that survive through winter then begin to be fed again from September until fledging, which can take place from November to the next year's January (Fernandes and Bardon et al. *in prep* - Chapter 3). As most of the mortality in this species occurs before fledging (Saraux et al. 2011), we use the survival of chicks until fledging as a proxy of recruitment in the population. A simplified schematic of the King penguin cycle is displayed in **Figure 11**.

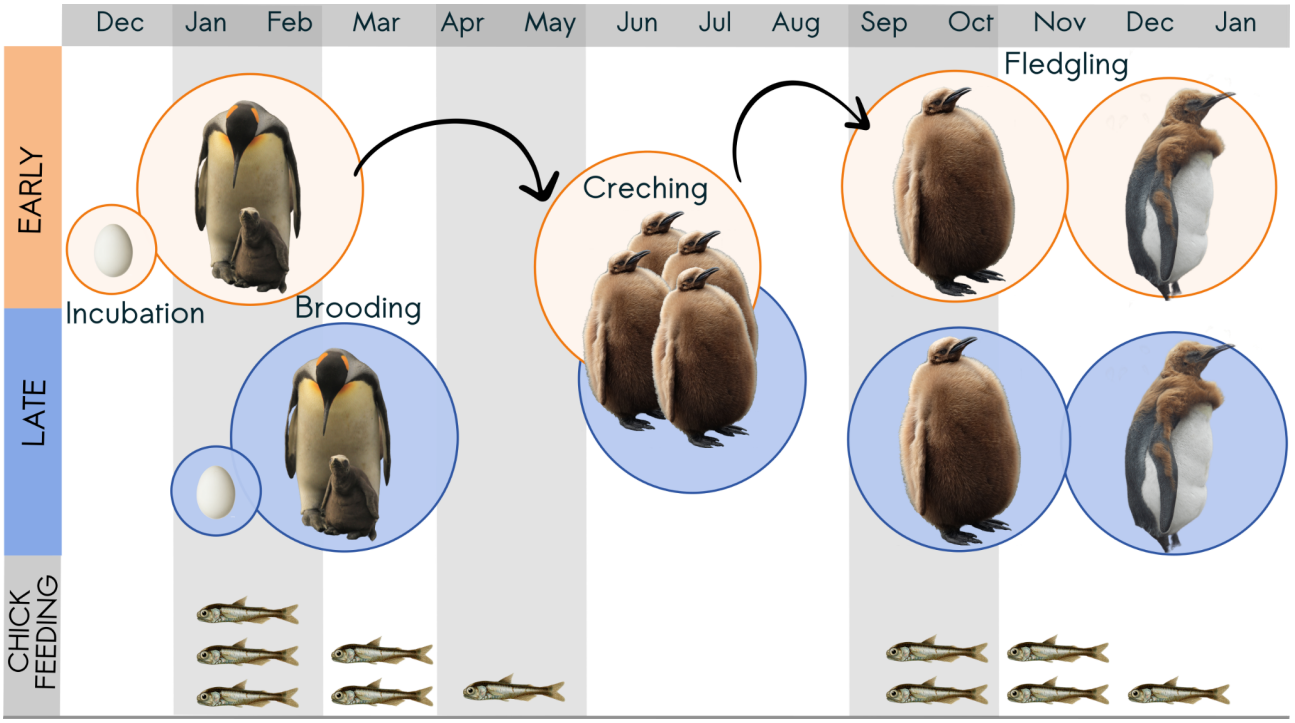


Figure 11. The King penguin breeding cycle. Representation of the year breeding cycle of the species, with the timing of each activity (i.e., incubation, brooding, crèching, and chick fledging) with the chick feeding frequency shown at the bottom. Early breeders are represented in orange, late breeders are in blue.

Because late-hatchlings are born closer to the seasonal APF myctophid drop, we considered being late-born as a mismatched condition with the environmental resources. Due to their delayed arrival in the colony, late breeders usually get the most peripheric available areas in the colony (Le Bohec et al. 2005; Descamps et al. 2005). Peripheral areas are more exposed to predation, although individuals suffer less from the interspecific stress of high density protected breeding spots (Descamps et al. 2009). Thus, the later an individual is born in the species, lower are the chances of survival (Olsson 1996; Weimerskirch et al. 1992; Le Bohec 2007; Stier et al. 2014). However, the stronger external pressures of late breeding could be acting as a stronger viability selection filter in the population. In this scenario, investigating the characteristics of surviving late-hatchlings may indicate traits that allow for survival even under the most unfavourable conditions.

Box 2. A bird's eye view on penguins evolutionary history

Penguins are flightless seabirds that belong to the order Sphenisciformes, sister to the Procellariiformes order (e.g., albatrosses, petrels, shearwaters). The Sphenisciformes order is composed of more than 50 extinct species, a.k.a. stem penguins, and a less diverse monophyletic extant clade with 19 species, also known as the "crown penguins" (Family Spheniscidae) (**Figure 12**) (Cole et al. 2022). The estimated origin of the stem penguins dates to the late Cretaceous, around 60 million years ago (Ma), while the

divergence of the crown group happened 40-45 Ma later, in the Neogene (**Figure 12**) (Vianna et al. 2020; Cole et al. 2022).

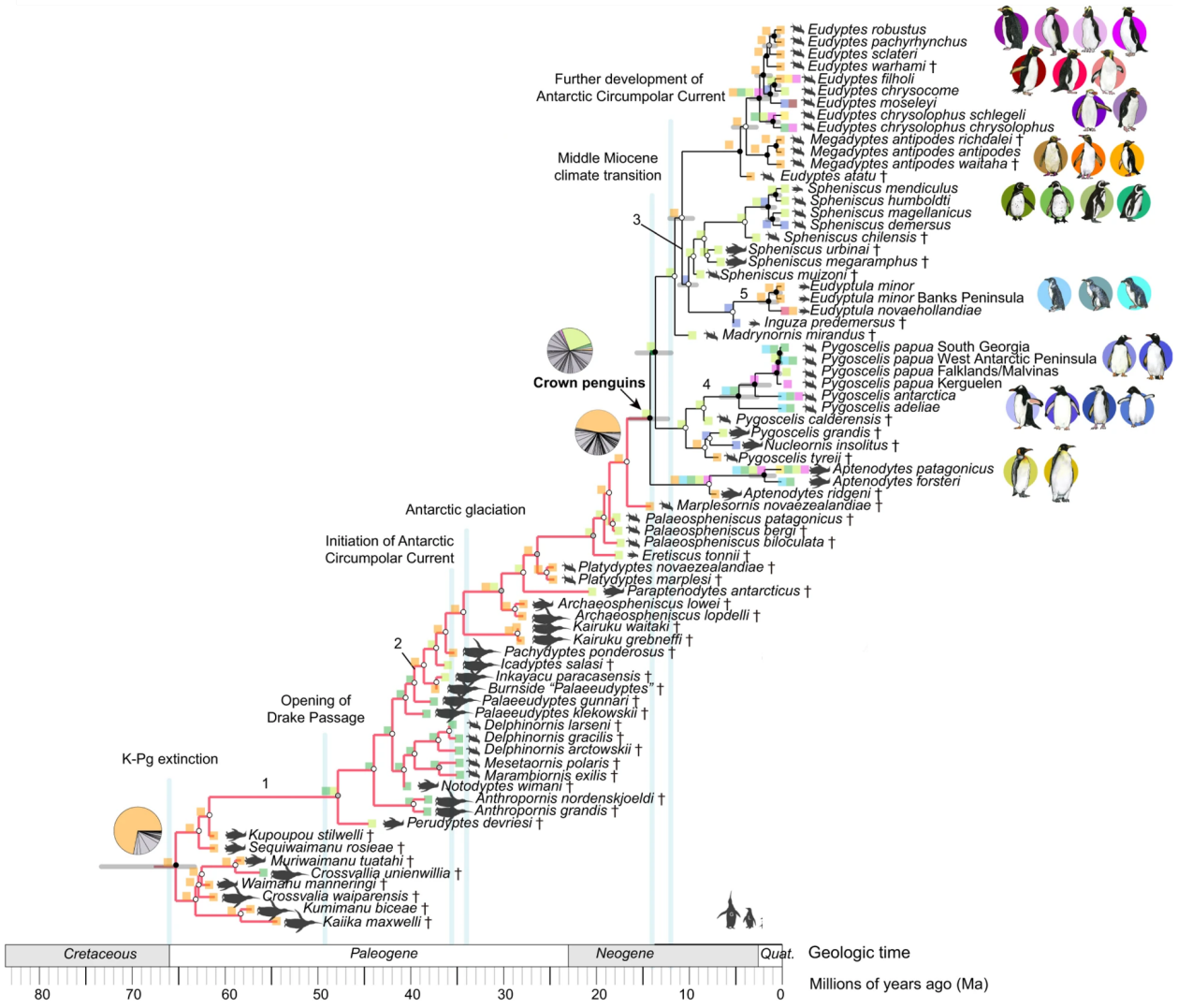


Figure 12. Phylogeny of Sphenisciformes adapted from Cole et al. (2022). Penguin silhouettes indicate body size. Coloured icons to the right of the tree represent the Crown species (including fossils from the crown group). Grey rectangles in the Crown penguins' nodes represent 95% confidence intervals of the estimated divergence times. Circles at the nodes are coloured according to posterior probability: black (>0.95), grey (0.75–0.95), white (<0.75). The single most probable ancestral range is indicated at each node using squares according to colours of locations in **Figure 13** with the exception of three key nodes (pie charts, grey represents multiple ranges). Node numbers (from 1 to 5) correspond to potential dispersion events from **Figure 13**.

Historical biogeographic reconstructions based on the extensive penguin fossil record estimate that the stem penguin clade originated in the Zealand region (**Figure 13**). Several speciation events would have occurred in this territory before new lineages colonised the southern areas of South America and the Antarctic Peninsula around 40 Ma (Cole et al. 2022). In comparison to the living species, stem penguins are characterised by several large-bodied species, including the largest known fossil recently discovered,

the *Kumimanu fordycei n. sp.* (Ksepka et al. 2023), not shown in the phylogeny from Cole et al. (2022) in Figure 12 due to its later discovery.

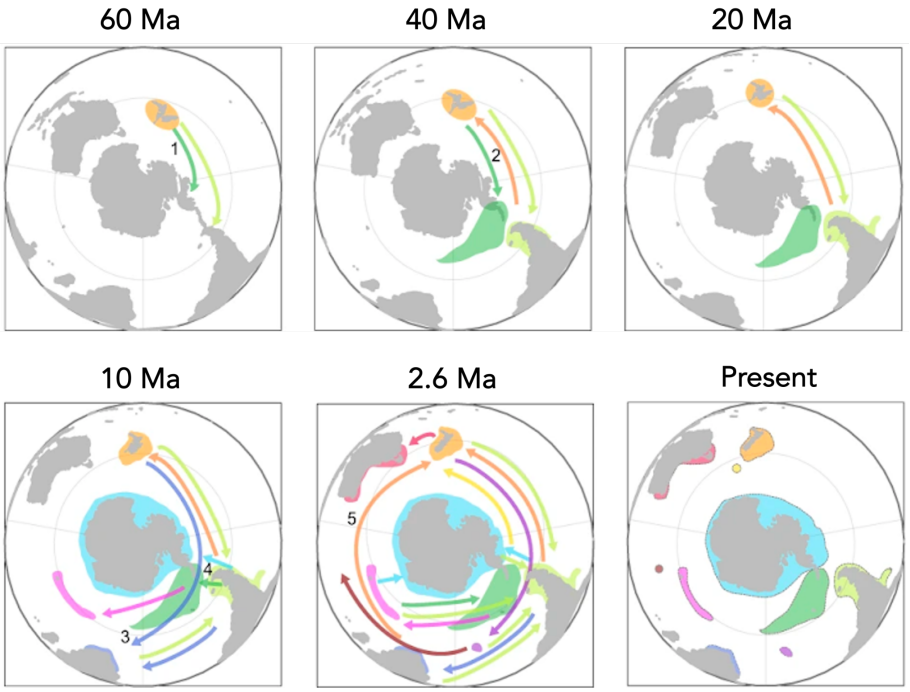


Figure 13. Hypothesis on the biogeographic history of penguins adapted from Cole et al. (2022). Stem penguins likely originated in the region of New Zealand around 60 million years ago, followed by a history of dispersion events (represented by the arrows). Numbers indicate major radiation events shown in Figure 12.

In a previous study, performed during my Master degree, we estimated that the origin of the crown penguin lineage to have taken place in the region of Australia-New Zealand during the Miocene (Vianna et al. 2020). Our biogeographic reconstructions based on the current species distributions in combination to evidence another recent study suggest that ocean currents, such as the Antarctic Circumpolar Current (ACC), played an essential role in the colonisation of the current areas in the Southern Hemisphere (Vianna et al. 2020; Cole et al. 2022) (**Figure 13**). In addition, ancient niche reconstructions propose that the ancestral of all Spheniscidae family initially occupied a more temperate environmental niche, and new lineages adapted to both warmer and colder environments (Vianna et al. 2020).

Even if crown penguin distribution is exclusive to the Southern Hemisphere (Borboroglu and Dee Boersma 2015), extant penguin species range from the equatorial Galápagos islands, in the case of the northernmost Galápagos penguin (*S. mendiculus*), to the coldest environments on Earth, in the case of the Antarctic Emperor (*A. forsteri*) and Adélie (*P. adeliae*) penguins. Apart from such extreme latitudinal ranges, most of the extant species breed in the sub-Antarctic region (i.e., roughly, between 46° and 60° south of the Equator), as it is the case of the king penguins, many crested *Eudyptes* species, the Yellow-eyed penguin (*M. antipodes*), and two other *Pygoscelis* penguins (Chinstrap and Gentoo) (**Figure 14**).

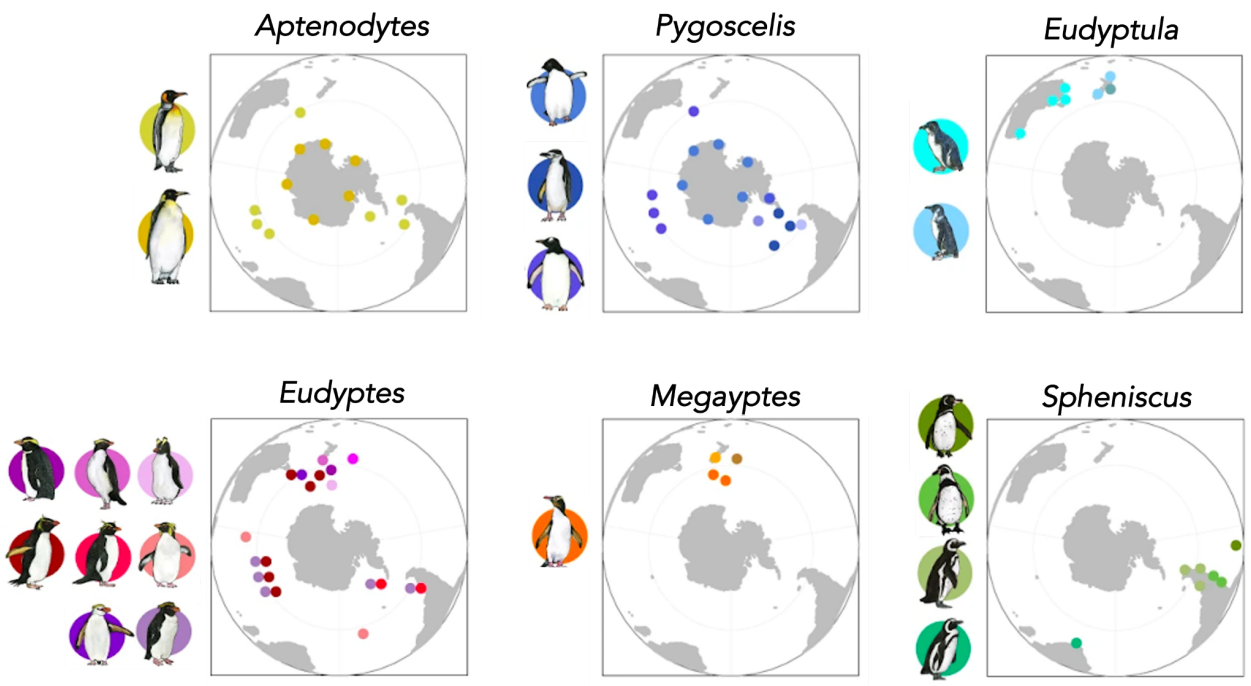


Figure 14. Extant penguin species current distribution (breeding range), adapted from Cole et al. (2022). We also maintained the Royal penguin (*Eudyptes chrysolophus schlegeli*) from the original image, although it is now considered a subspecies of the Macaroni penguin, due to the clear morphological difference between the two taxa.

Crown penguins are classified into six genera: *Aptenodytes, Pygoscelis, Eudyptula, Eudyptes, Megadyptes,* and *Spheniscus*. As mentioned before, the *Aptenodytes* genus contains the two largest living species, the King (*A. patagonicus*) and Emperor penguins (*A. forsteri*). *Pygoscelis* has three extant species, Adélie (*P. adeliae*), Chinstrap (*P. antarctica*), and Gentoo penguins (*P. papua*). *Eudyptula*, also known as the Little penguins, contains the two smallest living penguin species (*E. minor* and *E. novaehollandiae*) (Grosser et al. 2015).

The *Eudyptes* genus, a.k.a. crested penguins, is the most species rich, with seven species: the three rockhoppers, the Eastern (*E. filholi*), the Southern (*E. chrysocome*), and the Northern rockhopper penguins (*E. moseleyi*); the Macaroni penguin (*E. chrysolophus*), which hybridises with the Royal penguin (*E. chrysolophus schlegeli*), long considered as a separate species due to marked phenotypic differences; and three species endemic to New Zealand and its nearby islands: the Snares (*E. robustus*), the Fiordland (*E. pachyrhynchus*), and the Erect-crested penguins (*E. sclateri*). (Vianna et al. 2020; Cole et al. 2022). The *Megadyptes* genus contains a sole species, the Yellow-eyed penguin (*M. antipodes*), also endemic to New Zealand. Finally, the *Spheniscus*, a.k.a. banded penguins, contains four species: the equatorial Galápagos penguin (*S. mendiculus*), the South American Humboldt (*S. humboldti*) and Magellanic penguins (*S. magellanicus*), and the African penguin (*S. demersus*). Extant penguin distribution is shown in **Figure 14**.

The definition of the exact number of extant penguin species has long been a challenge, due to recent speciation of some lineages. Recent and rapid speciation events

leave signals of incomplete lineage sorting (ILS), in which the divergence between different parts of the genome do not coincide with the species divergence (see Vianna et al. 2020; Cole et al. 2022). This is the case of little penguins, (Grosser et al. 2015) and the three rockhopper species (Banks et al. 2006; Frugone et al. 2018). In the case of gentoo penguins (*Pygoscelis papua*), high genetic differentiation may lead to a future speciation event, but all populations are considered as one species at the moment (Clucas et al. 2018; Cole et al. 2019).

Another challenging factor for penguin species delimitation is due to interspecies hybridization in the wild. This is the case of Macaroni and Royal penguins (*Eudyptes chrysolophus spp.*), which are now considered as a single species with different phenotypes concerning the face colour (**Figure 14,** (Frugone et al. 2018). Hybridization events are also known in the *Spheniscus* genus between Humboldt and Magellanic penguins, but these are still considered as separate species (Hibbets et al. 2020). Such events leave introgression signals in the genome that can be confounding, especially when few markers are used (Frugone et al. 2018; Vianna et al. 2020).

Recent phylogenetic studies estimate that the first extant clade to diverge from the other lineages around 14 Ma was the large bodied *Aptenodytes* genera. The divergence between the two *Aptenodytes* species, the Emperor (*A. forsteri*) and the King penguins (*A. patagonicus*), on the other hand, was much more recent, estimated around 2 Ma (**Figure 12**, (Vianna et al. 2020; Cole et al. 2022). The *Aptenodytes* ancestor is also estimated to have occupied a more similar ecological niche to the King penguin's current niche, in terms of maximum and minimum surface water temperature thresholds (Vianna et al. 2020). Moreover, a recent study from our research group has shown that adaptations to cold present in the Emperor, but not in the King penguin, support selection to extreme cold conditions as a derivative state in the *Aptenodytes* clade (Pirri et al. 2022). This study identified 165 candidate genes under selection on the Emperor penguin that are likely related to cold adaptation, such as the temperature sensing gene, TRPM8 (Yin et al. 2018).

Even if the Emperor penguin harbours exclusive extreme cold adaptations in comparison to the King, these two species also share putative adaptations absent in the other extant taxa. Cole and collaborators (2022) detected a set of genes under positive selection that are unique to the *Aptenodytes* branch. Two genes, FIBB and ANO6, may allow for higher diving capacity, as the King and Emperor penguins are known to be the deepest divers among penguins (record of 343 metres and 552 seconds, and 564 metres and 1308 seconds, respectively) (Pütz and Cherel 2005; Pütz et al. 1998; Wienecke et al. 2007). Two other genes, CREB3L1 and SMARCAD1, are related to large body size, and could represent an inheritance from the large bodied extinct lineages that suffered selection relaxation in the other extant lineages (Cole et al. 2022).

Objectives

1. General aims

Mismatched phenotypes could always be present in a population as part of its variability. According to the MMH, mismatched phenotypes stand stronger selective pressures, so that individuals surviving under mismatched conditions should carry phenotypes that permit the overcoming of unfavourable early-life conditions. If ongoing environmental changes will increase the frequency of mismatches in populations (Kharouba and Wolkovich 2020), then the phenotypic traits that allowed mismatched individuals to survive in the past, may become the adaptive standing variation making a population thrive under the novel environmental conditions.

In the case of the King penguin, the most dramatic future scenario predicts that several colonies located north of the APF will be too far from the feeding ground during the reproductive season to remain viable (Cristofari et al. 2018). Such future pressure is analogous to that exerted by the currently mismatched phenology of late breeding, that is completing their development out of the peak of food resource availability. Hence, our main prediction here is that the variability held by mismatched individuals in the current population would increase the future adaptive potential of the population under rapid environmental changes. To investigate this question, we used genomes, transcriptomes, and life history data derived from king penguins born early (match) and late (mismatch) in the season (i.e., in and out of the peak of food resources).

In the first chapter, our aim was to measure the effects of selection on matched and mismatched genotypes during King penguin's first year of life. We expected that the stronger purifying selection acting on mismatched individuals would purge deleterious mutations more efficiently while maintaining higher levels of genetic variability than in the matched individuals.

In the second chapter, we investigate whether mismatched phenotypes could be adaptive or not at the population level. By sequencing the blood transcriptomes of individuals that survived until fledging (used in the first chapter), we analyse differential gene expression between the early- (match) and late-born (mismatch) groups. We expected late-born individuals to show plastic adaptations to faster growth and efficient energetic accumulation, acting as a predictive adaptive response (PAR) to future physiological challenges faced later in life.

In the third and final chapter, we explored the potential impacts of being born under mismatched conditions on early-life traits, return rates, and recruitment of individuals into the population, using a long-term life-history dataset on individuals born under matched and mismatched conditions. If genetic and plastic adaptations allow mismatched individuals to survive through the first winter, in which a minimum body condition is determinant for survival in both groups (Stier et al. 2014), a morphological catch-up with the matched group is expected at fledging. Moreover, if a body condition catch-up is observed, return rates and reproductive success should be similar between the two groups.

With the combination of these three studies, which investigate fitness from three different analytical perspectives (i.e., genetic, plastic, life history components), we intend to provide a further step on our understanding of the evolutionary potential of mismatched phenotypes in natural systems.

2. Hypotheses and predictions

a) Genetic determinants of selection in matched and mismatched phenotypes

If chicks mortality during the first winter is mostly due to natural selection, then we should observe a non-random distribution of genetic diversity among dead and survivors.

- If survival is mostly related to genetic load (Bertorelle et al. 2022), then we should observe a lower proportion of deleterious alleles in surviving individuals;

- If survival is mostly related to genome-wide genetic variation (Kardos et al. 2021), then we should observe higher heterozygosity, at least at non deleterious sites, in survival individuals;

- If the higher mortality of individuals born late in the season is caused by stronger selection exerted on mismatched phenotypes, then deleterious alleles should be removed more efficiently and heterozygosity at non deleterious sites should be higher in late surviving individuals than in early ones. When the environmental change trajectory is increasing the occurrence of mismatches, this outcome can represent the adaptive potential of the mismatch strategy.

b) Plastic response of mismatched phenotypes

If mismatched phenotypes, that is from chicks hatching after the peak of resources, stand stronger pressure to grow enough in a shorter period of time before the first winter, then a plastic response in terms of differential gene expression between early and late chicks soon after hatching should be observed.

- If differentially expressed genes are mainly related to stress response, then this condition could be detrimental to the fitness of the individuals (Sanghvi et al. 2021), resulting in a silver spoon effect (Graffen 1988);

- If differentially expressed genes are mainly related to pressures that will also be encountered in adult-life, such as efficiency in energy accumulation (Saastamoinen et al. 2010), this condition can act as a predictive adaptive response (Gluckman et al. 2005a).

c) Pre-fledging traits and post-fledging fitness of mismatched phenotypes

If late surviving individuals have genetic or plastic adaptations allowing e.g., faster growth (Stier et al. 2014) and efficient energy storage, a catch-up with a good, or at

least minimum, body condition should be observed at fledging, as energy reserve is an important determinant of survival (pre and post-fledging) in the species (Saraux et al. 2011). The energy allocation during the winter could result in a greater allocation to body reserve (i.e., body condition) than on growth (i.e., structural size) at fledging.

- If catching-up with body condition at fledging comes with a physiological cost that may generate carryover effects in adult-life, mismatched phenotypes should show lower return rates and a delayed first reproductive success in the early adult-life (Marcil-Ferland et al. 2013);

- If catching-up with body condition at fledging does not come with a cost, mismatched phenotypes should show equivalent return rates and first reproductive success compared to matched phenotypes. The harsh conditions faced by mismatched phenotypes in early-life would prepare individuals for similar constraints in adulthood, such as long fasting periods (Groscolas and Robin 2001), as a predictive adaptive response (Gluckman et al. 2005a).

General Material and Methods

1. Location and life history data

All samples and data used in this thesis come from individuals belonging to the King penguin colony of La Baie du Marin (here referred to as BDM), on Possession Island, Crozet Archipelago (46°24′27″S 51°45′27″E), and were collected in the framework of the Program 137 ANTAVIA from the French Polar Institute (IPEV). The colony of BDM is one of the most well studied King penguin colonies in the world, as several long-term studies that take place in the location since the 1970s (Barrat 1976; Weimerskirch et al. 1992; Jouventin and Lagarde 1995; Descamps et al. 2002; Gauthier-Clerc et al. 2000; Gauthier-Clerc et al. 2001; Gauthier-Clerc et al. 2002; Le Bohec et al. 2007; Le Bohec et al. 2008; Saraux et al. 2011; Bordier et al. 2014; Cristofari et al. 2015). The detailed description of the annual cycle of species, including previously unknown winter activities, was first completely characterised in this colony, owing to a monitoring system of underground antennas (Descamps et al. 2002; Gendner et al. 2005). This system allows the remote monitoring of individuals from a sub-colony of BDM called 'Antavia', which is a naturally enclosed zone with four passages to the sea, containing approximately 10,000 breeding pairs of penguins (**Figure 15**).

The four passageways of Antavia are equipped with an underground system of paired antennas, which captures and stores the entry and exit movements of more than 17,000 RFID-tagged penguins equipped with Radio Frequency Identification (RFID) tags since 1998 (detailed information can be found in Gendner et al. 2005). Hundreds of individuals that are annually equipped with RFID tags at fledging can be monitored throughout their lifetime, as they tend to come back to the same colony where they were born (i.e., philopatry) (Barrat et al. 1976; Bried & Jouventin 2001; Cristofari et al. 2015).

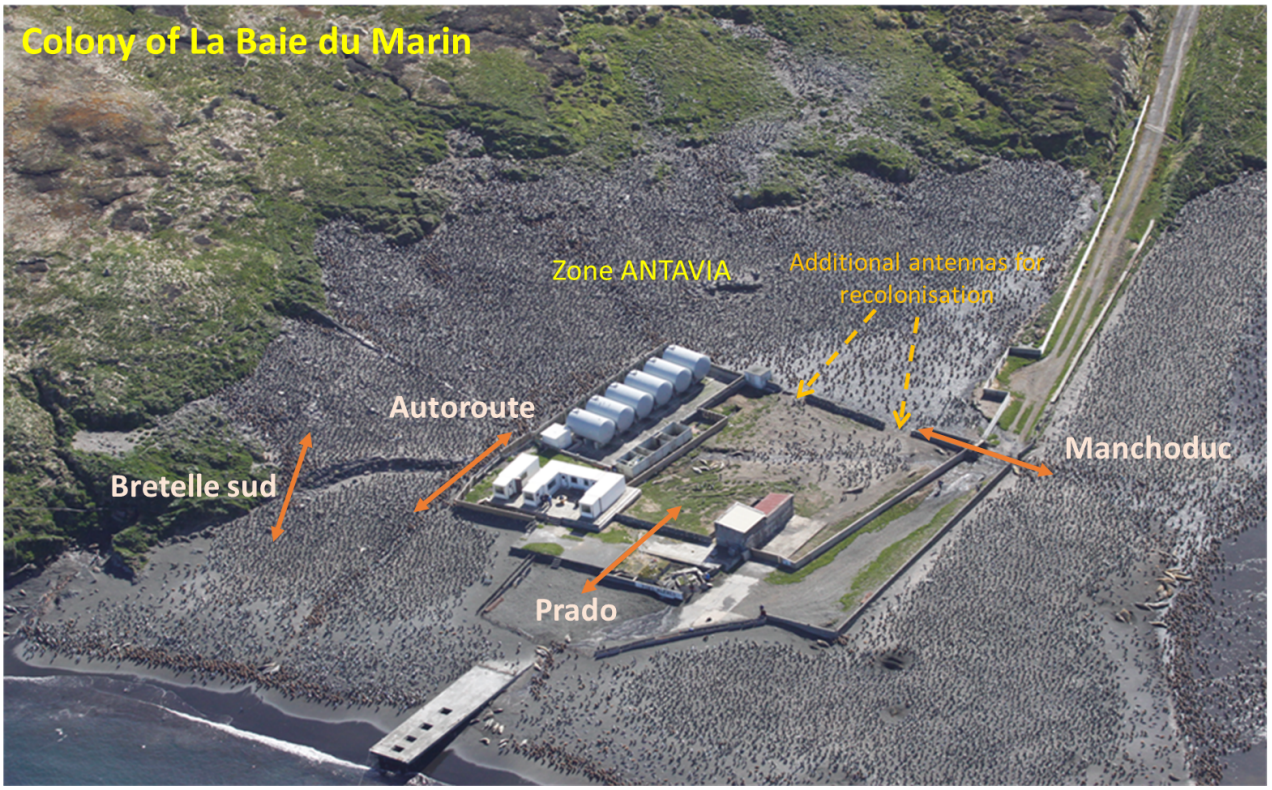


Figure 15. Sub-colony Antavia, Crozet Archipelago. The four passageways through which penguins exit the colony are represented in arrows: “Bretelle sud”, “Autoroute”, “Manchoduc”, and “Prado”.

The system stores the movements of each penguin as an entry in the sub-colory (i.e., breeding zone) or as an exit to the sea. Each individual’s entries and exits produce a characteristic pattern of periods in land and at sea, that are used to detect the individuals’ annual activities, such as breeding, moulting, parading, among others (**Figure 16**). Due to the stereotyped patterns of periods inland and at sea generated by king penguins’ during reproduction, it is possible to assess an individual’s breeding status at each season. More specifically, when an individual starts a reproductive attempt, it will show a pattern of ~15 days in land and ~15 days at sea, that will shorten along the following months until longer periods at sea will be observed. The described pattern corresponds to a typical beginning of cycle for a female, which takes the first shift at sea after laying the egg (Stonehouse 1960; Barrat 1976), while a male would have shown a first inland period of at least 20 days (**Figure 16**).

During incubation, males and females take shifts of approximately two weeks to forage at sea and incubate the egg. Once the chick is born, these shifts become shorter (i.e., adults spend shorter periods in land and at sea), until both parents stop making frequent returns to the colony during austral winter. In a characteristic successful breeding cycle, adults re-start making frequent returns to land for chick feeding around September. Around November-December, reproducing adults will stop doing round trips, and the chicks will fledge. These successful adults will then show a period of one to two weeks at sea (depending on the moment the chick fledges) before moulting, characterised by a ~15 days period in land. Finally, the cycle will restart if the adult attempts a new consecutive breeding in the following year. An example of an annual male cycle is shown in Figure 16.

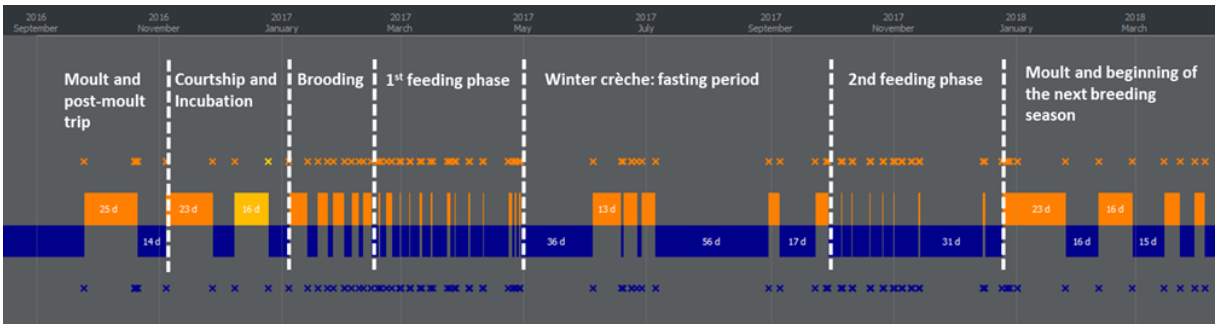


Figure 16. Visual representation of periods spent inside and outside the colony of Antavia for a given RFID-tagged individual in a successful breeding cycle, from Bardon et al. Accepted in *Methods in Ecology and Evolution* with minor revisions (Appendix). Orange and blue crosses correspond to RFID detections (outside antenna in blue, and inside antenna in orange). The periods outside the colony (in blue) and inside (in orange) are interpreted from the sequence of detections. Phases of the breeding cycle are translated from the patterns of an individual’s period spent inside and outside the colony, and are indicated by the white dashed lines. The duration of periods inside and outside of the colony is given in days (d).

The successful breeding cycle described above represents a year of data from one RFID-tagged individual in the sub-colony. It is also possible to detect breeding failures, as the cycle will show a break of the stereotyped pattern that can be interpreted as a

cessation of parental care. Moreover, this system permits the remote monitoring of an individual throughout its lifetime, as all years of each individual can be “cycled” and its activities can be interpreted. This massive amount of data can be used to estimate the fitness of thousands of individuals, by using information such as age of first breeding, total breeding success in life, and approximate age of death (i.e., when detections stop).

However, the massive amount of data generated by this system for more than 17,000 individuals also makes the interpretation of every penguin’s yearly activities humanly intractable. For this reason, owing to the long-term data, our research team has developed a deep learning algorithm to automatically interpret ecological features, such as breeding status and phenology, from RFID mark-recapture data (Bardon et al. *Accepted in MEE with minor revisions*, **Appendix, RFIDeep: unfolding the potential of deep learning for radio-frequency identification**).

Briefly, this methodology, coined as “RFIDeep” works in three main steps. First, a one-dimensional convolution neural network (1D-CNN) architecture was developed. Second, to account for variance in breeding phenology and technological limitations of field data acquisition (e.g., missing detections), a data augmentation step mimicking a shift in breeding dates and missing RFID detections (i.e., missing recaptures) was added. Third, to identify segments of the breeding activity used during classification, a visualisation tool was included, allowing users to understand what is usually considered a “black box” step of deep learning.

In order to train the algorithm and compare its efficacy in correctly identifying the penguin cycles, a set of manually performed cycles were used. Such cycles were performed by a human expert’s interpretation of penguin patterns, to which I have also contributed during the 3-years period of this thesis. The manual cycling allowed us to identify regions and patterns of the penguin’s cycle that are confounding for both the algorithm and a human expert, such as determining the date of failure. Further details about the development of this workflow can be found in the Appendix section.

In this thesis, we used data collected (morphological measurements and body mass) or generated by the RFIDeep (life history traits) each year between 2010 and 2022, from individuals born early and late in the season.

2. Sampling

The genomic and transcriptomic data used in this thesis comes from samples collected from early- and late-hatchlings from the Antavia sub-colony. Sampling was divided into three consecutive summer campaigns in Possession island, Crozet archipelago (missions of 5-6 months during the Austral summer). Genomes were generated from a single time point (2020) with individuals sampled ~3 weeks after hatching, while transcriptomic data includes chicks born at two consecutive years (2020 and 2021) and at two developmental stages (after hatching and before fledging) (**Table 1**).

Summer campaigns lasted from October/November to March/April of the following year. In the case of the King penguin, this period corresponds to both the end of the previous breeding season and the start of a new one. Therefore, because we collected samples of chicks around hatching (~3-weeks-old chicks in January and February) and recaptured the same chick at fledging (~11-months-old chicks in November, December and January), three sample campaigns were done in order to include individuals from two breeding seasons. I participated in two sampling campaigns, 2020 and 2022, while the 2021 campaign was performed solely by field assistants and colleagues, as it took place during the COVID-19 outbreak.

Table 1. Outline of the samples and data used in each chapter of the thesis.

	Chapter 1	Chapter 2	Chapter 3
Data type	Whole genomes (~20x coverage)	3'end RNA seq	Life history data; Morphometric data
Sample or Origin	Whole blood in Queen's Lysis Buffer	Whole blood in PAXgene® solution	RFID data + Measurements
Life stage	Hatching (~3-weeks-old)	Hatching (~3-weeks-old) + Fledging (~11-months-old)	Fledging (~11-months-old) + Juveniles Adults (7 years-old max)
Time span	2020	2020 + 2021	2010 - 2022

Penguins, as well as other sauropsids, contain nucleated blood cells (Chiari and Galtier 2011). Therefore, the use of the blood as a source of DNA requires a less invasive sampling than when sampling other tissues, and still provides enough amounts of DNA for sequencing. For transcriptome data (i.e., RNA sequencing), the blood was also expected to

provide a wide variety of transcripts, as it is a circulating tissue (Liew et al. 2006). We sampled the chicks' whole blood for both the genome and transcriptome next generation sequencing (NGS) and stored samples in respective buffers against nucleic acid degradation, as it is shown in **Table 1** and specified in the Material and Methods section of chapters 1 and 2.

Chapter 1:

Do few survivors persisting under mismatched conditions affect the fitness of the whole population?

Chapter 1

Do few survivors persisting under mismatched conditions affect the fitness of the whole population? *Ready to be submitted*

Authors:
Flávia A. N. Fernandes[1,2]
Paolo Gratton[3]
Lorena Ancona[1]
Gaël Bardon[2,4]
Pierre Carette[2]
Robin Cristofari[5]
Alessio Iannucci[6]
Piergiorgio Massa[1,7]
Josephine Paris[1]
Céline Le Bohec[2,5] & Emiliano Trucchi[1]

Affiliations:
1 - Marche Polytechnic University, Italy
2 - University of Strasbourg, France
3 - University of Rome Tor Vergata, Italy
4 - Scientific Center of Monaco, Monaco
5 - University of Turku, Finland
6 - University of Florence, Italy
7 - University of Bologna, Italy

Keywords
Match-mismatch; bet-hedging; genetic load; adaptive potential; seabirds.

Authors' contribution to this paper

This paper's idea was jointly conceived and designed by F. Fernandes, C. Le Bohec and E. Trucchi. Sample collection was done by F. Fernandes and P. Carette, with the aid of other field work assistants, during the king penguin breeding season of 2020, coordinated by C. Le Bohec in the French Southern and Antarctic Lands (TAAF). DNA extractions and other wet laboratory procedures, such as DNA quantification and quality assessment, were done by F. Fernandes. Sequencing libraries' preparation and genome sequencing were done in the facilities of the University of Florence by A. Iannucci. Post-sequencing pre-processing steps from trimming to variant calling, filtering and categorisation were performed by F. Fernandes. Analyses were performed by F. Fernandes, E. Trucchi, and P. Gratton. Results' discussion and interpretation were done by F. Fernandes, E. Trucchi and C. Le Bohec with the contributions of L. Ancona, G. Bardon, R. Cristofari, P. Massa, and J. Paris. Writing was done by F. Fernandes, with the contribution of J. Paris, C. Le Bohec, and E. Trucchi.

1. Introduction

In the late 1960s, the finding that species with seasonal variation in resource availability synchronise their most energetically demanding activities (e.g., reproduction) with the peak of resources (Cushing 1969) led to the development of the match/mismatch hypothesis (MMH) (Cushing 1974; Cushing and Saleem 1982; Cushing 1990). Although initially characterised in high latitude fishes, evidence for the MMH has since been documented in a variety of marine and terrestrial systems (Post and Forchhammer 2008; Nakazawa and Doi 2012; Plard et al. 2014; Doiron et al. 2015; Régnier et al. 2019; Ferreira et al. 2020). The MMH suggests that phenological mismatches (i.e., when a consumer shows a phenological asynchrony with its resource) results in reduced survival and reproductive success (i.e., fitness). Consequences of mismatch include poor body condition in offspring (Doiron et al. 2015), lower recruitment in the population (Reed et al. 2013), and impaired adult body condition (or even death) when physiological limits are surpassed (Marvelde et al. 2011; Thomas et al. 2001).

According to the MMH, phenological mismatches will only have an impact on fitness if two assumptions are met (Kharouba and Wolkovich 2020): 1) the fitness of the consumer mainly depends on the resource availability, and 2) the food resource has a degree of seasonality that limits the growth or reproduction of its consumer. However, the loss of fitness predicted by the MMH may be diminished if the production of mismatched phenotypes can contribute to population persistence in the long-term (Ghalambor et al. 2015; Leonard and Lancaster 2020; Petrullo et al. 2023).

A recent study showcased empirical evidence of the MMH in the form of behavioural adaptations to mismatch conditions in red squirrels (*Tamiasciurus hudsonicus*) (Petrullo et al. 2023). This research showed that the strategy of producing smaller clutches leads to increased offspring survival in both high- and low-food availability years, even if this leads to a smaller clutch production in highly productive years. Consequently, individuals that “play safe” by producing modest litter sizes under any condition had a higher lifetime fitness than individuals that increase litter size in years when conditions were favourable, even if there was a cost associated with producing less offspring when conditions were ideal.

This strategy, known as bet-hedging, seems to be especially advantageous for species which are subjected to variable selective pressures (Sæther and Engen 2015). In such cases, the maintenance of a phenotype that is assumed to be maladaptive under good conditions could actually promote long-term fitness in the overall population through a balance of costs and benefits when conditions are not favourable. Empirical studies, such as the one described above, have just begun to emerge, lending evidence to the idea that mismatched phenotypes can contribute to adaptation to variable conditions through phenotypic plasticity (i.e., when individual genotypes produce different phenotypes). However, whether adaptation to variable conditions proceeds solely via phenotypic plasticity is still an open question (Torda et al. 2017; Jordan et al. 2017).

In fact, mismatched phenotypes could contribute to long-term fitness via adaptive changes in the genetic background of a population (Beaumont et al. 2009). In particular,

they could increase population-level standing genetic variation, crucial to quick adaptation in unstable and unpredictable environments (Barrick and Lenski 2013). Besides this, mismatched phenotypes may also play as an intensified selective ground against moderately deleterious alleles due to their more negative effects in the mismatched individuals bearing them, thus limiting their segregation and fixation in the population. This bet-hedging model, which implies an evolutionary process favouring long-term higher population fitness via short-term lower individual fitness, has yet to be empirically evidenced.

To measure the potential evolutionary contribution of mismatched phenotypes, two fitness proxies can be used: genetic variability (measured via quantifying the level of heterozygosity); and/or genetic load (measured via quantifying the level of deleterious mutations). From a population genetics perspective, individual and mean population fitness have been commonly estimated using genetic variability (Lande and Shannon 1996; Saccheri et al. 1998; Bozzuto et al. 2019). The rationale is that low variability translates into low fitness because there is less raw genetic material for selection to act upon, and consequently, more limited bedrock for evolution to carve adaptations (Carvalho 1993; Kardos 2021). An alternative, which has only recently been empirically explored thanks to increased genomic data availability, is to estimate fitness using the accumulation of deleterious mutations (i.e., genetic load; reviewed in Bertorelle et al. 2022). A large body of emerging literature has shown that high levels of genetic load could be equally or more detrimental for population fitness than low levels of genetic variability (e.g., Rowe and Beebee 2003; Agrawal and Whitlock 2012; Benazzo et al. 2017; Robinson et al. 2019; Kyriazis et al. 2021; Mathur and DeWoody 2021).

Even if the relative importance of high genetic variability versus the effectiveness of deleterious allele purging for population persistence is hotly debated (Teixeira and Huber 2021; Kardos et al. 2021), their interplay in large populations is of particular concern. In fact, large populations are expected to have high levels of genetic diversity, but also high levels of masked genetic load (i.e., recessive deleterious mutations that are carried by individuals in heterozygosity without being expressed) (Mathur et al. 2021). Masked genetic load can be detrimental to fitness when it becomes realised load (i.e. deleterious mutations in homozygosity), especially in situations of population size fluctuations (Bertorelle et al. 2022).

Here, we aimed to estimate both genetic diversity and genetic load in the framework of the MMH, using a species that lives in the increasingly highly fluctuating and unpredictable environmental conditions of the Southern Ocean, the King penguin (*Aptenodytes patagonicus*). In fact, adaptive mismatch strategies can be fundamental for species to counteract the increasing asynchrony in natural systems caused by current and future climate change (Kharouba et al. 2018).

The King penguin provides an ideal model to study genetic footprints in the framework of the MMH for two main reasons. First, the reproductive success of this sub-Antarctic seabird is highly dependent on the seasonality of food resources; specifically, the seasonality of the position and the intensity of the Antarctic Polar Front (APF). The APF

is an upwelling zone where, during the austral summer season, adults feed on the species' main prey, mesopelagic lanternfishes (Myctophidae) (Adams and Klages 1987; Hindell 1988; Cherel et al. 1993).

The second reason concerns the breeding asynchrony characteristic of this species due to its long reproductive cycle (e.g., 12 to 14 months, Barrat 1976). Every year, two peaks of hatching take place during the austral summer, generating two groups of chicks (Barrat 1976): an early group, born at the peak of food availability (i.e., match), and a late group, born at the end of the food peak (i.e., mismatch) (Charrassin et al. 1998; Charrassin et al. 2002; Gauthier-Clerc et al. 2002). After a period of intensive feeding and parental care lasting approximately one month (brooding phase), chicks are left alone in the colony and fed occasionally during autumn and winter (crèching phase), until they fledge in the following summer (Barrat 1976). Previous studies have shown that late-born chicks have a higher mortality rate than early-born chicks before and during the first winter (see Fernandes and Bardon et al. in prep - Chapter 3; Weimerskirch et al. 1992; Stier et al. 2014) since they have less time to grow enough body mass until the beginning of winter. In this context, we expected that the high mortality rates in the late group during the first year of life would represent an indicator of stronger selective pressures.

Using this non-model species naturally submitted to contrasted environmental constraints, we analysed whole genomes of individuals born under matched (early-born chicks) and mismatched (late-born chicks) conditions to test for different patterns of genetic variability and genetic load. Considering that both genetic load and genetic variability are used as fitness proxies (Teixeira and Huber 2021; Kardos et al. 2021), we measured both variables in early- and late-born individuals that survived and did not survive until fledging. Our prediction was that survivors would harbour less deleterious alleles and higher heterozygosity in putatively neutral alleles than non-survivors. Because mismatched individuals are thought to be subjected to stronger viability selection, we expected this difference in genetic load and genetic variability between survivors and non-survivors to be stronger in the late group of chicks.

2. Material and Methods

2.1. Sampling design

Sampling took place in the colony of *La Baie du Marin* (here referred to as BDM), Possession Island, Crozet Archipelago (46°24′27″S 51°45′27″E), during the 2020 breeding season. King penguin chicks were sampled from a sub-colony of BDM, called 'Antavia', which represents a naturally enclosed zone with four passages used by the penguins to exit and enter the breeding area.

In order to study chicks born early and late in the same breeding season, we performed two separate sampling sessions by following the historically known hatching peaks of the species in the Possession Island (Descamps et al. 2002). The early session took place between January 25th and January 31st (N=200 early chicks), while the late session occurred between February 25th and March 4th (N=132 late chicks), both during the brooding stage (i.e., when chicks are still kept in the brooding patch of the parents during 3 to 4 weeks). Chicks weighing between 500 to 1000 grams were sampled in a time window that did not exceed 10 days for each group, to avoid sampling intermediate individuals.

Once a chick around the mass interval described above was spotted, it was captured and replaced in the parents' pouch with a heated dummy egg during the whole manipulation to reduce adult stress. After checking the chick's mass, we collected three drops of blood (approximately 100 µL) from the brachial vein using a 25 gauge needle and a microcapillary tube. The blood was immediately transferred to a 1,5 mL microcentrifuge tube containing 700 µL of Queen's Lysis Buffer (Seutin et al. 1991) and stored at -20°C until the DNA extraction. Finally, the chick was equipped with a number-coded plastic tag ("fish-tag", Floy Tag and MFG, Inc. Seattle, WA, USA) attached to the chick's upper-back for recapture at fledging (around 11 months later), before being returned to its parent. Chicks that were not recaptured at fledging were considered dead. All manipulations were approved by the French Ethics Committee (APAFIS#4897-2015110911016428) and the French Polar Environmental Committee (TAAF permit #2019-115 & 2019-129) and conducted in accordance with these guidelines.

2.2. DNA extraction and whole genome sequencing

In order to compare the strength of viability selection between the two phenological conditions, we randomly selected samples from early and late chicks that survived until fledging (i.e., recaptured 11 months later) and chicks that did not survive until fledging (i.e., not found in the colony after 11 months). In this way, we proceeded with the DNA extraction and genome sequencing of 40 chicks in total: 10 early survivors, 10 early non-survivors, 10

late survivors, 10 late non-survivors (**Table 1**). We also extracted RNA from the 20 survivor chicks (10 early and 10 late), as described in Fernandes et al. *in prep.* (see Chapter 2).

After overnight thawing, we extracted the DNA from the whole blood using the Invitrogen PureLink™ Genomic DNA Kit (Thermo Fisher Scientific), following the manufacturer's protocol for genomic DNA extraction from tissue. DNA was quantified with a spectrophotometer and a fluorometer, and molecule integrity was checked with an 1% agarose gel via electrophoresis. Samples were sent to the University of Florence for genomic library preparation using the IDT for Illumina DNA/RNA UD Indexes kit for 150 base paris (bp) paired-end reads. Library preparation and sequencing were performed with all 40 individuals in the same batch to avoid any potential batch effect. Whole genome sequencing was performed in the Illumina NovaSeq 6000 platform in two consecutive sequencing rounds with an expected coverage of 22X (considering the King penguin genome size of 1,25 Gb, Pan et al. 2019). Detailed information about sequenced samples and effective coverage achieved can be found in Table 1.

Table 1. Whole genome sequencing samples. First column shows the total number of sequenced individuals; second column the individual's ID; third column the phenological group in which the individual was born; fourth column, if the individuals survived after winter of not; fifth column, the individual's sex, determined after sequencing with SATC (see **2.3.3. Identification of sex-linked scaffolds and masking** below); sixth and seventh columns indicate the sequencing coverage in the first and second runs, respectively; and the eighth column has the final effective sequencing coverage for that individual.

	ID	Phenological group	Survivor	Sex (SATC)	Coverage run 1 (X)	Coverage run 2 (X)	Effective coverage (X)
1	KP2020_E584	Early	Yes	Male	17.25	13.46	24.56
2	KP2020_E594	Early	Yes	Male	11.96	13.79	20.60
3	KP2020_E143	Early	Yes	Male	11.20	12.73	19.14
4	KP2020_E147	Early	Yes	Female	11.84	14.04	20.70
5	KP2020_E595	Early	Yes	Female	11.21	13.49	19.76
6	KP2020_E599	Early	Yes	Male	11.63	14.15	20.62
7	KP2020_E580	Early	Yes	Male	10.47	13.28	18.99
8	KP2020_E153	Early	Yes	Male	10.03	13.23	18.60
9	KP2020_E169	Early	Yes	Male	12.99	14.05	21.63
10	KP2020_E190	Early	Yes	Female	11.20	13.40	19.68
11	KP2020_E578	Early	No	Female	7.18	10.54	14.18
12	KP2020_E583	Early	No	Female	8.38	11.93	16.25
13	KP2020_E586	Early	No	Female	9.83	12.37	17.76
14	KP2020_E590	Early	No	Female	9.72	12.47	17.75
15	KP2020_E593	Early	No	Male	11.80	13.22	20.01
16	KP2020_E596	Early	No	Male	9.39	12.11	17.20
17	KP2020_E144	Early	No	Female	11.66	13.43	20.07
18	KP2020_E170	Early	No	Female	8.07	11.19	15.41

	ID	Phenological group	Survivor	Sex (SATC)	Coverage run 1 (X)	Coverage run 2 (X)	Effective coverage (X)
19	KP2020_E173	Early	No	Female	10.51	12.58	18.47
20	KP2020_E587	Early	No	Male	12.89	14.10	21.59
21	KP2020_L151	Late	Yes	Male	11.77	13.45	20.18
22	KP2020_L008	Late	Yes	Female	15.63	13.89	23.61
23	KP2020_L019	Late	Yes	Female	16.28	13.61	23.92
24	KP2020_L033	Late	Yes	Male	16.19	13.70	23.91
25	KP2020_L034	Late	Yes	Female	14.54	13.86	22.71
26	KP2020_L044	Late	Yes	Male	14.63	13.94	22.85
27	KP2020_L053	Late	Yes	Female	12.30	13.24	20.43
28	KP2020_L058	Late	Yes	Female	17.06	13.93	24.79
29	KP2020_L059	Late	Yes	Female	10.86	13.32	19.34
30	KP2020_L062	Late	Yes	Male	10.22	13.51	18.98
31	KP2020_L158	Late	No	Female	11.00	13.12	19.29
32	KP2020_L159	Late	No	Female	11.49	13.50	20.00
33	KP2020_L160	Late	No	Female	8.95	11.56	16.40
34	KP2020_L161	Late	No	Male	13.56	14.40	22.36
35	KP2020_L152	Late	No	Male	10.41	13.10	18.81
36	KP2020_L153	Late	No	Male	13.40	14.23	22.10
37	KP2020_L155	Late	No	Male	11.96	13.12	20.06
38	KP2020_L156	Late	No	Male	9.61	12.33	17.54
39	KP2020_L157	Late	No	Female	10.40	12.98	18.70
40	KP2020_L002	Late	No	Male	12.34	13.41	20.60

2.3. Data processing

After receiving the raw sequence reads of the two libraries for the 40 genomes, we performed a first sequence quality check in FastQC v0.11.9 (Andrews 2010). Taking into account that each sample was sequenced twice in two separated flow cells, the preprocessing steps explained below were performed for sample's reads produced by each library separately. The two files for each sample were only merged prior to variant calling, after the deduplication step.

We started by trimming the paired-end reads to remove sequences with a Phred quality score lower than 15 using a sliding window approach in Trimmomatic v3.9 (Bolger et al. 2014). We decided not to be excessively stringent in quality score filtering at this stage to avoid loss of sequence information, as we also filtered for low mapping quality at the mapping step afterwards. Sequencing adapters had already been trimmed in the sequence facility after the demultiplexing step.

We then mapped the trimmed reads to the available reference genome of the king penguin, GCA_010087175.1 BGI_Apat.V1 (Pan et al. 2019) using BWA-MEM2 v2.2.1 software (Li 2013). BWA-MEM works by first producing a reference genome index, followed by the read's alignment to the genome and the production of a BAM (Binary Alignment/Map format) file per sample per library. We ran the BWA-MEM algorithm using the *-M* option to generate Picard tools compatible files. Alignments with mapping quality (MAPQ) below 10 were removed with SAMtools v1.12 (Li et al. 2009) using the *view -q* 10 option. In summary, mapping quality is calculated by the formula $-10(\log_{10}P)$, where P is the probability that the mapping position is wrong, rounded to the nearest integer.

After mapping reads to the reference genome, BAM files were sorted with *samtools sort* (Li et al. 2009) followed by sequence deduplication with the MarkDuplicates tool in Picard (http://broadinstitute.github.io/picard/). The MarkDuplicates algorithm identifies artifactual duplications (e.g., from PCR library construction) by comparing the 5' extremity of both paired-end reads and differentiating the primary and the duplicate reads by ranking reads by the sums of base-quality scores. All duplicates (i.e., artifactual or not) were tagged with REMOVE_SEQUENCING_DUPLICATES=true option. Finally, we merged the two BAM files of the different libraries for a sample using Picard tools AddOrReplaceReadGroups and MergeSamFiles, and indexed the merged BAM files with *samtools index*.

2.3.1. Variant calling

Once we had a unique BAM file per sample (i.e., 40 alignment files), we called the genetic variants with the Genome Analysis Toolkit (GATK) v4.2.2.0 using the *Germline short variant discovery* pipeline for single nucleotide polymorphisms (SNPs) and insertions-deletions (indels) (McKenna et al. 2010; DePristo et al. 2011). This pipeline starts with a dictionary preparation of the reference genome using CreateSequenceDictionary GATK tool. Then, the variant calling is performed in three steps: variant calling per sample, consolidation of all variants in one database, and a final joint genotyping of the database.

The first step was done using the HaplotypeCaller tool, which calls SNPs and indels simultaneously, producing a GVCF file per sample. Next, the contents of each sample's GVCF were stored in a 2D array with the GenomicsDBImport tool. The GenomicsDB array contains information about the genomic position of each SNP/indel (columns) per sample (rows), meaning that the cells of the array contain call data for each sample at a given genomic position. Finally, the GenotypeGVCFs tool was used for the joint genotyping of the database, by calling SNPs and indels across all samples. In this final step, an output VCF (Variant Call Format) file was generated per scaffold, as the King penguin reference genome is assembled at a scaffold level.

2.3.2. SNP filtering

After variant calling, we proceed with the filtering of VCF files to remove unwanted information, such as indels, short and sex-linked scaffolds, as well as to filter out sources of potential error due to the variant calling, such as due to low sequence depth regions. We first removed indels with the SelectVariants tool in GATK, by filtering VCF files with the option *--select-type-to-include SNP* to keep SNPs uniquely. Once we had VCF files that contained SNP information uniquely, we eliminated scaffolds that were shorter than 100 Kb. This was done by first annotating the scaffold's length with the *faidx* command in SAMtools (Li et al. 2009), which indexes the reference genome FASTA file, giving each scaffold's length in the second column of the output list. We then sorted this list by the scaffolds lengths and removed scaffolds shorter than 100 Kb. We proceeded with the filtering and later analysis only with the VCF files from scaffolds present in this list.

The next filtering step consisted of a quality filtering of the SNPs with the VariantFiltration tool in GATK (McKenna et al. 2010; DePristo et al. 2011). We used GATK's standard variant filtering, which filters for SNP call quality, strand bias, and mapping quality considering the reference and alternate alleles. Briefly, the quality filtering removed SNPs with a call quality lower than 30 (QUAL < 30.0), and quality normalised by allele depth (AD) lower than 2 (QD < 2.0). QD is calculated by HaplotypeCaller and GenotypeGVCFs by using the QUAL/AD of heterozygous samples in the reference genome. GATK's strand bias filtering is based on the estimate of how much one DNA strand is favoured over the other during sequencing. We removed SNPs with strand bias greater than 60 (FS > 60.0 estimated by the Fisher's Exact Test).

We used two mapping quality (MQ) filters: the first removes SNPs with median mapping quality lower than 40 for reads supporting each alternative allele (MQ < 40.0); the second removes SNPs with mapping quality rank sum test of reference versus alternative reads lower than -12.5 (MQRankSum < -12.5). Negative MQRankSum values indicate that the reads supporting the alternate allele have a lower mapping quality than those supporting the reference allele. Finally, we filtered the SNPs based on a GATK test called rank sum that evaluates whether there is evidence of bias in the position of alleles within the reads between the reference and alternate alleles (ReadPosRankSum < -8.0). Negative ReadPosRankSum values indicate that the alternate allele is more often found at the ends of the reads than the reference allele.

We also used VCFtools v0.1.17 (Danecek et al. 2011) to filter for genotype's minimum depth per individual, maximum mean depth per loci, and maximum number of alleles per loci. The *--minDP 3* option keeps individuals that have at least 3 reads for that genotype. The *--max-meanDP 50* keeps only loci with a maximum mean count of 50 around all individuals. This filter is used to avoid high mean depths, which can be indicative of paralogs or repeated regions. The last used filter, *--max-alleles 2*, keeps only biallelic loci.

2.3.3. Identification of sex-linked scaffolds and masking

The removal of sex-linked chromosomes (i.e., chromosomes linked to sex determination) is an important step before population genetic analysis, as these

chromosomes evolve in a different manner compared to autosomes (Zhou et al. 2014; Makova 2019). As the reference genome of the King penguin is at scaffold level, we identified and removed sex-linked scaffolds from the subsequent analysis. We used a sex identification approach based on the depth of coverage per sample, known as the Sex Assignment Through Coverage (SATC) method (Nursyifa et al. 2022) in R.

We first produced a file with the depth coverage information per sample based on the BAM alignment files using SAMtools *idxstats* (Li et al. 2009). Then, we proceeded with SATC, which identifies sex-linked scaffolds in four main internal steps. It first normalises the depth of each scaffold within each sample. Second, normalised sequencing depths are projected in a two-dimensional principal component analysis (PCA). Third, SATC clusters the samples on the top PCs using Gaussian mixtures clustering. At the last step, the sex and sex-linked scaffolds are identified from the clustering and the sequencing depth (**Appendix Chapter 1, Figure 1**).

In practice, SATC produces a list of "sex-linked" and "XZ-linked" scaffolds, with the sex-linked nomenclature used to define scaffolds that might not be exclusively from a sex-chromosome according to the genome annotation (Nursyifa et al. 2022). For simplicity, we will refer to both "sex-linked" and "XZ-linked" scaffolds as sex-linked. In addition, SATC also outputs a table with the inferred sex of individuals based on the median depth of the sex scaffolds (i.e., homomorphic or heteromorphic sex). The sex identification of sequenced individuals can be found in **Table 1** in section **2.2. DNA extraction and whole genome sequencing.** Note that, in birds, females are the heterogametic sex, with the ZW genotype, while males are homogametic, with the ZZ genotype (Ohno et al. 1964; Susumu Ohno 1967).

After the removal of sex-linked scaffolds from the dataset, we masked the VCFs to remove repeated sequences with VCFtools *--mask* (Danecek et al. 2011). This option takes as input the VCF files and a FASTA-like file that indicates the loci to be filtered from the reference genome assembly. We produced the FASTA-like genome file by substituting the repetitive regions to be masked (i.e., lower case loci) by the integer 1 and the non-repetitive loci to be kept (i.e., upper case loci) were substituted with a zero. VCFtools *--mask* was then run, filtering out regions with scores higher than zero in the VCF files.

2.4. SNP effect annotation and missing data filtering

In order to measure how early and late chicks accumulate genetic load and variability, we annotated the SNPs based on their predicted fitness effect using SnpEff v5.1 (Cingolani et al. 2012). SnpEff uses the genome annotation of coding sequences (cds) and proteins to predict the impact of each SNP in the protein sequence. By doing so, it classifies SNPs into four main categories of predicted impact: HIGH, MODERATE, LOW, and MODIFIER (Cingolani et al. 2012).

Based on the SnpEff annotation, HIGH impact SNPs are generally related to loss of protein function through the generation of a stop codon, the elimination of a start codon, occur in loci that may alter protein conformation (e.g., AA regions inside the same protein),

change of a rare amino acid, among other factors. This set of SNPs are usually the most infrequent ones and here are considered as highly deleterious. MODERATE impact SNPs are usually related to non-synonymous changes in the coding sequence, like a codon insertion or deletion, and an amino acid change, and hence tend to be less harmful than the HIGH impact SNPs. LOW impact SNPs almost always involve a synonymous codon change that can also be in start and stop codons, and are expected to affect an individual's fitness even less than moderate SNPs. Finally, MODIFIER SNPs are the most numerous and consist of mutations for which there is no predicted impact, as they are usually in non-coding regions (e.g., intronic and intergenic) or non-coding genes. Because MODIFIER SNPs can also include mutations inside genes (intronic), we further filtered this category by removing SNPs located at least 50 Kb from the closest gene (which should be free of linkage to the closest genes, Balakrishnan and Edwards 2009). Therefore, this MODIFIER set of SNPs comprised the putatively neutral mutations in our dataset.

This type of SNP-effect prediction represents a reasonable approximation of the impact of each SNP on individual fitness. Indeed, the actual outcome of each SNP on our individuals could only be assessed with empirical evidence. Given this, the subsequent analyses did not focus on the particular impact of each SNP, but on the cumulative impact of each class of SNPs in the genomes of early and late individuals that survived and died during the first year of life. In any case, for ease of discussion, we will also refer to each category as highly, moderately, low, and non-deleterious when referring to HIGH, MODERATE, LOW, and MODIFIER impact categories, respectively.

After categorising our total set of SNPs into the four described SnpEff predicted classes, we merged the VCFs of each category together in one file containing all individuals with MergeVcfs in Picard tools. Thereby, we concluded this part of our genomic pipeline with four VCF files, each one containing the HIGH, MODERATE, LOW, and MODIFIER effect SNPs in our 10 early survivors, 10 early non-survivors, 10 late survivors, and 10 late non-survivors.

A large amount of missing data (i.e., SNPs absent in many samples) can decrease analysis power, or even bias the results if missing data are not evenly distributed in the dataset (O'Leary et al. 2018; Wright et al. 2019). Our missing data filtering strategy aimed to eliminate substantial amounts of missing data while accounting for a balance of missingness within each group (i.e., early survivors, early non-survivors, late survivors, and late non-survivors). To calculate these means, we divided our filtering into two steps. First, we filtered the VCFs for a maximum missing count (MAC) of 2 within each group (option *--max-missing-count* in VCFtools, Danecek et al. 2011). Then, we refiltered for a MAC of 8 across all individuals without group separation.

In practice, we redivided each SNP category VCF into four VCFs, per penguin group. For example, for the high impact SNP, we divided the VCF file with all 40 individuals into four VCFs, each one containing the 10 individuals of each penguin group. After the MAC filtering of 2 counts in each penguin group's VCF with VCFtools *--max-missing-count*, we merged the filtered VCFs into one file per SNP category with BCFtools v1.12 *merge*

(Danecek et al. 2021). Then, we refiltered the VCF with all individuals for a MAC of 8 counts. After this last filtering step, we proceeded with the analyses using the four VCF files per SNP category.

2.5. Analyses

We started our analysis by doing a Principal Component Analysis (PCA) for each SNP category (i.e., HIGH, MODERATE, LOW, and MODIFIER) using SNPRelate v1.20.1 (Zheng et al. 2012) in Bioconductor, in order to explore individual genetic clustering in our dataset. Due to the large number of detected modifier SNPs (i.e., 12 million), we selected a random subset of those SNPs for the PCA using the command *vcfrandomsample* in the vcflib software (Garrison et al. 2022). We used the option -r 0.008 to indicate the fraction of variants we wanted to retain from the original VCF file, which corresponds to approximately 100,000 SNPs out of the 12 million total MODIFIER SNPs.

Then, we calculated genetic divergence between our groups by estimating weighted Weir and Cockerham F_{ST} statistics (Weir and Cockerham 1984) using VCFtools *--weir-fst-pop* (Danecek et al. 2011) across all four sets of SNPs. We calculated the F_{ST} per site between survivors and non-survivors within the early (N=20) and late (N=20) groups. Weir and Cockerham's F_{ST} calculation consists of an analysis of covariance (ANOVA) approach that estimates between population variance while correcting for various population sizes (Weir and Cockerham 1984). Although F_{ST} statistics range from 0 to 1, where values closer to zero represent higher population admixture, Weir and Cockerham's F_{ST} calculation in VCFtools can also generate negative values when there is higher genetic variation within the population than between populations. In such cases, negative values can be considered as zeros for ease of interpretation (i.e., absence of population structure).

We also plotted the mean and median pairwise F_{ST} per comparison against a random distribution of F_{ST} to correct for any bias that might have occurred due to small sample sizes (i.e., N=10 per group because of the limited number of late survivors). For this, we performed 1000 iterations of the F_{ST} per site calculation between random groups of individuals from the 40 samples (each iteration was performed with a random set of individuals) with a custom Python script. We then calculated the *P-values* between the mean and median F_{ST} in the true groups against the random distribution.

Even if F_{ST} is a widely used summary statistic for the determination of population divergence, particular attention must be given to the interpretation of low F_{ST} values. Low F_{ST} values reflect situations in which the observed allele frequency differences are inferior to differences that arise by accident (Neigel 2002). In any case, our main objective was not to detect global population structure, but to determine whether the two hatching groups have accumulated genetic variability differently. In order to detect such fine-scale differentiation between survivors and non-survivors in the early and late groups, we used a

somehow rawer summary statistic that allows for the direct comparison of the allele frequencies between groups, the absolute Allele Frequency Difference (AFD) (Berner 2019).

The AFD is proposed as an alternative for measuring population differentiation and should be more sensitive than the F_{ST} in cases where population structure is weaker (Berner 2019). The AFD calculation relies on a simple formula, in which the sum of minor alleles in one group is subtracted by the sum of minor alleles in another group:

$$AFD = \sum min\_alleles_{group1} - \sum min\_alleles_{group2}$$

In our case, we calculated the AFD for the minor allele in all four SNP sets. We used the minor allele frequency since our main focus was on the segregation of deleterious mutations (MODERATE and HIGH in particular) between groups, and these types of mutations are usually found at low frequencies in the population (Sunyaev et al. 2001). Also for this reason, we did not filter for a minimum allele frequency. As described above for the F_{ST} estimates, we also tested the AFD calculated in true early and late groups against a random distribution, using 1000 iterations of random groupings of individuals.

We make an observation regarding the biological meaning of using the minor allele for the AFD. For the HIGH impact SNPs, we considered the minor allele as an equivalent of the derived allele that should be highly deleterious, due to the low frequency at which these alleles are found in the population, as explained above. However, for the MODERATE, LOW, and MODIFIER SNPs, the use of the minor allele as a proxy of deleteriousness does not hold, which is obvious in the case of LOW and MODIFIER. In our analyses, the use of the minor allele in the less deleterious SNP categories was done as a matter of standardisation in the AFD calculation, as the AFD formula demands a focal allele to be used for comparison.

In order to test whether early and late chicks show different patterns of segregation between dead and survivor individuals for polymorphic sites in each category, we compared the distribution of genotype counts (i.e., heterozygous, homozygous for the minor and major alleles) per individual among the four groups. In particular, we performed linear regressions per SNP category in R (R Development Core Team, 2022), fitting a linear model (*lm*) in which genotype counts are predicted by both survival (i.e., survivor or non-survivor) and phenological group (i.e., early or late) (model fit_1 = genotype ~ group*survival). We then tested this model against a null model, in which genotype counts are not predicted by these factors (model fit_0 = genotype ~ 1), through an analysis of variance (ANOVA) in R. In cases where the ANOVA was significant, we decomposed our model in order to detect the contribution of survival and phenological group separately (model fit_2 = genotype ~ group + survival), performing a chi-squared test (*chisq*) to test for significance.

3. Results

3.1. Absence of global genetic structure

The SnpEff categorisation predicted 12,107,720 MODIFIER, 94,837 LOW, 53,219 MODERATE, and 649 HIGH impact SNPs in the king penguin genomes in our dataset. Independently of the SNP category, no global separation or clustering pattern was seen among early and late individuals that survived or not after winter with the PCA (**Figure 1**). Even if a few samples appear to be distant from the main cluster in the PCA plots (**Figure 1**), the variance in both principal components (PC1 and PC2) is so low that they cannot be considered as outlier individuals (maximum PC variance is lower than 1.3 % in the HIGH impact PCA). This can also be observed with the standard deviation bars.

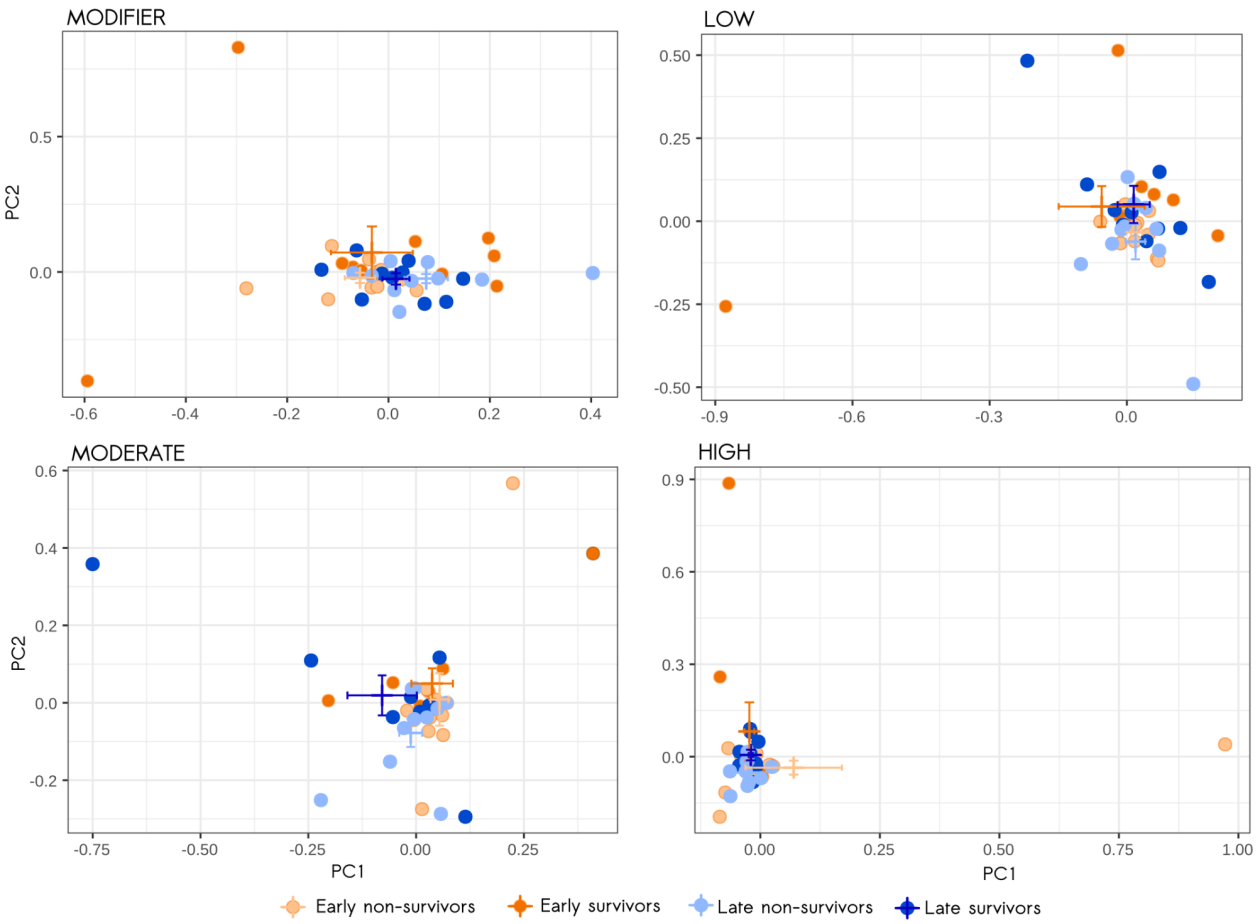


Figure 1. Principal component analysis (PCA) per SNP category in the 40 penguin chicks: early survivors (dark orange), early non-survivors (light orange), late survivors (dark blue), late non-survivors (light blue) per SNP category (modifier, low, moderate, and high impact SNPs). MODIFIER SNPs are a random subset of 96,615 SNPs from the total 12 Gb SNPs in this category. Bars represent the standard deviation in each group.

In line with these results, mean F_{ST} estimates also showed no evidence of genetic divergence between survivors and non-survivors in either early- or late-born individuals in

all SNP categories (**Figure 2**). Mean F_{ST} per site in all comparisons were extremely low, with the maximum value of 0.0102 between late survivors and non-survivors in the MODIFIER SNPs. This particular F_{ST} comparison was also the only statistic with a *P-value* < 0.05 when tested against a random distribution of F_{ST} (*P-value*=0.047, **Figure 2**). However, this result should be interpreted with caution, as the *P-value* can slightly oscillate depending on the iteration of random sampling, and was not always significant when we strictly used a *P-value* threshold < 0.05.

In addition, as briefly mentioned in the Material and Methods, small F_{ST} values should be interpreted with caution, as this statistic is less sensitive to slight changes in allele frequencies when population differentiation is low. In other words, the relationship between F_{ST} and allele frequency changes is non-linear, and it may inflate F_{ST} values when population differentiation is high, while disproportionately deflating F_{ST} when population differentiation is low (Berner 2019), as is our case. For this reason, we cannot assume any level of genetic differentiation between survivors and non-survivors in the early and late groups by these results.

As the King penguin is a largely panmictic species (Cristofari et al. 2018), the lack of global genetic differentiation is in line with our expectations. In fact, early and late breeders are known to come from an overlapping set of individuals that start reproducing early or late depending on their success in the previous season (Stonehouse 1960; Barrat 1976) (i.e., early breeders that successfully fledge a chick in one season will start late in the next season, or will skip a breeding year).

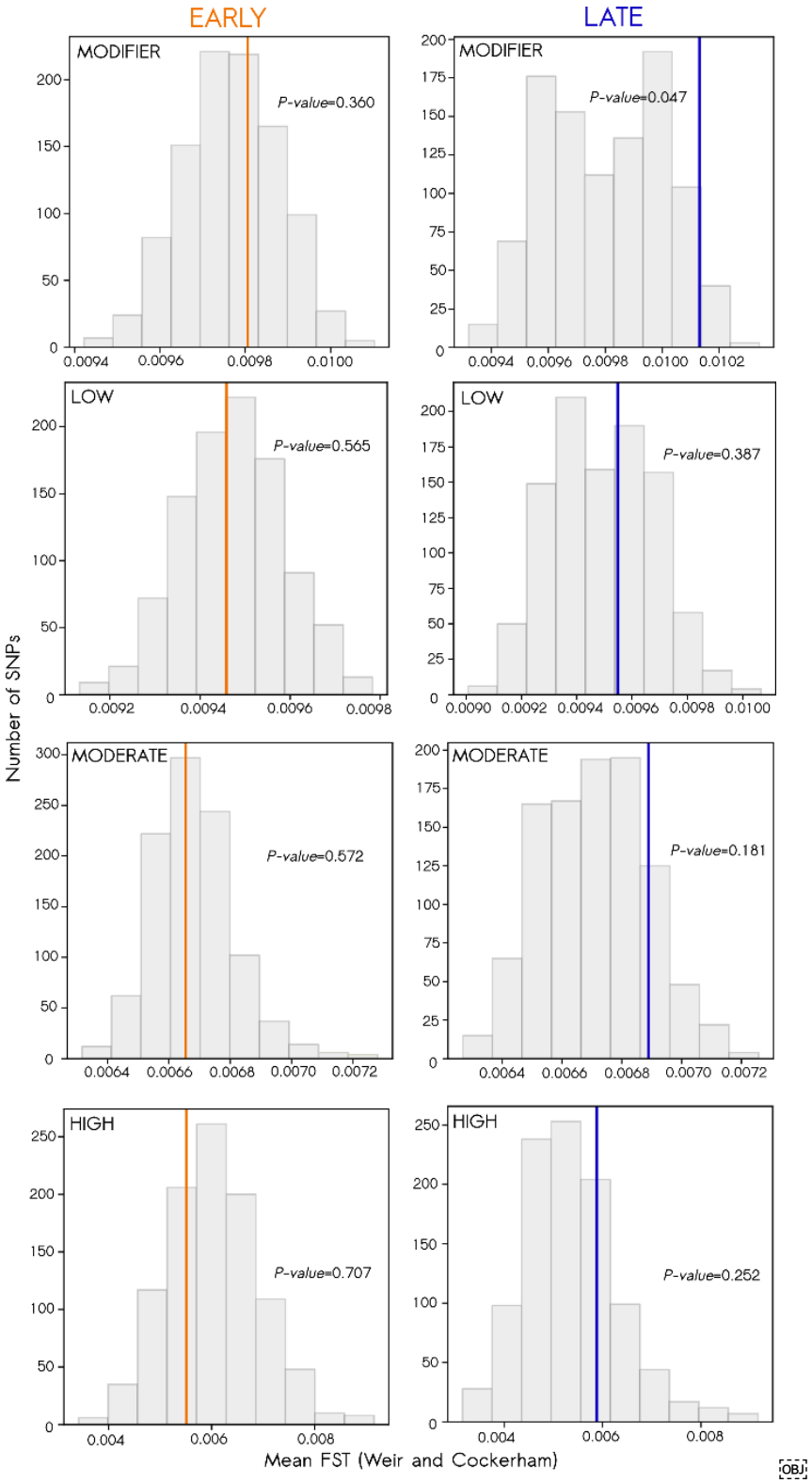


Figure 2. Mean F_{ST} per site between survivors and non-survivors in early and late groups tested against 1000 random grouping iterations. From top to bottom, graphs represent the mean F_{ST} in MODIFIER, LOW, MODERATE, and HIGH impact SNPs. Histograms represent the 1000 iterations of F_{ST} calculations per site with random groupings of individuals. Orange and blue lines represent the mean F_{ST} per site between survivors and non-survivors in the early and late groups, respectively. *P-values* represent the probability that the mean F_{ST} between early and late survivors and non-survivors was found by chance, in the random sampling of individuals.

3.2. Genetic variability and load

We calculated the allele frequency differences (AFD) of minor alleles between survivors and non-survivors within the early and late groups, testing the mean AFDs against a random distribution. Our hypothesis here was that minor alleles are deleterious, which is a good approximation in the case of HIGH effect SNPs, and we tested whether survivors have less minor alleles. A *P-value* < 0.05 is supportive of a statistical difference in this direction. However, a *P-value* > 0.95 would be supportive of a statistically higher proportion of minor alleles in the survivors. Interestingly, we found significant AFD in the HIGH impact SNPs between early survivors and non-survivors (*P-value*=0.009), but not between late survivors and non-survivors (*P-value*=0.625) (**Figure 3**). The negative mean AFD in the early group (-0.0052) is indicative of a higher frequency of highly deleterious alleles in non-survivor early individuals. Late non-survivors do not seem to have a higher frequency of these mutations when compared to late survivors. The other SNPs categories did not produce any significant AFD between early and late survivors and non-survivors (*P-values* > 0.05) (**Figure 3**). In the case of MODIFIER SNPs, we observe a slightly significant result in the opposite direction (higher proportion of minor alleles in the survivors *P-value* = 0.045).

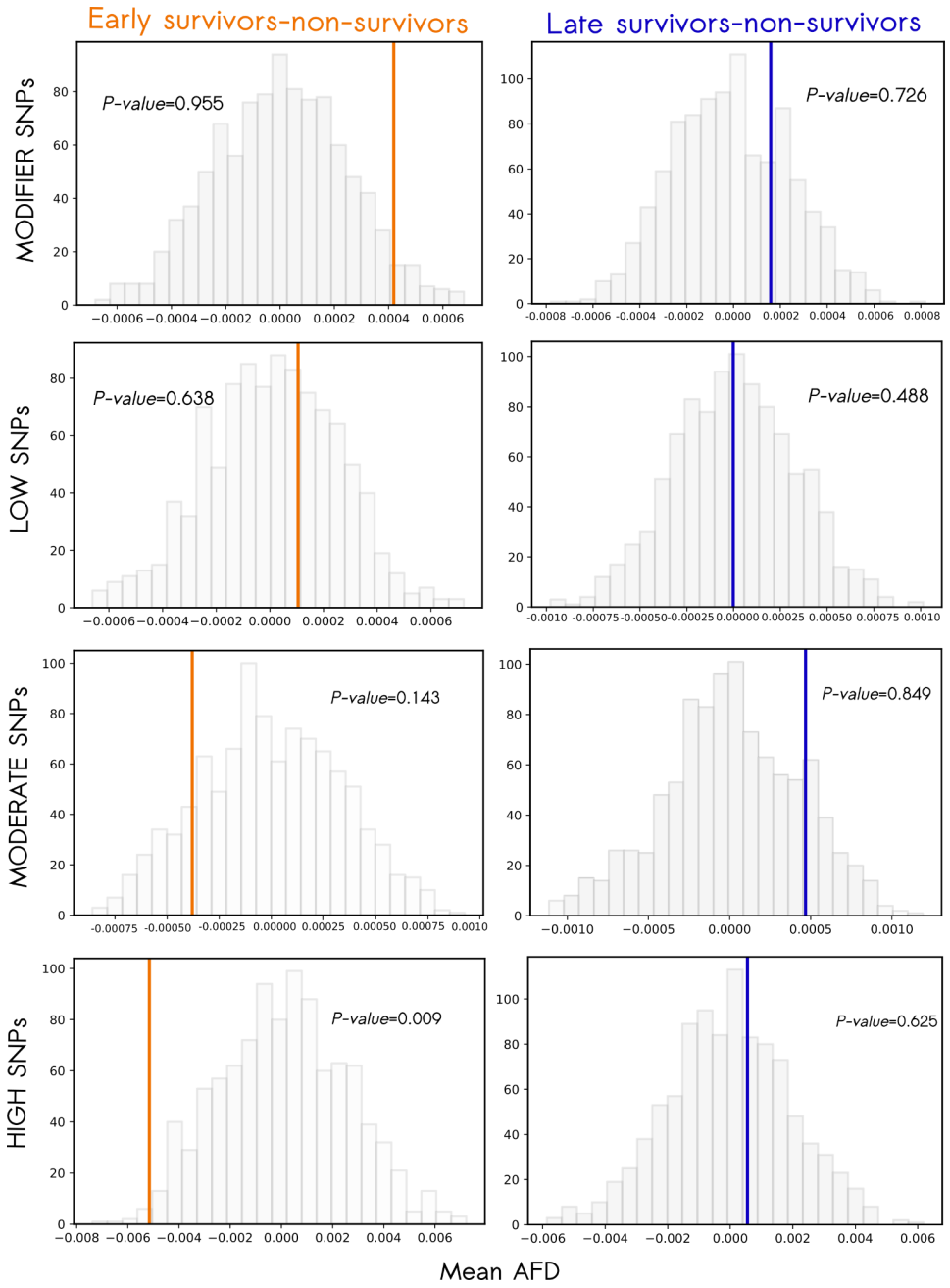


Figure 3. Allele frequency differences (AFD) of minor alleles between survivors and non-survivors of early and late chicks. From top to bottom, graphs represent the mean F_{ST} in MODIFIER, LOW, MODERATE, and HIGH impact SNPs. Histograms represent 1000 iterations of AFD calculations with random groupings of individuals. Orange and blue lines represent the mean AFD between survivors and non-survivors in the early and late groups, respectively. *P-values* represent the probability that the mean AFD between early and late survivors and non-survivors was found by chance, in the random sampling of individuals.

To investigate how the different SNPs are distributed within groups in more detail, we compared the allele counts per genotype in each SNP category (**Figure 4**). As can be seen in Figure 4 and in Table 2, some of the observed genotype counts can be related to the group and/or the survival of the chick (i.e., significant ANOVA and linear regression results, **Table 2**). All *chisq* values mentioned in the text below make reference to the results in **Table 2**, unless otherwise explicitly indicated.

MODIFIER SNPs (i.e., likely non-deleterious variation) represented the only SNP category in which all genotype counts could be predicted by survival, and in some cases also by group (**Table 2**). More specifically, heterozygous genotypes were significantly more prevalent in survivor individuals than in individuals that did not survive until fledging (surv *chisq*=0.001975) independently of the group (group *chisq*=0.126521) (**Figure 4b**). Accordingly, homozygous genotypes for both minor and major alleles were more frequent in non-survivors (**Figure 4c**). In this case, the quantity of homozygous sites for the major allele is only related to survival (surv *chisq*=0.002531) (**Figure 4a**), while for the minor allele, both survival and group have an impact on genotype counts (group *chisq*=0.031475; surv *chisq*=0.004789) (**Figure 4c**).

In other words, late-born chicks showed a lower accumulation of homozygous sites for the minor allele than early-born chicks, even though no major assumption can be made about the minor allele in this category. The effects of MODIFIER SNPs minor or major alleles cannot be known from our unpolarized data, and especially when no other information about the sequence is known (e.g., conservation level). Minor alleles could represent new mutations that can still have a frequency increase in the population due to positive selection, or even mutations that we still do not have information about the impact (e.g., in a regulatory intragenic region) (Park et al. 2011). In general, this SNP category should be characterised by neutral mutations, or mutations whose impact on fitness is not known.

For these reasons, no conclusive interpretation can be drawn regarding the major and minor alleles in homozygosity in this category for the moment. As will be explained in the **Conclusions and perspectives section** of this chapter we are currently producing a multi-species alignment to acquire SNP conservation scores in order to have a better understanding of the relative importance of each region. In any case, what our results show so far is that survivor individuals from both early and late groups were less homozygous than the non-survivor individuals from their respective groups.

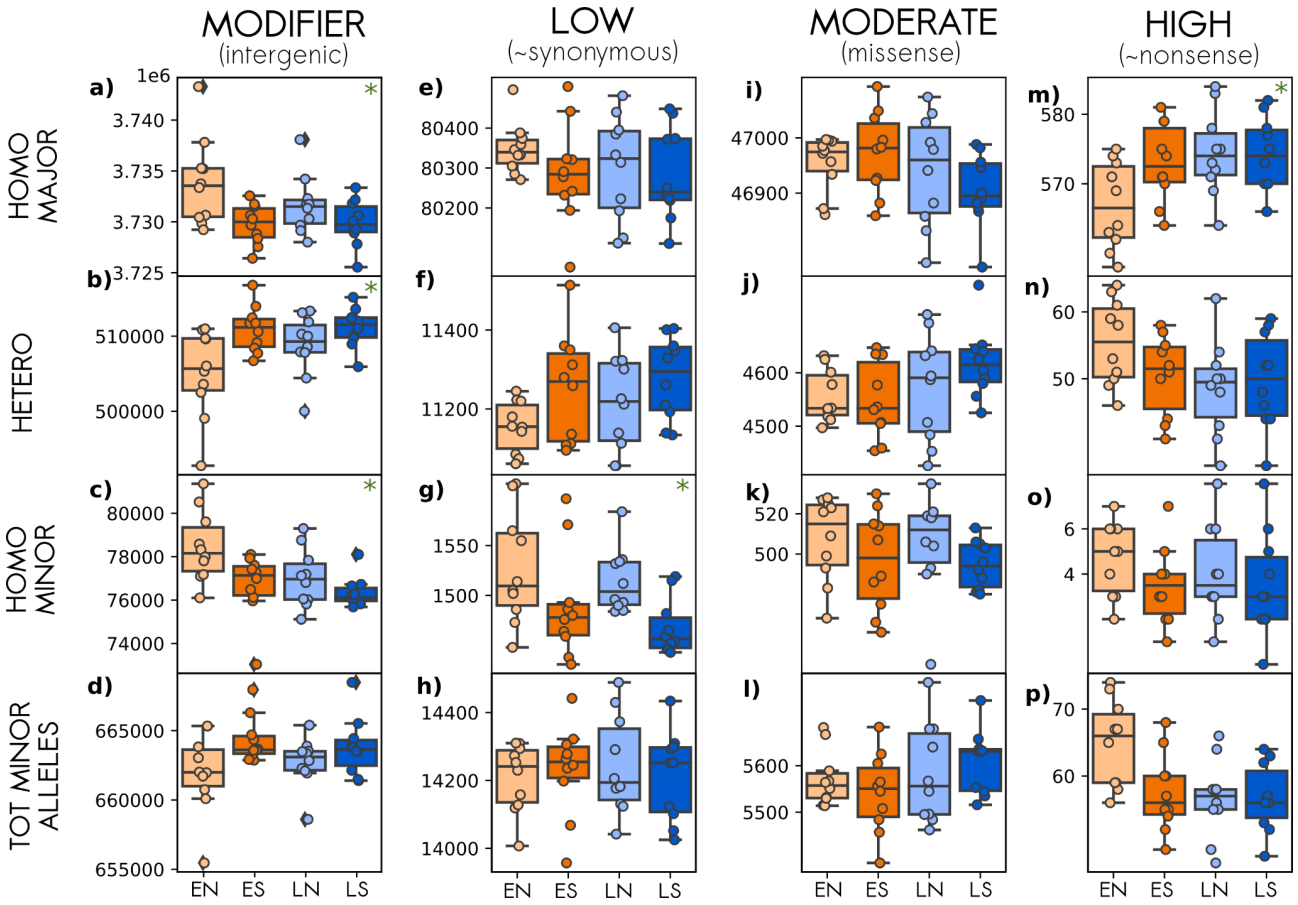


Figure 4. Distribution of minor alleles based on SnpEff predicted categories. Boxplots of the counts of SNP by category, according to SnpEff, in each of the four groups of chicks: EN early non-survivors (in light orange), ES early survivors (in dark orange), LN late non-survivors (in light blue), LS late survivors (in dark blue). From left to right: MODIFIER, LOW, MODERATE, and HIGH impact SNPs. The four bottom plots display the total number of alleles of each SNP category within all genotypes in each group; the bottom roll shows all the minor allele counts (contained in both heterozygous and homozygous for the minor allele genotype); the second bottom row shows the SNPs in homozygous genotypes for the minor allele; the third bottom row, the SNPs present in heterozygous genotypes; and the top row, the SNPs in homozygous genotypes for the major allele. Green asterisks are indicative of significant ANOVA tests of phenological groups combined with survival (*P-value* < 0.05, **Table 2**).

Table 2. *P-values* of the ANOVA of genotype counts prediction by a combination of phenological group and survival per SNP category. The ANOVA test was performed between the linear model in which genotype counts are predicted by the combination of being an early- or late-born and surviving or not-surviving (fit_1 = genotype ~ group*surv), and a model in which genotype counts are not predicted by those factors (fit_0 = genotype ~ 1). *P-values* < 0.05 (*) suggest that the two models are significantly different. When *P-values* were significant (in bold), we performed another linear regression model separating the effects of group and survival (fit_2 = genotype ~ group+surv). The contribution of group and survival to genotype counts was calculated with a chi-squared test (*chisq*).

	MODIFIER	LOW	MODERATE	HIGH
Homozygous major allele	**0.0127*** Group: 0.255771 **Surv: 0.002531**	0.551	0.2431	**0.02814*** **Group: 0.03038** Surv: 0.14478
Heterozygous	**0.006736*** Group 0.126521	0.08792	0.1652	0.1271

	Surv: 0.001975			
Homozygous minor allele	**0.006003*** **Group: 0.031475** **Surv: 0.004789**	**0.02509*** Group: 0.209561 **Surv: 0.0029515**	0.4645	0.4837

After observing significantly higher heterozygosity of MODIFIER SNPs in survivors, we wondered whether such heterozygosity would be concentrated in specific genomic regions, or if it was a widespread pattern throughout the genome. To explore this, we explored the heterozygosity difference between the survivors and non-survivors in 5 Kb windows for all SNPs in both groups (**Figure 5**). We observe that peaks of differences in heterozygosity are apparently uniformly distributed along the genomic windows, for both early and late. This widespread heterozygosity, mostly caused by the numerous MODIFIER SNPs, indicates that genetic variability is not concentrated in specific regions.

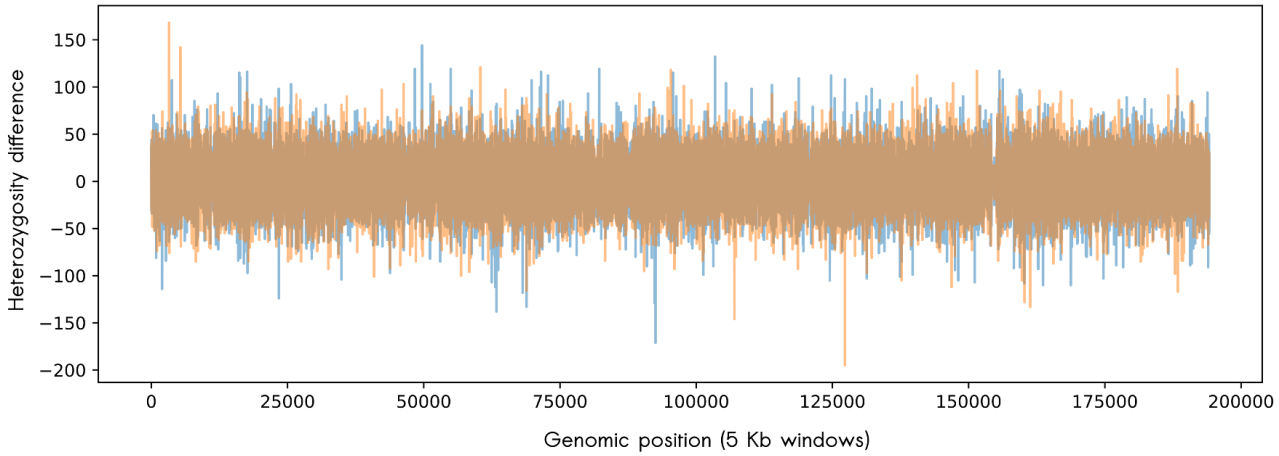


Figure 5. Genome wide heterozygosity by 5 Kb window. Difference between survivors and non-survivors heterozygosity per 5 Kb window in the genome with all SNPs. Orange and blue bars represent early- and late-born individuals, respectively. Positive values represent higher heterozygosity in survivors, while negative values represent higher heterozygosity in non-survivor.

The LOW impact SNPs showed a similar pattern to the MODIFIER SNPs, with an apparent accumulation of heterozygous genotypes in surviving individuals, and homozygous genotypes in non-surviving individuals (**Figure 4e-g**). Despite this trend, only the homozygous genotypes for the minor allele were significantly predicted by survival (surv *chisq*=0.0029515) (**Figure 4g**). Even though LOW impact SNPs are located inside genes, these are mainly synonymous sites where neither the major nor the minor allele cannot be considered as the deleterious. Consequently, for the moment, we interpret the LOW impact SNP results similarly to the MODIFIER SNP results where heterozygosity is taken as a proxy of genetic diversity.

On the other hand, MODERATE impact SNPs did not show any statistically significant relationships between genotype counts, group and survival (**Table 2**). Visual comparison of genotypes' distributions suggests that survivor individuals (from both early and late groups) have an apparent lower accumulation of homozygous genotypes for the minor allele

(**Figure 4k**). However as mentioned, minor alleles at MODERATE impact sites cannot be confidently considered as deleterious. As suggested by their category which includes mainly non-synonymous sites, moderately deleterious mutations have lower selection coefficients and are affected by less intense purifying selection so that some of them can also reach high frequency in the population. Therefore, we cannot approximate the MODERATE SNPs' impact by the frequency in which they are detected in our dataset. Moreover, this dataset can still be filtered through sequence polarisation, by using the derived alleles as an approximation of the most deleterious alleles (Grossen et al. 2020).

Finally, we detected a general accumulation of HIGH impact SNPs (mostly nonsense mutations) in early non-survivors in comparison to all other groups (**Figure 4p**), as had also been observed in the AFD analysis. In other words, early non-survivors harbour higher quantities of minor alleles expected to be highly deleterious. Such alleles are both in heterozygous genotypes (i.e. masked load), which will contain one copy of the allele, and in homozygous genotypes for the minor allele (i.e. realised load), containing two copies of this allele.

Even if not statistically significant, the prevalence of deleterious alleles in the early non-survivor group seems to be mainly due to its occurrence in heterozygous genotypes (**Figure 4n**) than in homozygous genotypes for the minor allele (**Figure 4n**) (HETERO group *chisq*=0.0701; HOMO MINOR group *chisq*=0.5554, not shown in **Table 2**). Indeed, due to the likely highly deleteriousness of these SNPs, the quantity of homozygous SNPs for this allele is extremely low even for non-survivors (i.e., maximum 7 loci).

The only statistically significant comparison was in the HIGH impact SNP category for the homozygous genotypes of the major allele, which should be the non deleterious one (*P-value*=0.02814, **Figure 4m**). When decomposing the factors possibly causing the difference in counts for this genotype, we detected a significant relationship only with the phenological group (group *chisq*=0.03038), but not with survival (surv *chisq*=0.14478), even though early non-survivors show the most diverging distribution of homozygous genotypes for the major allele (supposedly non-deleterious). In other words, early non-survivors have an apparent higher frequency of deleterious alleles compared to early survivors and late individuals. Furthermore, late non-survivor individuals do not seem to harbour higher levels of highly deleterious alleles compared to late individuals that survive (at least until fledging). In fact, early survivors, late survivors and non-survivors seem to have a similar mean of total highly deleterious alleles across all genotype categories. This was surprising, as our initial hypothesis was that late-born individuals that survive until fledging would be subject to stronger purifying selection, which would purge highly deleterious alleles from the population more efficiently than in the early group. In spite of our predictions, our results suggest that late-born individuals, whether surviving through winter or not, already start with lower levels of deleterious mutations in comparison to early-born individuals (i.e., lower levels of genetic load in the late compared to the early group).

4. Discussion

The main objective of this study was to investigate whether individuals born in mismatch with resources can help to purge deleterious mutations from the population while maintaining high genetic variability. Of particular interest was the widespread genomic effect of breeding under mismatch in a large natural population, which can hold high levels of masked genetic load. To this end, we evaluated the distribution of SNPs with different predicted effects from the whole genomes of 40 king penguin chicks under matched or mismatched conditions that either survived or did not survive the first year of life before fledging. We expected individuals that survive through the first year of life to show lower levels of deleterious mutations and higher levels of neutral genetic variability than non-survivors, independently of being born in match or mismatch with conditions. Because late-born individuals are expected to suffer from stronger selective pressures in this first year (higher mortality rates, Olsson 1996; Weimerskirch et al. 1992; Stier et al. 2014), we expected this pattern to be stronger in the late (mismatched) compared to early (matched) group.

4.1. Late-born individuals have lower genetic load

Differently from our prediction on genetic load, highly deleterious alleles were more prevalent in non-survivors of the early-born group, but not in non-survivors from the late-born group. However, this was due to the fact that late chicks that did not survive had similar lower levels of deleterious mutations than all survivor chicks, from both early and late groups. Our initial expectation, that the late-hatchling would more efficiently segregate deleterious mutations, was based on the fact that early- and late-breeders (parents) are not considered as separate groups of adults, as a successful early-breeder in year *n* can be a late-breeder in year *n*+1 (Descamps et al. 2002). Thus, we had assumed early and late chick's alleles to have originated from the same genetic pool, and, consequently, that early- and late-hatchlings would be born with an analogous baseline of deleterious mutations. In this scenario, from all chicks born in the late group, those with a higher accumulation of deleterious mutations would have not been able to outlast the intense selective pressure of winter fasting. In spite of this, what we observe is that even late individuals that do not survive until fledging already start with a lower baseline of highly deleterious mutations.

This suggests that individuals with higher genetic load are not even born in mismatched conditions in this species, while matched conditions allow for a more heterogeneous group to be generated. In this case, we suggest that mismatch imposes stronger selection before chick hatching, acting pre-fecundity (i.e., adults with higher levels of deleterious mutations do not try breeding late) or post-fecundity (i.e., eggs with high genetic load foetus do not hatch). Individuals breeding in match with resources, on the other hand, do not seem to suffer from such pre-hatching selective pressures with such strong intensity.

Furthermore, independently of hatching phenology, our results indicate that viability selection in the King penguin plays an important role in removing highly deleterious mutations from the population under both match and mismatch. If we exclusively consider the selective pressures during the first year of a chicks' life, we can conclude that breeding under mismatched conditions selects for breeders with lower genetic load.

However, because our dataset is composed of a unique sampling year, and king penguin reproductive success is dependent on environmental conditions, whether purifying selection can be less stringent in both matched and mismatched groups in years when food is more abundant (e.g., depending on the oceanographic oscillations that influence the position of the APF, Freeman et al. 2016), is still an open question. The inclusion of genomic data from individuals born in different years with variable chick survival rates could elucidate the real impact of the recruitment of late chicks to the population's genetic load. The inclusion of time-series data is planned in the next steps following this project.

4.2. Lower homozygosity in mismatched genotypes

Regarding the distribution of genetic variability on survivors and non-survivors, our results support our initial prediction that individuals that survive until fledging harbour higher levels of heterozygosity in putatively neutral alleles (MODIFIER SNPs) than non-survivors. This suggests that heterozygosity levels are likely related to survival through the first year. Complementary to this, high levels of homozygous genotypes for these putatively neutral alleles were related to mortality.

As has been long posited by classic population genetics theory, neutral genetic variability is expected to have an impact on population level fitness, through the increase of the adaptive potential of the species (Willi et al. 2006; Kardos et al. 2021). At the same time, the main impact of high homozygosity levels is related to inbreeding depression (i.e., reduction in fitness due to the combination of alleles identical by descendent), caused by recessive deleterious alleles that combine more often in small populations (Charlesworth and Willis 2009). However, we are not aware of previous studies showing such a strong direct impact of homozygosity of nearly neutral alleles in individual mortality, and therefore fitness, in a large population.

More information is needed to know if all MODIFIER intragenic mutations are indeed neutral or nearly neutral. We are working on the generation of conservation scores (e.g., GERP scores, Cooper et al. 2005) from a multispecies bird alignment (Feng et al. 2020) to obtain a second measurement of SNP impact. Sequence conservation scores provide clearer information about sequence neutrality than the categorisation predicted from sequence annotation, as in SnpEff, although such scores can be less clear for defining highly deleterious mutations (Grossen et al. 2020). Independent of this, we are also aware that even apparently neutral intergenic loci can have an impact on fitness, such as gene expression regulation, RNA editing, and protein folding (Chamary et al. 2006).

In fact, non-survivors of both groups have an accumulation of homozygous genotypes for the minor allele, which could be potentially slightly more deleterious (Kido et

al. 2018). Additionally, the whole early group harbours more of these genotypes than the whole late group, which could mean a lower accumulation of genetic load in the late group if the minor alleles are actually more deleterious. However, as mentioned above, this is merely a speculation, as we cannot know the real effect of minor and major alleles at the moment. In any case, the overall strikingly high amount of putatively neutral genotypes that differentiate survivors and non-survivors in our data suggests that there could be a more widespread fitness effect of intergenic heterozygosity than previously thought.

4.3. Genetic load and variability as predictors of survival

Conservation genetics is a realm of population genetics focused on the use of evolutionary and molecular genetics applied to species conservation (Frankham 2010). Even with the increasing number of conservation genetics studies concerned with the capacity of species or populations to persist under rapid climate change (Benazzo et al. 2017; Bozzuto et al. 2019; Robinson et al. 2019; Grossen et al. 2020; Kardos and Luikart 2021), there is still no consensus about the use of genetic variability or genetic load as a proxy of fitness and extinction risk (Teixeira and Huber 2021; Kardos et al. 2021). This is probably due to the complexity of life history strategies, demographic histories and interactions in natural systems. Such complexity makes it difficult, if not impossible, to use a generalisable proxy that would be adequate to most species. However, the use of different genetic markers and statistics can also lead to discordant conclusions that may inflate this discussion.

For example, Teixeira and Huber (2021) have argued that nucleotide diversity (π), a commonly used measure of genetic variability, should not be used as a predictor of extinction risk. They demonstrate that some Critically Endangered species according to the International Union for Conservation of Nature (IUCN) Red List status, such as the gorilla (*Gorilla gorilla*), can show similar or even higher levels of nucleotide diversity than Least Concern species, such as the common minke whale (*Balaenoptera acutorostrata*). On the other hand, other Least Concern species, such as the naked-mole rat (*Heterocephalus glaber*), can have equally low levels of π than some Critically Endangered species, such as the West African chimpanzee (*Pan troglodytes verus*). The authors argue that low nucleotide diversity is mostly a consequence of small effective population sizes, in which genetic drift acts more strongly than selection. In this case, lower genetic diversity is just a reflection of demographic history and is not representative of the impact of selection on population fitness. The opposite can also be true, as species considered as Critically Endangered can keep higher levels of genetic diversity in a larger core population, while having peripheral populations suffering from rapid loss of genetic variability, such as in the case of gorillas (van der Valk et al. 2018).

Indeed, nucleotide diversity may not be a suitable proxy for species risk assessment, but mostly if demographic history is not taken into account. Actually, π is known to be highly affected by population demography, as it depends on the total amount of alleles in the population (Coop 2020). In addition, there is a lag of time between demographic decline

and genetic variability loss, so the higher levels of genetic diversity in endangered species could decrease drastically in a few generations, as demonstrated by (Kardos et al. 2021). These authors defend that genome-wide variation should not be taken for granted when assessing extinction risk. Kardos et al. claim that criticism over the use of genetic variability, and especially π, as a proxy is mostly unjustified, as opposers tend to judge it without considering the factors that lead to the observed genetic variability, such as demography, number of sampled individuals and how much they are actually representative of the population.

Finally, Kardos et al. (2021) also advocate for the use of genome-wide variability, as a broader set of widespread loci can be more easily used for conservation decision making than a smaller amount of deleterious loci. In this troubled scenario, our results actually give hints that both measures can be indicative of fitness in a wild population. In our study, we detected that both a genome-wide lack of heterozygosity and a higher accumulation of highly deleterious mutations are related to early-life mortality in the king penguin. Moreover, considering mortality as a component that drags individual fitness to zero (Orr 2009), we can suggest that both high genetic variability and low genetic load seem to be proxies of fitness. Whether one of those factors can be predictive of the other is still an open question in which we are currently working on. Our future objective is to test if individuals with higher heterozygosity at neutral loci also harbour lower amounts of genetic load.

Apart from winter fasting, a significant amount of king penguin chick mortality is also due to predation, which is responsible for 48% of deaths in the studied colony (Descamps et al. 2005). Although giant petrels and skuas tend to attack smaller chicks at the periphery of the colony and creches, which usually correspond to late chicks (Le Bohec et al. 2005; Descamps et al. 2005), specific individual predation can also happen by random chance. From what we observed in our study, the genetic composition of the chick plays a greater role than random chance in determining mortality. This cements the hypothesis that, even though many external factors can be related to fitness reduction and extinction risk, such as habitat loss and predation pressure, genetic composition can be crucial to species persistence (Spielman et al. 2004).

Although we cannot currently use our results to confidently conclude that these patterns can be generalised, as the observed patterns need to be refined (work in progress) and also tested in other species, it gives a first step towards the comprehension of genetic mechanisms that can allow populations to avoid extinction in the future. We advocate for the use of other large natural populations, which have not passed through recent detrimental demographic events, such as bottlenecks, as a way of reaching more generalisable conclusions. By understanding general genetic composition patterns that can lead to mortality in a large population, proper reintroduction measures can also be taken in smaller endangered populations to avoid extinction. Moreover, even currently large populations are subjected to rapid declines, which can even lead to genetic meltdowns in the case of rapid inbreeding and realisation of previously masked load (Bertorelle et al. 2022). Thus, studying the genetic composition of large populations can also assist their proper conservation in the future.

The King penguin is currently considered as a Least Concern species by the IUCN Red List, as it is composed of a large panmictic population (BirdLife International 2023). Yet, a previous study using ecological niche modelling has estimated that climate change might push the Antarctic Polar Front southwards, farther from the swimming capacity of the species during the breeding season (Cristofari et al. 2018). Such APF displacement would lead to the disappearance of all colonies from nine out of the fifteen islands the species currently breeds in, including the studied colony in the Crozet Archipelago, before 2100. In this scenario, we believe that the late breeding strategy could serve as a way out of extinction, as it keeps individuals capable of breeding out of the peak of resources, while harbouring high genetic variability and low genetic load.

5. Conclusions and perspectives

Although reproducing in mismatch with environmental resources is mostly known for its negative impacts on individual fitness, many studies have suggested the potential that adaptive mismatched phenotypes can have for population persistence under climate change (Stenseth and Mysterud 2002; Charmantier et al. 2008; Visser et al. 2012). Our study represents some of the first evidence that mismatched genotypes can also contribute to fitness in a natural population, through the purging of highly deleterious alleles and of high levels of homozygosity from the population. A clearer interpretation of the fitness impact of the different mutation classes is still needed to give a better picture on how moderately and slightly deleterious mutations are segregating as mismatched genotypes. However, the widespread signal detected in the genotypes of survivors and non-survivors indicate that mortality, and consequently, fitness, has a strong genetic component in this species. Moreover, we expect that these results represent the first of more studies to examine the genetic background of large natural populations before it is too late to protect them from extinction.

Among our short-term perspectives, which include analyses that are already being done, we first intend to polarise SNP data in order to better interpret the moderately and slightly deleterious mutations (MODERATE and LOW impact SNPs from the SnpEff annotation). To do this, we are using GATK v.4.2.2.0 (McKenna et al. 2010; DePristo et al. 2011) to call SNPs of 12 individuals from three closely related penguin species to the King penguin, the Emperor, Adelie and Gentoo penguins. The output of these analyses can be used to polarise the alleles into an ancestral or derived state.

Also regarding an improved annotation of potential SNP impact on fitness, we have mapped the king penguin genome to a 363 bird species genome alignment using cactus (Armstrong et al. 2020). This is so we can use another approach for the prediction of deleterious mutations based on sequence conservation with GERP++ (Cooper et al. 2005) or PhyloP (Siepel et al. 2005). In order to test whether high heterozygosity for neutral loci can predict levels of deleterious mutations, we plan to use a linear model of regression to test dependency of genotypes within the same individual. Finally, as a longer term perspective, we intend to include more genomes from early and late individuals born in different years, in order to see whether our observed signals change based on oscillating environmental conditions, such as the position of the APF.

References

Adams, N. J., and N. T. Klages. 1987. “Seasonal Variation in the Diet of the King Penguin (Aptenodytes Patagonicus) at sub-Antarctic Marion Island.” *Journal of Zoology* 212 (2): 303–24.

Agrawal, Aneil F., and Michael C. Whitlock. 2012. “Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant.” *Annual Review of Ecology, Evolution, and Systematics* 43 (1): 115–35.

Andrews, S. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data” [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Armstrong, Joel, Glenn Hickey, Mark Diekhans, Ian T. Fiddes, Adam M. Novak, Alden Deran, Qi Fang, et al. 2020. “Progressive Cactus Is a Multiple-Genome Aligner for the Thousand-Genome Era.” *Nature* 587 (7833): 246–51.

Balakrishnan, Christopher N., and Scott V. Edwards. 2009. “Nucleotide Variation, Linkage Disequilibrium and Founder-Facilitated Speciation in Wild Populations of the Zebra Finch (Taeniopygia Guttata).” *Genetics* 181 (2): 645–60.

Barrat, A., 1976. “Quelques aspects de la biologie et de l'écologie du Manchot royal (Aptenodytes patagonicus) des îles Crozet.”

Barrick, Jeffrey E., and Richard E. Lenski. 2013. “Genome Dynamics during Experimental Evolution.” *Nature Reviews. Genetics* 14 (12): 827–39.

Beaumont, Hubertus J. E., Jenna Gallie, Christian Kost, Gayle C. Ferguson, and Paul B. Rainey. 2009. “Experimental Evolution of Bet Hedging.” *Nature* 462 (7269): 90–93.

Benazzo, Andrea, Emiliano Trucchi, James A. Cahill, Pierpaolo Maisano Delser, Stefano Mona, Matteo Fumagalli, Lynsey Bunnefeld, et al. 2017. “Survival and Divergence in a Small Group: The Extraordinary Genomic History of the Endangered Apennine Brown Bear Stragglers.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (45): E9589–97.

Berner, Daniel. 2019. “Allele Frequency Difference AFD¯An Intuitive Alternative to FST for Quantifying Genetic Population Differentiation.” *Genes* 10 (4). https://doi.org/10.3390/genes10040308.

Bertorelle, Giorgio, Francesca Raffini, Mirte Bosse, Chiara Bortoluzzi, Alessio Iannucci, Emiliano Trucchi, Hernán E. Morales, and Cock van Oosterhout. 2022. “Genetic Load: Genomic Estimates and Applications in Non-Model Animals.” *Nature Reviews. Genetics* 23 (8): 492–503.

BirdLife International. 2023. Species factsheet: Aptenodytes patagonicus. Downloaded from http://www.birdlife.org on 16/02/2023.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20.

Bozzuto, Claudio, Iris Biebach, Stefanie Muff, Anthony R. Ives, and Lukas F. Keller. 2019. “Inbreeding Reduces Long-Term Growth of Alpine Ibex Populations.” *Nature Ecology & Evolution* 3 (9): 1359–64.

Carvalho, G. R. 1993. “Evolutionary Aspects of Fish Distribution: Genetic Variability and Adaptation.” *Journal of Fish Biology* 43 (sa): 53–73.

Chamary, J. V., Joanna L. Parmley, and Laurence D. Hurst. 2006. “Hearing Silence: Non-Neutral Evolution at Synonymous Sites in Mammals.” *Nature Reviews. Genetics* 7 (2): 98–108.

Charlesworth, Deborah, and John H. Willis. 2009. “The Genetics of Inbreeding Depression.” *Nature Reviews. Genetics* 10 (11): 783–96.

Charmantier, Anne, Robin H. McCleery, Lionel R. Cole, Chris Perrins, Loeske E. B. Kruuk, and Ben C. Sheldon. 2008. “Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population.” *Science* 320 (5877): 800–803.

Charrassin, J-B, C. A. Bost, K. Pütz, J. Lage, T. Dahier, T. Zorn, and Y. Le Maho. 1998. “Foraging Strategies of Incubating and Brooding King Penguins Aptenodytes Patagonicus.” *Oecologia* 114 (2): 194–201.

Charrassin, J-B, Y. Le Maho, and C-A Bost. 2002. “Seasonal Changes in the Diving Parameters of King Penguins (Aptenodytes Patagonicus).” *Marine Biology* 141 (3): 581–89.

Cherel, Y., C. Verdon, and V. Ridoux. 1993. “Seasonal Importance of Oceanic Myctophids in King Penguin Diet at Crozet Islands.” *Polar Biology* 13 (5). https://doi.org/10.1007/bf00238362.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. “A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; Iso-2; Iso-3.” *Fly* 6 (2): 80–92.

Coop, Graham. 2020. “Population and Quantitative Genetics” Davis: University of California.

Cooper, Gregory M., Eric A. Stone, George Asimenos, NISC Comparative Sequencing Program, Eric D. Green, Serafim Batzoglou, and Arend Sidow. 2005. “Distribution and Intensity of Constraint in Mammalian Genomic Sequence.” *Genome Research* 15 (7): 901–13.

Cristofari, Robin, Xiaoming Liu, Francesco Bonadonna, Yves Cherel, Pierre Pistorius, Yvon Le Maho, Virginie Raybaud, Nils Christian Stenseth, Céline Le Bohec, and Emiliano Trucchi. 2018. “Climate-Driven Range Shifts of the King Penguin in a Fragmented Ecosystem.” *Nature Climate Change* 8 (3): 245–51.

Cushing, D. H. 1969. “The Regularity of the Spawning Season of Some Fishes.” *ICES Journal of Marine Science: Journal Du Conseil* 33 (1): 81–92.

Cushing, D.H., 1974. "The possible density-dependence of larval mortality and adult mortality in fishes." The Early Life History of Fish” JHS, 103-111.

Cushing, D. H. 1990. “Plankton Production and Year-Class Strength in Fish Populations: An Update of the Match/Mismatch Hypothesis.” In *Advances in Marine Biology*, edited by J. H. S. Blaxter and A. J. Southward, 26:249–93. Academic Press.

Cushing, J. M., and M. Saleem. 1982. “A Predator Prey Model with Age Structure.” *Journal of Mathematical Biology* 14 (2): 231–50.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab008.

DePristo, Mark A., Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data.” *Nature Genetics* 43 (5): 491–98.

Descamps, Sebastien, M. Gauthier-Clerc, J. P. Gendner, and Y. Le Maho. 2002. “The Annual Cycle of Unbanded King Penguins Aptenodytes Patagonicus on Possession Island (Crozet)” 2 (January): 87–98.

Descamps, Sébastien, Michel Gauthier-Clerc, Céline Le Bohec, Jean-Paul Gendner, and Yvon Le Maho. 2005. “Impact of Predation on King Penguin Aptenodytes Patagonicus in Crozet Archipelago.” *Polar Biology* 28 (4): 303–10.

Doiron, Madeleine, Gilles Gauthier, and Esther Lévesque. 2015. “Trophic Mismatch and Its Effects on

the Growth of Young in an Arctic Herbivore." *Global Change Biology* 21 (12): 4364–76.

Feng, Shaohong, Josefin Stiller, Yuan Deng, Joel Armstrong, Qi Fang, Andrew Hart Reeve, Duo Xie, et al. 2020. "Dense Sampling of Bird Diversity Increases Power of Comparative Genomics." *Nature* 587 (7833): 252–57.

Ferreira, A. S. A., L. C. Stige, A. B. Neuheimer, B. Bogstad, N. Yaragina, I. Prokopchuk, and J. M. Durant. 2020. "Match-Mismatch Dynamics in the Norwegian-Barents Sea System." *Marine Ecology Progress Series* LFC (April): 1–14.

Frankham, Richard. 2010. "Challenges and Opportunities of Genetic Approaches to Biological Conservation." *Biological Conservation* 143 (9): 1919–27.

Freeman, Natalie M., Nicole S. Lovenduski, and Peter R. Gent. 2016. "Temporal Variability in the Antarctic Polar Front (2002-2014)." *Journal of Geophysical Research, C: Oceans* 121 (10): 7263–76.

Garrison, Erik, Zev N. Kronenberg, Eric T. Dawson, Brent S. Pedersen, and Pjotr Prins. 2022. "A Spectrum of Free Software Tools for Processing the VCF Variant Call Format: Vcflib, Bio-Vcf, cyvcf2, Hts-Nim and Slivar." *PLoS Computational Biology* 18 (5): e1009123.

Gauthier-Clerc, M., Y. Le Maho, Y. Clerquin, C. A. Bost, and Y. Handrich. 2002. "Seabird Reproduction in an Unpredictable Environment: How King Penguins Provide Their Young Chicks with Food." *Marine Ecology Progress Series* 237: 291–300.

Ghalambor, Cameron K., Kim L. Hoke, Emily W. Ruell, Eva K. Fischer, David N. Reznick, and Kimberly A. Hughes. 2015. "Non-Adaptive Plasticity Potentiates Rapid Adaptive Evolution of Gene Expression in Nature." *Nature* 525 (7569): 372–75.

Grossen, Christine, Frédéric Guillaume, Lukas F. Keller, and Daniel Croll. 2020. "Purging of Highly Deleterious Mutations through Severe Bottlenecks in Alpine Ibex." *Nature Communications* 11 (1): 1001.

Hindell, Mark A. 1988. "The Diet of the King Penguin Aptenodytes Patagonicus at Macquarie Island." *The Ibis* 130 (2): 193–203.

Jordan, Rebecca, Ary A. Hoffmann, Shannon K. Dillon, and Suzanne M. Prober. 2017. "Evidence of Genomic Adaptation to Climate in Eucalyptus Microcarpa: Implications for Adaptive Potential to Projected Climate Change." *Molecular Ecology* 26 (21): 6002–20.

Kardos, Marty. 2021. "Conservation Genetics." *Current Biology: CB* 31 (19): R1185–90.

Kardos, Marty, Ellie E. Armstrong, Sarah W. Fitzpatrick, Samantha Hauser, Philip W. Hedrick, Joshua M. Miller, David A. Tallmon, and W. Chris Funk. 2021. "The Crucial Role of Genome-Wide Genetic Variation in Conservation." *Proceedings of the National Academy of Sciences of the United States of America* 118 (48). https://doi.org/10.1073/pnas.2104642118.

Kardos, Marty, and Gordon Luikart. 2021. "The Genetic Architecture of Fitness Drives Population Viability during Rapid Environmental Change." *The American Naturalist* 197 (5): 511–25.

Kharouba, Heather M., Johan Ehrlén, Andrew Gelman, Kjell Bolmgren, Jenica M. Allen, Steve E. Travers, and Elizabeth M. Wolkovich. 2018. "Global Shifts in the Phenological Synchrony of Species Interactions over Recent Decades." *Proceedings of the National Academy of Sciences of the United States of America* 115 (20): 5211–16.

Kharouba, Heather M., and Elizabeth M. Wolkovich. 2020. "Disconnects between Ecological Theory and Data in Phenological Mismatch Research." *Nature Climate Change* 10 (5): 406–15.

Kido, Takashi, Weronika Sikora-Wohlfeld, Minae Kawashima, Shinichi Kikuchi, Naoyuki Kamatani, Anil Patwardhan, Richard Chen, et al. 2018. "Are Minor Alleles More Likely to Be Risk Alleles?" *BMC Medical Genomics* 11 (1): 3.

Kyriazis, Christopher C., Robert K. Wayne, and Kirk E. Lohmueller. 2021. "Strongly Deleterious Mutations Are a Primary Determinant of Extinction Risk due to Inbreeding Depression."

Evolution Letters 5 (1): 33–47.

Lande, Russell, and Susan Shannon. 1996. “THE ROLE OF GENETIC VARIATION IN ADAPTATION AND POPULATION PERSISTENCE IN A CHANGING ENVIRONMENT.” *Evolution; International Journal of Organic Evolution* 50 (1): 434–37.

Le Bohec, Céline, Michel Gauthier-Clerc, and Yvon Le Maho. 2005. “The Adaptive Significance of Crèches in the King Penguin.” *Animal Behaviour* 70 (3): 527–38.

Leonard, Aoife M., and Lesley T. Lancaster. 2020. “Maladaptive Plasticity Facilitates Evolution of Thermal Tolerance during an Experimental Range Shift.” *BMC Evolutionary Biology* 20 (1): 47.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Makova, Kateryna. 2019. “Young Sex Chromosomes Discovered in White-Eye Birds.” *Peer Community in Evolutionary Biology*, June, 100073.

Marvelde, Luc, Simone L. Webber, Harro A. J. Meijer, and Marcel E. Visser. 2011. “Mismatched Reproduction Is Energetically Costly for Chick Feeding Female Great Tits.” *Functional Ecology* 25 (6): 1302–8.

Mathur, Samarth, and J. Andrew DeWoody. 2021. “Genetic Load Has Potential in Large Populations but Is Realized in Small Inbred Populations.” *Evolutionary Applications* 14 (6): 1540–57.

Mathur, Samarth, John Tomeček, Luis Tarango-Arámbula, Robert Perez, and Andrew DeWoody. 2021. “An Evolutionary Perspective on Contemporary Genetic Load in Threatened Species to Inform Future Conservation Efforts.” *Authorea Preprints*, June. https://doi.org/10.22541/au.162495929.94655412/v1.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303.

Nakazawa, Takefumi, and Hideyuki Doi. 2012. “A Perspective on Match/mismatch of Phenology in Community Contexts.” *Oikos* 121 (4): 489–95.

Neigel, Joseph E. 2002. “Is Fst Obsolete?” *Conservation Genetics* 3 (2): 167–73.

Nursyifa, Casia, Anna Brüniche-Olsen, Genis Garcia-Erill, Rasmus Heller, and Anders Albrechtsen. 2022. “Joint Identification of Sex and Sex-Linked Scaffolds in Non-Model Organisms Using Low Depth Sequencing Data.” *Molecular Ecology Resources* 22 (2): 458–67.

Ohno, S., C. Stenius, L. C. Christian, W. Becak, and M. L. Becak. 1964. “CHROMOSOMAL UNIFORMITY IN THE AVIAN SUBCLASS CARINATAE.” *Chromosoma* 15 (August): 280–88.

Ohno, Susumu. 1967. *Sex Chromosomes and Sex-Linked Genes*. Springer Science & Business Media.

O’Leary, Shannon J., Jonathan B. Puritz, Stuart C. Willis, Christopher M. Hollenbeck, and David S. Portnoy. 2018. “These Aren’t the Loci You'e Looking for: Principles of Effective SNP Filtering for Molecular Ecologists.” *Molecular Ecology*, July. https://doi.org/10.1111/mec.14792.

Olsson, Olof. 1996. “Seasonal Effects of Timing and Reproduction in the King Penguin: A Unique Breeding Cycle.” *Journal of Avian Biology* 27 (1): 7–14.

Orr, H. Allen. 2009. “Fitness and Its Role in Evolutionary Genetics.” *Nature Reviews. Genetics* 10 (8): 531–39.

Pan, Hailin, Theresa L. Cole, Xupeng Bi, Miaoquan Fang, Chengran Zhou, Zhengtao Yang, Daniel T. Ksepka, et al. 2019. “High-Coverage Genomes to Elucidate the Evolution of Penguins.”

GigaScience 8 (9). https://doi.org/10.1093/gigascience/giz117.

Park, Ju-Hyun, Mitchell H. Gail, Clarice R. Weinberg, Raymond J. Carroll, Charles C. Chung, Zhaoming Wang, Stephen J. Chanock, Joseph F. Fraumeni Jr, and Nilanjan Chatterjee. 2011. "Distribution of Allele Frequencies and Effect Sizes and Their Interrelationships for Common Genetic Susceptibility Variants." *Proceedings of the National Academy of Sciences of the United States of America* 108 (44): 18026–31.

Petrullo, Lauren, Stan Boutin, Jeffrey E. Lane, Andrew G. McAdam, and Ben Dantzer. 2023. "Phenotype-Environment Mismatch Errors Enhance Lifetime Fitness in Wild Red Squirrels." *Science* 379 (6629): 269–72.

Plard, Floriane, Jean-Michel Gaillard, Tim Coulson, A. J. Mark Hewison, Daniel Delorme, Claude Warnant, and Christophe Bonenfant. 2014. "Mismatch between Birth Date and Vegetation Phenology Slows the Demography of Roe Deer." *PLoS Biology* 12 (4): e1001828.

Post, Eric, and Mads C. Forchhammer. 2008. "Climate Change Reduces Reproductive Success of an Arctic Herbivore through Trophic Mismatch." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1501): 2369–75.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reed, Thomas E., Stephanie Jenouvrier, and Marcel E. Visser. 2013. "Phenological Mismatch Strongly Affects Individual Fitness but Not Population Demography in a Woodland Passerine." *The Journal of Animal Ecology* 82 (1): 131–44.

Régnier, T., F. M. Gibb, and P. J. Wright. 2019. "Understanding Temperature Effects on Recruitment in the Context of Trophic Mismatch." *Scientific Reports* 9 (1): 15179.

Robinson, Jacqueline A., Jannikke Räikkönen, Leah M. Vucetich, John A. Vucetich, Rolf O. Peterson, Kirk E. Lohmueller, and Robert K. Wayne. 2019. "Genomic Signatures of Extensive Inbreeding in Isle Royale Wolves, a Population on the Threshold of Extinction." *Science Advances* 5 (5): eaau0757.

Rowe, Graham, and Trevor J. C. Beebee. 2003. "Population on the Verge of a Mutational Meltdown? Fitness Costs of Genetic Load for an Amphibian in the Wild." *Evolution; International Journal of Organic Evolution* 57 (1): 177–81.

Saccheri, Ilik, Mikko Kuussaari, Maaria Kankare, Pia Vikman, Wilhelm Fortelius, and Ilkka Hanski. 1998. "Inbreeding and Extinction in a Butterfly Metapopulation." *Nature* 392 (6675): 491–94.

Sæther, Bernt-Erik, and Steinar Engen. 2015. "The Concept of Fitness in Fluctuating Environments." *Trends in Ecology & Evolution* 30 (5): 273–81.

Seutin, Gilles, Bradley N. White, and Peter T. Boag. 1991. "Preservation of Avian Blood and Tissue Samples for DNA Analyses." *Canadian Journal of Zoology*, February. https://doi.org/10.1139/z91-013.

Siepel, Adam, Gill Bejerano, Jakob S. Pedersen, Angie S. Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, et al. 2005. "Evolutionarily Conserved Elements in Vertebrate, Insect, Worm, and Yeast Genomes." *Genome Research* 15 (8): 1034–50.

Spielman, Derek, Barry W. Brook, and Richard Frankham. 2004. "Most Species Are Not Driven to Extinction before Genetic Factors Impact Them." *Proceedings of the National Academy of Sciences of the United States of America* 101 (42): 15261–64.

Stenseth, Nils Chr, and Atle Mysterud. 2002. "Climate, Changing Phenology, and Other Life History Traits: Nonlinearity and Match-Mismatch to the Environment." *Proceedings of the National Academy of Sciences of the United States of America* 99 (21): 13379–81.

Stier, Antoine, Vincent A. Viblanc, Sylvie Massemin-Challet, Yves Handrich, Sandrine Zahn, Emilio R. Rojas, Claire Saraux, et al. 2014. "Starting with a Handicap: Phenotypic Differences between

Early- and Late-born King Penguin Chicks and Their Survival Correlates." *Functional Ecology* 28 (3): 601–11.

Stonehouse, Bernard. 1960. "The King Penguin... of South Georgia. I. Breeding Behaviour and Development." HM Stationery Office.

Sunyaev, S., V. Ramensky, I. Koch, W. Lathe 3rd, A. S. Kondrashov, and P. Bork. 2001. "Prediction of Deleterious Human Alleles." *Human Molecular Genetics* 10 (6): 591–97.

Teixeira, João C., and Christian D. Huber. 2021. "The Inflated Significance of Neutral Genetic Diversity in Conservation Genetics." *Proceedings of the National Academy of Sciences of the United States of America* 118 (10). https://doi.org/10.1073/pnas.2015096118.

Thomas, D. W., J. Blondel, P. Perret, M. M. Lambrechts, and J. R. Speakman. 2001. "Energetic and Fitness Costs of Mismatching Resource Supply and Demand in Seasonally Breeding Birds." *Science* 291 (5513): 2598–2600.

Torda, Gergely, Jennifer M. Donelson, Manuel Aranda, Daniel J. Barshis, Line Bay, Michael L. Berumen, David G. Bourne, et al. 2017. "Rapid Adaptive Responses to Climate Change in Corals." *Nature Climate Change* 7 (9): 627–36.

Valk, Tom van der, Edson Sandoval-Castellanos, Damien Caillaud, Urbain Ngobobo, Escobar Binyinyi, Radar Nishuli, Tara Stoinski, et al. 2018. "Significant Loss of Mitochondrial Diversity within the Last Century due to Extinction of Peripheral Populations in Eastern Gorillas." *Scientific Reports* 8 (1): 6551.

Visser, Marcel E., Luc te Marvelde, and Marjolein E. Lof. 2012. "Adaptive Phenological Mismatches of Birds and Their Food in a Warming World." *Journal of Ornithology / DO-G* 153 (1): 75–84.

Weimerskirch, H., J. C. Stahl, and P. Jouventin. 1992. "The Breeding Biology and Population Dynamics of King Penguins Aptenodytes Patagonica on the Crozet Islands." *The Ibis* 134 (2): 107–17.

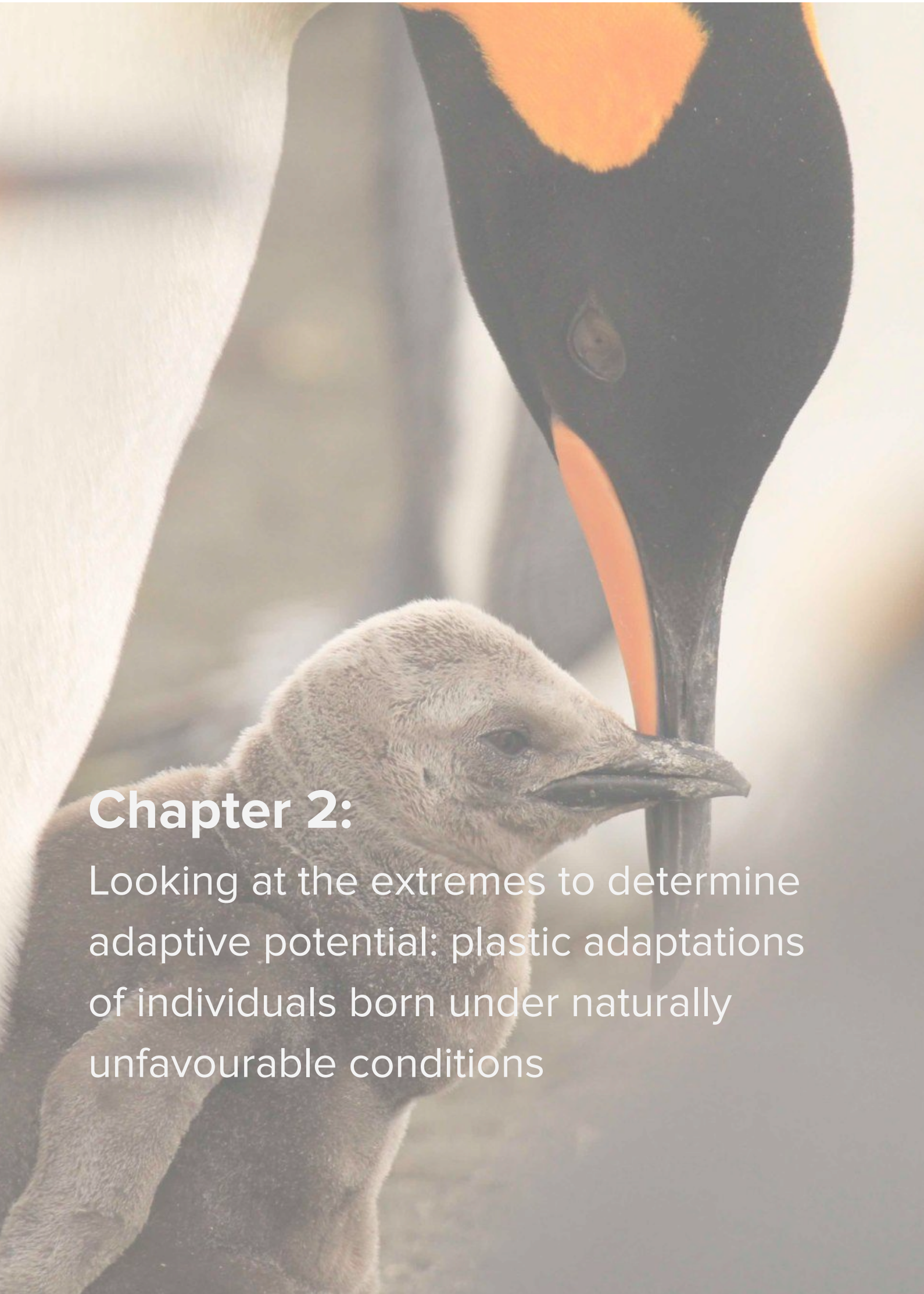
Weir, B.S. and Cockerham, C.C., 1984. "Estimating F-statistics for the analysis of population structure." Evolution, 1358-1370.

Willi, Yvonne, Josh Van Buskirk, and Ary A. Hoffmann. 2006. "Limits to the Adaptive Potential of Small Populations." *Annual Review of Ecology, Evolution, and Systematics* 37 (1): 433–58.

Wright, Belinda, Katherine A. Farquharson, Elspeth A. McLennan, Katherine Belov, Carolyn J. Hogg, and Catherine E. Grueber. 2019. "From Reference Genomes to Population Genomics: Comparing Three Reference-Aligned Reduced-Representation Sequencing Pipelines in Two Wildlife Species." *BMC Genomics* 20 (1): 453.

Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics* 28 (24): 3326–28.

Zhou, Qi, Jilin Zhang, Doris Bachtrog, Na An, Quanfei Huang, Erich D. Jarvis, M. Thomas P. Gilbert, and Guojie Zhang. 2014. "Complex Evolutionary Trajectories of Sex Chromosomes across Bird Taxa." *Science* 346 (6215): 1246338.



Chapter 2:

Looking at the extremes to determine adaptive potential: plastic adaptations of individuals born under naturally unfavourable conditions

Chapter 2

Looking at the extremes to determine adaptive potential: plastic adaptations of individuals born under naturally unfavourable conditions *Ready to be submitted*

Authors:
Flávia A. N. Fernandes[1,2]
Samuele Greco[3]
Lorena Ancona[1]
Gaël Bardon[2,4]
Pierre Carette[2]
Robin Cristofari[5]
Marco Gerdol[3]
Elodie Paciello[2]
Federica Pirri[1]
Josephine Paris[1]
Céline Le Bohec[2,4] & Emiliano Trucchi[1]

Affiliations:
1 - Marche Polytechnic University, Italy
2 - University of Strasbourg, France
3 - University of Trieste, Italy
4 - Scientific Center of Monaco, Monaco
5 - University of Turku, Finland

Keywords
Predictive adaptive response; transcriptome; adaptive potential; phenotypic plasticity; seabirds.

Authors' contribution to this paper

This paper's idea was jointly conceived and designed by F. Fernandes, C. Le Bohec and E. Trucchi. Sample collection was done by F. Fernandes, G. Bardon, R. Cristofari, P. Carette, and E. Paciello, with the aid of other field work assistants, during the king penguin breeding seasons of 2020 and 2021, coordinated by C. Le Bohec in the French Southern and Antarctic Lands (TAAF). RNA extractions and other wet laboratory procedures, such as RNA quantification and quality assessment, were done by F. Fernandes. Post-sequencing pre-processing steps from trimming to table of read counts' preparation were performed by F. Fernandes. S. Greco and M. Gerdol contributed to the RUVseq normalisation and quality control for outliers. F. Pirri, S. Greco, and M. Gerdol provided the transcriptome assembly for the transcriptome-mapping pipeline. All analyses were performed by F. Fernandes. Results' discussion and interpretation were done by F. Fernandes, E. Trucchi and C. Le Bohec with the contributions of L. Ancona, G. Bardon, R. Cristofari, S. Greco, M. Gerdol, and J. Paris. Writing was done by F. Fernandes, with the contribution of J. Paris, C. Le Bohec, and E. Trucchi.

1. Introduction

The exposure to stressful environmental conditions can have direct impacts on individual fitness (i.e., survival and reproduction) (Marak et al. 2003; Crino and Breuner 2015; Watson et al. 2017). These fitness impacts may be even more accentuated when stressful conditions occur during early-life, when individuals have not yet fully completed their development and are still largely inexperienced (Lindström 1999). Many studies have demonstrated that early-life restrictions, such as limited food availability, can impair an individual's survival and performance during adulthood (Taborsky 2006; Mugabo et al. 2010; Millon et al. 2011; Hamel et al. 2009; Hayward et al. 2013). Such detrimental impacts of early developmental constraints on adult survival and performance are known as the silver-spoon effect (Grafen 1988).

Poor early-life conditions that lead to silver-spoon effects can affect individuals at different stages of their lifetime. As an example, a study on a long-lived shorebird showed that individuals born in better quality areas (i.e., nesting areas closer to feeding ones) had higher juvenile survival, and adult survival and reproductive success than individuals born in low quality areas (Van de Pol et al. 2006). Another example involves lower survival and reproductive rates of human populations that faced famine during childhood (Hayward et al. 2013).

Aside from the apparent negative effects of being born under non-optimal conditions on individual fitness, challenging conditions may generate adaptations that can be useful later in the individual's life, or even increase the fitness at the population level (Ghalambor et al. 2007). In this context, an alternative hypothesis to the silver-spoon effects is the predictive adaptive response (PAR), which posits that stress experienced in early-life can provide phenotypic advantages to individuals when re-exposed to stress during adulthood (Bateson et al. 2014). However, this hypothesis may only hold true if the constraints faced in early-life are similar to adult-life conditions (Gluckman et al. 2005). Additionally, stressful conditions will only be able to lead to unique adaptations if the pressures fall within the reaction norm limits (i.e., the sensitivity of individuals with similar genotypes to a specific environmental variability) of the population (Schlichting and Pigliucci 1998).

Considering the increasing unpredictability of environmental conditions due to the rapid global changes, populations able to produce a higher range of phenotypes under different environmental conditions (i.e., phenotypic plasticity) are expected to have a quicker response and, therefore, higher chances of persisting under new pressures (Aubin-Horth and Renn 2009). Furthermore, for species which are highly dependent on the seasonality of the resources (e.g., food peaks), phenological plasticity can play a key role in adaptation to asynchrony (De Lisle et al. 2022). However, other studies show possible limitations in the potential of adaptive plasticity under unpredictable conditions. More specifically, when environmental cues are not reliable, plasticity may be limited by the levels of genetic diversity in the population (Oostra et al. 2018).

To better understand the role of phenotypic plasticity in adaptation to environmental asynchrony, we investigated the potential plastic phenological responses of a species exposed to naturally contrasting and seasonal environmental conditions, the King penguin

(*Aptenodytes patagonicus*). This sub-Antarctic species is characterised by two peaks of egg-laying producing two main groups of chicks that are born and grow under different levels of stress, generated by the seasonality of food availability which deteriorates throughout the season (Descamps et al. 2002; Stier et al. 2014). One of the main early-life challenges for this species is overcoming the first long winter fast before fledging (i.e., from June to September), as chicks are fed less often by the adults, which have to forage farther from breeding grounds during that period (Descamps et al. 2002). Thus, chicks that survive until fledging must build up sufficient energy reserves until the beginning of winter fasting, being that this pressure is more critical for late-born individuals (Stier et al. 2014).

Although late chicks that survive through winter are capable of growing faster than early chicks in a shorter period of time, the stress of fast growth is expected to come with the cost of producing high levels of reactive oxygen species (ROS) (Stier et al. 2014). In this scenario, whether the unfavourable early-life conditions faced by late-born individuals generates a burden (silver spoon hypotheses) or prepares individuals for pressures of adult-life (PAR), such as fasting periods during moulting and breeding (Cherel et al. 1988a,b; Groscolas & Cherel 1992), is still an open question.

Here, we explore this matter by using the blood transcriptome of individuals born under naturally distinct stressful conditions: early-hatchlings born in January (matched conditions) versus late-hatchlings born in February). If some late-hatchlings survive until fledging even with less time to grow until the first winter fasting, we expected the phenological groups to show plastic differences (i.e., gene expression) already a few days after hatching. Because the pressures posed by fasting are also present in adult-life, our hypothesis is that the plastic response of late chicks could act as a PAR, and not as silver spoon effects, preparing individuals for the pressures faced later in life. We expect to find signals of genes and pathways regulating faster growth in the late chicks in comparison to early chicks, as well as genes and pathways related the potential metabolic costs of fast growth (e.g., higher oxidative stress (OS); (Geiger et al. 2012).

Furthermore, until the moment of writing, our study is the first to sequence and characterise the blood transcriptome of the King penguin, meaning this work represents the first foray into the analysis of potentially differentially expressed genes between two different developmental stages. Finally, we demonstrate that the blood transcriptome can provide a valuable resource for inferring the physiological status of individuals.

2. Material and Methods

2.1. Sampling design

Sampling was carried out in the king penguin colony of *La Baie du Marin* (here referred to as BDM), Possession Island, Crozet Archipelago (46°24′27″S 51°45′27″E), during the breeding seasons of 2020 and 2021. More specifically, king penguin chicks were sampled from the sub-colony Antavia, which represents a breeding area in the southern part of BDM that is naturally enclosed with four passages used by the penguins to reach for the sea (Gendner et al. 2005). Restricting the sampling to this sub-colony allows for a more efficient monitoring of chicks during their first year on land.

To investigate the effects of different intensities of early-life stressors on gene expression, and whether individuals are able to catch-up to the conditions required to fledge, early- and late-born chicks were sampled at two time-points: *ca.* 3 weeks after hatching and *ca.* 2 weeks before fledging. To ensure a homogenous developmental stage between the early and late sampling sessions, king penguin chicks captured at hatching weighed between 500 to 1000 grams and were sampled during the brooding stage (i.e., when chicks are still kept in the brooding patch of the parents). This mass threshold was also used as a guiding procedure to avoid mixing early and late chicks when sampling late chicks.

A total of 69 early and 72 late chicks were captured at hatching in 2020. From these, 39 early and 10 late chicks were recaptured at fledging. In 2021, 40 early and 60 late chicks were captured at hatching, from which 28 early and 22 late were recaptured at fledging. Because late chicks have a higher mortality rate (Olsson 1996; Weimerskirch et al. 1992; Stier et al. 2014), we sampled more chicks from this group to increase the probability of having at least ~10 chicks that survived to fledging. As described in section **2.2. Wet laboratory and RNA sequencing** of this chapter, the selection of the survivors group selected for RNA sequencing was based on blood sample and RNA extraction quality.

Captures at hatching occurred within a time window of 7 to 10 days, starting at the end of January for early chicks and at the end of February for late chicks (see **Table 1** for specific dates). Chicks were sampled for blood at the brachial vein using a 25-gauge needle and a microcapillary tube. We directly transferred 3 to 9 drops of blood into a 1.5 mL microcentrifuge tube prepared with aliquots of the PAXgene® Blood RNA Solution following the manufacturer's recommended ratio of 2.76 blood:solution. Tubes were inverted 10 times and freezing procedures followed the manufacturer's instructions, with final freezing at -80°C until processing in the laboratory. Prior to release, chicks were also marked with a small external plastic pin (Fishtag, Floytag) stamped with a unique number for individual recognition during and after winter (**Figure 1a**).

Chicks were then followed through the winter and surviving individuals were recaptured at fledging (i.e., at the end of their first moult, at ~ 11 months old) (**Figure 1c**).

Recaptures at fledging occurred from early November to February of the following year (see **Table 1** for specific dates), when a second blood sampling was taken following the same procedure described above. After the blood sampling, we removed the fishtag from the chicks and marked the individuals with subcutaneous Radio Frequency Identification (RFID) tags, to monitor them remotely throughout their life (detailed in Fernandes and Bardon et al. *in prep* - Chapter 3). All manipulations were approved by the French Ethics Committee (APAFIS#4897-2015110911016428) and the French Polar Environmental Committee (TAAF permit #2019-115 & 2019-129) and conducted in accordance with its guidelines.

Table 1. Capture dates of hatching and fledging chicks in 2020 and 2021 breeding seasons. Note that although the 2020 chicks were recaptured at fledging, these samples are not included in the analyses, due to an absence of RNA integrity for sequencing.

	Early hatching	**Late hatching**	**Early fledging**	**Late fledging**
2020	25/01/2020 - 31/01/2020	25/02/2020 - 04/03/2020	30/10/2020 - 22/01/2021	08/11/2020 - 14/01/2021
2021	22/01/2021 - 25/01/2021	26/02/2021 - 03/03/2021	21/10/2021 - 22/01/2022	01/11/2021 - 20/12/2021

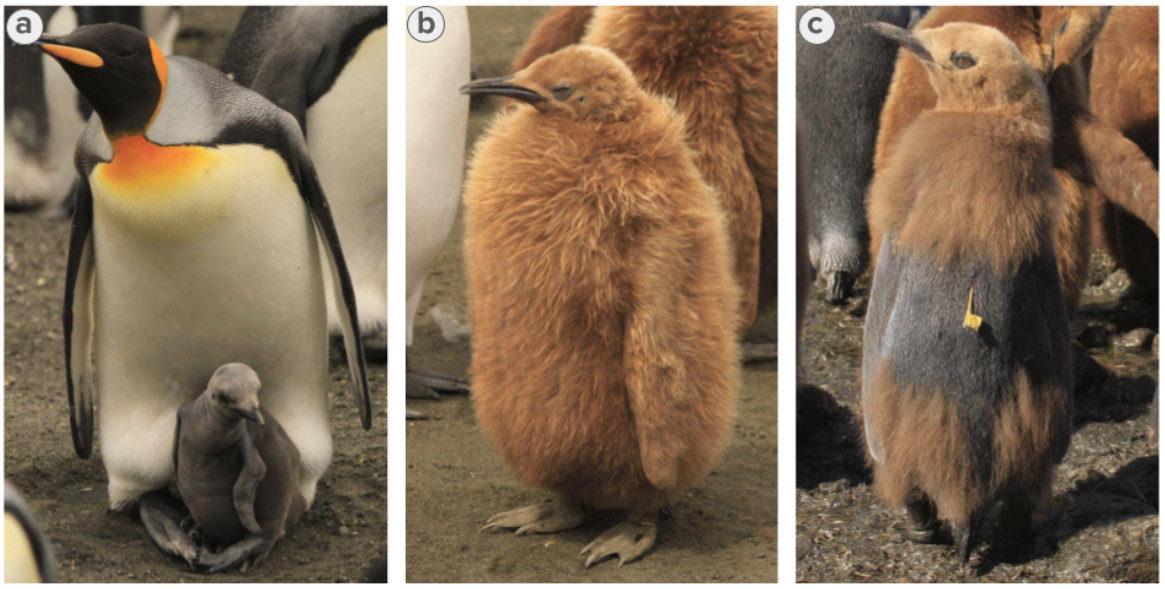


Figure 1. Developmental stages of king penguin chicks. a) King penguin chick with one parent at the end of the brooding phase. The chick in the photo represents the approximate maximum individual size sampled at hatching; b) Chick at the beginning of the first moult; c) Fledging chick a few days before the end of the first moult, with a visible yellow fish-tag.

2.2. Wet laboratory and RNA sequencing

RNA extraction from the whole blood was performed using the PAXgene® Blood RNA kit following the manufacturer's protocol, with one modification after the overnight thawing: we transferred the samples from the 1.5 mL microcentrifuge tube to a 15 mL falcon.

This modification allowed us to perform the first centrifugation steps using a swing-rotor and the addition of 4 mL of RNAse-free water to wash out the PAXgene® Blood RNA solution after the first centrifugation step, as indicated in the manufacturer's protocol. RNA quantification was performed with a spectrophotometer and a fluorometer, and RNA integrity was checked with an agarose gel electrophoresis before samples were sent for sequencing. Although chicks were sampled at both hatching and fledging in the two years, 2020 fledging samples showed intense marks of RNA degradation in the electrophoresis gel, and were excluded from sequencing and posterior analyses.

After RNA extraction and a first quality check, 20 samples from 2020 (10 early, 10 late hatching chicks) and 24 samples from 2021 (6 early and 6 late chicks, at hatching and fledging) (**Table 2**) were shipped to the sequencing company, *BMR genomics s.r.l.* (Padova), where a final 2100 Bioanalyzer quality check was made. Sequencing libraries were prepared using the QuantSeq 3' mRNA-Seq Library Prep Kit, which captures the 3' end extremity of the mRNA transcripts' polyadenylated tail. QuantSeq is a RNA sequencing strategy that allows direct quantification of gene expression, as each read would represent a transcript count (Moll et al. 2014). Sequencing was performed on an Illumina NovaSeq 6000 platform, aiming for over 5 million single-end reads of 75 base pairs (bp) per sample.

Table 2. Details of samples for RNA sequencing. ID: the name of the individuals; Category: the phenological category in which the individual was born; Year: the year in which the individual was born; Hatching RNA: if the chick's RNA was sequenced at hatching (around 3 weeks old); Fledging RNA: if the chick's RNA was sequenced at fledging (around 11 months old); Sex: the sex of the individuals was determined using SATC (Nursyifa et al. 2022), see section **2.3.3. Identification of sex-linked scaffolds and masking** in Fernandes et al. *in prep.* (Chapter 1) for details; and Genome Sequenced: if the genome of the individuals was sequenced (same individuals as in Fernandes et al. *in prep.* Chapter 1).

	ID	Phenological group	Year	Hatching RNA	Fledging RNA	Sex	Genome Sequenced
1	KP2020_E584	Early	2020	Yes	No	Male	Yes
2	KP2020_E594	Early	2020	Yes	No	Male	Yes
3	KP2020_E143	Early	2020	Yes	No	Male	Yes
4	KP2020_E147	Early	2020	Yes	No	Female	Yes
5	KP2020_E595	Early	2020	Yes	No	Female	Yes
6	KP2020_E599	Early	2020	Yes	No	Male	Yes
7	KP2020_E580	Early	2020	Yes	No	Male	Yes
8	KP2020_E153	Early	2020	Yes	No	Male	Yes
9	KP2020_E169	Early	2020	Yes	No	Male	Yes
10	KP2020_E190	Early	2020	Yes	No	Female	Yes
11	KP2020_L151	Late	2020	Yes	No	Male	Yes
12	KP2020_L008	Late	2020	Yes	No	Female	Yes
13	KP2020_L019	Late	2020	Yes	No	Female	Yes
14	KP2020_L033	Late	2020	Yes	No	Male	Yes
15	KP2020_L034	Late	2020	Yes	No	Female	Yes

	ID	Phenological group	Year	Hatching RNA	Fledging RNA	Sex	Genome Sequenced
16	KP2020_L044	Late	2020	Yes	No	Male	Yes
17	KP2020_L053	Late	2020	Yes	No	Female	Yes
18	KP2020_L058	Late	2020	Yes	No	Female	Yes
19	KP2020_L059	Late	2020	Yes	No	Female	Yes
20	KP2020_L062	Late	2020	Yes	No	Male	Yes
21	KP2021_E551	Early	2021	Yes	Yes	NA	No
22	KP2021_E553	Early	2021	Yes	Yes	NA	No
23	KP2021_E569	Early	2021	Yes	Yes	NA	No
24	KP2021_E570	Early	2021	Yes	Yes	NA	No
25	KP2021_E601	Early	2021	Yes	Yes	NA	No
26	KP2021_E608	Early	2021	Yes	Yes	NA	No
27	KP2021_L160	Late	2021	Yes	Yes	NA	No
28	KP2021_L161	Late	2021	Yes	Yes	NA	No
29	KP2021_L164	Late	2021	Yes	Yes	NA	No
30	KP2021_L168	Late	2021	Yes	Yes	NA	No
31	KP2021_L173	Late	2021	Yes	Yes	NA	No
32	KP2021_L208	Late	2021	Yes	Yes	NA	No

2.3. Quality control and initial processing of RNAseq data

After receiving the raw 3'end RNA reads, we performed a first quality control step with FastQC v0.11.9 (Andrews 2010). As we detected a drop in sequencing quality at the last 3 bases in all individuals, we removed all tails in 3 base pairs, as well as both 3' and 5' extremities if mean base quality was below Q15 (i.e., a probability higher than 96.84% that the base was correctly called) with the software *fastp* v0.20.1 (Chen et al. 2018). We also used the *fastp* options "-g" and "-x", in order to eliminate possible polyG and polyX tails, which could be sequencing artefacts from Illumina NextSeq/NovaSeq or tails from mRNA-Seq reads, respectively.

Once the reads were trimmed, we followed two different workflows to generate gene count tables for the posterior differential gene expression (DGE) analyses: (I) we aligned reads to the reference genome of the species, followed by gene quantification, and (II) mapped reads to a target transcriptome of the species combined with coding sequences from the genome, followed by transcript quantification. We will hereafter refer to the data generated from the first approach as *genome-aligned* data and to the second approach and *transcriptome-mapped* data.

In our case, the transcriptome mapping strategy included an extra step after DGE analyses, in which the different transcript isoforms were assigned to a gene. Because our

study did not focus on the expression of different isoforms, we assigned all different transcripts to genes so we could make a direct comparison with the genome-aligned data. The specific pipeline performed in each approach is detailed below.

2.3.1. Aligning and mapping RNA-seq reads to the reference genome and transcriptome

Both genome-alignment and transcriptome-mapping strategies are widely used in DGE analysis in the literature (Stark et al. 2019). Each approach has its own advantages and limitations, and depends on the resource availability for the studied species. The alignment of RNA-seq data to a high-quality reference genome has the advantage of controlling for gene boundaries more finely than when mapping reads to a target transcriptome. This is because genome annotations contain information on the position of exons and introns, while reference transcriptomes may contain incomplete genes or different portions of it due to mRNA splicing, which can generate aberrant transcripts during mapping (Stark et al. 2019).

However, aligning RNA-Seq data to a reference genome first requires the availability of a high-quality genome of the species, which was not the case for most non-model species in the past decade (Gilad et al. 2009). In this scenario, mapping to a transcriptome assembly seems to be an appropriate cheaper alternative, considering that at the time of this work, transcriptome sequencing was less costly and time consuming than generating a high-quality reference genome for the species (Alvarez et al. 2015). Mapping RNA-seq reads to a reference transcriptome may also allow the discovery of new unannotated transcripts, especially when the reference genome of the species is incomplete (Stark et al. 2019). It can also be useful when the focus of the study is on the differential transcript isoform expression, instead of gene expression.

The main downside of using a transcriptome mapping is in the case where the reference transcriptome available for the species was produced from samples of different tissues compared to the analysed one. This is the case of the available transcriptome assembly of the king penguin, which was built using samples that did not include the blood (i.e., it contains transcripts from brain, kidney, liver, pectoral muscle, and skin) (Pirri 2022). In this case, especially if there is a strongly tissue-specific expression, there should be a significant amount of unmapped reads that will be lost (Bentz et al. 2019).

As we have both an available reference genome and an assembled transcriptome for the King penguin, we decided to use the two workflows in order to maximise the detection of genes. For the generation of the genome-aligned data, we aligned our single-end RNA-seq reads to the available reference genome of the King penguin (GCA_010087175.1 BGI_Apat.V1) (Pan et al. 2019) using a splice-aware aligner, STAR v2.7.9a (Dobin et al. 2013). STAR generated a BAM (binary alignment and map) file per sample, already sorted by coordinate thanks to the *--outSAMtype BAM SortedByCoordinate* option. BAM files were then indexed with *samtools index* in samtools (Danecek et al. 2021).

The next step of the genome-alignment pipeline consisted of assigning read counts to genomic features (i.e., genes), based on the alignment files and the genome annotation of the species. The quantity of reads that map to each genomic feature was computed with HTSeq-count (Anders et al. 2014). HTSeq-count considers a feature of RNA-Seq data as a gene (i.e., a union of the gene's exons), and generates a table of counts per feature/gene for each sample as output. This raw table of gene counts was then used as input for the DGE analyses.

For the transcriptome-mapping approach, we mapped the RNA-seq reads to the assembled transcriptome of the species, which was generated from samples of five different tissues of king penguin chicks (i.e., brain, kidney, liver, pectoral muscle, and skin) (Pirri 2022), using the alignment-free method of Salmon v1.4.0 (Patro et al. 2017). The salmon workflow has two main steps: the generation of a reference transcriptome index, followed by the direct quantification of the reads that map to each transcript.

Because our samples originate from the chick's whole blood and not from the same tissues used in the assembled transcriptome, we complemented Salmon's transcriptome index with coding sequences (cds) from the genome of the species, using salmon's *-g* argument. After running the quantification step with *salmon quant*, we merged each sample's transcript quantification file in a unique table using the *salmon quantmerge* command. Note that, different from the genome-alignment pipeline, this pipeline generates a table of counts per transcripts, and not per gene. We assigned transcripts to genes using the transcriptome annotation only after the DE analyses, which was performed at a transcript level.

2.3.2. RNA-Seq data normalisation

Once RNA-seq reads were aligned to the king penguin reference genome and mapped to the assembled transcriptome, we had data on the raw gene or transcript counts per individual, respectively. A normalisation of these raw count tables prior to DE analyses is needed, because RNA-Seq raw counts are subjected to different types of biases (Bullard et al. 2010; Dillies et al. 2013). Such biases can be due to differences in sequencing depth between regions (i.e., high coverage regions will automatically have more reads than low coverage regions), RNA composition (i.e., a highly expressed gene in a sample will skew the distribution of counts for the other genes in that sample if not normalised), and gene length (i.e., longer genes will have more reads mapped to it than shorter genes), although this last bias is not a concern when using 3'-seq RNA data (Moll et al. 2014).

We therefore used the RUVSeq package in Bioconductor (Risso et al. 2014) to normalise both gene and transcript counts prior to the DGE analysis, to ensure accurate comparisons of gene expression between samples. RUVSeq Removes Unwanted Variation (RUV), which is useful to control for batch and library preparation effects, as well as other forms of noise that can be present in RNA-Seq data. RUVSeq uses the rationale that generalised linear models (GLM, used in the DE analyses) will take the regression of

RNA-Seq read counts on both the covariates of interest (i.e., the conditions of being early or late) and the unwanted variation (e.g., batch, library effect) as input. The normalisation acts by adjusting the DE analysis model to the estimated unwanted variation factors.

We used the *RUVg* approach inside RUVSeq, which estimates the unwanted variation factors based on negative control genes that are assumed not to be influenced by the covariates of interest. As we did not have a set of candidate genes to use as a negative control (e.g., housekeeping genes or spike-in controls in the case of the zebrafish, *Danio rerio*; Risso et al. 2014), we used empirically estimated genes that were not differentially expressed in our dataset when contrasting the early- and late-born conditions, as suggested in the RUVSeq manual. In practice, the read counts of these genes are expected to be constant among samples, so any deviations from the nominal fold-changes would indicate nuisance in the data.

This procedure requests a first analysis of DE in DESeq2 without RUVSeq normalisation, in order to select a list of empirical genes with the least significant DE (i.e., genes with the highest *P-value*s) and generates a *W* matrix of factors of unwanted variation. *W* consists of a *n x k* matrix, where *n* is the number of samples and *k* the number of unwanted factors we want to remove, that is set by the user. We set *k*=2, after seeing that it generated slightly more DEGs than *k*=1 in the 2020 data, and higher values of *k* did not change the results, as reported by (Gagnon-Bartsch and Speed 2012) in a discussion on *k* choice. Once we had obtained the normalised count data we performed a real analysis of DE, implementing RUVSeq's unwanted variation matrix in the DE design.

2.4. Blood transcriptome characterization

To our knowledge, this is the first study to sequence the blood transcriptome of the King penguin. For this reason, we first made a general characterisation of the expressed genes in this tissue. Our initial idea was to characterise the genes and pathways at each developmental stage separately, but after noticing that the great majority of genes are shared between hatching and fledging chicks (see **Results** section), we proceeded with a unique pathway characterisation.

After read count normalisation, we filtered both the genome-aligned genes and transcriptome-mapped transcripts separately to keep genes/transcripts with at least 5 read counts in at least 3 individuals in all the 31 hatching chicks (i.e., early and late, from 2020 and 2021, excluding a 2021 outlier, as it will be shown in the Results section) and the 12 fledging chicks of 2021. Once we had a table of genes and transcripts that passed the count filter at each developmental stage, for the transcriptome-mapped data, we assigned the transcripts to genes based on the species transcriptome annotation (Pirri 2022), and removed redundant genes. We finally merged the transcriptome-mapped genes with the genome-aligned genes table.

To perform a characterisation of the pathways present in the blood transcriptome of the king penguin chicks, we performed a Gene Ontology (GO) enrichment of all genes

detected at both stages. Gene Ontology (GO) for biological processes was conducted in PANTHER v14 (Protein Analysis Through Evolutionary Relationships) (Mi et al. 2019). PANTHER uses Fisher's exact tests to determine the significance of the number of genes assigned to that biological process category of GO (i.e., GO term) (*P-value* < 0.05), and then corrects the multiple comparisons for false discovery rate (FDR < 0.05) (Mi et al. 2019). The Gene Ontology database used for analysis was the doi:10.5281/zenodo.6799722 Released in 2022-07-01, and we used the chicken (*Gallus gallus*) gene set as a background.

However, significant GO terms can still be too numerous and specific for a clear understanding of the biological functions taking place in such a complex circulating tissue. For this reason, we also assigned GO terms to higher level biological summaries, called GOslims, using the GSEABase v1.60.0 package in Bioconductor (Morgan et al. 2022). GSEABase gives broader biological functions to a set of GO terms, organising them in a parent-offspring set.

2.5. Differential gene expression analyses and gene ontology enrichment

The differential gene expression (DGE) analyses between early and late-born chicks were performed with the R package DESeq2 (Love et al. 2014) using both the genome-aligned and transcriptome-mapped data. We will use the DGE acronym for simplification, even though the DE analyses with the transcriptome-mapped data were done at the transcript level instead of the gene level.

Following the standard DESeq2 pipeline for DGE, we started by filtering the table of counts to keep genes if they had at least 5 read counts in at least 3 individuals, as described in section **2.4. Blood transcriptome characterization**. We did not set a minimum threshold for the differences in gene expression (i.e., log2 fold change, which represents log-ratio of a gene or a transcript's expression values in two different conditions), as we were expecting small variation between early and late groups given that our samples came from the same type of tissue, the same population, and at the same developmental stage (Uebbing et al. 2016).

We ran the DGE analysis implementing the normalisation of RUVSeq, described in section **2.3.2. RNA-Seq data normalisation** and the log fold change (LFC) shrinking procedure in DESeq2. The LFC shrinking is a method used to reduce the strong variance and noise of weakly expressed genes (i.e., with low counts, but high LFC). In brief, the shrinkage procedure contracts LFC estimates towards zero when the counts for that gene are low throughout the samples, when dispersion is high (i.e., some individuals have higher counts than others out of the contrast sampling design), or when there are few degrees of freedom (Zhu et al. 2019).

The threshold to consider a gene or transcript as differentially expressed was an adjusted *P-value* for false discovery rate (FDR) smaller than 0.05. Only after the DE analyses, DE transcripts were assigned to genes based on the transcriptome annotation of

the species (Pirri 2022). We checked for overlapping genes using both genome-alignment and transcriptome-mapping approaches, by merging both DEGs lists and highlighting redundant genes from each analysis.

For a general functional annotation of the most up and downregulated genes, we used the human genome in the database GeneCards (Stelzer et al. 2016; Safran et al. 2021) with the intent of gaining a potentially predictive function of the genes. Finally, as described in the section above, **2.4. Blood transcriptome characterization**, DEGs were assigned to GO terms with PANTHER v14 (Mi et al. 2019) and GO slims, when possible, using GSEABase v1.60.0 (Morgan et al. 2022). Both the individual gene functional annotation and GO information are predictions that are limited and thus must be interpreted with prudence, as the real effect on the studied organisms cannot be known only from these analyses. In any case, due to the exploratory nature of this study, such predictions serve as hints of the main processes that can be further investigated in detail if considered relevant.

For sample data visualisation, we produced Principal Component Analysis (PCA) graphs using a variance stabilising transformation (VST) of the count data, which removes experiment-wide trend variation (Tibshirani 1988; Huber et al. 2003; Anders and Huber 2010). We transformed the count data using the DESeq2 function *vst* with the *blind=FALSE* argument, so the already estimated dispersions based on the design formula of early vs. late groups can be used in the transformation. We also plotted the Multidimensional Scaling (MDS) plot for the most differentially expressed DEGs in each condition (with lower *padj* values). The MDS plot, which shows the two dimensions that explain the greatest variance between samples (Cox and Cox 2008), was used to visualise the up and down-regulation of genes in early and late groups.

After a first round of analysis, we detected an unexpected sex bias in our data (see section **3.3. Sex and interannual variation**), which led to the removal of all 2021 samples from the early versus late-born chicks comparison. Given this, the DGE experimental design of early and late groups was exclusively run on 2020 hatching chicks. However, because 2021 samples contain the two developmental chick stages, we still used the 2021 dataset to perform an analysis of DGE between hatching and fledging chicks. In order to detect pathways that could be more active at each condition, we also enriched the GOs and GOslims of the DEGs between developmental stages, as explained above for the early and late DEGs, and for the blood characterisation.

3. Results

3.1. Description of the King penguin's blood transcriptome

Results regarding control steps, including detection of overrepresented sequences, outlier samples, and general noise in the data, can be found in the Appendix of Chapter 2. The results below are those obtained after removing these sources of noise.

3.1.1. Hatching and fledging chick's transcriptome description

The hatching chick's blood transcriptome was characterised by a total of 9488 genes, from which 1236 aligned exclusively to the genome annotation, 4850 mapped exclusively the transcriptome annotation, and 3402 genes were detected both with the alignment to the reference genome and transcriptome (**Table 3**). The fledging chick's blood transcriptome was characterised by 9591 genes, 1268 of which originated from the genome alignment, 4873 from the transcriptome mapping, and 3450 were detected by both approaches (**Table 3**). These first results showed that the transcriptome mapping contributed to more than 50% of the detected genes in comparison to the genome alignment.

Table 3. Number of expressed genes in the blood of king penguin chicks at hatching and fledging. Genes considered in this table had at least 5 read counts in at least 3 individuals per developmental stage. We separate the total number of genes detected both by the genome alignment and transcriptome mapping approaches, and exclusively by each approach.

	Hatching	Fledging
Total genes detected	**9488**	**9591**
Genome + Transcriptome annotations	3402	3450
Genome annotation exclusively	1236	1268
Transcriptome annotation exclusively	4850	4873

From all genes expressed in chicks at hatching (9488 genes) and fledging (9591 genes), around 95% were shared between the two developmental stages (9045 genes). Only 443 genes were exclusively expressed at hatching (4.7% of genes) and 546 at fledging (5.7% of genes). Because these results suggested that blood gene expression is majorly characterised by a constant set of genes during the first year of life in this species (**Figure 2**), we proceeded with the characterisation of the main biological processes present in the king penguin blood by considering the species as a whole (i.e., not differentiating by developmental stage). We then performed a DGE analysis to identify if the up and

down-regulated genes were related to specific biological processes (section **3.1.3. Expression differences between developmental stages**).

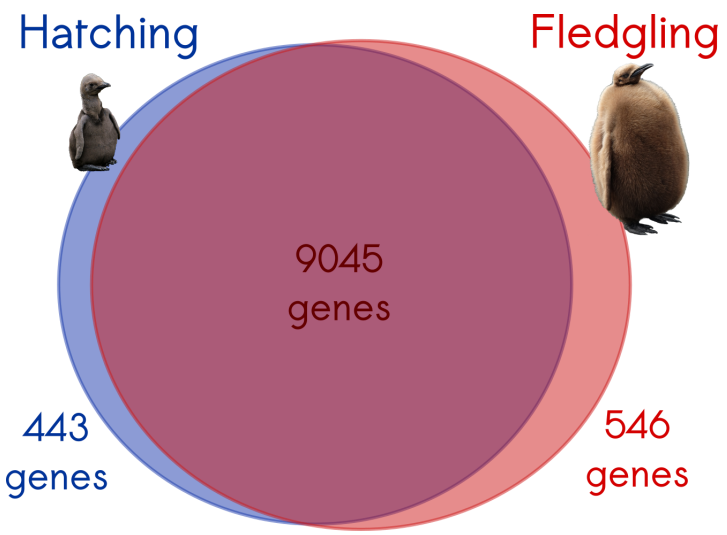


Figure 2. Venn diagram with the number of genes expressed in King penguin whole blood at hatching (blue) and fledging (red), with an overlap of 9045 genes in both developmental stages (maroon). Illustrative figure, not in scale.

A total of 346 GO terms were enriched for the 10,034 genes present in the blood of the king penguin chicks (hatching and fledging together). After submitting hundreds of significant GO terms to GSEABase, our GO set was reduced to 34 GOslims (**Table 4**). The main enriched GO slims identified included immune system processes (26 GO terms), signalling (22 GO terms), and anatomical structure development (17 GO terms). Other processes, such as protein modification (11 GO terms), protein catabolic processes (8 GO terms), cell differentiation, and nervous system processes (both 9 GO terms) were among the most represented.

Table 4. GO slims of the king penguin chick's blood transcriptome at hatching and fledging. GOslim: GO slim annotation ID of a subset of GO terms enriched for the genes present in the chick's blood; Count: number of GO terms contained in the GO slim; Terms: name of the GOslim term.

GOslim	Count	Term
GO:0002376	26	Immune system process
GO:0023052	22	Signalling
GO:0048856	17	Anatomical structure development
GO:0036211	11	Protein modification process
GO:0030154	9	Cell differentiation
GO:0050877	9	Nervous system process
GO:0030163	8	Protein catabolic process
GO:0016192	7	Vesicle-mediated transport
GO:0016071	6	mRNA metabolic process
GO:0006629	4	Lipid metabolic process

GO:0007155	4	Cell adhesion
GO:0042254	4	Ribosome biogenesis
GO:0006091	3	Generation of precursor metabolites and energy
GO:0006281	3	DNA repair
GO:0006325	3	Chromatin organisation
GO:0006399	3	tRNA metabolic process
GO:0061024	3	Membrane organisation
GO:0000278	2	Mitotic cell cycle
GO:0006351	2	Transcription, DNA-templated
GO:0006886	2	Intracellular protein transport
GO:0012501	2	Programmed cell death
GO:0065003	2	Protein-containing complex assembly
GO:0098542	2	Defence response to other organism
GO:0002181	1	Cytoplasmic translation
GO:0006260	1	DNA replication
GO:0006310	1	DNA recombination
GO:0006457	1	Protein folding
GO:0006914	1	Autophagy
GO:0007005	1	Mitochondrion organisation
GO:0030198	1	Extracellular matrix organisation
GO:0034330	1	Cell junction organisation
GO:0051604	1	Protein maturation
GO:0140053	1	Mitochondrial gene expression
GO:1901135	1	Carbohydrate derivative metabolic process

Among the GOslim for immune system processes, many GO terms were associated with the lymphocyte and B cell mediated immunity, as well as cytotoxic T cell differentiation, which are related to the adaptive immune system (**Appendix Chapter 2, Table S1**). The second most representative GOslim was signalling, which was due to the high quantity of GO terms related to synaptic processes and the immune system. Many signalling GO terms are immune system GOs (6 out of 22 signalling GO terms), and therefore, the overrepresentation of the signalling may be partly due to its implication in immune response processes (**Appendix Chapter 2, Table S1**). The third most represented GOslim, anatomical structure development, is likely related to the early stage of development of chicks, with GO terms predominantly related to organ and anatomic structure morphogenesis and nervous system development. Moreover, the anatomical structure development GOslim was highly similar to the cell differentiation GOslim, with the former sharing 7 out of 9 GO terms with the latter.

3.1.2. Expression differences between developmental stages

Although most of the gene transcripts found in the blood are shared by chicks at hatching and fledging, the regulation of expression of such genes can still differ between the two developmental stages. To explore this matter, we performed a differential gene expression (DGE) analysis between the hatching and fledging stages. A total of 4392 differentially expressed genes (DEGs) were detected between the two stages, from which 2343 were upregulated at hatching, and 2049 at fledging.

The GO enrichment of the 4392 DEGs generated 292 significant GO terms (**Appendix Chapter 2, Table S2**), which can be summarised into 31 GOslims (**Table 5**). The top five mostly represented GOslims were mitotic cell cycle (27 GO terms), immune system process (23 GO terms), signalling (19 GO terms), chromosome segregation (18 GO terms), and nucleobase-containing small molecule metabolic process (14 GO terms).

Table 5. GO slims from DEGs between hatching and fledging stages. GOslim: GO slim annotation ID of a subset of GO terms enriched for the genes present in the chick's blood; Count: number of GO terms contained in the GO slim; Term: name of the GOslim term.

GOSlim	Count	Term
GO:0000278	27	Mitotic cell cycle
GO:0002376	23	Immune system process
GO:0023052	19	Signalling
GO:0007059	18	Chromosome segregation
GO:0055086	14	Nucleobase-containing small molecule metabolic process
GO:0140014	11	Mitotic nuclear division
GO:0006091	10	Generation of precursor metabolites and energy
GO:0048856	8	Anatomical structure development
GO:0050877	7	Nervous system process
GO:0022414	6	Reproductive process
GO:0065003	6	Protein-containing complex assembly
GO:0006260	5	DNA replication
GO:1901135	5	Carbohydrate derivative metabolic process
GO:0006281	4	DNA repair
GO:0007010	4	Cytoskeleton organisation
GO:0000910	3	Cytokinesis
GO:0002181	3	Cytoplasmatic translation
GO:0006310	3	DNA recombination
GO:0006325	3	Chromatin organisation
GO:0007155	3	Cell adhesion
GO:0042254	3	Ribosome biogenesis
GO:0140013	3	Meiotic nuclear division
GO:0012501	2	Programmed cell death

GO:0016192	2	Vesicle-mediated transport
GO:0030154	2	Cell differentiation
GO:0098542	2	Defence response to other organism
GO:0007005	1	Mitochondrion organisation
GO:0030163	1	Protein catabolic process
GO:0036211	1	Protein modification process
GO:0048870	1	Cell motility
GO:0055085	1	Transmembrane transport

The most upregulated genes in the hatching group were primarily related to mitotic and cell division processes, such as the CENPF, CLEC19A, CKAP2, SMC2, SPAG5, and NUSAP1 genes. The CENPF gene (centromere protein F, log2FC=6.53; *padj*=4.17E-08) codes for a protein that is part of the centromere-kinetochore complex, and plays a potential role in the regulation of skeletal mitogenesis and cell differentiation during embryogenesis. The CLEC19A gene (C-type lectin domain containing 19A, log2FC=6.36; *padj*=4.58E-19) codes for a protein that composes the extracellular matrix, which is one of the major components of most tissues and organs (studied in zebrafish, *Danio rerio,* (Nauroy et al. 2018)). The CKAP2 (cytoskeleton associated protein 2, log2FC=6.29; *padj*=3.94E-20) is a gene that codes for the cytoskeleton-associated protein that stabilises microtubules. SMC2 (structural maintenance of chromosome 2, log2FC=6.06; *padj*=1.82E-21) is related to chromosome condensation, and SPAG5 (sperm associated antigen 5, log2FC=5.82; *padj*=2.21E-08) and NUSAP1 (nucleolar and spindle associated protein 1, log2FC=5.73; *padj*=5.21E-16) are related to chromosome segregation.

The majority of genes that were identified as significantly upregulated in the fledging group (and therefore downregulated in the hatching group) were involved in protein ubiquitination and deubiquitination. Among them, the ATE1 (arginyltransferase 1, log2FC=4.29; *padj*=1.18E-08) is a gene related to ubiquitin-dependent protein degradation, and OTUD1 (OTU deubiquitinase 1, log2FC=3.93; *padj*=5.11E-05) codes for an deubiquitinating enzyme. Overall, our results showed that, even if a majoritarily equal set of genes is expressed in the blood of king penguin chicks at both hatching and fledging, the up and down regulation of such genes, and their related pathways, vary according to developmental stage.

3.2. Sex and interannual variation OBJ

After whole genome sequencing, we were able to sex the 2020 chicks through the identification of sex scaffolds in the genome with SATC (Nursyifa et al. 2022) (method described in Fernandes et al. *in prep.* (Chapter 1), section **2.3.3. Identification of sex-linked scaffolds and masking** and **Table 3. RNA sequencing samples** in the **2.2. Wet laboratory and RNA sequencing** Materials and Methods section. The sex of 2021 chicks remained

unknown and was not initially taken into account until the detection of sex-bias patterns that will be exposed below.

3.2.1. Sex-bias in gene expression

We detected a clear separation between males and females from 2020 in the principal component analysis (PCA) of normalised gene counts (**Figure 3**). This separation was clearer in the transcriptome-mapped dataset (**Figure 3b**). Male-female segregation in PC1 (21% of variance) was even stronger than early-late separation, even if early and late groups were also slightly detached within the sexes.

When plotting the PCA using the genome-aligned data, the separation between sexes was less distinct (**Figure 3a**): we found an overlap between males and females, but not as strong as in the transcriptome-mapped data. We also observed a slight separation between early and late individuals (**Figure 3a**), and the genome-aligned data appeared to be less sex-biassed than the transcriptome-mapped data.

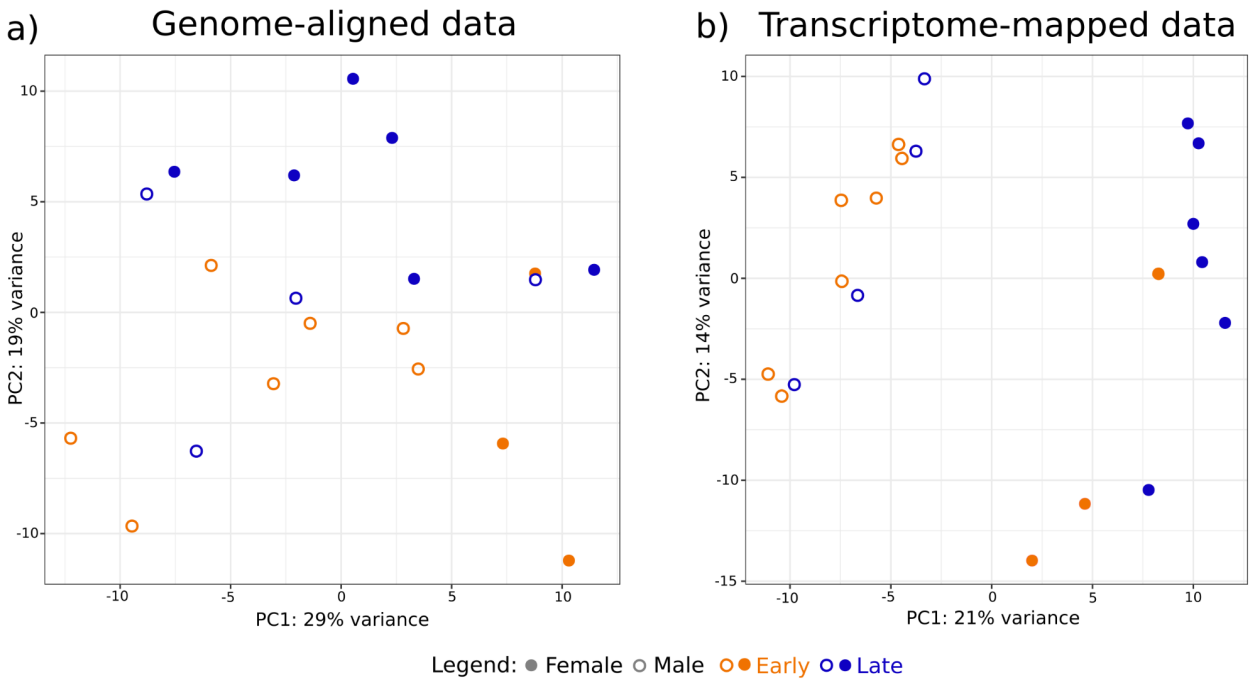


Figure 3. Principal component analysis (PCA) of normalised genes and transcript read counts of 2020 chicks at hatching according to sex. a) Transformed gene counts of the 3'-end RNA-Seq data aligned to the genome; b) Transformed transcript counts of the 3'-end RNA-Seq data mapped to the transcriptome.

The sex separation observed in the PCA with the transcriptome-mapped data may simply reflect the higher number of genes mapped to the transcriptome compared to the genome (see section **3.2.1. The contribution of genome-alignment and transcriptome-mapping pipelines**). However, if this was the only reason, a more pronounced sex separation could have been seen in the genome-aligned data if more

reads had been mapped to the genome annotation, simply because of the higher probability of including sex-biassed DEGs in the PCA.

We also detected 339 DEGs shared between the sexes in both genome-aligned and transcriptome-mapped datasets. We found 13 GO terms that were enriched for the DEGs shared between sexes (**Appendix Chapter 2, Table S3**), with the alcohol biosynthetic process as the most representative one (fold enrichment = 7.39). Although we did not detect DEGs that could be directly linked to this pathway in either sexes, this could reflect a difference in preen oil composition between males and females.

From the 339 significant DEGs identified between the sexes, 75 genes were upregulated in the females and 264 were upregulated in males. Most of the detected DEGs originate from the transcriptome quantification in both sexes: 59 from the transcriptome, 8 from the genome, plus 8 detected by both methods in females; while 173 from the transcriptome, 44 from the genome, plus 47 detected by both methods in males. When looking at the predicted function of the most differentially expressed genes in males, most upregulated genes were related to insulin uptake, such as NLN (neurolysin, log2FC=2.77, *padj*=0.0001) and PIK3R1 (phosphoinositide-3-kinase regulatory subunit 1, isoform log2FC=5.95, *padj*=6.35E-15), and microtubule formation, such as TBB6 (tubulin beta 6 class V, log2FC=2.27, *padj*=0.016) and TPX2 (TPX2 microtubule nucleation factor, isoform log2FC=5.75, *padj*=0.0001). The most upregulated gene from the genome-aligned dataset was the CCNB1 (cyclin B1, log2FC=3.02, *padj*=0.003), which codes for a regulatory protein likely involved in mitosis.

On the other hand, the main upregulated genes in females were related to platelet regulation, such as BLOC1S5 (biogenesis of lysosomal organelles complex 1 subunit 5, isoform log2FC=22.68, *padj*=4.63E-12) and DMTN (dematin actin binding protein, isoform log2FC=12.36, *padj*=0.001). Another female DEG, GPBP1 (GC-rich promoter binding protein 1, isoform log2FC=10.71, *padj*=1.48E-59), is related to the development of atherosclerosis.

3.2.2. Prediction of chick sex using gene expression data

As such a clear separation was detected between 2020 males and females, we decided to perform a PCA of 2020 and 2021 hatching chick gene expression data together, to see whether a separation would also be seen in 2021 chicks, from which we had no sex information. Surprisingly, a separation between males and females was apparent from PC1 (16% of variance), even if there was no clustering of individuals from different years (**Figure 4**). From the 11 chicks born in 2021, five individuals clustered on the left side of the PCA with 2020 females, while the other six clustered at the side of 2020 males. Due to this evident separation, we considered the five 2021 individuals on the left of the PCA as potential females and the six individuals to the right as potential males. If this is indeed the case, all 2021 early individuals are females and all late individuals are males. In order to define the sex of 2021 individuals with certainty, a more accurate methodology will be required, for

example via the development of a molecular sexing probe developed from genomic regions specific to the heterogametic sex (females in birds, ZW).

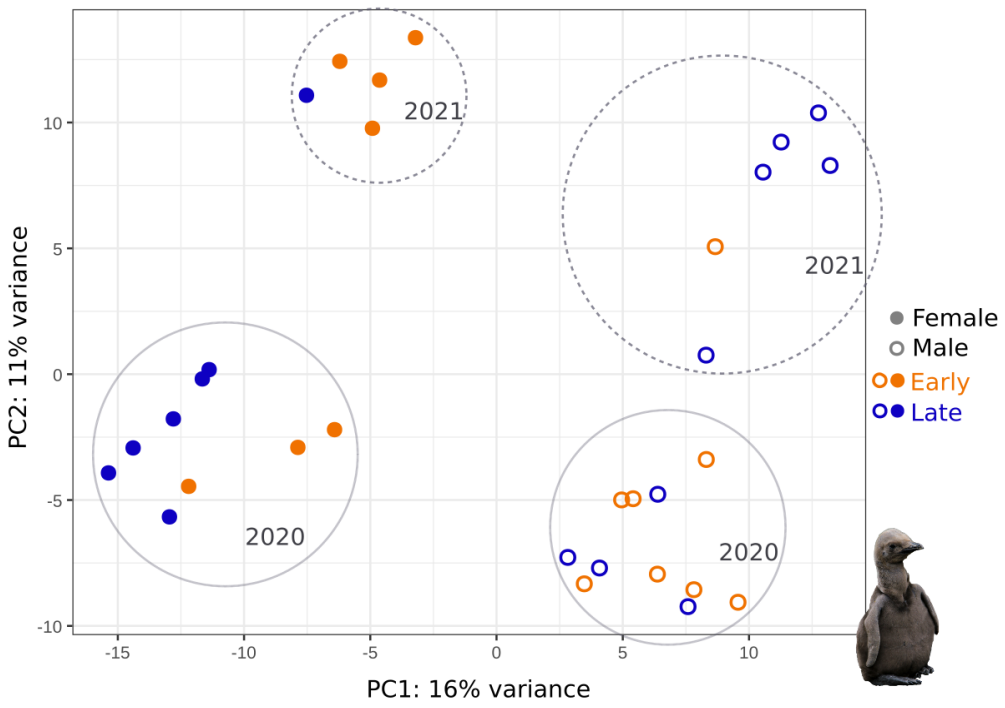


Figure 4. Principal component analysis (PCA) of the normalised transcript counts of hatching chicks from 2020 and 2021.

3.3. Early and late differential gene expression

The main objective in this study was to investigate if individuals born outside the peak of food resources show evidence of adaptation at the gene expression level that might allow them to survive under more stressful conditions. More specifically, we expected to detect gene expression patterns in chicks born later in the season that would shed light on the winter survival of these individuals (e.g., adaptation to faster growth). To this end, we performed DGE analyses on early- and late-hatching chicks from 2020.

Due to an apparent sex-bias in 2021 data observed above, 2021 chicks were not included in this part of the analysis. Even if the 2020 group of early- and late-born chicks was composed of an unbalanced mixture of sexes (i.e., early group N=3 females and N=7 males; late group N= 6 females, N=4 males) we considered that any source of sex-biasses in the 2020 data was removed with RUVSeq normalisation. Moreover, essentially all early-late DEGs were detected in the genome-aligned dataset (results below), which was less affected by the sex bias, as shown in section **3.3.1. Sex-bias in gene expression**. Thus, we considered that most of the sex-biassed regions were removed from this analysis.

We detected a total of 133 DEGs between early and late-born chicks, with both the genome-aligned and transcriptome-mapped datasets: 120 DEGs originated from the genome alignment, while a minority of 13 DEGs (i.e., 13 DE transcripts from 13 different

genes) originating from the transcriptome mapping. From the 133 DEGs, 67 were upregulated in the late group, and 66 were upregulated in the early group (FDR < 0.05). All DEGs log2 fold changes (log2FC) were lower than 1.4 (i.e., all differences in gene expression increased by fewer than 3 times in one group compared to the other; **Figure 5**).

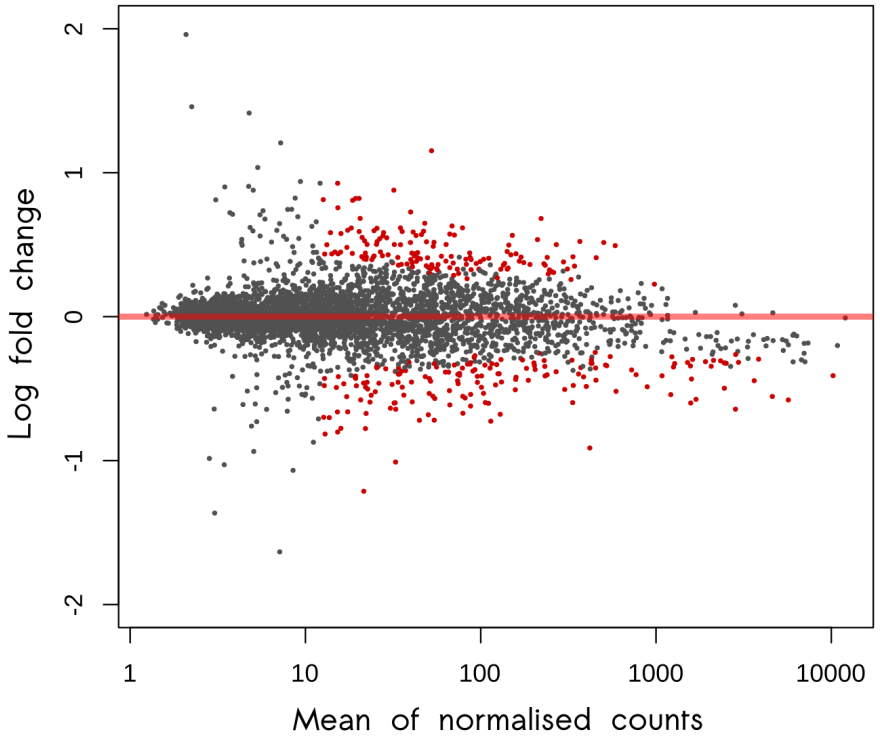


Figure 5. MA plot of log2 fold changes of the normalised shrunk counts of genome-aligned data in DESeq2. Red dots represent DEGs between early (negative values) and late (positive values) conditions.

Among the most upregulated genes in the late group (**Appendix Chapter 2, Table S4**), some were predicted as tumour suppressors, such as the APC (APC regulator of WNT signalling pathway, log2FC=1.28, *padj*=0.0025), MXD4 (MAX dimerization protein 4, log2FC=0.95, *padj*=0.0017) (**Figure 6a and 6b**), and MINDY3 (MINDY lysine 48 deubiquitinase 3, log2FC=0.91, *padj*=0.003). More specifically, APC and another upregulated gene in late-born chicks, CSNK1A1 (casein kinase 1 alpha 1, log2FC=0.93, *padj*=0.0012), negatively regulate the canonical Wnt pathway. Apart from apparent upregulation of tumour suppression, late-born chicks also overexpressed genes involved in protein ubiquitination, such as BFAR (bifunctional apoptosis regulator, log2FC=0.97, *padj*=0.018).

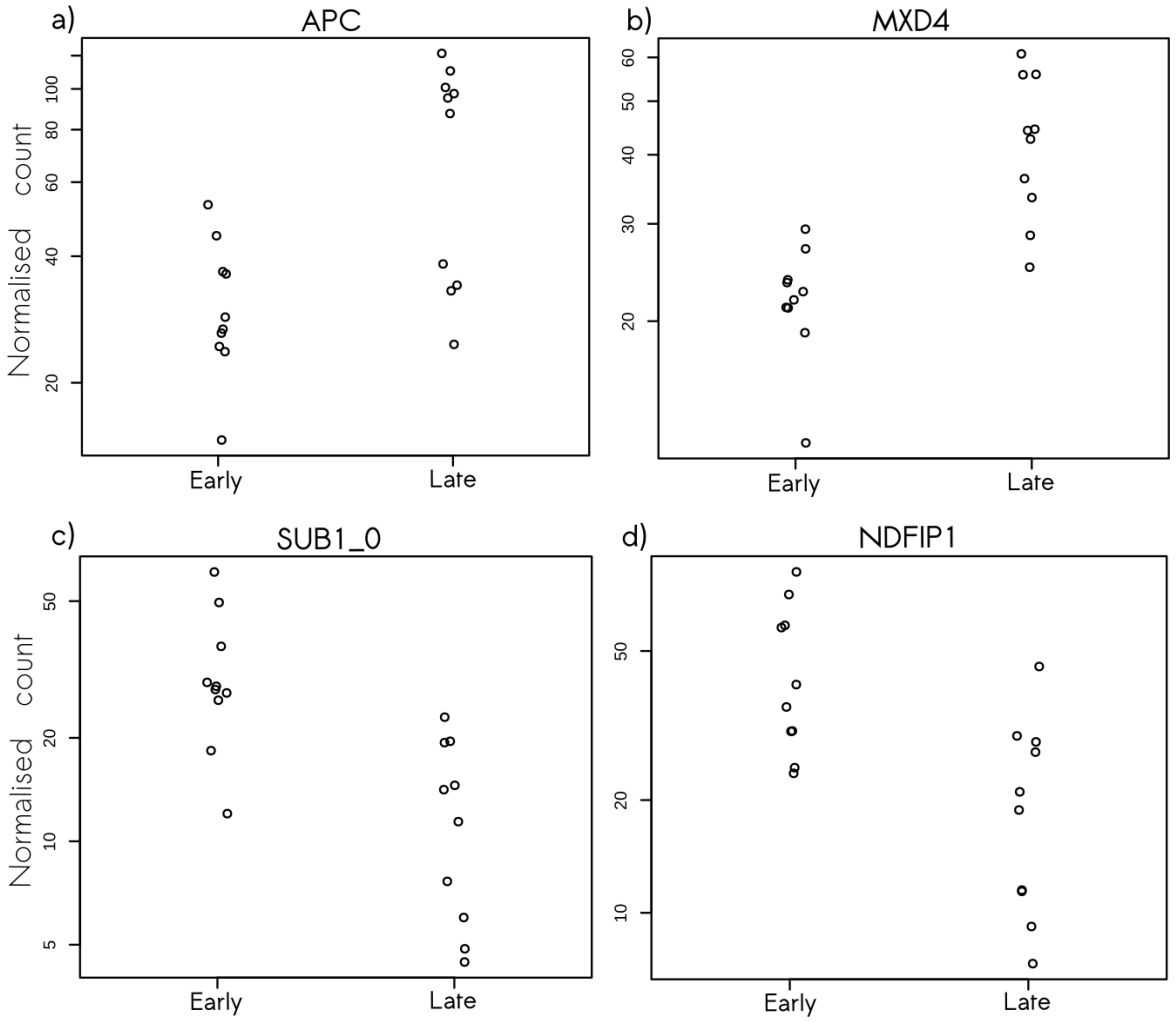


Figure 6. Multidimensional scaling plots (MDS) of some of the most differentially expressed genes (DEGs) between early- and late-born chicks in the genome-alignment approach. a) and b) show the MDS for the top DEGs upregulated in late-born chicks; c) and d) represent the top DEGs upregulated in early-born chicks.

Among the most downregulated genes in the late group (i.e., upregulated in the early group), we identified genes with functions related to the production of reactive oxygen species (ROS), but also related to cellular responses to oxidative stress, the NOX4 (NADPH oxidase 4, log2FC=0.97, *padj*=0.028) and SUB1 (SUB1 regulator of transcription, log2FC=1.36, *padj*=0.0033, **Figure 6c**) genes (Yu et al. 2016), respectively. We also detected DEGs related to protein synthesis, elongation, and marking. These included the MRPS27 gene (mitochondrial ribosomal protein S27, log2FC=0.98, *padj*=0.0288), which aids protein synthesis within the mitochondrion, and the EIF2AK1 gene (eukaryotic translation initiation factor 2 alpha kinase 1, log2FC=0.98, *padj*=0.023) that downregulates protein synthesis in response to stress. The RPL22L1 gene (ribosomal protein L22 like 1, log2FC=1.05, *padj*=0.0089) is a potential structural constituent of the ribosome, and could be related to peptide elongation, while the NDFIP1 gene (Nedd4 family interacting protein 1, log2FC=1.16, *padj*=0.0089, **Figure 6d**) is a target protein for ubiquitination and reduces insulin secretion.

Table 6. GO slims from DEGs between early- and late-born chicks. GOslim: GO slim annotation ID of a subset of GO terms enriched for the genes present in the chick's blood; Count: number of GO terms contained in the GO slim; Terms: name of GOslim term.

GOslim	Count	Term
GO:0030163	4	Protein catabolic process
GO:0036211	2	Protein modification process
GO:0002181	1	Cytoplasmic translation
GO:0002376	1	Immune system process

Regarding the biological processes in which the 133 DEGs between early- and late-born chicks are involved, we found 37 significantly enriched GO terms (**Appendix Chapter 2, Table S5**), mostly related to protein catabolic and modification processes (GOslims GO:0030163 and GO:0036211, **Table 6**). When taking a closer look at the GO terms that were grouped into these two main GOslims, we observed that protein catabolic processes were mostly linked to ubiquitination, while protein modification was linked to GO terms of protein nitrosylation (**Appendix Chapter 2, Table S5**). The genes related to protein ubiquitination upregulated in the late group included the BFAR (log2FC=0.97, *padj*=0.018), NUB1 (log2FC=0.48, *padj*=0.009), SKP1 (log2FC=0.50, *padj*=0.043), MARCH6 (gene log2FC=0.42, *padj*=0.025; isoform log2FC=0.68, *padj*=0.009), USP7 (log2FC=0.68, *padj*=0.033), and UBE3A (log2FC=0.51, *padj*=0.025), while the downregulated ones were ADRM1 (log2FC=0.49, *padj*=0.036), NDFIP1 (log2FC=1.16, *padj*=0.008), and PSMB3 (log2FC=0.48, *padj*=0.028). Only two genes were related to protein nitrosylation, the NOS2 (log2FC=0.68, *padj*=0.017), upregulated in late-born chicks, and the GAPDH (log2FC=0.75, *padj*=0.038), upregulated in early-born chicks.

4. Discussion

The main objective of this study was to investigate whether individuals born in and out of the peak food resources exhibit different early-life gene expression profiles that could be tied to survival capacities under different selective pressures. To address this, we analysed the blood transcriptomes of early- and late-born king penguin chicks that survived harsh winter conditions (including long fasting periods impairing growth) on land before fledging about 10 months later. Our study is the first to sequence the blood transcriptome of the King penguin, so we also characterised the genes and pathways that may be involved during the growth stage prior to fledging.

4.1. The King penguin blood transcriptome

The whole blood transcriptome of king penguin chicks at hatching and fledging is characterised by a large number of genes common to both developmental stages. The most representative biological processes in which these genes are involved are immune defence, especially in adaptive immune defence. The adaptive immune system consists of the acquired response against specific pathogens, different from the innate immune system, which is the first barrier of defence of the body against foreign bodies (Vivier and Malissen 2005). The adaptive immune system takes longer to develop than the innate immune system (Klasing et al. 1998). For this reason, the overrepresentation of active pathways related to the adaptive immune system that is also observed in chicks at hatching may indicate maternal transfer of antibodies (Grindstaff et al. 2003).

Immune system processes are commonly detected in blood transcriptomes, and for that reason, are extensively used to study response profiles to specific infectious agents in different vertebrate species (Zhao et al. 2014; Videvall et al. 2015; Z. Li et al. 2018; Jiminez et al. 2021). Moreover, king penguins from the study colony are known to be exposed to some pathogens as adults, including viruses such as influenza A (Gauthier-Clerc et al. 2002; Chang et al. 2009), and bacteria such as *Borrelia burgdorferia*, the Lyme disease agent transmitted by ticks *Ixodes uriae* infesting penguin colonies (Gauthier-Clerc et al. 1999; Gauthier-Clerc et al. 2003). Even though we did not find any signals of specific infections in the sequenced individuals, the blood transcriptome could be used, through dedicated studies, to better understand the response to pathogenic infections in the species.

Because the transcriptome is a snapshot of the individual's physiological conditions at the moment of sampling, gene regulation is expected to change throughout development (Cardoso-Moreira et al. 2019). In accordance with this, although hatching and fledging chicks have a similar set of circulating transcripts, different genes are up and downregulated at each stage. Hatching chicks exhibit upregulation of genes related to

mitosis and cell division processes, while fledging chicks upregulate genes involved in protein ubiquitination and deubiquitination. The overall upregulation of genes related to the mitotic process in the hatching chicks may be a consequence of the intense growth period, and thus cell division, to which newborns are passing through. Such upregulation of genes related to mitosis is usually more studied in earlier stages of development, such as cell division during foetal stages (Vesterlund et al. 2011; Lefebvre and Lécuyer 2018). However, it can also be present at later stages, such as when individuals experience other pressures that can lead to cell division, like body mass gain (Frias-Soler et al. 2020).

In the case of the king penguin, chick growth can be divided into two phases during the first year of life: pre- and post-winter growth (Barrat 1976; (Descamps et al. 2002). Briefly, chicks go through a first period of growth, from the moment of hatching to the beginning of the first winter, when they are regularly fed by the parents (i.e., from January/February to May/June). During winter, chicks stop growing and lose significant amounts of body mass, as they are fed less frequently or, in some cases, spend the whole winter without being nourished (Cherel and Le Maho 1985). When regular feeding restarts around September, individuals that survive through the winter then resume a second phase of growth until parental care ceases, leading to body mass loss and moulting of the chick before its departure to the sea, i.e., fledgling (Corbel et al. 2008). In this context, the efficient accumulation of body mass before the first winter is one of the main factors that will determine king penguin chick survival through this fasting period. A previous study has demonstrated that early- and late-born chicks that survive until fledgling already show a faster growth in the first 10 days of life, in comparison to chicks that do not survive the first year before fledging (Stier et al. 2014). Therefore, such a strong post-hatch, pre-winter growth demand may explain why cell division is apparently more active in the surviving hatchlings compared to fledglings that have already completed the second growth phase and are entering a second fasting period.

Regarding the upregulation of genes related to protein ubiquitination and deubiquitination in fledging chicks compared to hatching chicks, it is known that protein ubiquitination is a post-translational process widespread among eukaryotes that determines protein fate, usually leading to its degradation, which can be reversed through the deubiquitination process (Wang et al. 2022). We propose that the upregulation of genes related to protein post-translational modifications could be related to the allocation of amino acids caused by a combination of fasting and moulting in fledging chicks. Fledglings are sampled at the end of their first moult, which starts when they stop being fed by the parents and undergo a short fasting period, which can take around 16 days before leaving the colony for the sea (Corbel et al. 2008). As previously described in the king penguin and other penguin species that undergo drastic moult (i.e., all feathers are replaced at the same time), fasting while moulting leads to an increased use of and turnover of amino acids through energy metabolism, increasing the production of uric and nitrogen excretion (Cherel et al. 1994). Moreover, such a combination of moulting and fasting may lead to a higher reuse of available proteins, which could upregulate the cell machinery related to protein ubiquitination/deubiquitination in the fledging chicks.

Overall, these results show that although the majority of gene transcripts in the blood of king penguin chicks at hatching and fledging are the same, the regulation of these genes differ according to pressures associated with each developmental stage.

4.2. The relative contribution of the reference genome and transcriptome

From a methodological point of view, we found that a higher number of genes were detected when mapping our RNA sequencing reads to the assembled transcriptome of the species with the cds increment than when aligning to the reference genome. Although the assembly of the king penguin transcriptome was produced using five tissues that did not include the blood (i.e., brain, liver, kidney, pectoral muscle, and skin, Pirri 2022), a higher proportion of transcripts mapped to it. Given this, we suggest that the blood transcriptome has a low specificity compared to transcriptomes from other tissues, such as the brain (Bentz et al. 2019). The permeability of blood in other tissues may provide blood transcripts to the transcriptome assembly, as has been shown before in humans (Azevedo et al. 2021) and in other birds (Bentz et al. 2019). As an example, the two haemoglobin subunit genes, HBA and HBB, characteristic of the blood transcriptome, are also present in the transcriptome annotation of the other tissues. Therefore, the low tissue-specificity of blood may result in high detection of transcripts even with a reference transcriptome generated from other tissues.

The limitation of the genome alignment pipeline for 3' reads data could be due to the fact that many 3' reads were not assigned to a feature (i.e., gene) at the *htseq-count* step, even for genes present in the genome annotation, which is expected to be mostly complete (only 3.20% missing BUSCO) (Pan et al. 2019). This is because the alignment of 3'end reads to a gene depends on the detection of the 3' untranslated region (UTR), which is absent for many genes in current published genome annotations (Fernandez-Valverde et al. 2015; Lawson et al. 2020). A possible solution to this feature detection problem, apart from generating a more complete genome annotation for the species of interest (Fernandez-Valverde et al. 2015; Lawson et al. 2020), is to increase the feature length in HTSeq to include the UTR regions. In any case, we acknowledge that using the two complementary approaches allowed us to access a richer set of genes than if we used each approach exclusively.

4.3. The unexpected sex-bias in chick blood transcripts

The King penguin is a species characterised by the absence of sexual dimorphism (i.e., absence of visual cues that differentiate males and females) (Nicolaus et al. 2007), apart from slight body size and acoustic variation in the calls between sexes at the adult

stage (Kriesell et al. 2018). In any case, such slight differences between males and females cannot be detected at an immature stage, and even in adults, they can lead to inaccurate conclusions if appropriate tools are not used (i.e., recording calls and analysing these with specific software).

Nevertheless, we did not initially expect a gene expression sex-bias in our sampling design. Most published studies focused on bird sex-biassed gene expression use tissues that are directly related to sex differentiation, such as gonads and the brain, in sexually mature individuals (Vicoso et al. 2013; Mueller et al. 2016; Bentz et al. 2019; He et al. 2022). More importantly, our sampling design is also limited by the small number of surviving chicks after winter, especially in the late group. Therefore, we did not have the possibility of accounting for a better sex-balance within the early and late groups for our DGE analyses. Finally, we did not expect sex-bias patterns in the transcriptome of king penguin chicks, as such bias is expected to be less prevalent at early stages of development (Mank et al. 2010; Perry et al. 2014). However, we detected a clear sex separation between males and females in the gene expression PCA, as well as differentially expressed genes between sexes.

Males showed a much higher number of upregulated genes in comparison to females (264 in males, 75 in females), and we suspect that this could be related to the possible absence or incompleteness of dosage compensation on birds (Itoh et al. 2007). Dosage compensation consists on the expression silencing of one sexual chromosome in the homogametic sex (i.e., XX females in mammals, ZZ males in birds), as the heterogametic sex is usually characterised by the deterioration of the Y and W chromosomes (Disteche 2012). Several studies have previously demonstrated that birds do not effectively perform this mechanism leading to males that have a higher overall gene expression than females (Ellegren et al. 2007; Vicoso et al. 2013; Mueller et al. 2016; Ramstad et al. 2016). The higher number of differentially expressed genes in males in our study indicate that the king penguin may also ineffectively perform or even lack a dosage compensation mechanism. However, a clear conclusion can only be taken through the quantification of differentially expressed genes linked to each sex chromosome. Even if this was not the focus of our study, this analysis could be done in the future using the sex-linked scaffolds identified in Fernandes et al. *in prep.* (Chapter 1) for gene mapping and quantification.

Regarding the biological processes in which differentially expressed genes between the sexes are likely involved, alcohol biosynthesis processes were the most representative. This pattern could be related to the preen oil composition, which can vary between sexes and developmental stages (Grieves et al. 2022). The preen oil, produced by the uropygial gland of birds (Johnston 1988; Moreno-Rueda 2017), is composed of a combination of waxes and volatiles (such as alcohols) (Campagna et al. 2012), and it has been proposed to play a role in sexual selection and/or olfactory camouflage (Grieves et al. 2022). Although a previous study on the composition of volatile compounds from the uropygial gland of king penguin adults did not detect a differences between sexes (Gabirot et al. 2018), this should be confirmed in newborn individuals, as our results indicate a sex differentiation in alcohol biosynthesis. In our case, as samples are from chicks exposed to terrestrial predation (Le Bohec et al. 2005; Descamps et al. 2005); this could also be an olfactory camouflage strategy that differs between sexes and could to be further tested and investigated.

Among the upregulated genes in females, GPBP1 (GC-rich promoter binding protein 1) is related to the development of atherosclerosis, a disease characterised by the thickening or hardening of arteries caused by a buildup of plaque in the inner lining of an artery, and also known to affect birds (Moghadasian 2002). This disease is related to the accumulation of plasma cholesterol in other bird species of economic interest (Bavelaar and Beynen 2004). Moreover, the upregulation of such genes in the females could suggest a sex-related propensity for the development of atherosclerosis in this species, as it has already been shown by studies in psittacine birds (Beaufrère et al. 2013).

To our knowledge, only two other studies have previously detected sex-biassed expression in the blood transcriptome of birds, one in kiwis (*Apteryx spp*, Ramstad et al. 2016) and another in the European blackbird (*Turdus merula*, Franchini et al. 2017). As has been also suggested by the other studies, the expression profiles of males and females could even be used for individual sex identification, as we have shown in our study, by the prediction of 2021 chick's sex. The next step will be to perform a clustering analysis to detect whether the sex differentiation pattern observed in the PCAs of gene expression can be used as predictors of individual sex.

4.4. Plastic adaptations to stress and energy accumulation in late chicks

To investigate whether being born out of the peak of food resources triggers adaptive plastic responses, acting as a PAR, we performed a differential gene expression analysis between early- and late-born chicks of 2020. Our results showed a slight differentiation between the expression patterns of the two groups, with a set of 133 differentially expressed genes.

As our analyses use samples from same-age individuals in a wild environment, some level of overlap could be expected, as a consequence of a higher heterogeneity of responses generated by natural conditions in comparison to more homogeneous laboratory experiments (Krishnan et al. 2020). This can be seen by the fact that even the most differentially expressed genes between early and late chicks showed low log fold changes. Additionally, a slight overlap between the groups in the PCA, especially when using the genome-aligned data. In this case, the use of more biological replicates could clarify the observed patterns, but even with a sample size of 10 individuals per group we were already able to detect differences between the two groups.

As mentioned above, the actual function of up and downregulated genes and their interactions was assumed according to the literature mostly based on human or other model organisms, but also based on the chicken (*Gallus gallus*), which is phylogenetically closer to the penguins, and therefore, any inferred gene function and synteny should be more relevant for our study (Ellegren 2010).

Regarding the biological processes to which the differentially expressed genes between early- and late-born chicks were involved, the most representative ones were

protein catabolic and modification processes. More specifically, many genes were related to ubiquitin-dependent protein catabolism. Ubiquitin-dependent protein catabolism involves the degradation of proteins through ubiquitin marking, which is specially important for ensuring correct signalling, cell fate and functioning at the beginning of development (Rape 2017). Among the upregulated genes related to ubiquitin-dependent protein catabolism in the late group, USP7 (ubiquitin-specific protease 7) has been found to be related to food efficiency uptake in chickens (Prakash et al. 2021). In this study, the authors demonstrated that, when subjected to the same amount of food intake during the same period of time, individuals capable of increasing body weight more efficiently showed the upregulation of this gene. In king penguins, Stier et al. (Stier et al. 2014) have shown that late chicks which survive the first winter already have higher body mass at 10 days of life in comparison to early survivor chicks. If the overexpression of the USP7 gene has a similar effect in king penguins as has been evidenced in the chicken study detailed above, it could represent a promising candidate gene for plastic adaptation to the faster growth pressure late chicks are subjected to before the first winter.

In addition to the USP7, one of the most upregulated genes in the late group, MXD4 (MAX dimerization protein 4), is physically located close to two quantitative trait loci (QTL) related to abdominal fat weight in chickens (Sun et al. 2013). Although we do not have information about this gene's location in the king penguin genome, the MXD4 could also be a target for further investigation of body mass accumulation by late chicks if the region shows a significant level of synteny with the chicken genome. MXD4 is also known for being a tumour suppressor, recently discovered to be regulated by p53, another key tumour suppressor gene, under stressful conditions (Coronel et al. 2021). Along with the MXD4, other tumour suppressor genes were upregulated in the late group, such as the APC and CSNK1A1 genes, which are involved in cancer and embryonic development and are highly conserved among metazoans (Croce and McClay 2008). Tumour suppressors are mostly studied for protecting the organism from uncontrolled cell proliferation characteristic of cancers (Cooper 2000). Additionally, they are also essential to protect the DNA strand from external stressors, such as high amounts of reactive oxygen species (ROS) (Vurusaner et al. 2012).

Although king penguin chicks, whether early- or late-born, are exposed to intense growth pressure before the first winter (Descamps et al. 2002), the higher growth pressure in the late group has been shown to generate higher amounts of ROS and shorter telomeres (Stier et al. 2014). In this case, even if late-born chicks seem to produce more ROS than early chicks, our results suggest that late-born chicks that survive until fledgling may have a more active tumour suppression machinery, which could be restraining the harmful effects of ROS accumulation, such as slow growth (Alonso-Alvarez et al. 2007) or death (Stier et al. 2014). Furthermore, the upregulation of tumour suppressor genes in the late survivors could even be related to the lower accumulation of highly deleterious mutations detected in this group shown in Fernandes et al. *in prep.* (Chapter 1).

According to our initial hypothesis, late-born individuals seem to show adaptations to rapid accumulation of energetic reserves, through the overexpression of genes potentially

related to higher efficiency of body mass accumulation and abdominal fat (USP7 and MXD4, respectively). Additionally, even if faster growth can generate high amounts of ROS, late chicks upregulate genes putatively related to the tumour suppression machinery, which can avoid major damage to DNA integrity. Overall, these results indicate that late chicks express genes indicative of plastic adaptation to stressful conditions that are not equally expressed in early chicks at hatching, and such plasticity could contribute to their survival through the first winter.

The breeding cycle of king penguins also generates pressures of efficient body mass accumulation during adult-life, especially during the pre-reproductive season. If adults successfully fledge a chick, they have around 15 days of interval until moulting, followed by the beginning of the next breeding season (Descamps et al. 2002). Individuals have to use this short time period to forage at sea and accumulate enough body mass to endure the whole 15 days of moulting, which is highly energetically demanding (Cherel et al. 1994). Apart from this, during incubation and chick rearing, breeding success also depends on the foraging efficiency of parents, as one partner needs to stay on land fasting and feeding the chick while the other leaves to the sea (Descamps et al. 2002).

Considering that king penguin adults also have the pressure of efficiently accumulating body mass in short periods of foraging time during the breeding season and between breeding seasons, the gene expression patterns detected in late-born chicks could be considered as a predictive adaptive response (PAR). Late chicks suffer from a much higher mortality through the first year (Olsson 1996; Weimerskirch et al. 1992; Stier et al. 2014), and therefore, show individual detrimental costs of being born late, in a sort of short-term silver spoon effect. However, the gene expression pattern of surviving late-born chicks, which are downregulated in the early group, could be considered as a PAR. Moreover, whether such differences in gene expression are also present during fledgling and through the individual's lifetime is still a matter of study.

Finally, considering the increasing unpredictability on resource availability caused by climate change, adaptations to a more asynchronous environment can be a way for species to escape extinction (Kharouba and Wolkovich 2020). In this scenario, even if late-born chicks may show a higher mortality during the first year of life, late individuals that survive until fledging harbour plastic adaptations that could be close to the reaction norms limits of the species. Therefore, if late-born individuals are indeed equally performant to early-born individuals in adulthood, in terms of survival and reproduction, we can conclude that the late breeding strategy enhances the adaptive potential of the King penguin species.

5. Conclusion and perspectives

In this study, we characterised the King penguin blood transcriptome, using samples from chicks at hatching and fledging. Although most genes are expressed at both developmental stages, the regulation of these between the groups differ, probably according to growth pressures at hatching, during the winter fasting, and during moulting just before fledgling. We also detected a sex-biassed in the blood gene expression of the sampled chicks. We suggest that the blood transcriptome could be used for the sexing of individuals in this species independently of developmental stage, although a proper clustering analysis is still needed to be certain of the reliability of such a method. Finally, we detected genes indicative of adaptations to faster growth and tumour suppression in the late-born chicks. Although the long-term consequences of being born outside of ideal conditions is still an open question for this species, the gene expression patterns of these chicks at hatching could be a predictive adaptive response to efficient accumulation of body mass, also important for adult king penguins during moulting and the breeding season.

Differences between years will be explored in the coming months by the sequencing of a third year of early- and late-born chicks at hatching and fledging (i.e., 2022 born chicks). We will also strengthen the 2021 analyses with more samples to compensate for the sex-bias (i.e., including more 2021 late females and early males, at both hatching and fledging).

References

Alonso-Alvarez, C., S. Bertrand, B. Faivre, and G. Sorci. 2007. “Increased Susceptibility to Oxidative Damage as a Cost of Accelerated Somatic Growth in Zebra Finches.” *Functional Ecology* 21 (5): 873–79.

Alvarez, Mariano, Aaron W. Schrey, and Christina L. Richards. 2015. “Ten Years of Transcriptomics in Wild Populations: What Have We Learned about Their Ecology and Evolution?” *Molecular Ecology* 24 (4): 710–25.

Anders, Simon, and Wolfgang Huber. 2010. “Differential Expression Analysis for Sequence Count Data.” *Genome Biology* 11 (10): R106.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2014. “HTSeq – A Python Framework to Work with High-Throughput Sequencing Data.” *bioRxiv*. https://doi.org/10.1101/002824.

Andrews, S. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data” [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aubin-Horth, Nadia, and Susan C. P. Renn. 2009. “Genomic Reaction Norms: Using Integrative Biology to Understand Molecular Mechanisms of Phenotypic Plasticity.” *Molecular Ecology* 18 (18): 3763–80.

Azevedo, Tiago, Giovanna Maria Dimitri, Pietro Lió, and Eric R. Gamazon. 2021. “Multilayer Modelling of the Human Transcriptome and Biological Mechanisms of Complex Diseases and Traits.” *NPJ Systems Biology and Applications* 7 (1): 24.

Barrat, A., 1976. “Quelques aspects de la biologie et de l'écologie du Manchot royal (Aptenodytes patagonicus) des îles Crozet.”

Bateson, Patrick, Peter Gluckman, and Mark Hanson. 2014. “The Biology of Developmental Plasticity and the Predictive Adaptive Response Hypothesis.” *The Journal of Physiology* 592 (11): 2357–68.

Bavelaar, F. J., and A. C. Beynen. 2004. “The Relation between Diet, Plasma Cholesterol and Atherosclerosis in Pigeons, Quails and Chickens.” 2004. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7a3664f5925312d548f3a970bb2b2612cc57ef12.

Beaufrère, Hugues, Mélanie Ammersbach, Drury R. Reavill, Michael M. Garner, J. Jill Heatley, Nobuko Wakamatsu, Javier G. Nevarez, and Thomas N. Tully. 2013. “Prevalence of and Risk Factors Associated with Atherosclerosis in Psittacine Birds.” *Journal of the American Veterinary Medical Association* 242 (12): 1696–1704.

Bentz, Alexandra B., Gregg W. C. Thomas, Douglas B. Rusch, and Kimberly A. Rosvall. 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): 15849.

Bullard, James H., Elizabeth Purdom, Kasper D. Hansen, and Sandrine Dudoit. 2010. “Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments.” *BMC Bioinformatics* 11 (February): 94.

Campagna, Sylvie, Jérôme Mardon, Aurélie Celerier, and Francesco Bonadonna. 2012. “Potential Semiochemical Molecules from Birds: A Practical and Comprehensive Compilation of the Last 20 Years Studies.” *Chemical Senses* 37 (1): 3–25.

Cardoso-Moreira, Margarida, Jean Halbert, Delphine Valloton, Britta Velten, Chunyan Chen, Yi Shao, Angélica Liechti, et al. 2019. “Gene Expression across Mammalian Organ Development.” *Nature* 571 (7766): 505–9.

Chang, Chung-Ming, Camille Lebarbenchon, Michel Gauthier-Clerc, Céline Le Bohec, David Beaune,

Yvon Le Maho, and Sylvie van der Werf. 2009. “Molecular Surveillance for Avian Influenza A Virus in King Penguins (Aptenodytes Patagonicus).” *Polar Biology* 32 (4): 663.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.

Cherel, Y., and Y. Le Maho. 1985. “Five Months of Fasting in King Penguin Chicks: Body Mass Loss and Fuel Metabolism.” *The American Journal of Physiology* 249 (4 Pt 2): R387–92.

Cherel, Y., Robin, J.P., Walch, O., Karmann, H., Netchitailo, P.. and Le Maho, Y., 1988a. Fasting in king penguin. I. Hormonal and metabolic changes during breeding. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 254(2), pp.R170-R177.

Cherel, Y., Leloup, J. and Le Maho, Y.V.O.N., 1988b. Fasting in king penguin. II. Hormonal and metabolic changes during molt. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 254(2), pp.R178-R184.

Cherel, Yves, Bullet And, and Frederic Freby. 1994. “Daily Body-Mass Loss and Nitrogen Excretion During Molting Fast of Macaroni Penguins.” *The Auk* 111 (111): 447–453492.

Cooper, Geoffrey M. 2000. *Tumor Suppressor Genes*. Sinauer Associates.

Corbel, Hélène, Francis Morlon, and René Groscolas. 2008. “Is Fledging in King Penguin Chicks Related to Changes in Metabolic or Endocrinal Status?” *General and Comparative Endocrinology* 155 (3): 804–13.

Coronel, Luis, Konstantin Riege, Katjana Schwab, Silke Förste, David Häckes, Lena Semerau, Stephan H. Bernhart, Reiner Siebert, Steve Hoffmann, and Martin Fischer. 2021. “Transcription Factor RFX7 Governs a Tumor Suppressor Network in Response to p53 and Stress.” *Nucleic Acids Research* 49 (13): 7437–56.

Cox, Michael A. A., and Trevor F. Cox. 2008. “Multidimensional Scaling.” In *Handbook of Data Visualization*, edited by Chun-Houh Chen, Wolfgang Härdle, and Antony Unwin, 315–47. Berlin, Heidelberg: Springer Berlin Heidelberg.

Crino, Ondi L., and Creagh W. Breuner. 2015. “Developmental Stress: Evidence for Positive Phenotypic and Fitness Effects in Birds.” *Journal of Ornithology / DO-G* 156 (1): 389–98.

Croce, Jenifer C., and David R. McClay. 2008. “Evolution of the Wnt Pathways.” *Methods in Molecular Biology* 469: 3–18.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab008.

De Lisle, Stephen P., Maarit I. Mäenpää, and Erik I. Svensson. 2022. “Phenotypic Plasticity Is Aligned with Phenological Adaptation on Both Micro- and Macroevolutionary Timescales.” *Ecology Letters* 25 (4): 790–801.

Descamps, Sebastien, M. Gauthier-Clerc, J. P. Gendner, and Y. Le Maho. 2002. “The Annual Cycle of Unbanded King Penguins Aptenodytes Patagonicus on Possession Island (Crozet)” 2 (January): 87–98.

Descamps, Sébastien, Michel Gauthier-Clerc, Céline Le Bohec, Jean-Paul Gendner, and Yvon Le Maho. 2005. “Impact of Predation on King Penguin Aptenodytes Patagonicus in Crozet Archipelago.” *Polar Biology* 28 (4): 303–10.

Dillies, Marie-Agnès, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, et al. 2013. “A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput RNA Sequencing Data Analysis.” *Briefings in Bioinformatics* 14 (6): 671–83.

Disteche, Christine M. 2012. “Dosage Compensation of the Sex Chromosomes.” *Annual Review of Genetics* 46 (September): 537–60.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21.

Ellegren, Hans. 2010. “Evolutionary Stasis: The Stable Chromosomes of Birds.” *Trends in Ecology & Evolution* 25 (5): 283–91.

Ellegren, Hans, Lina Hultin-Rosenberg, Björn Brunström, Lennart Dencker, Kim Kultima, and Birger Scholz. 2007. “Faced with Inequality: Chicken Do Not Have a General Dosage Compensation of Sex-Linked Genes.” *BMC Biology* 5 (September): 40.

Fernandez-Valverde, Selene L., Andrew D. Calcino, and Bernard M. Degnan. 2015. “Deep Developmental Transcriptome Sequencing Uncovers Numerous New Genes and Enhances Gene Annotation in the Sponge Amphimedon Queenslandica.” *BMC Genomics* 16 (1): 387.

Franchini, Paolo, Iker Irisarri, Adam Fudickar, Andreas Schmidt, Axel Meyer, Martin Wikelski, and Jesko Partecke. 2017. “Animal Tracking Meets Migration Genomics: Transcriptomic Analysis of a Partially Migratory Bird Species.” *Molecular Ecology* 26 (12): 3204–16.

Frias-Soler, Roberto Carlos, Lilian Villarín Pildaín, Liviu G. Pârâu, Michael Wink, and Franz Bairlein. 2020. “Transcriptome Signatures in the Brain of a Migratory Songbird.” *Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics* 34 (June): 100681.

Gabirot, Marianne, Bruno Buatois, Carsten T. Müller, and Francesco Bonadonna. 2018. “Odour of King Penguin Feathers Analysed Using Direct Thermal Desorption Discriminates between Individuals but Not Sexes.” *The Ibis* 160 (2): 379–89.

Gagnon-Bartsch, Johann A., and Terence P. Speed. 2012. “Using Control Genes to Correct for Unwanted Variation in Microarray Data.” *Biostatistics* 13 (3): 539–52.

Gauthier-Clerc, Michel, Nicolas Eterradossi, Didier Toquin, Michèle Guittet, Grégoire Kuntz, and Yvon Le Maho. 2002. “Serological Survey of the King Penguin, Aptenodytes Patagonicus, in Crozet Archipelago for Antibodies to Infectious Bursal Disease, Influenza A and Newcastle Disease Viruses.” *Polar Biology* 25 (4): 316–19.

Gauthier-Clerc, Michel, Stephan Mangin, Céline Le Bohec, Jean-Paul Gendner, and Yvon Le Maho. 2003. “Comparison of Behaviour, Body Mass, Haematocrit Level, Site Fidelity and Survival between Infested and Non-Infested King Penguin Aptenodytes Patagonicus by Ticks Ixodes Uriae.” *Polar Biology* 26 (6): 379–82.

Gauthier-Clerc, M., B. Jaulhac, Y. Frenot, C. Bachelard, H. Monteil, Y. Le Maho, and Y. Handrich. 1999. “Prevalence of Borrelia Burgdorferi (the Lyme Disease Agent) Antibodies in King Penguin Aptenodytes Patagonicus in Crozet Archipelago.” *Polar Biology* 22 (2): 141–43.

Geiger, Sylvie, Maryline Le Vaillant, Thomas Lebard, Sophie Reichert, Antoine Stier, Yvon LE Maho, and Francois Criscuolo. 2012. “Catching-up but Telomere Loss: Half-Opening the Black Box of Growth and Ageing Trade-off in Wild King Penguin Chicks.” *Molecular Ecology* 21 (6): 1500–1510.

Gendner, Jean-Paul, Michel Gauthier-Clerc, Celine Le Bohec, Sebastien Descamps, and Yvon Le Maho. 2005. “A New Application for Transponders in Studying Penguins.” *Journal of Field Ornithology* 76 (2): 138–42.

Ghalambor, C. K., J. K. McKAY, S. P. Carroll, and D. N. Reznick. 2007. “Adaptive versus Non-Adaptive Phenotypic Plasticity and the Potential for Contemporary Adaptation in New Environments.” *Functional Ecology* 21 (3): 394–407.

Gilad, Yoav, Jonathan K. Pritchard, and Kevin Thornton. 2009. “Characterizing Natural Variation Using next-Generation Sequencing Technologies.” *Trends in Genetics: TIG* 25 (10): 463–71.

Gluckman, Peter D., Mark A. Hanson, Hamish G. Spencer, and Patrick Bateson. 2005. “Environmental Influences during Development and Their Later Consequences for Health and

Disease: Implications for the Interpretation of Empirical Studies." *Proceedings. Biological Sciences / The Royal Society* 272 (1564): 671–77.

Grafen, Alan. 1988. "On the uses of data on lifetime reproductive success." Reproductive success 454-485.

Grieves, Leanne A., Marc Gilles, Innes C. Cuthill, Tamás Székely, Elizabeth A. MacDougall-Shackleton, and Barbara A. Caspers. 2022. "Olfactory Camouflage and Communication in Birds." *Biological Reviews of the Cambridge Philosophical Society* 97 (3): 1193–1209.

Grindstaff, Jennifer L., Edmund D. Brodie 3rd, and Ellen D. Ketterson. 2003. "Immune Function across Generations: Integrating Mechanism and Evolutionary Process in Maternal Antibody Transmission." *Proceedings. Biological Sciences / The Royal Society* 270 (1531): 2309–19.

Groscolas, R. and Cherel, Y., 1992. How to molt while fasting in the cold: the metabolic and hormonal adaptations of emperor and king penguins. Ornis Scandinavica, pp.328-334.

Hamel, Sandra, Jean-Michel Gaillard, Marco Festa-Bianchet, and Steeve D. Côté. 2009. "Individual Quality, Early-Life Conditions, and Reproductive Success in Contrasted Populations of Large Herbivores." *Ecology* 90 (7): 1981–95.

Hayward, Adam D., Ian J. Rickard, and Virpi Lummaa. 2013. "Influence of Early-Life Nutrition on Mortality and Reproductive Success during a Subsequent Famine in a Preindustrial Population." *Proceedings of the National Academy of Sciences of the United States of America* 110 (34): 13886–91.

Huber, Wolfgang, Anja von Heydebreck, Holger Sueltmann, Annemarie Poustka, and Martin Vingron. 2003. "Parameter Estimation for the Calibration and Variance Stabilization of Microarray Data." *Statistical Applications in Genetics and Molecular Biology* 2 (April): Article3.

Itoh, Yuichiro, Esther Melamed, Xia Yang, Kathy Kampf, Susanna Wang, Nadir Yehya, Atila Van Nas, et al. 2007. "Dosage Compensation Is Less Effective in Birds than in Mammals." *Journal of Biology* 6 (1): 2.

Jiminez, Janelle, Edouard Timsit, Karin Orsel, Frank van der Meer, Le Luo Guan, and Graham Plastow. 2021. "Whole-Blood Transcriptome Analysis of Feedlot Cattle With and Without Bovine Respiratory Disease." *Frontiers in Genetics* 12 (March): 627623.

Johnston, D.W., 1988. Morphological atlas of the avian uropygial gland. British Museum (Natural History).

Kharouba, Heather M., and Elizabeth M. Wolkovich. 2020. "Disconnects between Ecological Theory and Data in Phenological Mismatch Research." *Nature Climate Change* 10 (5): 406–15.

Kriesell, Hannah J., Thierry Aubin, Víctor Planas-Bielsa, Marine Benoiste, Francesco Bonadonna, Hélène Gachot-Neveu, Yvon Le Maho, et al. 2018. "Sex Identification in King Penguins*Aptenodytes Patagonicus*through Morphological and Acoustic Cues." *The Ibis* 160 (4): 755–68.

Krishnan, Jaya, Jenna L. Persons, Robert Peuß, Huzaifa Hassan, Alexander Kenzior, Shaolei Xiong, Luke Olsen, Ernesto Maldonado, Johanna E. Kowalko, and Nicolas Rohner. 2020. "Comparative Transcriptome Analysis of Wild and Lab Populations of Astyanax Mexicanus Uncovers Differential Effects of Environment and Morphotype on Gene Expression." *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 334 (7-8): 530–39.

Lawson, Nathan D., Rui Li, Masahiro Shin, Ann Grosse, Onur Yukselen, Oliver A. Stone, Alper Kucukural, and Lihua Zhu. 2020. "An Improved Zebrafish Transcriptome Annotation for Sensitive and Comprehensive Detection of Cell Type-Specific Genes." *eLife* 9 (August). https://doi.org/10.7554/eLife.55792.

Le Bohec, Céline, Michel Gauthier-Clerc, and Yvon Le Maho. 2005. "The Adaptive Significance of

Crèches in the King Penguin." *Animal Behaviour* 70 (3): 527–38.
Lefebvre, Fabio Alexis, and Éric Lécuyer. 2018. "Flying the RNA Nest: Drosophila Reveals Novel Insights into the Transcriptome Dynamics of Early Development." *Journal of Developmental Biology* 6 (1). https://doi.org/10.3390/jdb6010005.
Lindström, J. 1999. "Early Development and Fitness in Birds and Mammals." *Trends in Ecology & Evolution* 14 (9): 343–48.
Li, Zan, Xiumei Liu, Jinxiang Liu, Kai Zhang, Haiyang Yu, Yan He, Xubo Wang, Jie Qi, Zhigang Wang, and Quanqi Zhang. 2018. "Transcriptome Profiling Based on Protein-Protein Interaction Networks Provides a Core Set of Genes for Understanding Blood Immune Response Mechanisms against Edwardsiella Tarda Infection in Japanese Flounder (Paralichthys Olivaceus)." *Developmental and Comparative Immunology* 78 (January): 100–113.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
Marak, Hamida B., Arjen Biere, and Jos M. M. Van Damme. 2003. "Fitness Costs of Chemical Defense in Plantago Lanceolata L.: Effects of Nutrient and Competition Stress." *Evolution; International Journal of Organic Evolution* 57 (11): 2519–30.
Mi, Huaiyu, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D. Thomas. 2019. "PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools." *Nucleic Acids Research* 47 (D1): D419–26.
Millon, Alexandre, Steve J. Petty, Brian Little, and Xavier Lambin. 2011. "Natal Conditions Alter Age-Specific Reproduction but Not Survival or Senescence in a Long-Lived Bird of Prey." *The Journal of Animal Ecology* 80 (5): 968–75.
Moghadasian, Mohammed H. 2002. "Experimental Atherosclerosis: A Historical Overview." *Life Sciences* 70 (8): 855–65.
Moll, Pamela, Michael Ante, Alexander Seitz, and Torsten Reda. 2014. "QuantSeq 3′ mRNA Sequencing for RNA Quantification." *Nature Methods* 11 (12): i – iii.
Moreno-Rueda, Gregorio. 2017. "Preen Oil and Bird Fitness: A Critical Review of the Evidence." *Biological Reviews of the Cambridge Philosophical Society* 92 (4): 2131–43.
Morgan M, Falcon S, Gentleman R. 2022. "GSEABase: Gene set enrichment data structures and methods." R package version 1.60.0.
Mueller, Jakob C., Heiner Kuhl, Bernd Timmermann, and Bart Kempenaers. 2016. "Characterization of the Genome and Transcriptome of the Blue Tit Cyanistes Caeruleus: Polymorphisms, Sex-Biased Expression and Selection Signals." *Molecular Ecology Resources* 16 (2): 549–61.
Mugabo, M., O. Marquis, S. Perret, and J. F. LE Galliard. 2010. "Immediate and Delayed Life History Effects Caused by Food Deprivation Early in Life in a Short-Lived Lizard." *Journal of Evolutionary Biology* 23 (9): 1886–98.
Nicolaus, Marion, Céline Le Bohec, Paul M. Nolan, Michel Gauthier-Clerc, Yvon Le Maho, Jan Komdeur, and Pierre Jouventin. 2007. "Ornamental Colors Reveal Age in the King Penguin." *Polar Biology* 31 (1): 53–61.
Nursyifa, Casia, Anna Brüniche-Olsen, Genis Garcia-Erill, Rasmus Heller, and Anders Albrechtsen. 2022. "Joint Identification of Sex and Sex-Linked Scaffolds in Non-Model Organisms Using Low Depth Sequencing Data." *Molecular Ecology Resources* 22 (2): 458–67.
Olsson, Olof. 1996. "Seasonal Effects of Timing and Reproduction in the King Penguin: A Unique Breeding Cycle." *Journal of Avian Biology* 27 (1): 7–14.
Oostra, Vicencio, Marjo Saastamoinen, Bas J. Zwaan, and Christopher W. Wheat. 2018. "Strong Phenotypic Plasticity Limits Potential for Evolutionary Responses to Climate Change." *Nature Communications* 9 (1): 1005.

Pan, Hailin, Theresa L. Cole, Xupeng Bi, Miaoquan Fang, Chengran Zhou, Zhengtao Yang, Daniel T. Ksepka, et al. 2019. “High-Coverage Genomes to Elucidate the Evolution of Penguins.” *GigaScience* 8 (9). https://doi.org/10.1093/gigascience/giz117.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” *Nature Methods* 14 (4): 417–19.

Pirri, Federica. 2022. “Comparative genomics and transcriptomics reveal the genetic basis for adaptation to the extreme Antarctic environment in the Emperor penguin (*A. forsteri*)” PhD thesis. University of Padova.

Prakash, A., V. K. Saxena, Ravi Kumar, S. Tomar, M. K. Singh, and Gagandeep Singh. 2021. “Differential Gene Expression in Liver of Colored Broiler Chicken Divergently Selected for Residual Feed Intake.” *Tropical Animal Health and Production* 53 (3): 403.

Ramstad, Kristina M., Hilary C. Miller, and Gabriel Kolle. 2016. “Sixteen Kiwi (Apteryx Spp) Transcriptomes Provide a Wealth of Genetic Markers and Insight into Sex Chromosome Evolution in Birds.” *BMC Genomics* 17 (May): 410.

Rape, Michael. 2017. “Ubiquitylation at the Crossroads of Development and Disease.” *Nature Reviews. Molecular Cell Biology* 19 (1): 59–70.

Risso, Davide, John Ngai, Terence P. Speed, and Sandrine Dudoit. 2014. “Normalization of RNA-Seq Data Using Factor Analysis of Control Genes or Samples.” *Nature Biotechnology* 32 (9): 896–902.

Safran, Marilyn, Naomi Rosen, Michal Twik, Ruth BarShir, Tsippi Iny Stein, Dvir Dahary, Simon Fishilevich, and Doron Lancet. 2021. “The GeneCards Suite.” In *Practical Guide to Life Science Databases*, edited by Imad Abugessaisa and Takeya Kasukawa, 27–56. Singapore: Springer Nature Singapore.

Schlichting, C.D. and Pigliucci, M., 1998. Phenotypic evolution: a reaction norm perspective. Sinauer associates incorporated.

Stark, Rory, Marta Grzelak, and James Hadfield. 2019. “RNA Sequencing: The Teenage Years.” *Nature Reviews. Genetics* 20 (11): 631–56.

Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, et al. 2016. “The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 54 (June): 1.30.1–1.30.33.

Stier, Antoine, Vincent A. Viblanc, Sylvie Massemin-Challet, Yves Handrich, Sandrine Zahn, Emilio R. Rojas, Claire Saraux, et al. 2014. “Starting with a Handicap: Phenotypic Differences between Early- and Late-born King Penguin Chicks and Their Survival Correlates.” *Functional Ecology* 28 (3): 601–11.

Sun, Yanfa, Guiping Zhao, Ranran Liu, Maiqing Zheng, Yaodong Hu, Dan Wu, Lei Zhang, Peng Li, and Jie Wen. 2013. “The Identification of 14 New Genes for Meat Quality Traits in Chicken Using a Genome-Wide Association Study.” *BMC Genomics* 14 (July): 458.

Taborsky, Barbara. 2006. “The Influence of Juvenile and Adult Environments on Life-History Trajectories.” *Proceedings. Biological Sciences / The Royal Society* 273 (1587): 741–50.

Tibshirani, Robert. 1988. “Estimating Transformations for Regression Via Additivity and Variance Stabilization.” *Journal of the American Statistical Association* 83 (402): 394–405.

Uebbing, Severin, Axel Künstner, Hannu Mäkinen, Niclas Backström, Paulina Bolivar, Reto Burri, Ludovic Dutoit, et al. 2016. “Divergence in Gene Expression within and between Two Closely Related Flycatcher Species.” *Molecular Ecology* 25 (9): 2015–28.

Van de Pol, Martijn, Leo W. Bruinzeel, Dik Heg, Henk P. Van der Jeugd, and Simon Verhulst. 2006. “A

Silver Spoon for a Golden Future: Long-Term Effects of Natal Origin on Fitness Prospects of Oystercatchers (Haematopus Ostralegus)." *The Journal of Animal Ecology* 75 (2): 616–26.

Vesterlund, Liselotte, Hong Jiao, Per Unneberg, Outi Hovatta, and Juha Kere. 2011. "The Zebrafish Transcriptome during Early Development." *BMC Developmental Biology* 11 (May): 30.

Vicoso, Beatriz, Vera B. Kaiser, and Doris Bachtrog. 2013. "Sex-Biased Gene Expression at Homomorphic Sex Chromosomes in Emus and Its Implication for Sex Chromosome Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 110 (16): 6453–58.

Videvall, Elin, Charlie K. Cornwallis, Vaidas Palinauskas, Gediminas Valkiūnas, and Olof Hellgren. 2015. "The Avian Transcriptome Response to Malaria Infection." *Molecular Biology and Evolution* 32 (5): 1255–67.

Vivier, Eric, and Bernard Malissen. 2005. "Innate and Adaptive Immunity: Specificities and Signaling Hierarchies Revisited." *Nature Immunology* 6 (1): 17–21.

Vurusaner, Beyza, Giuseppe Poli, and Huveyda Basaga. 2012. "Tumor Suppressor Genes and ROS: Complex Networks of Interactions." *Free Radical Biology and Medicine* 52 (1): 7–18.

Wang, Jiayu, Qi Zhou, Jinli Ding, Tailang Yin, Peng Ye, and Yan Zhang. 2022. "The Conceivable Functions of Protein Ubiquitination and Deubiquitination in Reproduction." *Frontiers in Physiology* 13 (July): 886261.

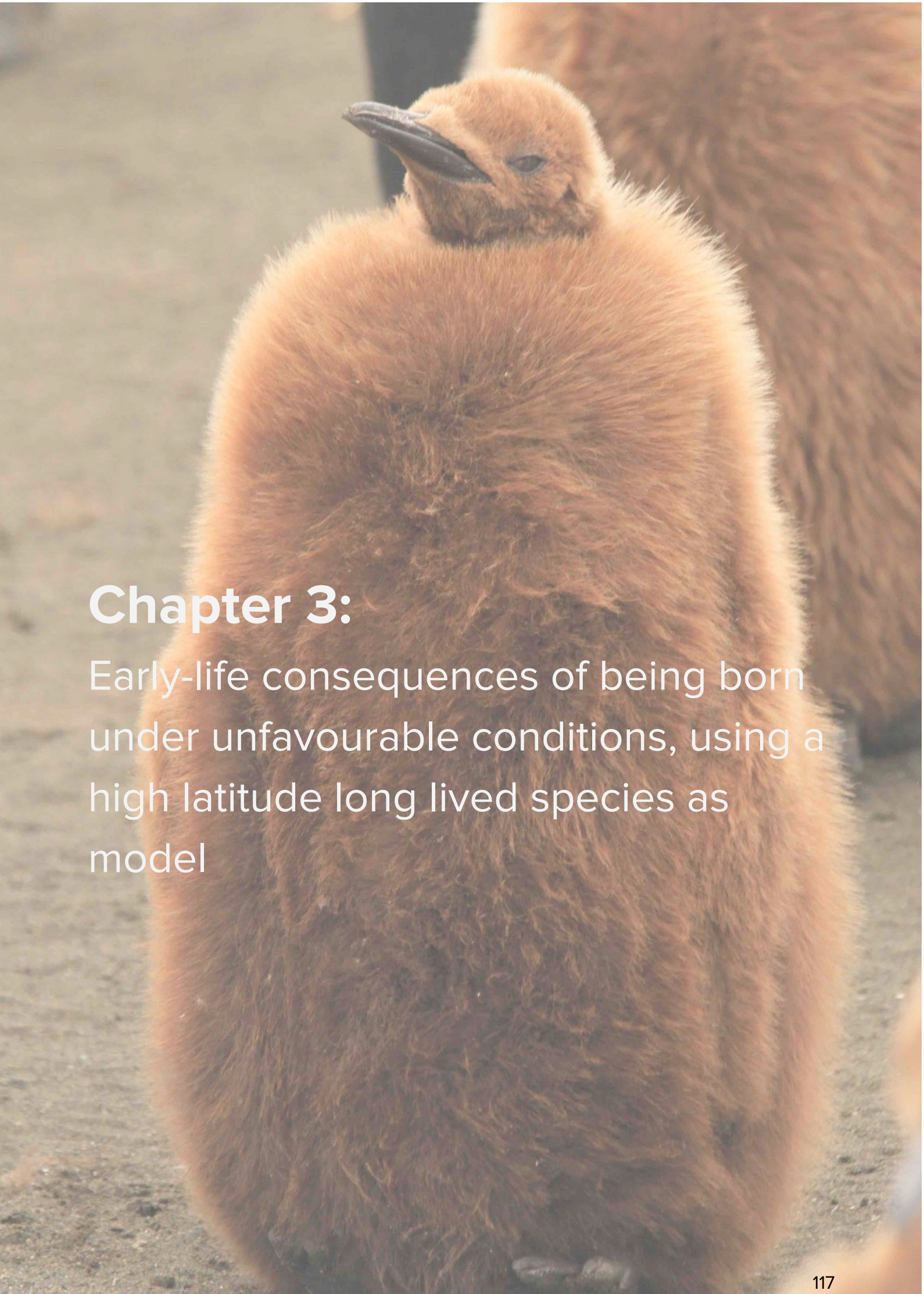
Watson, Hannah, Elin Videvall, Martin N. Andersson, and Caroline Isaksson. 2017. "Transcriptome Analysis of a Wild Bird Reveals Physiological Responses to the Urban Environment." *Scientific Reports* 7 (March): 44180.

Weimerskirch, H., J. C. Stahl, and P. Jouventin. 1992. "The Breeding Biology and Population Dynamics of King Penguins Aptenodytes Patagonica on the Crozet Islands." *The Ibis* 134 (2): 107–17.

Yu, Lijian, Hong Ma, Xincai Ji, and Michael R. Volkert. 2016. "The Sub1 Nuclear Protein Protects DNA from Oxidative Damage." *Molecular and Cellular Biochemistry* 412 (1-2): 165–71.

Zhao, Feng, Chao Yan, Xuan Wang, Yang Yang, Guangyin Wang, Wenhui Lee, Yang Xiang, and Yun Zhang. 2014. "Comprehensive Transcriptome Profiling and Functional Analysis of the Frog (Bombina Maxima) Immune System." *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 21 (1): 1–13.

Zhu, Anqi, Joseph G. Ibrahim, and Michael I. Love. 2019. "Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences." *Bioinformatics* 35 (12): 2084–92.



Chapter 3:

Early-life consequences of being born under unfavourable conditions, using a high latitude long lived species as model

Chapter 3:

Early-life consequences of being born under unfavourable conditions, using a high latitude long lived species as model

In preparation

Authors:
Flávia A. Nitta Fernandes[1, 2] & Gaël Bardon[1,3]
Robin Cristofari[4]
Denis Allemand[3]
Marine Benoiste[1]
Nicolas Chatelain[1]
Julien Courtecuisse[1]
Rebecca Garcia[1]
Benoit Vallas[5]
Emiliano Trucchi[2] & Céline Le Bohec[1,3]

Affiliations:
1 - University of Strasbourg, France
2 - Marche Polytechnic University, Italy
3 - Scientific Center of Monaco, Monaco
4 - University of Turku, Finland
5 - Terres Australes et Antarctiques Françaises (TAAF), France

Keywords
Match-mismatch; pre-fledging traits; post-fledging traits; recruitment; seabirds, polar regions; silver spoon hypothesis; predictive adaptive response hypothesis

Authors' contribution to this paper

This paper's idea was jointly conceived and designed by F. Fernandes, G. Bardon, C. Le Bohec and E. Trucchi. The time series data collection (measurements, RFID implantation, field observations) was done by F. Fernandes, G. Bardon, R. Cristofari, M. Benoiste, R. Garcia, B. Vallas, E. Trucchi, and C. Le Bohec, with the aid of other field work assistants from 2010 to 2022, coordinated by C. Le Bohec in the French Southern and Antarctic Lands (TAAF). The automatic monitoring system used in this study was developed or consolidated by N. Chatelain, J. Courtecuisse, C. Le Bohec, and their respective research groups, with the posterior contributions of G. Bardon, R. Cristofari, and C. Le Bohec on the development of the pipeline for biological interpretation of information stored by the system. Analyses were performed by G. Bardon and C. Le Bohec. Results' discussion and interpretation were done by F. Fernandes, G. Bardon, and C. Le Bohec with the contributions of R. Cristofari and E. Trucchi. Writing was done by F. Fernandes, G. Bardon, C. Le Bohec with the contribution of E. Trucchi and D. Allemand.

1. Introduction

In natural ecosystems, most organisms synchronise the onset of reproduction and migration with the peak of environmental resources (Price 1988; Williams et al. 2017). According to life history theory, this match between most energetically costly activities and the period of highest resource abundance is known as the match-mismatch hypothesis (MMH) (Cushing 1974; Cushing and Saleem 1982; Cushing 1990). In the MMH, a "mismatch" refers to a reduction in the fitness of the individuals phenologically outside of the optimal synchrony. In this case, individuals' fitness is strongly dependent on resource availability and the resource-consumer system is characterised by a high degree of seasonality (Kharouba and Wolkovich 2020), which is usually the case of high latitude species. The mortality rates of individuals born out of the peak of food resources can therefore be high (Durant et al. 2005; Post and Forchhammer 2008; Schenk et al. 2018) and/or carryover consequences are expected to affect survival and reproduction later in life (Wilson et al. 2021).

Carryover effects arise when an individual's previous condition influences their performance in a later stage (O'Connor et al. 2014). Poor birth conditions can, for instance, impair body condition at the juvenile stage, and juvenile body condition, in its turn, can lead to lower survival and reproductive performance in adulthood. For example, Olympia oysters (*Ostrea lurida*) reared under acidic environments have slower growth rates than conspecifics reared under a neutral pH, showing a carryover effect from bad larvae condition in the juvenile stage (Hettinger et al. 2012). Another study on the post-fledging survival of greater sage-grouse (*Centrocercus urophasianus*) showed that chick body condition positively affects survival after fledging (Blomberg et al. 2014).

Carryover effects can persist through the individual's lifetime and even through generations (Salinas et al. 2013; Burton and Metcalfe 2014), remaining as a life history trait in the population (Van Allen and Rudolf 2016). In this way, the positive or detrimental early-life conditions can affect later fitness and population dynamics in what is known as the *silver spoon* effect (Grafen 1988). Silver spoon effects have been largely studied in natural systems (Taborsky 2006; Mugabo et al. 2010; Millon et al. 2011; Hamel et al. 2009; Hayward et al. 2013; Pigeon et al. 2017), and can be especially important for population dynamics and species persistence under variable conditions (Song et al. 2019).

On the other hand, if a strategy is maintained within a species or population, it is expected to provide evolutionary advantages, or at least, not to be detrimental to the point of being removed by selection (Crespi 2000). In the context of the silver spoon hypothesis, if a poor early developmental condition will negatively affect lifetime fitness, it should be removed from the population or occur at a low frequency. However, many natural systems show a regular production of individuals under unfavourable conditions, such as under mismatch with the peak of resources (Stenseth and Mysterud 2002; Singer and Parmesan 2010; Plard et al. 2014; Doiron et al. 2015). The widespread persistence of mismatch in nature can be explained by the fact that these strategies provide some advantage to the population, by producing individuals able to persist through harsh situations that could also

be encountered in adult-life (i.e., predictive adaptive response, PAR) (Monaghan 2008). Apparent “maladaptive” strategies (e.g., mismatch) can become adaptive with the increase of environmental unpredictability due to climate changes, as such strategies can increase fitness under variable conditions (Kharouba et al. 2018). Thus, measuring fitness components, such as survival and reproductive success, of individuals born under mismatched conditions through different years can bring to light the potential adaptive response of such strategies.

In this study, we tested whether individuals naturally born under mismatched conditions show carryover effects that can reduce fitness (survival and first reproductive years after fledging). We used a time-series of mark-recapture data from a long-lived high latitude species, the King penguin (*Aptenodytes patagonicus*). Two phenological peaks of hatching take place in king penguin colonies every breeding season, generating two groups of chicks, early and late, which are born in match and mismatch with the peak of resources, respectively (Weimerskirch et al. 1992; Descamps et al. 2002).

Late-born chicks are known for having lower survival rates during the first year of life and, more specifically, during the first winter (Weimerskirch et al. 1992; Olsson 1996; Stier et al. 2014). This is because, by hatching later in the season, individuals have less time to grow and accumulate enough energy reserves before winter, a period when food resources become scarce and adults have to forage farther aways from breeding colonies (Bost et al. 2004). More specifically, during the winter, the king penguin’s main foraging ground during the reproductive season, the Antarctic Polar Front (APF), becomes deeper and inaccessible for adults, as it surpasses the diving capacity of the species (Charrassin et al. 1998). As a consequence, chicks fast for long periods and individuals that are not large enough by the beginning of winter have lower chances of surviving (Stier et al. 2014).

Even though the higher pre-fledging mortality of late-born king penguins is well described in literature (Weimerskirch et al. 1992; Olsson 1996; Stier et al. 2014), no previous study has assessed the post-fledging fitness effects on late-born individuals that survive after fledging. Therefore, whether late-born individuals are able to catch-up with the early born individuals in terms of body condition, survival and reproduction after fledging is still an open question. In this context, our main goal was to test whether hatching date had an impact on fitness in this species, in terms of survival in the first years of life and age of first reproductive success. Our hypothesis was that late-born individuals would show similar post-fledging survival and reproductive success in comparison to early-born conspecifics, as most of the mortality and, therefore, selection, happens during the first winter (Saraux et al. 2011).

Additionally, because chick survival in the first year of life is highly dependent on body condition, we expected that late-born individuals that survive until fledgling would do so by catching-up in terms of body size and condition with early-born conspecifics. As late chicks are born around one month after early chicks, we also expected this morphological catch-up to take place due to a later fledging by the late-born chicks, so both groups would have the same time to grow from hatching to fledging. Investigating life history traits underlying a successful breeding out of the peak of resources could help understanding

how species may be able to cope with non-ideal conditions, which will be more common under current climate changes (Kharouba et al. 2018).

2. Material and Methods

2.1. Long-term monitoring

Our study was conducted in the King penguin colony of La Baie du Marin (here referred to as BDM), on Possession Island, Crozet Archipelago (46°24′27″S 51°45′27″E). from a sub-colony of BDM called 'Antavia', which represents a naturally enclosed zone with four passages used by the penguins to exit and enter the breeding area (Gendner et al. 2005). The four passageways of Antavia are equipped by underground systems of paired antennas, which capture and store the entry and exit movements of penguins equipped with Radio Frequency Identification (RFID) tags, as specified in the **General Material and Methods** section **1. Location and life history data**. This movement information allows the remote monitoring of each penguin's main activities during the lifetime, such as return to the colony, breeding attempts, among others. More detailed information about the data gathered by this system can also be found in Bardon et al. *Accepted with minor revisions* (see **Appendix General Material and Methods**).

From 2010 to 2022, per year, ca. 200 early-born (January) and ca. 200 late-born (February) king penguin chicks were captured and marked about 3 weeks after hatching (i.e. chicks between 500g and 1.3 kg maximum). At this stage, chicks are temporarily tagged with a small external plastic pin (Fishtag, Floytag), stamped with a unique number for individual recognition during and after the winter. In order to avoid overlapping between the phenological groups, a period of two weeks separated the captures of the last early chick and the first late chick (early group tagged before the 5th of February; late group tagged after the 17th of February of each year). Chicks that survived after winter were then recaptured on the edge of the colony ca. 2 weeks before fledging, at the end of moulting (i.e., between November and January). During recapture, each chick was measured and implanted with a passive RFID-tag under the skin of their leg. Finally, the Fishtag was removed and individuals released without any other external mark. Annual early fledging groups were defined as chicks that fledged before December 8th of a given year (estimated according to the mean distribution of the departures), while late fledging group were leaving for the first time their natal colony after this date.

2.2. Individual traits: morphological measurements and body condition

Measurements of flipper length (± 1 mm), beak length (± 1 mm), and body mass (kg) were collected at fledging (N=1489 chicks), and for some birds in adulthood (N=75 males). Analyses of adult birds' measurements were performed only on males due to the sample

size of females. Flipper and beak lengths represent good proxies of the penguin's structural size and are known to be highly correlated (Fahlman et al. 2006). We established a Structural Size Index (SSI) using the first component of the principal component analysis between flipper and beak, as previously described in Saraux et al. (2011). The following equation, obtained from measures collected at fledging, was used to calculate the SSI at fledging:

$$SSI_{fledging} = PC1 = 0.30 * Beak + 0.95 * Flipper$$

The SSI equation was also computed on adult male measurements. The following equation was obtained:

$$SSI_{adults} = PC1 = 0.33 * Beak + 0.94 * Flipper$$

Body mass is highly variable in this species, reflecting differences in nutritional status as well as structural size (Saraux et al. 2011). In fact, an individual can have a high mass because it has a large structural size or because it is carrying metabolised energetic reserves in the form of fat or protein (Dobson 1992). When calculating the individual's energy store through its body mass, we must correct for structural body size. Therefore, we used an Ordinary Least Squares (OLS) regression residuals of body mass on structural size to provide a better reflection of the actual energy stores of the animal (Schulte-Hostedde et al. 2005; Saraux et al. 2011; Bordier et al. 2014). Individuals with positive residuals are considered to be in better body condition (BC) and have higher energy storage than individuals with negative residuals (Jakob et al. 1996; Schulte-Hostedde et al. 2001).

2.3. Life history determination

To test early-life effects on pre-fledging traits, we estimated: I) the differential survival of chicks during the first winter on land; II) the differential growth of chicks during the first winter on land; and III) the fledging date of early- and late-hatchlings. Fledging dates were defined as the first detection by the antennas in the two months following the RFID-tagging that was followed by no detection for at least a month. This pattern represents the chick's first departure to the sea after the end of moulting. It allowed us to remove the birds that were never detected (i.e., died between sampling and fledging or RFID tag was nonfunctional) and the birds that would have come back during the summer of their fledging. In the last case, several detections after the potential fledging date generate noise and may confound the real fledging date.

To test early-life effects on post-fledging traits, we evaluated post-fledging return rates, pattern (i.e., dates and age of first returns), and differential growth of early- and late-hatchlings. We also estimated the breeding propensity and the age at first breeding for both groups to investigate how early-life conditions affect the recruitment into the breeding

population. For those means, we used the movement data from monitored penguins from the 'Antavia' system and the reproduction patterns described in **General Material and Methods** section **1. Location and life history data**.

2.5. Statistics

Generalised Linear Models (GLMs) were fitted with Gaussian distribution for fledging and return dates, flipper length, beak length, structural size and body condition. Poisson distributions were used for age at first breeding and age at first successful breeding. Binomial distributions were used for survival, return, and breeding rates. As not all variables were available for all individuals, we computed separate models for different independent variables. Model estimates, standard deviation (Std), and the significance (*P-values* < 0.05) of the explanatory variables were given according to Type-II Anova. All statistics were computed using the R v4.0.3 statistical environment (R Development Core Team, 2022).

3. Results

3.1. Early-life effects on winter survival and growth before fledging

3.1.1. Differential survival of chicks during the first winter of growth on land

Hatching date and year had a significant effect on the survival probability of chicks during the first winter of growth on land (GLM – Binomial: *P-value* < 0.001; **Figure 1; Table 1**). Early-hatched chicks survived better the winter compared to late-hatched chicks, and the winter chick survival probability varied significantly between years, with lower winter mortality after 2015 for both early and late-hatched chicks. The interaction between hatching date and year was also significant (*P-value* < 0.001), meaning that in some years the effect of hatching date was stronger.

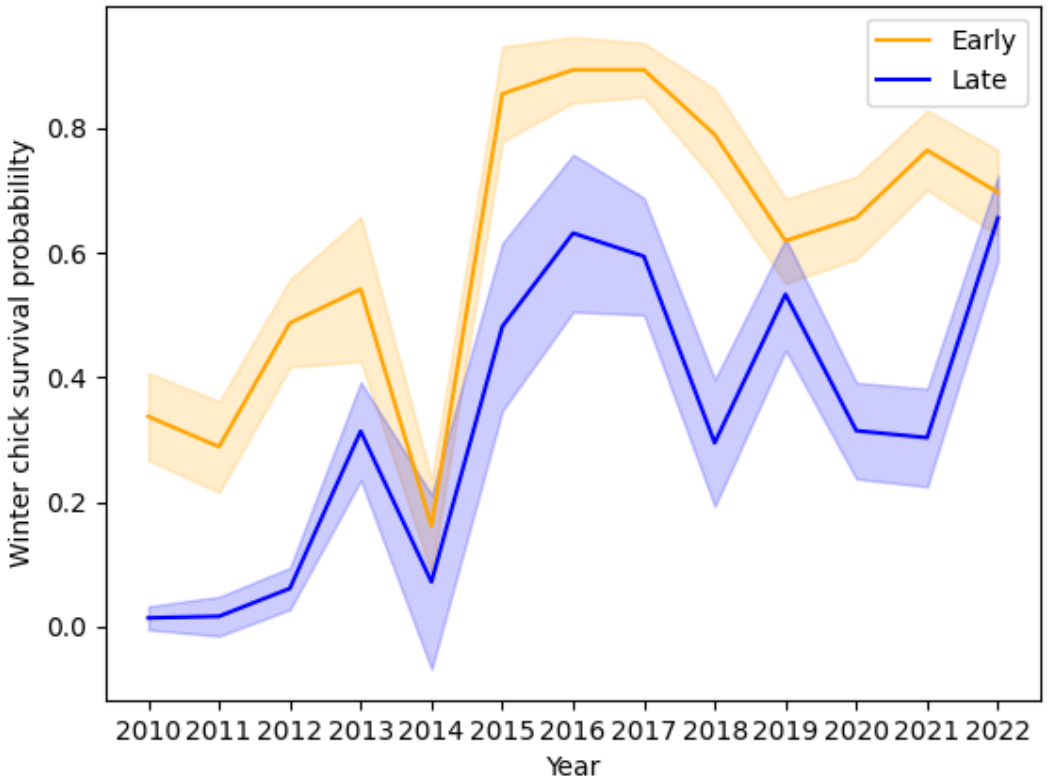


Figure 1. Winter chick survival probability of Early-hatched chicks (in orange) and Late-hatched chicks (in blue) according to the year (cohort). Orange and blue areas around lines represent the 95% confidence intervals.

Table 1. Model estimates, standard deviation (Std), *P-values* for all response variables. The factor Early/Late is reported for level 'Late' in reference to the level 'Early'. The *P-value* corresponds to z value for binomial models and t value for others. Pr(>F) or Pr(>Chisq) gives the significance of the variable according to Type-II Anova. Estimates, standard deviation and *P-values* are not given for categorical variables and their interaction (grey cells). Values in bold indicate variables with significant effect (under 0.05) in the models. SSI refers to Structural Size Index and BC to body condition.

Response variables	Type of model	Explanatory variables	Estimate	Std	*P-value*	Pr(>F) or Pr(>Chisq)
Winter Survival	Binomial	Intercept	-0.431	0.098	<0.001	
		Early/Late	**-1.778**	0.291	<0.001	**<0.001**
		Year				**<0.001**
		Year*Early/Late				**<0.001**
SSI at fledging	Linear	Intercept	-2.1806	1.3783	0.1139	
		Early/Late	**-4.1681**	0.6581	<0.001	**<0.001**
		Year				**<0.001**
BC at fledging	Linear	Intercept	-1.32434	0.1831	<0.001	
		Early/Late	-0.16725	0.0874	0.0560	0.0560
		Year				**<0.001**
Fledging date	Linear	Intercept	77.00	2.826	<0.001	
		Early/Late	**12.50**	8.238	0.1295	**<0.001**
		Year				**<0.001**
		Year*Early/Late				**<0.001**
Post-Fledging return at 2 y-o	Binomial	Intercept	-0.30061	0.4542	0.5081	
		Early/Late	-0.20743	0.1226	0.0907	0.0912
		Year				**0.0333**
		Fledging date	-0.00722	0.0038	0.0604	0.0553
Post-Fledging return at 2 to 4 y-o	Binomial	Intercept	0.2342	0.1723	0.1739	
		Early/Late	4.7845	166.1940	0.9770	0.8935
		Year				**0.0013**
		Year*Early/Late				**0.0146**
Adult flipper length	Linear	Intercept	327.750	5.057	<0.001	
		Early/Late	-6.124	3.109	0.0531	0.0531
		Year				0.3285
Adult beak length	Linear	Intercept	129.2500	2.4372	<0.001	
		Early/Late	-1.4097	1.4985	0.350	0.3503
		Year				0.1107
Adult SSI	Linear	Intercept	1.03	5.2284	0.8442	
		Early/Late	-6.244	3.2147	0.0564	0.0564
		Year				0.24623
Differential size of flipper length (adult - fledging)	Linear	Intercept	10.75	2.3414	<0.001	
		Early/Late	-0.5625	1.4443	0.6982	0.6982
		Year				**0.00224**
Differential size of beak length (adult - fledging)	Linear	Intercept	20.75	2.9196	<0.001	
		Early/Late	1.6097	1.8010	0.3747	0.3747
		Year				**<0.001**
Breeding propensity	Binomial	Intercept	-0.5194	0.1702	0.0023	
		Early/Late	0.1391	0.1143	0.2236	0.2224
		Year				**<0.001**

Age of 1st breeding attempt	Poisson	Intercept	1.437217	0.1119	<0.001	
		Early/Late	0.007026	0.0623	0.910	0.9103
		Year				0.7980
Young breeding success	Binomial	Intercept	-1.10737	0.2040	<0.001	
		Early/Late	-0.12799	0.1364	0.3481	0.3485
		Year				**<0.001**
Age at 1st breeding success	Poisson	Intercept	1.860752	0.1054	<0.001	
		Early/Late	0.049098	0.0757	0.5168	0.5181
		Year				**0.0002**

3.1.2. Differential growth of chicks during the first winter on land

Hatching date and year had a significant effect on the developmental winter growth of the chicks (GLM – Gaussian: *P-value* < 0.001; **Figure 2**; **Table 1**). Overall, the structural size (SSI, flipper and beak lengths) of early-hatched chicks was significantly higher than late-hatched chicks at fledging. Body condition (BC) at fledging was also greater for early-hatched chicks compared to late-hatched chicks, even if it was not strictly significant (*P-value* = 0.056). Structural size and body condition at fledging varied significantly between years, with globally a better overall condition after 2016. However, in this case, most of the SSI and BC data also concentrates after 2016.

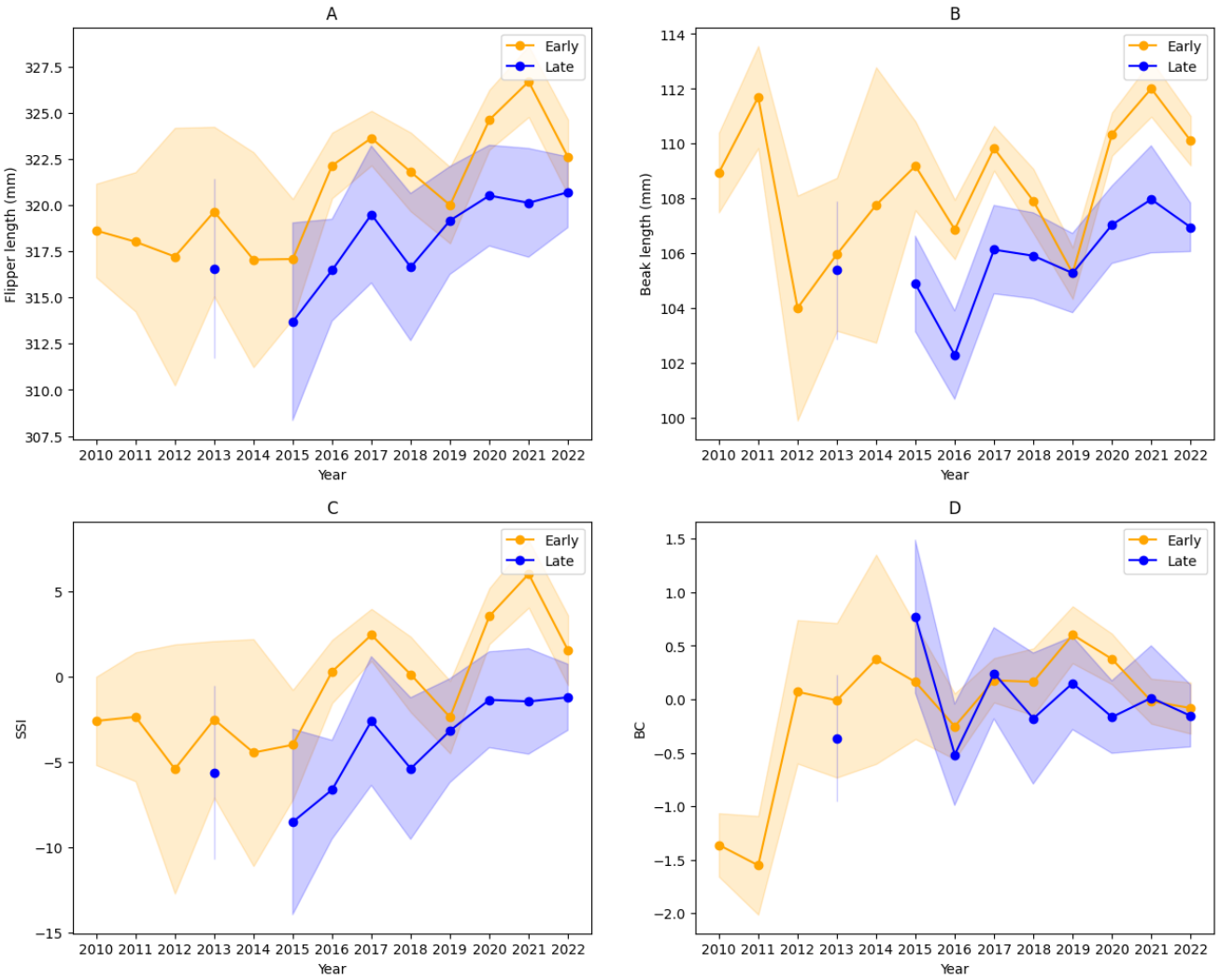


Figure 2. Morphological traits at fledging (at the end of moult) of Early-hatched chicks (in orange) and Late-hatched chicks (in blue) according to the year (cohort). A. Flipper length, B. Beak length, C. Structural Size Index (SSI), and D. Body Condition (BC). Orange and blue areas around lines represent the 95% confidence intervals. Late-hatched chick data is missing from 2010 to 2012 due to the absence of morphological information. Missing data in 2014 is due to the absence of late-survivor chicks at fledging.

3.1.3. Fledging date of Early- and Late-hatchlings

Hatching date and year had a significant effect on the fledging dates of chicks (GLM – Gaussian: *P-value* < 0.001; **Figure 3; Table 1**). Overall, early-hatched chicks left their natal colony earlier than late-hatched chicks (in average 12.5 ± 8.2 days). Fledging date varied significantly between years, with globally an earlier departure after 2015. The interaction between hatching date and year was also significant (*P-value* < 0.001), meaning that in some years early and late-hatched chicks fledged synchronously (e.g., 2016 and 2020).

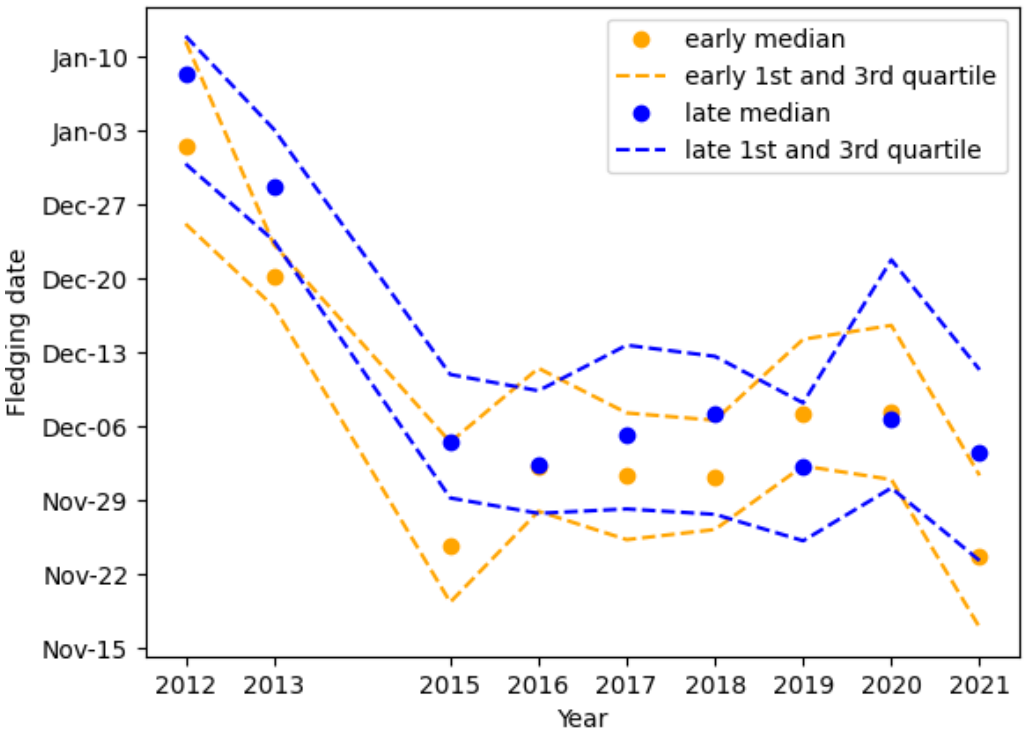


Figure 3. Fledging date of early-hatched chicks (in orange) and late-hatched chicks (in blue) according to the year (cohort). Dashed lines represent the 1st and 3rd quartiles of the distribution. Missing data in 2014 is due to the absence of late-survivor chicks at fledging.

3.2. Early-life effects on post-fledging traits

3.2.1. Post-fledging return rate and pattern of early- and late-hatchlings

While year had an effect on the post-fledging return of juveniles (GLM – Binomial: *P-value* = 0.033; **Figure 4; Table 1**), hatching date appeared to have a slight effect on the post-fledging return rate at age 2 (*P-value* = 0.091), although the trend was not significant, with early-hatched chicks returning earlier to their natal colony compared to late-hatched chicks, especially before 2017 (**Figures 4 and 5**). Post-fledging return rate varied significantly between years, and the interaction between hatching date and year when all age classes were all pooled together was also significant (*P-value* = 0.014). This means that in some years early and late-hatched chicks returned synchronously (i.e., both at age 3 instead of age 2) to their natal colony (e.g., 2017).

We did not observe differences in post-fledging return patterns according to the hatching date of the chicks (**Figure 4**). Overall, juveniles that did not come back to their natal colony before 3 years old, arrived earlier in the summer season compared to the juveniles that came back the year following their first departure, at 2 years of age. Two peaks of return were observed in early-hatched and late-hatched birds coming back one year after fledging, while in the second year, there was one main early return peak (**Figure**

4). Finally, post-fledging return rates one year after fledging of late-hatched chicks that were able to fledge early tended to be similar to early-hatched early-fledging chicks (**Figure 6**).

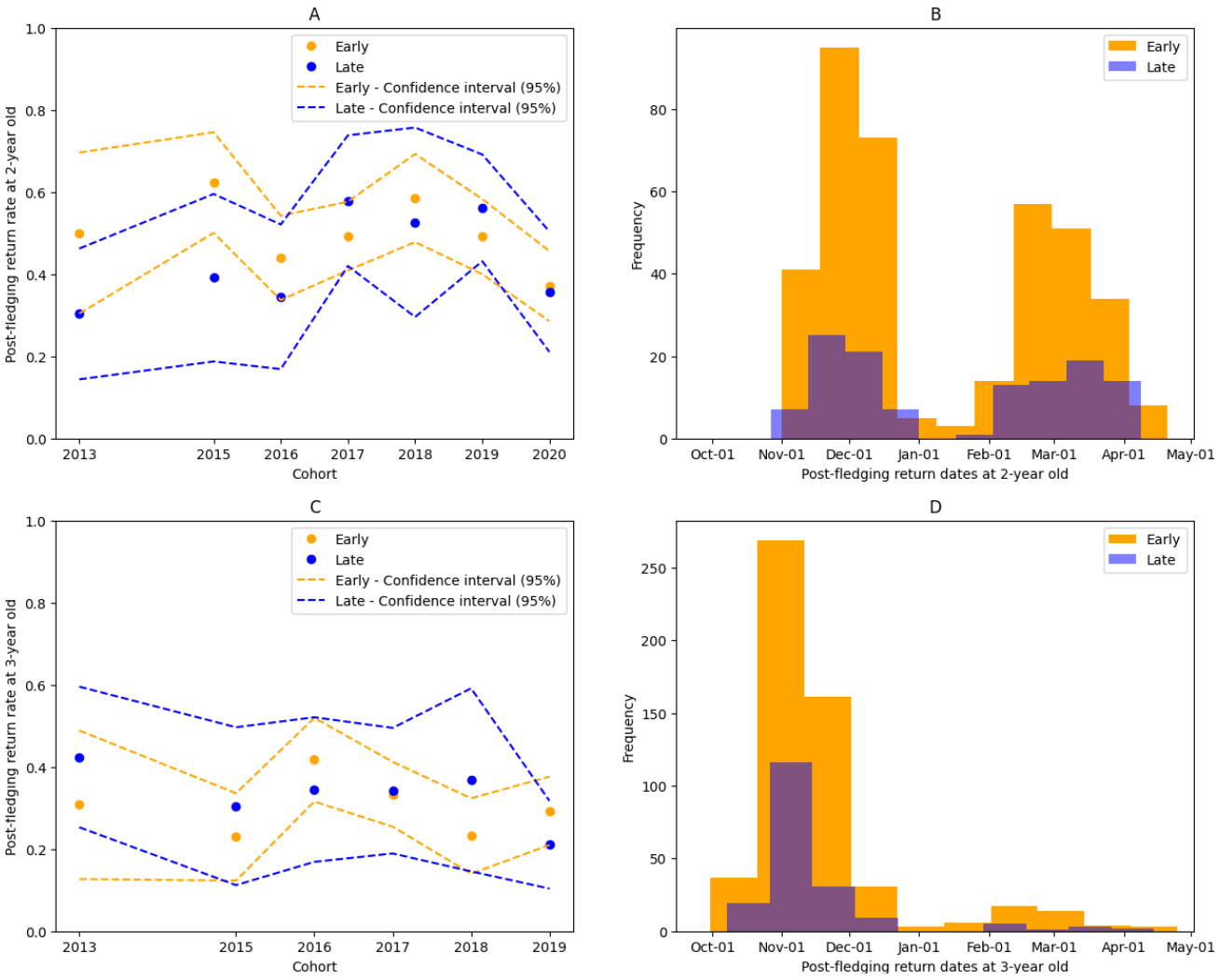


Figure 4. Post-fledging return rates (left panels, A and C) and patterns (right panels, B and D) at 2-years old (top, A and B) and 3-years old (bottom, C and D) of early-hatched chicks (in orange) and late-hatched chicks (in blue).

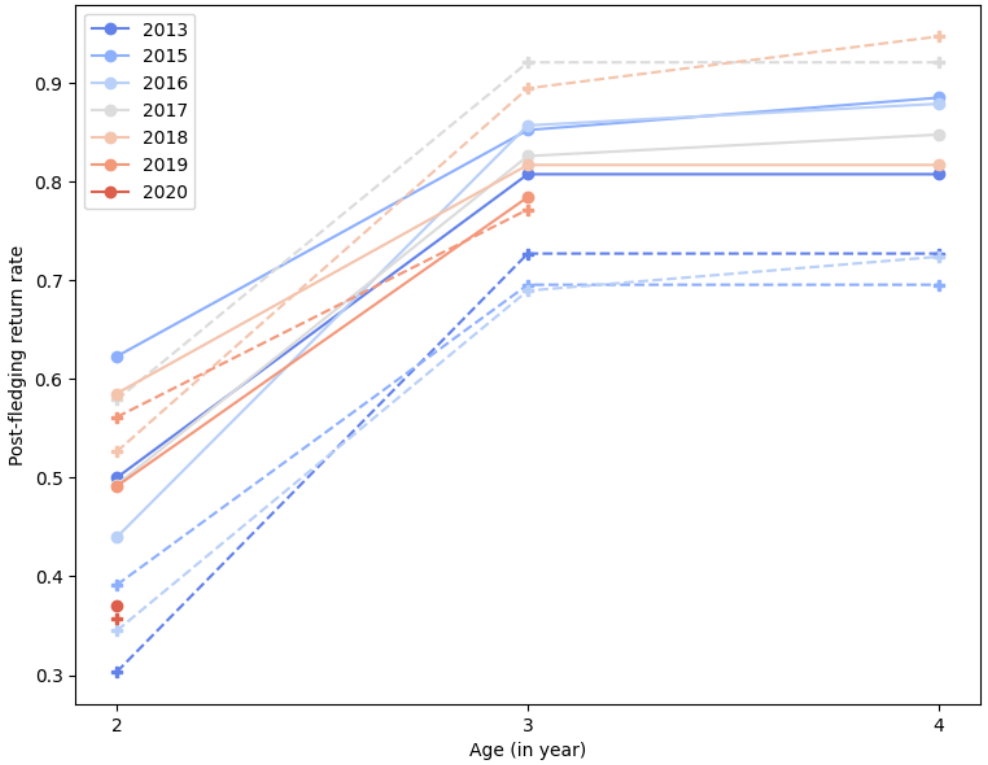


Figure 5. Post-fledging return rates of early-hatched chicks (dots and solid lines) and late-hatched chicks (crosses and dashed lines) at 2, 3, and 4-years old according to the year (cohort).

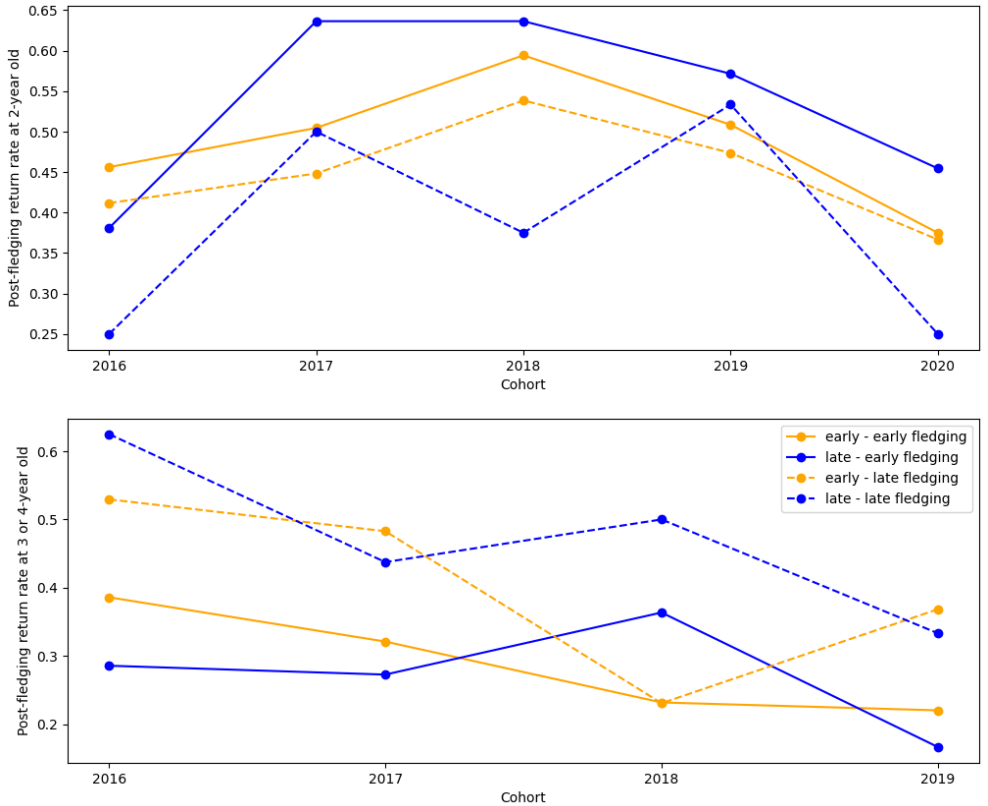


Figure 6. Post-fledging return rates of early-hatched chicks (in orange) and late-hatched chicks (in blue) at 2-years old (top) and 3+4-years old (bottom) according to the year (cohort) and according to the fledging date.

3.2.2. Post-fledging differential growth of Early- and Late-hatchlings of chicks

Hatching date appeared to have a slight effect on some adult morphological traits (GLM – Gaussian: *P-value* = 0.053; **Table 1**; note that this analysis was performed only on males for which the sample size at hatching/fledging and in adulthood was sufficient), although the trend was not significant. Early-hatched individuals had longer flippers (327.75 ± 5.05 mm) at adulthood compared to late-hatched individuals (6.12 ± 3.11 mm smaller). Overall, flipper length in adulthood did not differ between years, while beak length tended to differ according to the cohort, although it was not significant (**Table 1**). When using the SSI (i.e., pooling individual morphological traits), differences between early- and late-born groups were still observed, with a trend of individuals from the early-born group being bigger than those from the late-born group. The differential length of flipper and beak between adult stage and fledging stage did not differ between early-hatched individual and late-hatched individual (**Table 1**).

3.3. Recruitment into the breeding population

3.3.1. Breeding propensity and age at first breeding

Year had a significant effect on the breeding propensity (GLM – Binomial: *P-value* < 0.001; **Figure 7; Table 1**), unlike the hatching date that appeared to have no effect. Age at first breeding attempts did not differ between early and late-hatched chicks (**Figure 8; Table 1**), although the distribution of the age at first breeding attempt tended to be shifted to an earlier attempt.

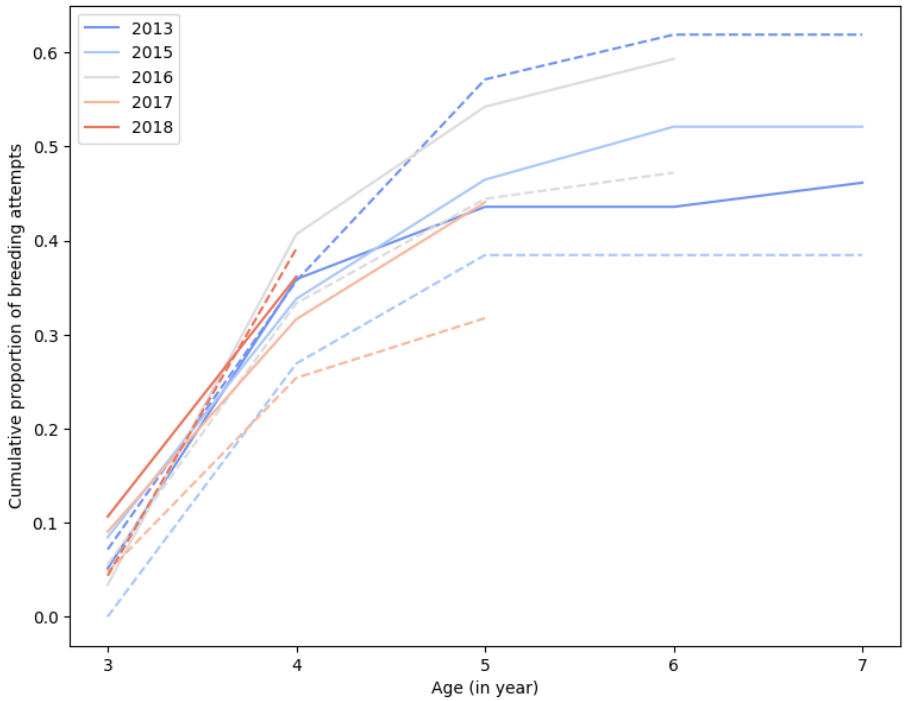


Figure 7. Cumulative proportion of breeding attempts of early-hatched chicks (line) and late-hatched chicks (dash) between 3 and 7-years old according to the year (cohort).

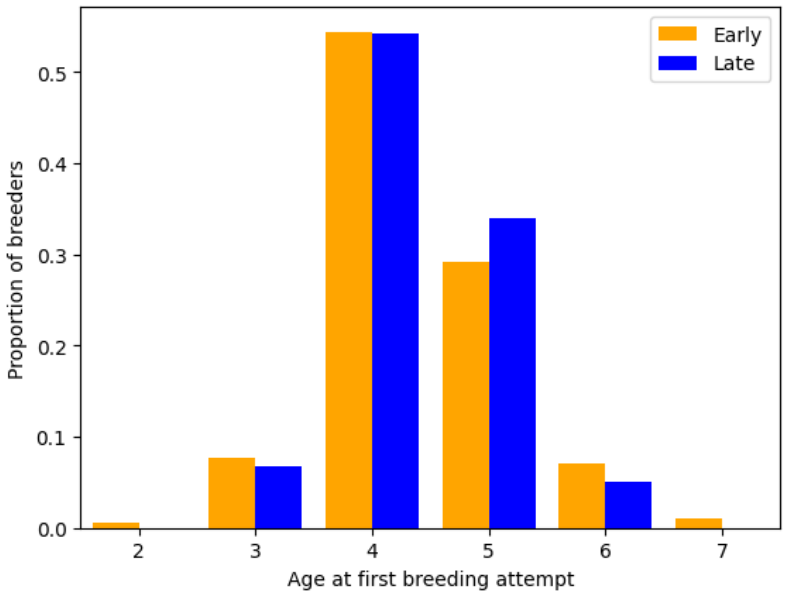


Figure 8. Age at first breeding attempts of early-hatched chicks (orange) and late-hatched chicks (blue) from 7 cohorts (2010 to 2016 combined).

3.3.2. Breeding success

Hatching date had no significant effect on the breeding success of the first breeding attempts (e.g. at age 5 or lower) (GLM – Binomial: *P-value* = 0.348; **Figure 9; Table 1**). Early-hatched chicks tended to perform better earlier in life compared to late-hatched

chicks, although the effect was not significant. Breeding success of young birds and age at first breeding varied significantly between years (GLM – Binomial and Poisson: *P-value* < 0.001; **Figure 9; Table 1**).

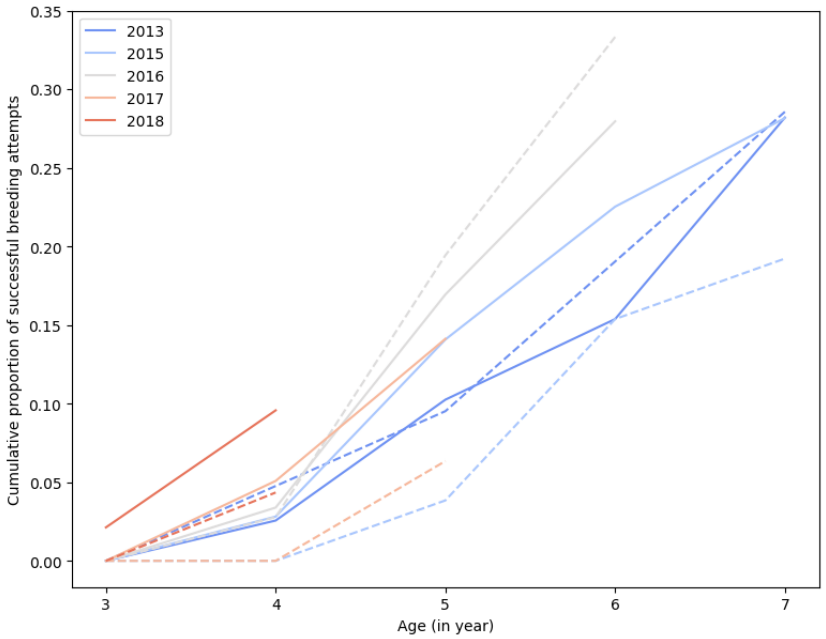


Figure 9. Cumulative proportion of successful breeding attempts of early-hatched chicks (solid line) and late-hatched chicks (dashed line) between 3 and 7-years old according to the year (cohort).

4. Discussion

In this study, we investigated whether individuals born and raised under mismatch with environmental resources showed carryover effects on fitness at fledging and in the first years of adulthood, using a wild population of king penguins as a model. Our expectation was that late-born chicks that survive through the first winter may catch-up with early-born conspecifics at fledging, in terms of minimum body conditions (i.e. energy reserves) and/or size to be able to fledge, leading to potentially similar return rates/patterns and reproductive success in the first years of adulthood between both phenological groups. Overall, our results indicated an equal return probability and reproductive success between early- and late-born individuals in the first years after fledging, even without a complete morphological catch-up by the late-born group. Even with smaller body size, late-hatchlings fledged with similar body conditions as early-hatchlings. We also observed a strong interannual variability both in pre-fledging survival and post-fledging return, as well as in the recruitment of breeders in the population.

4.1. Interannual variability in pre-fledging survival and strong winter selection on late-born chicks

Our results indicated a higher probability of winter survival in early-born chicks compared to late-born chicks, while this pattern was stronger in some years than others. Overall, the average survival rate from 2010 to 2014 was lower for both early and late groups in comparison to the period of 2015 to 2022. Higher late chick mortality before fledging has already been described in literature (vanHeezik et al. 1993; Stier et al. 2014), but our results indicated that the difference in winter mortality between phenological groups was strongly subjected to interannual variability. Our results supported that late-born individuals were subjected to higher selective pressure than early-born, but also that such selective pressure was stronger for both groups depending on the year.

In the investigated time-window, both early- and late-born chicks fledged later in the years preceding 2015, with mean fledging dates around late November/early December after 2015 compared to the previous late December/early January. We also found that 2012 and 2013 were characterised by lower winter survival rates when compared to more recent years, specially after 2016. Both lower winter survival rates and later fledging dates could reflect poor environmental conditions (e.g., lower food availability), pushing parents to make longer foraging trips (Bost et al. 2015) during those years.

King penguin reproductive success is known to be influenced by climate variability, being negatively affected by abnormally warmer ocean temperatures, a.k.a. negative southern oscillation index (SOI) (Le Bohec et al. 2008; Bost et al. 2015). Previous studies have shown that during years of negative SOI, species' foraging grounds appeared to be pushed southwards, reducing in turn the survival of chicks (Le Bohec et al. 2008; Pascoe et

al. 2022) and adults from Crozet and Macquarie islands (Le Bohec et al. 2008). However, our results showed a higher chick survival probability during a period of warmer southern ocean temperatures (negative SOI), from 2016 to 2022 (Zhang et al. 2022), compared to 2012/2013. Although such global climatic oscillations, as well as local sea surface temperature (SST), are considered as good proxies of the APF position, the direct impact of such climatic events on myctophid fish availability and abundance is still unknown (Bost et al. 2015). In fact, recent studies have shown patterns of population increase in other king penguin colonies despite warming events (Foley et al. 2018; Brisson-Curadeau et al. 2023).

The first case concerns colonies located south of the APF, in South Georgia island on the Atlantic sector of the Southern Ocean. These colonies are expected to be less affected by the predicted southward shift of the APF (Cristofari et al. 2018). Moreover, it has been suggested that king penguin population increase in South Georgia since 1883 might be the outcome of the disruption of the complex trophic relationship between king penguins, its main prey during summer (the myctophid fish, Olsson and North 1997; Raclot et al. 1998), the Antarctic krill, and other kill predators. The decline of large krill predators, such as whales and seals, due to intensive hunting during the 20th century (Laws 1973; Branch et al. 2004), could have led to a regional increase in myctophid in the last decades, favouring king penguin foraging in the region (Foley et al. 2018).

The second study compared the effect of warmer sea surface temperatures on the King penguin breeding success and chick mortality in our studied colony in the Crozet archipelago, in contrast to a close-by colony also located in the Indian Ocean sector in the Kerguelen islands (Ratmanoff's colony, 49°14′33″S, 70°33′40″E) (Brisson-Curadeau et al. 2023). Regardless of the proximity between the two colonies (ca. 1500 Km of distance, BirdLife International), king penguins from Kerguelen are more closely located to the APF, as the Kerguelen-Heard Plateau forms a southern barrier, which prevents the front from moving polewards (Park et al. 2014). Consequently, the authors showed that years of warmer temperatures had a positive effect on Kerguelen's chick survival in contrast to what is observed in Crozet (Le Bohec et al. 2008; Bost et al. 2015). Warmer years did not seem to affect foraging distances of adults from Kerguelen, and warmer winter air temperatures likely reduced the thermoregulation cost of chicks, resulting in better survival (Brisson-Curadeau et al. 2023). Interestingly, survival curves from Kerguelen chicks during the studied period showed a similar pattern compared to our study (see Figure 5 from Brisson-Curadeau et al. 2023 and our study's **Figure 1**).

We therefore suggest that the higher chick survival observed in Crozet after 2015 may be a short-term response to warmer winters. Indeed, warmer air temperatures during the winter could limit the costs associated with offspring thermoregulation during winter, as it was suggested by Brisson-Curadeau et al. (2023) for Kerguelen chicks. Adults from Crozet may be able to cope, to some extent, with the higher foraging efforts of swimming farther to the APF during warmer summer seasons in the short-term, as it has already been seen in extreme climate dipole events (Bost et al. 2015). However, the long-term effect of such higher foraging efforts during the summer season but also potentially during the winter, may have negative consequences in adult survival, which can lead to population's decline (Le Bohec et al. 2008).

Another non-exclusive hypothesis is that warmer years could lead to a shift in summer diet to other types of prey that can be found outside the APF, such as squids (*Moroteuthis ingens*), which are also known to be in the species' diet during winter (Cherel et al. 1996). Moreover, to feed their chick during winter, king penguins forage predominantly on small- and medium-sized juvenile squids that spawned in the previous spring (Cherel and Weimerskirch 1999) in the outer shelf, upper slope, and oceanic areas in the close vicinity of Crozet Islands (Cherel et al. 1993). Some studies demonstrated a strong influence of environmental conditions on squid growth and maturation, with individuals hatched in warmer temperatures growing faster and maturing earlier compared to those hatched in colder waters (Forsythe 2004; Pecl and Jackson 2008; Pang et al. 2022). Therefore, warmer sea temperatures may increase food availability close by the colony to feed the chicks. In combination with lower thermoregulatory costs, this could have a positive effect on chick survival in the short-term. However, the lower energetic value of squid compared to myctophids (Raclot et al. 1998) could impact the supply efficiency of adults (for their own energetic needs and those of their chicks) and increase the rearing costs of chicks, which could have, over a longer time, consequences on the survival of adults.

Our next step is now to select and incorporate local and global on-land and at-sea environmental variables, such as the ambient or sea surface temperature (SST), the SOI or the Southern annular mode (SAM), or even the winter position of the Antarctic marginal ice zone (MIZ), which corresponds to an expected foraging area for adults during winter (Bost et al. 2004). In addition, analyses of stable isotopes on archive samples will help us to investigate a possible annual shift in summer and winter diets.

4.2. Late-born chicks fledge smaller, but with equal body conditions as early-born chicks

Late-hatchlings fledged later than early-born ones, with smaller body size but similar body condition. Although late-born chicks fledged later, this did not completely offset the delay compared to the early ones. Hence, the shorter time-window to grow, especially before the winter fasting and from the end of winter until fledging, may explain their smaller body size. However, late-born fledging chicks appeared to be able to grow to a sufficient threshold that allowed them to survive the winter, and to reach the same body condition as early-hatchlings through energy accumulation over the last growing period. We suggest that this might be the result of plastic adaptations of late-hatchlings, for example, via the upregulation of genes related to efficient energy accumulation (e.g., USP7 and MXD4), as shown in Fernandes et al. (*in prep.* - Chapter 2). Another study has shown physiological responses at hatching of late chicks that survive through winter, such as high corticosterone levels and higher body mass at 10 days of life (Stier et al. 2014). The same study showed that high body mass 10 days after hatching was a strong predictor of survival until fledging in both early and late groups.

Overall, the fact that late-born chicks fledged with similar energy reserves but smaller body sizes than early-born chicks underlines that energy reserve accumulation a couple of weeks before fledging and up to an optimal body mass (see Jebb et al. 2021) is the critical component to fledge. It also points out that producing smaller chicks for late breeders should be an advantage in several ways: it is faster to produce (while the match with resources availability is almost over), it is less costly to raise (i.e. lower energy demand) under harsh environmental conditions (like the harsh austral winter), and, as the reproduction is likely to fail, for the parents that may face poorer condition/nutritional stress (linked to the costs of previous successful breeding), producing a less costly smaller chick may avoid impairing future fitness (i.e. the 'cost of reproduction hypothesis', Williams 1966). This assumption is supported by the sex-biassed allocation strategy depending on yearly environmental conditions that has been observed in king penguins (Bordier et al. 2014). The higher costs of producing the larger sex (i.e. male in king penguins) appeared to favour the production bias towards the smaller one (i.e. female) under harsher conditions yearly, but also seasonally as the late-hatched group was female-bias while the early one was male-bias (Figure 2 in Bordier et al. 2014). Individual sex information is still unavailable in our study, but the next step will also be to validate this seasonally sex-bias over several years

4.3. Return rates do not depend on hatching date, but vary among years

King penguin juveniles spend from one to three years out of their birth colony after fledging, when around ¾ of juveniles return to the natal colony (Saraux et al. 2011). In spite of fledging at smaller sizes, late-born individuals did not show differences in return rates in the following years compared to early-born individuals. These results indicate that, independently of catching-up in terms of size, individuals born in mismatch with resources were equally likely to return to natal colonies in the first years of adulthood in this species. This result indicates that the harsh conditions to which late survivors were exposed in their first year of growth on land may trigger responses that can be beneficial as adults. Thus, late-hatchling traits can be considered as a predictive adaptive response (PAR), instead of carryover effects in their early years of life.

Depending on the location of their breeding colony, king penguin adults have to swim for hundreds of kilometres to reach the main foraging grounds (Bost et al. 1997; Bost et al. 2015), which is also the case for juveniles and non-breeders (Orgeret et al. 2019). A proportion of juveniles do not survive until recruitment in the population, probably because those individuals are less efficient in swimming and accumulating energy reserves than the ones that survive and breed regularly (Saraux et al. 2011). In addition, during the yearly moulting and reproduction (more specifically during incubation and brooding), individuals have to fast for weeks while waiting for the new feathers to grow (Cherel et al. 1994; Gauthier-Clerc et al. 2002) or for the partner to return from the sea (Weimerskirch et al. 1992; Groscolas and Robin 2001), respectively. This means that successful adults have to

manage their energy reserves properly in order not to die of starvation or hypothermia, or abandon the egg/chick during breeding. We suggest that the ability to efficiently accumulate energy reserves before the first winter, prepare the late-born chicks to face physiological challenges in adulthood (such as fasting in the colony during breeding and moulting), as a PAR.

Our results showed a highly interannual variability in the juvenile return rate, which is most likely related to the highly variable environmental conditions/food availability between years. This result was consistent with those found previously, which indicate that interannual variability of the return rate (from 68% to 87% depending on the year, Saraux et al. 2011) was related to body condition, but not to structural size. In contrast to what is observed in king penguin adults (Le Bohec et al. 2008), this study showed that juveniles returned at higher rates in warmer years, due to the fact that juveniles forage in more subtropical areas while adults are foraging in the APF (Saraux et al. 2011).

Our next step is to incorporate into our models on-land and at-sea environmental variables experienced by juveniles during their first year of growth on land and their first years at sea before returning to their colony to better understand the early-life selective pressures in this species. As already mentioned earlier, climate change may not show detrimental effects at the population level in the short-term. However, because King penguin foraging grounds are projected to be affected by warming (Le Bohec et al. 2008; Bost et al. 2015; Cristofari et al. 2018), lagged and long-term effects on population dynamics are expected in the near future (Le Bohec et al. 2008).

4.4. Returning dates and recruitment in the population

King penguin juveniles are also known to show individual variation in the age of first return to the colony (i.e., returning date) (Saraux et al. 2011). Our results showed a higher proportion of individuals coming back at 3 and 4 years old (from 66% to 95% of return) than at 2 years old (from 30% to 61%), in accordance with Saraux et al. (2011). Early-hatchlings tended to return at a younger age compared to late-hatchlings, even if this trend was not statistically significant. This trend may be blurred by the fact that not all individuals that hatch early will fledge early, and vice-versa. Indeed, we observed some late-hatchlings that fledged early in the season, having an extremely short period of development, while some early-hatchlings fledged late, as their parents did not succeed to effectively rear them in a timely window. How this rapid developmental rate will affect the fitness of late-hatchlings, and if this is a result of plastic adaptations to faster growth and body mass accumulation or driven exclusively by parental feeding efficiency are still open questions.

Independently on the hatching date, we found that chicks that fledged later in the season came back later to their natal colony (i.e., at 3 years of age), while early-fledglings returned a year younger. The pattern of late-fledglings returning at older ages has also been previously observed in the studied colony (Saraux et al. 2011). Juvenile activity at sea

has only started to be explored in recent years. For instance, a study of 2-year old king penguins showed evidence of exploratory foraging behaviour throughout a wide area to the western side of the APF to the Antarctic pack ice (Orgeret et al. 2019). A more recent study has shown that juvenile's foraging performance in the first months at sea, in terms of diving depth, is lower than adults (Enstipp et al. 2021). Thus, juveniles pass through an acclimation and training period, which likely facilitates their ability of swimming in cold waters and diving deeper in adulthood. Moreover, penguins leave the colony for the first time with a low body condition after moulting, meaning that the ability of efficient fat accumulation is essential to stand the cold polar waters (Enstipp et al. 2019). We therefore suggest that the potentially poorer conditions of the late-fledglings at the beginning of their first winter at sea (as they had less time to replenish their reserves before winter since they left the colony later) may impede them in terms of foraging efficiency during the winter, resulting in a delayed return to the colony.

As in Saraux et al. 2011, we also observed two peaks of arrival to the colony at 2 years old, while a single peak was characteristic of the return at 3 years old. The bimodal arrival at 2 years old also happened later in the summer season (December and March) when compared to the unique earlier peak of arrival at 3 years old (November). We still do not have a proper explanation to these patterns, apart from that the colony might be too crowded in December/January so that the juveniles would stay out of the breeding colony to avoid the aggressiveness of the breeders.

Finally, our results showed that some early-born individuals attempted to breed at 2 years old, while the majority of the birds (early- or late-born) attempted to breed for the first time at 4 and 5 years of age. The mean age at first breeding of king penguins has been estimated at 6 years old (Weimerskirch et al. 1992), and recent work on unbanded birds has shown that they attempt to breed, even successfully, at a younger age, i.e., average age of 5 years (Le Bohec 2007; Kriesell et al. 2021; this study). In our study, early-born individuals tended to achieve a first successful breeding attempt at a younger age than late-born chicks, although the effect was not significant. Our long-term monitoring of these early- and late-hatched chicks started 10 years ago for a species that can live up to 30 years in the wild (Gauthier–Clerc et al. 2004), thus the long-term effects of early-life success in reproduction can still not be estimated from our dataset. However, numerous studies on long-lived birds have shown that breeding success can improve with experience (e.g., Lewis et al. 2006; Limmer and Becker 2009; Limmer and Becker 2010; Zhang et al. 2015). Therefore, attempting to breed at a young age to gain experience (learning the best place to reproduce and/or the courtship behaviours, or meeting potential future partners) may result in an earlier success. Their delayed secondary sexual characteristics (Nicolaus et al. 2007; Kriesell et al. 2021), such plumage and morphological differences, signalling a noncompetitive status to adults, would lead to lower aggression from adults, while learning.

There may be a possible trade-off between the costs of attempting a potentially unsuccessful early breeding attempt and the benefits of gaining experience. However, a high investment in reproduction early in life could be balanced by associated costs, such as reducing the probability of survival (Clutton-Brock 1984), accelerating senescence in later

life (Nussey et al. 2006; Spagopoulou et al. 2020), and lifetime fitness (Spagopoulou et al. 2020). Thus, only high quality individuals may have the ability to adopt this strategy (Fay et al. 2016). In our case, we expect that early- and late-hatched chicks that reach the condition to fledge early in the season (a proxy of good body condition before the next winter), would have this capacity, with late-hatched chick potentially better equipped to face the environmental threats of a changing world (i.e. PAR). In a longer timeframe, we will investigate whether these birds have a better lifetime reproductive success (LRS, Clutton-Brock 1988).

References

BirdLife International (2023) Endemic Bird Areas factsheet: Kerguelen and Crozet islands. Downloaded from http://www.birdlife.org on 09/05/2023.

Blomberg, Erik J., James S. Sedinger, Daniel Gibson, Peter S. Coates, and Michael L. Casazza. 2014. "Carryover Effects and Climatic Conditions Influence the Postfledging Survival of Greater Sage-Grouse." Ecology and Evolution 4 (23): 4488–99.

Bohec, Céline Le, Joël M. Durant, Michel Gauthier-Clerc, Nils C. Stenseth, Young-Hyang Park, Roger Pradel, David Grémillet, Jean-Paul Gendner, and Yvon Le Maho. 2008. "King Penguin Population Threatened by Southern Ocean Warming." Proceedings of the National Academy of Sciences 105 (7): 2493–97.

Bordier, Célia, Claire Saraux, Vincent A. Viblanc, Hélène Gachot-Neveu, Magali Beaugey, Yvon Le Maho, and Céline Le Bohec. 2014. "Inter-Annual Variability of Fledgling Sex Ratio in King Penguins." PloS One 9 (12): e114052.

Bost, C. A., J. B. Charrassin, Y. Clerquin, Y. Ropert-Coudert, and Y. Le Maho. 2004. "Exploitation of Distant Marginal Ice Zones by King Penguins during Winter." *Marine Ecology Progress Series* 283: 293–97.

Bost, C. A., J. Y. Georges, C. Guinet, Y. Cherel, K. Pütz, J. B. Charrassin, Y. Handrich, T. Zorn, J. Lage, and Y. Le Maho. 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.

Bost, Charles A., Cedric Cotté, Pascal Terray, Christophe Barbraud, Cécile Bon, Karine Delord, Olivier Gimenez, et al. 2015. "Large-Scale Climatic Anomalies Affect Marine Predator Foraging Behaviour and Demography." *Nature Communications* 6 (October): 8220.

Branch, Trevor A., Koji Matsuoka, and Tomio Miyashita. 2004. "Evidence for Increases in Antarctic Blue Whales Based on Bayesian Modelling." *Marine Mammal Science* 20 (4): 726–54.

Brisson-Curadeau, Émile, Kyle Elliott, and Charles-André Bost. 2023. "Contrasting Bottom-up Effects of Warming Ocean on Two King Penguin Populations." *Global Change Biology* 29 (4): 998–1008.

Burton, Tim, and Neil B. Metcalfe. 2014. "Can Environmental Conditions Experienced in Early Life Influence Future Generations?" *Proceedings. Biological Sciences / The Royal Society* 281 (1785): 20140311.

Charrassin, J-B, C. A. Bost, K. Pütz, J. Lage, T. Dahier, T. Zorn, and Y. Le Maho. 1998. "Foraging Strategies of Incubating and Brooding King Penguins Aptenodytes Patagonicus." *Oecologia* 114 (2): 194–201.

Cherel, Y., J. B. Charrassin, and E. Challet. 1994. "Energy and Protein Requirements for Molt in the King Penguin Aptenodytes Patagonicus." *The American Journal of Physiology* 266 (4 Pt 2): R1182–88.

Cherel, Y., V. Ridoux, and P. G. Rodhouse. 1996. "Fish and Squid in the Diet of King Penguin Chicks, Aptenodytes Patagonicus, during Winter at Sub-Antarctic Crozet Islands." *Marine Biology* 126 (4): 559–70.

Cherel, Y., C. Verdon, and V. Ridoux. 1993. "Seasonal Importance of Oceanic Myctophids in King Penguin Diet at Crozet Islands." *Polar Biology* 13 (5). https://doi.org/10.1007/bf00238362.

Cherel, Y., and H. Weimerskirch. 1999. "Spawning Cycle of Onychoteuthid Squids in the Southern Indian Ocean:new Information from Seabird Predators." *Marine Ecology Progress Series* 188: 93–104.

Clutton-Brock, T. H. 1984. "Reproductive Effort and Terminal Investment in Iteroparous Animals." *The*

American Naturalist 123 (2): 212–29.

Clutton-Brock, 1988. “Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems,” June. https://books.google.it/books?hl=en&lr=&id=ovzZjx84sccC&oi=fnd&pg=PP11&ots=zaZFZ1U4NL&sig=uwsxtkrmENZzAMJS0lz6_qtINLo&redir_esc=y.

Crespi, B. J. 2000. “The Evolution of Maladaptation.” *Heredity* 84 (Pt 6) (June): 623–29.

Cristofari, Robin, Xiaoming Liu, Francesco Bonadonna, Yves Cherel, Pierre Pistorius, Yvon Le Maho, Virginie Raybaud, Nils Christian Stenseth, Céline Le Bohec, and Emiliano Trucchi. 2018. “Climate-Driven Range Shifts of the King Penguin in a Fragmented Ecosystem.” *Nature Climate Change* 8 (3): 245–51.

Cushing, D.H., 1974. "The possible density-dependence of larval mortality and adult mortality in fishes." The Early Life History of Fish” JHS, 103-111.

Cushing, D. H. 1990. “Plankton Production and Year-Class Strength in Fish Populations: An Update of the Match/Mismatch Hypothesis.” In *Advances in Marine Biology*, edited by J. H. S. Blaxter and A. J. Southward, 26:249–93. Academic Press.

Cushing, J. M., and M. Saleem. 1982. “A Predator Prey Model with Age Structure.” *Journal of Mathematical Biology* 14 (2): 231–50.

Descamps, Sebastien, M. Gauthier-Clerc, J. P. Gendner, and Y. Le Maho. 2002. “The Annual Cycle of Unbanded King Penguins Aptenodytes Patagonicus on Possession Island (Crozet)” 2 (January): 87–98.

Dobson, F. S. 1992. “Body Mass, Structural Size, and Life-History Patterns of the Columbian Ground Squirrel.” *The American Naturalist* 140 (1): 109–25.

Doiron, Madeleine, Gilles Gauthier, and Esther Lévesque. 2015. “Trophic Mismatch and Its Effects on the Growth of Young in an Arctic Herbivore.” *Global Change Biology* 21 (12): 4364–76.

Durant, Joël M., Dag Ø. Hjermann, Tycho Anker-Nilssen, Grégory Beaugrand, Atle Mysterud, Nathalie Pettorelli, and Nils Chr Stenseth. 2005. “Timing and Abundance as Key Mechanisms Affecting Trophic Interactions in Variable Environments.” *Ecology Letters* 8 (9): 952–58.

Enstipp, Manfred R., Charles-André Bost, Céline Le Bohec, Caroline Bost, Robin Laesser, Yvon Le Maho, Henri Weimerskirch, and Yves Handrich. 2019. “The Dive Performance of Immature King Penguins Following Their Annual Molt Suggests Physiological Constraints.” *The Journal of Experimental Biology* 222 (Pt 20). https://doi.org/10.1242/jeb.208900.

Enstipp, Manfred R., Charles-André Bost, Céline Le Bohec, Nicolas Chatelain, Henri Weimerskirch, and Yves Handrich. 2021. “The Early Life of King Penguins: Ontogeny of Dive Capacity and Foraging Behaviour in an Expert Diver.” *The Journal of Experimental Biology* 224 (12). https://doi.org/10.1242/jeb.242512.

Fahlman, A., Halsey, L.G., Butler, P.J., Jones, D.R., Schmidt, A., Durand, S., Froget, G., Bost, C.A., Woakes, A.J., Duchamp, C. and Handrich, Y., 2006. Accounting for body condition improves allometric estimates of resting metabolic rates in fasting king penguins, Aptenodytes patagonicus. Polar Biology, 29, pp.609-614.

Fay, R., Barbraud, C., Delord, K. and Weimerskirch, H., 2016. Variation in the age of first reproduction: different strategies or individual quality?. Ecology, 97(7), pp.1842-1851.

Foley, C. M., T. Hart, and H. J. Lynch. 2018. “King Penguin Populations Increase on South Georgia but Explanations Remain Elusive.” *Polar Biology* 41 (6): 1111–22.

Forsythe, J. W. 2004. “Accounting for the Effect of Temperature on Squid Growth in Nature: From Hypothesis to Practice.” *Marine and Freshwater Research* 55 (4): 331–39.

Gauthier–Clerc, M., J-P Gendner, C. A. Ribic, W. R. Fraser, E. J. Woehler, S. Descamps, C. Gilly, C. Le Bohec, and Y. Le Maho. 2004. “Long–term Effects of Flipper Bands on Penguins.” *Proceedings*

of the Royal Society of London. Series B: Biological Sciences 271 (suppl_6): S423–26.

Gauthier-Clerc, M., Y. Le Maho, J-P Gendner, and Y. Handrich. 2002. “Moulting Fast and Time Constraint for Reproduction in the King Penguin.” *Polar Biology* 25 (4): 288–95.

Gendner, Jean-Paul, Michel Gauthier-Clerc, Celine Le Bohec, Sebastien Descamps, and Yvon Le Maho. 2005. “A New Application for Transponders in Studying Penguins.” *Journal of Field Ornithology* 76 (2): 138–42.

Grafen, Alan. 1988. "On the uses of data on lifetime reproductive success." Reproductive success 454-485.

Groscolas, R., and J. P. Robin. 2001. “Long-Term Fasting and Re-Feeding in Penguins.” *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 128 (3): 645–55.

Hamel, Sandra, Jean-Michel Gaillard, Marco Festa-Bianchet, and Steeve D. Côté. 2009. “Individual Quality, Early-Life Conditions, and Reproductive Success in Contrasted Populations of Large Herbivores.” *Ecology* 90 (7): 1981–95.

Hayward, Adam D., Ian J. Rickard, and Virpi Lummaa. 2013. “Influence of Early-Life Nutrition on Mortality and Reproductive Success during a Subsequent Famine in a Preindustrial Population.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (34): 13886–91.

Hettinger, Annaliese, Eric Sanford, Tessa M. Hill, Ann D. Russell, Kirk N. S. Sato, Jennifer Hoey, Margaux Forsch, Heather N. Page, and Brian Gaylord. 2012. “Persistent Carry-over Effects of Planktonic Exposure to Ocean Acidification in the Olympia Oyster.” *Ecology* 93 (12): 2758–68.

Jakob, Elizabeth M., Samuel D. Marshall, and George W. Uetz. 1996. “Estimating Fitness: A Comparison of Body Condition Indices.” *Oikos* 77 (1): 61–67.

Jebb, Alexandra H. M., Daniel T. Blumstein, Pierre Bize, and Julien G. A. Martin. 2021. “Bigger Is Not Always Better: Viability Selection on Body Mass Varies across Life Stages in a Hibernating Mammal.” *Ecology and Evolution* 11 (7): 3435–45.

Kharouba, Heather M., Johan Ehrlén, Andrew Gelman, Kjell Bolmgren, Jenica M. Allen, Steve E. Travers, and Elizabeth M. Wolkovich. 2018. “Global Shifts in the Phenological Synchrony of Species Interactions over Recent Decades.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (20): 5211–16.

Kharouba, Heather M., and Elizabeth M. Wolkovich. 2020. “Disconnects between Ecological Theory and Data in Phenological Mismatch Research.” *Nature Climate Change* 10 (5): 406–15.

Kriesell, Hannah Joy, Thierry Aubin, Víctor Planas-Bielsa, Quentin Schull, Francesco Bonadonna, Clément Cornec, Yvon Le Maho, Laura Troudet, and Céline Le Bohec. 2021. “How King Penguins Advertise Their Sexual Maturity.” *Animal Behaviour* 177 (July): 253–67.

Laws, R.M., 1973. Population increase of fur seals at South Georgia. Polar Record, 16(105), pp.856-858.

Le Bohec, Céline. 2007. “Stratégies D’histoire De Vie D’un Oiseau Longévif: Le Manchot Royal (Aptenodytes patagonicus)” PhD thesis. University of Strasbourg.

Lewis, Sue, Sarah Wanless, David A. Elston, Meg Duhr Schultz, Elizabeth Mackley, Michelle Du Toit, Jenny G. Underhill, and Mike P. Harris. 2006. “Determinants of Quality in a Long-Lived Colonial Species.” *The Journal of Animal Ecology* 75 (6): 1304–12.

Limmer, Bente and Peter H. Becker. 2009. “Improvement in Chick Provisioning with Parental Experience in a Seabird.” *Animal Behaviour* 77 (5): 1095–1101.

Limmer, Bente and Peter H. Becker. 2010. “Improvement of Reproductive Performance with Age and Breeding Experience Depends on Recruitment Age in a Long-Lived Seabird.” *Oikos* 119 (3): 500–507.

Millon, Alexandre, Steve J. Petty, Brian Little, and Xavier Lambin. 2011. “Natal Conditions Alter

Age-Specific Reproduction but Not Survival or Senescence in a Long-Lived Bird of Prey." *The Journal of Animal Ecology* 80 (5): 968–75.

Monaghan, Pat. 2008. "Early Growth Conditions, Phenotypic Development and Environmental Change." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1497): 1635–45.

Mugabo, M., O. Marquis, S. Perret, and J. F. LE Galliard. 2010. "Immediate and Delayed Life History Effects Caused by Food Deprivation Early in Life in a Short-Lived Lizard." *Journal of Evolutionary Biology* 23 (9): 1886–98.

Nicolaus, Marion, Céline Le Bohec, Paul M. Nolan, Michel Gauthier-Clerc, Yvon Le Maho, Jan Komdeur, and Pierre Jouventin. 2007. "Ornamental Colors Reveal Age in the King Penguin." *Polar Biology* 31 (1): 53–61.

Nussey, D.H., Kruuk, L.E., Donald, A., Fowlie, M. and Clutton-Brock, T.H., 2006. The rate of senescence in maternal performance increases with early-life fecundity in red deer. Ecology Letters, 9(12), pp.1342-1350.

O'Connor, C.M., Norris, D.R., Crossin, G.T. and Cooke, S.J., 2014. "Biological carryover effects: linking common concepts and mechanisms in ecology and evolution." Ecosphere 5(3): 1-11.

Olsson, Olof. 1996. "Seasonal Effects of Timing and Reproduction in the King Penguin: A Unique Breeding Cycle." *Journal of Avian Biology* 27 (1): 7–14.

Olsson, O. and North, A.W., 1997. Diet of the king penguin Aptenodytes patagonicus during three summers at South Georgia. Ibis, 139(3), pp.504-512.

Orgeret, F., C. Péron, M. R. Enstipp, K. Delord, H. Weimerskirch, and C. A. Bost. 2019. "Exploration during Early Life: Distribution, Habitat and Orientation Preferences in Juvenile King Penguins." *Movement Ecology* 7 (October): 29.

Pang, Yumeng, Chih-Shin Chen, Tomohiko Kawamura, and Yoko Iwata. 2022. "Environmental Influence on Life-History Traits in Male Squid Uroteuthis Edulis with Alternative Reproductive Tactics." *Marine Biology* 169 (3): 33.

Park, Young-Hyang, Isabelle Durand, Elodie Kestenare, Gilles Rougier, Meng Zhou, Francesco d'Ovidio, Cédric Cotté, and Jae-Hak Lee. 2014. "Polar F Ront around the K Erguelen I Slands: An Up-to-date Determination and Associated Circulation of Surface/subsurface Waters." *Journal of Geophysical Research, C: Oceans* 119 (10): 6575–92.

Pascoe, Penelope, Ben Raymond, Noel Carmichael, and Julie McInnes. 2022. "The Current Trajectory of King Penguin (Aptenodytes Patagonicus) Chick Numbers on Macquarie Island in Relation to Environmental Conditions." *ICES Journal of Marine Science: Journal Du Conseil* 79 (7): 2084–92.

Pecl, Gretta T., and George D. Jackson. 2008. "The Potential Impacts of Climate Change on Inshore Squid: Biology, Ecology and Fisheries." *Reviews in Fish Biology and Fisheries* 18 (4): 373–85.

Pigeon, Gabriel, Marco Festa-Bianchet, and Fanie Pelletier. 2017. "Long-Term Fitness Consequences of Early Environment in a Long-Lived Ungulate." *Proceedings. Biological Sciences / The Royal Society* 284 (1853). https://doi.org/10.1098/rspb.2017.0222.

Plard, Floriane, Jean-Michel Gaillard, Tim Coulson, A. J. Mark Hewison, Daniel Delorme, Claude Warnant, and Christophe Bonenfant. 2014. "Mismatch between Birth Date and Vegetation Phenology Slows the Demography of Roe Deer." *PLoS Biology* 12 (4): e1001828.

Post, Eric, and Mads C. Forchhammer. 2008. "Climate Change Reduces Reproductive Success of an Arctic Herbivore through Trophic Mismatch." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1501): 2369–75.

Price, T., Kirkpatrick, M. and Arnold, S.J., 1988. Directional selection and the evolution of breeding date in birds. Science, 240(4853), pp.798-799.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raclot, T., R. Groscolas, and Y. Cherel. 1998. “Fatty Acid Evidence for the Importance of Myctophid Fishes in the Diet of King Penguins, Aptenodytes Patagonicus.” *Marine Biology* 132 (3): 523–33.

Salinas, Santiago, Simon C. Brown, Marc Mangel, and Stephan B. Munch. 2013. “Non-Genetic Inheritance and Changing Environments.” *Non-Genetic Inheritance* 1 (January). https://doi.org/10.2478/ngi-2013-0005.

Saraux, Claire, Vincent A. Viblanc, Nicolas Hanuise, Yvon Le Maho, and Céline Le Bohec. 2011. “Effects of Individual Pre-Fledging Traits and Environmental Conditions on Return Patterns in Juvenile King Penguins.” *PloS One* 6 (6): e20407.

Schenk, Mariela, Jochen Krauss, and Andrea Holzschuh. 2018. “Desynchronizations in Bee-Plant Interactions Cause Severe Fitness Losses in Solitary Bees.” *The Journal of Animal Ecology* 87 (1): 139–49.

Schulte-Hostedde, A. I., J. S. Millar, and G. J. Hickling. 2001. “Evaluating Body Condition in Small Mammals.” *Canadian Journal of Zoology* 79 (6): 1021–29.

Schulte-Hostedde, Albrecht I., Bertram Zinner, John S. Millar, and Graham J. Hickling. 2005. “Restitution of Mass–size Residuals: Validating Body Condition Indices.” *Ecology* 86 (1): 155–63.

Singer, Michael C., and Camille Parmesan. 2010. “Phenological Asynchrony between Herbivorous Insects and Their Hosts: Signal of Climate Change or Pre-Existing Adaptive Strategy?” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 365 (1555): 3161–76.

Song, Zitan, Yuqi Zou, Canshi Hu, Yuanxing Ye, Chao Wang, Baoping Qing, Jan Komdeur, and Changqing Ding. 2019. “Silver Spoon Effects of Hatching Order in an Asynchronous Hatching Bird.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 30 (2): 509–17.

Spagopoulou, Foteini, Céline Teplitsky, Martin I. Lind, Stéphane Chantepie, Lars Gustafsson, and Alexei A. Maklakov. 2020. “Silver-Spoon Upbringing Improves Early-Life Fitness but Promotes Reproductive Ageing in a Wild Bird.” *Ecology Letters* 23 (6): 994–1002.

Stenseth, Nils Chr, and Atle Mysterud. 2002. “Climate, Changing Phenology, and Other Life History Traits: Nonlinearity and Match-Mismatch to the Environment.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (21): 13379–81.

Stier, Antoine, Vincent A. Viblanc, Sylvie Massemin-Challet, Yves Handrich, Sandrine Zahn, Emilio R. Rojas, Claire Saraux, et al. 2014. “Starting with a Handicap: Phenotypic Differences between Early- and Late-born King Penguin Chicks and Their Survival Correlates.” *Functional Ecology* 28 (3): 601–11.

Taborsky, Barbara. 2006. “The Influence of Juvenile and Adult Environments on Life-History Trajectories.” *Proceedings. Biological Sciences / The Royal Society* 273 (1587): 741–50.

Van Allen, Benjamin G., and Volker H. W. Rudolf. 2016. “Carryover Effects Drive Competitive Dominance in Spatially Structured Environments.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (25): 6939–44.

vanHeezik, Y. M., P. J. Seddon, C. J. Du Plessis, and N. J. Adams. 1993. “Differential Growth of King Penguin Chicks in Relation to Date of Hatching.” *Colonial Waterbirds* 16 (1): 71–76.

Weimerskirch, H., J. C. Stahl, and P. Jouventin. 1992. “The Breeding Biology and Population Dynamics of King Penguins Aptenodytes Patagonica on the Crozet Islands.” *The Ibis* 134 (2): 107–17.

Williams, George C. 1966. “Natural Selection, the Costs of Reproduction, and a Refinement of Lack’s Principle.” *The American Naturalist* 100 (916): 687–90.

Williams, C.T., Klaassen, M., Barnes, B.M., Buck, C.L., Arnold, W., Giroud, S., Vetter, S.G. and Ruf, T., 2017. Seasonal reproductive tactics: annual timing and the capital-to-income breeder continuum. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1734), p.20160250.

Wilson, Samantha M., Thomas W. Buehrens, Jennifer L. Fisher, Kyle L. Wilson, and Jonathan W. Moore. 2021. “Phenological Mismatch, Carryover Effects, and Marine Survival in a Wild Steelhead Trout Oncorhynchus Mykiss Population.” *Progress in Oceanography* 193 (April): 102533.

Zhang, He, Oscar Vedder, Peter H. Becker, and Sandra Bouwhuis. 2015. “Age-Dependent Trait Variation: The Relative Contribution of within-Individual Change, Selective Appearance and Disappearance in a Long-Lived Seabird.” *The Journal of Animal Ecology* 84 (3): 797–807.

Zhang, Liping, Thomas L. Delworth, Xiaosong Yang, Fanrong Zeng, Feiyu Lu, Yushi Morioka, and Mitchell Bushuk. 2022. “The Relative Role of the Subsurface Southern Ocean in Driving Negative Antarctic Sea Ice Extent Anomalies in 2016–2021.” *Communications Earth & Environment* 3 (1): 1–9.

Synthesis and discussion

General conclusions

In the framework of the match/mismatch hypothesis (MMH), individuals can be generated inside or outside the peak of environmental resources (match and mismatch, respectively) in natural systems (Cushing 1974; Cushing and Saleem 1982; Cushing 1990). Most individuals tend to reproduce under the peak of resources (match), as this strategy generally results in higher offspring survival and fitness (e.g., Dunn 2004; Reed et al. 2013; Doiron et al. 2015). Nevertheless, populations naturally harbour individuals that breed in mismatch with the peak of resources, such as in the case of the King penguin (Descamps et al. 2002). If such mismatched strategy is expected to negatively affect individual fitness, then why is it often found in natural populations?

Our main hypothesis was that individuals born and surviving under mismatched conditions could have even higher fitness than individuals born under match conditions under future environmental change scenarios. More specifically, some individuals from the mismatched group could have higher fitness than the mean fitness of the matched group, while this difference would be intensified in more future challenging environmental conditions. Given that environmental changes, such as climate changes, are increasing the frequency of mismatches (Kharouba et al. 2018), the higher fitness of surviving mismatched individuals could indeed contribute to the adaptive response of the King penguin to climate change.

In the case of the King penguin, individuals born in mismatch with environmental conditions, here represented by individuals born later in the season, have less time to grow and accumulate energy reserves until the beginning of the austral winter (Weimerskirch et al. 1992; Stier et al. 2014). Such late-born individuals face early-life conditions that may be analogous to the forecasted scenario for the studied population in the next decades, as the main foraging ground of the species during summer (the Antarctic Polar Front, APF), is predicted to move southwards, farther away from many breeding colonies (Péron et al. 2012; Cristofari et al. 2018) (**Figure 17**).

In this thesis, we measured fitness under three different perspectives: through a genetic component, in which we expected the stronger selection on mismatched individuals to result in fitter genotypes, with lower genetic load and higher genetic variability; through a plastic component, by which we expected to find signals of predictive adaptive response (PAR) in early-life mismatched phenotypes that could reflect later pressures of adult-life; and through a life history component, which would be characterised by similar return and first reproductive patterns if mismatched individuals were able to catch-up with matched offspring at fledging, via the predicted genetic and plastic adaptations.

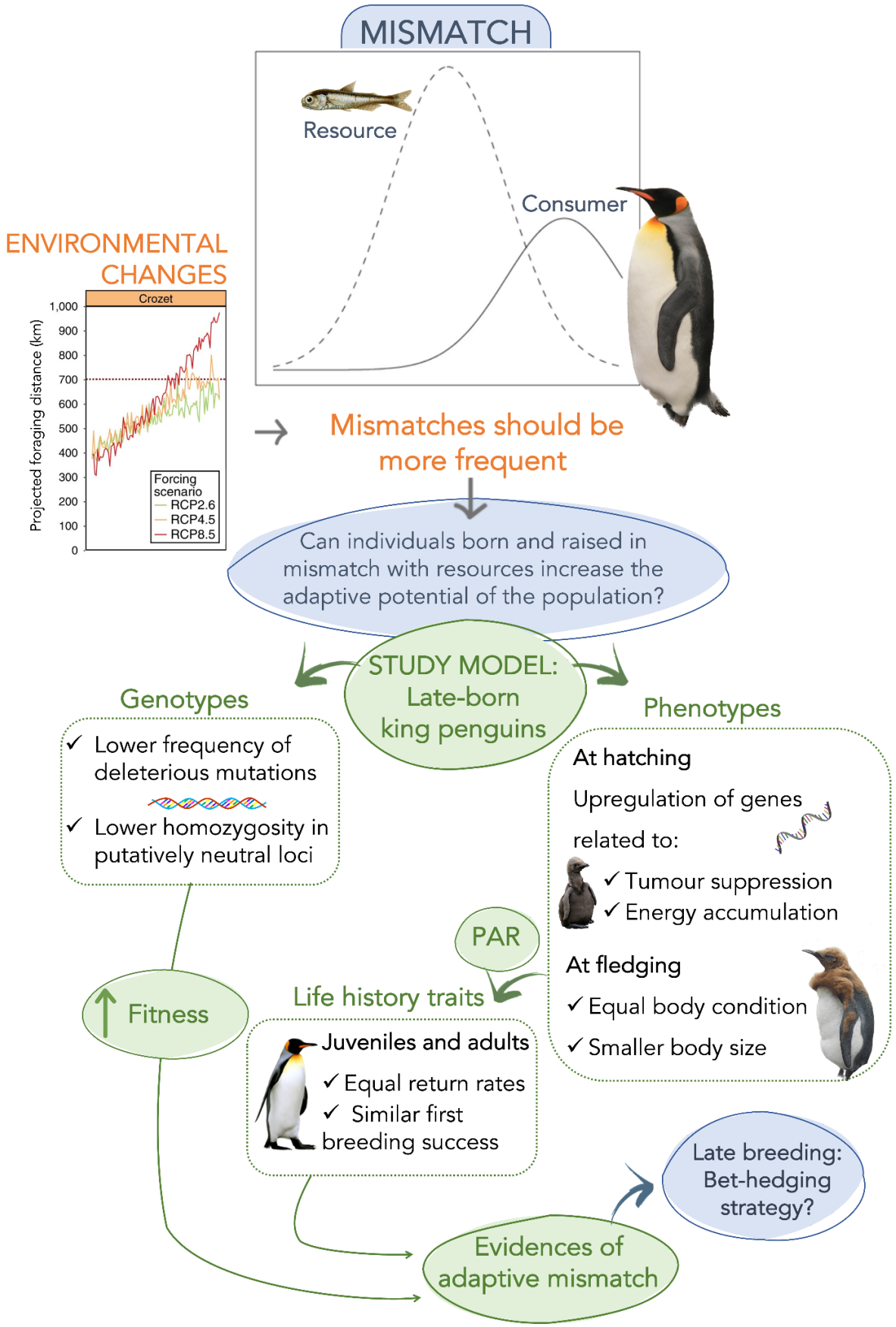


Figure 17. Schema of the main results of this thesis. Mismatch graphic representation adapted from Kharouba and Wolkovich (2020). Curves of resource and consumer abundance (King penguin and myctophid fish) are merely illustrative. Projections of King penguin foraging distance from the breeding colonies of Crozet,

projected according to three greenhouse gas concentration trajectories, RCP2.6, RCP4.5 and RCP8.5, were taken from Cristofari et al. (2018). The horizontal red line in the projections graph represents the 700 km projected distance of the APF to the archipelago, beyond which no successful breeding is expected. Items in green represent results and conclusions taken from our study. PAR stands for predictive adaptive response.

Our first prediction was that the poor early-life conditions faced by mismatched king penguins could act as a strong selection filter, reflected by the high mortality rates of late-born chicks during the first winter (Olsson 1996; Weimerskirch et al. 1992; Stier et al. 2014; Fernandes and Bardon et al. *in prep* - Chapter 3), which would purge deleterious alleles more efficiently from the population. At the same time, because deleterious mutations occur in low frequencies at the population level (Bertorelle et al. 2022), we did not expect this genetic purging to affect global genetic variability in neutral loci. In fact, as higher levels of genetic variability are also considered as a proxy of better fitness (Kardos et al. 2021), we expected survivors to harbour higher levels of heterozygosity (i.e., lower homozygosity) compared to non-survivors, and potentially even more in mismatched *versus* matched ones.

We detected signals indicative of stronger purifying selection acting on late chicks that survived through the first winter, which showed a lower accumulation of highly deleterious mutations in comparison to the early group. Yet, the higher genetic load detected in early-born individuals was due to the early individuals that did not survive, while in late-born individuals harboured lower frequencies of deleterious mutation independently on survival. Moreover, the accumulation of highly deleterious mutations' seemed to be related to chick mortality in the early group, but not in the late group, which already started with a lower genetic load threshold. Thus, our conclusion so far is that the whole late group may have mechanisms that more efficiently purge deleterious alleles from the population.

Now if we look at breeding adults, late breeders are known to be individuals that have been successful in year *n-1*, and arrived later for the breeding season of year *n*. Early breeders, on the other hand, include those successful late breeders in year *n-1*, but also individuals that start early in the season because of failed or non-breeding in year *n-1* (Descamps et al. 2002). Thus, the late group of breeders is expected to be composed of high quality breeders, which are also contained in the early group. However, the early group is much more heterogeneous, also holding individuals of lower reproductive success. Indeed, our results showed that the early-hatchlings were much more heterogeneous in terms of deleterious mutations than the late group, supporting this parental quality hypothesis. Further investigation on the genetic composition and breeding success of breeders of each phenological group can bring light into the role of parental quality to viability selection in the population.

Genetic diversity, in terms of heterozygosity at putatively neutral loci, did not seem to differ greatly between phenological groups, in contrast to our expectations. However, late-born chicks showed significant lower levels of homozygosity for the minor allele of putatively neutral loci. Even though the real fitness effect of the less frequent MODIFIER allele is unknown in our dataset, being less homozygous (for any MODIFIER allele) was significantly related to survival. Accordingly, lower levels of heterozygosity were mostly related to general mortality, as individuals that did not survive in both phenological groups

had lower amounts of heterozygosity for presumably neutral alleles. This genetic diversity pattern may indicate that mortality can be related to the overall heterozygosity carried by an individual independent of the phenological group.

Even though there is currently a debate about the relative importance of genetic diversity and genetic load to fitness (Teixeira and Huber 2021; Kardos et al. 2021), our results indicated that both factors can be related to mortality, and therefore, to a total reduction of individual fitness. Moreover, whether one of those factors can be predictive of the other is still an open question. Our future objective is to test if higher heterozygosity for neutral loci is correlated with a lower amount of highly deleterious mutations. If this correlation holds true, both proxies could be used to measure fitness and thus for extinction risk assessment in wild populations. Of course, these hypotheses would also have to be tested in other species before any major conclusions can be drawn.

Although poor early-life conditions may act as a strong selective filter for individuals that will be recruited in the population, they can also express carryover effects on survivors at later developmental stages (i.e., silver spoon effects, Grafen 1988). On the other hand, if the pressures acting on individuals during development are analogous to stressful situations that will also be faced in adulthood, the expressed phenotypes (e.g., gene expression, morphological traits) can act as a predictive adaptive response (PAR) (Gluckman et al. 2005; Monaghan 2008). Our second prediction was that king penguins born in mismatch with resources would show signals of PAR soon after hatching, as previous studies have shown the capacity of late-survivor chicks to grow faster than early-survivors at 10 days of age (Stier et al. 2014). Also in line with the PAR hypothesis, we predicted that mismatched fledging phenotypes would also present signals of catching-up with the early group, in terms of body condition, which would allow individuals to equally perform in the first post-fledging years and later in life.

In addition to the low genetic load, late individuals also showed evidence of adaptation to efficient growth and tumour suppression at the gene expression level. These could represent plastic adaptation to allow faster growth until the first winter, a period when poor body condition leads to high mortality in the species (Stier et al. 2014). Because faster growth usually generates high quantities of ROS (Stier et al. 2014), the upregulation of tumour suppressor genes could act as a protection against DNA strand damage caused by oxidative molecules (Vurusaner et al. 2012).

In fact, if further evidence of an effective tumour suppression machinery in late chicks could be evidenced, this could be a possible molecular mechanism that leads to the reduced accumulation of deleterious mutations in this group. If the same genes were found to be upregulated in late chicks in other years, in a replicate of the 2020 data shown here, this would give a better indication of target genes to be further explored in order to supplement a better understanding of the mechanisms that inhibit uncontrolled cell growth in penguins. To this end, we intend to sequence the blood transcriptome of a third year of individuals, born in 2022, as well as supplementing 2021 data with more individuals to balance the sex-bias inside each phenological group.

Finally, no major carryover effects were detected in the early adult-life of individuals born under mismatch when compared to individuals born in match with resources. Even if late-born individuals fledged with lower body size in comparison to early-born conspecifics, this did not seem to influence return rates and, therefore, survival, in the first years after fledging. In addition, individuals from both phenological groups generally started their first breeding attempts at the same age. Information about survival and reproductive success through the whole adult-life is still needed to reach clear conclusions about the lifetime impacts of mismatch conditions at birth.

Throughout the three studies included in this thesis, we provide the first evidence that late breeding may have an adaptive advantage in king penguins, in stark contrast to what has previously been postulated (Weimerskirch et al. 1992; Stier et al. 2014). Indeed, late chick mortality is much higher during the first year of life, but this seems to come at the cost of selecting the "best quality breeders" that can attempt to breed late in the season. Because conditions are less favourable for late-season breeding, the selective pressure on survivor chicks is also high, but this does not seem to negatively affect them, at least until maturity is reached. Moreover, we here propose that the late breeding may have been maintained in the species as a type of bet-hedging strategy. In this context, a set of good quality breeders take a chance on breeding late, under less favourable conditions, instead of skipping a year. Those individuals which succeed end up generating offspring with similar fitness compared to the early breeders that successfully fledge a chick.

The bet-hedging strategy derives from the logic of "not putting all of your eggs in the same basket", and can be of great value when conditions are unpredictable (Seger and Brockmann 1987). From a theoretical perspective, under variable environmental conditions, bet-hedging will optimise a population's geometric mean fitness (w_{gm}, more sensitive to variance) at the expense of the reduction of its arithmetic mean fitness (w_{am}) (Simons 2011). If individual fitness is measured at specific years with good environmental conditions, a non-bet-hedging strategy will provide higher fitness. However, this will lead to a high lifetime variance in success, as non-bet-hedgers will experience several years of minimal fitness, while bet-hedgers will have constant yearly fitness. It then follows that if we measure total fitness at the end of a non-bet-hedger lifetime, it will be lower. This is because the low number of years with maximum success in non-bet-hedgers will not pay off for the more frequent year of minimal success. A simplified representation of the impact of bet-hedging and non-bet-hedging strategies on the arithmetic and geometric fitness can be visualised in **Figure 18**, from (Simons 2011).

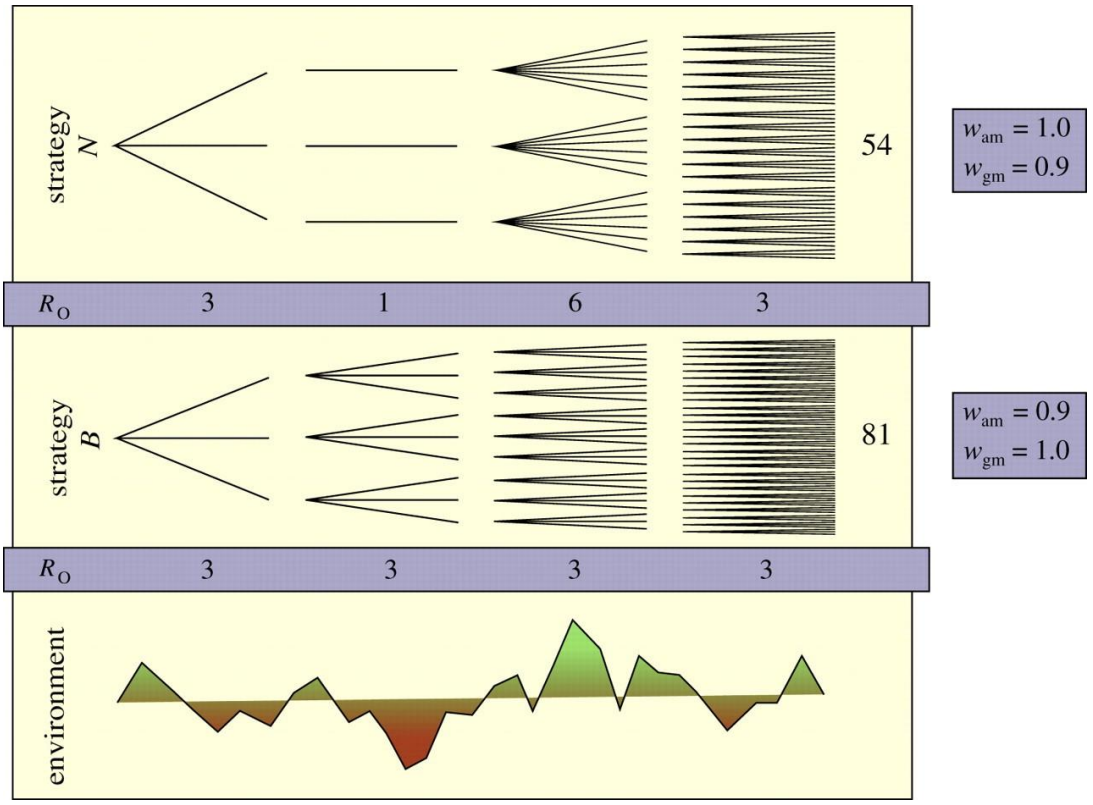


Figure 18. Representation of the bet-hedging strategy, from (Simons 2011). At the top and middle, a theoretical representation of the reproductive success under a non-bet-hedging (N) and a bet-hedging strategy (B), respectively. The bottom image represents environmental fluctuations (favourable conditions in green, unfavourable conditions in red). The numbers in the purple bars represent the number of offspring generated under each strategy in four generations, and the numbers at the end of yellow boxes represent the total offspring after four generations. The offspring number after each generation remains constant in the bet-hedging strategy, while it varies according to the environmental conditions in the non-bet-hedging strategy (i.e., unfavourable conditions lead to lower success, favourable conditions lead to higher success). The expected arithmetic mean fitness (w_{am}) and geometric mean fitness (w_{gm}) at each strategy are displayed in purple boxes at the right of the figure. While the bet-hedging strategy leads to lower w_{am}, its w_{gm} and offspring number across generations are higher than in the non-bet-hedging strategy under variable environmental conditions.

Considering that bet-hedgers can endure some degree of mismatch generated by environmental variation (Sæther and Engen 2015), bet-hedging strategies could then provide species with the ability to more rapidly counteract new conditions posed by climate change (Villa Martín et al. 2019). As it was outlined throughout this thesis, climate change is already increasing mismatches in the wild, by affecting the synchrony of many consumer-resource systems, specially in bird species (Stevenson and Bryant 2000; Visser and Holleman 2001; Dunn 2004; Both et al. 2009; Keogan et al. 2018). Under this scenario, understanding the mechanisms that allow bet-hedgers to cope with changes under mismatched conditions, could inform the adaptive potential of species under future conditions.

A further step from our study would be to test whether late-breeding in king penguins can be considered as a bet-hedging strategy. Preliminary simulation results on the study system have shown evidence that the late-breeding strategy (i.e., two consecutive years of breeding, early and late, with a third year of pause) increases the geometrical

fitness of the population when compared to strategies of always breeding early (Massa 2021).

The results presented in this thesis represent an exploratory survey of how individuals born under mismatch with environmental conditions can contribute to the persistence of a population, through genetic, plastic, and life history components. To overcome the limitations of some of our predictions, especially regarding the exact functions of upregulated genes observed in our species, we aim to provide target genes that can be further explored under another specific study design. Natural systems are highly heterogeneous and noisier when compared to laboratory conditions (Krishnan et al. 2020). However, exploring selective pressures in the wild can lead to more accurate predictions about the real species response to different pressures. Ultimately, considering future climate change scenarios, in which the king penguin is also predicted to be subjected to drastic habitat losses (Cristofari et al. 2018), mismatched adaptations, such as the ones detected here, could help to avoid extinction under unpredictable and variable environments.

References

Adamo, Nasrat, Nadhir Al-Ansari, and Varoujan Sissakian. 2021. “Review of Climate Change Impacts on Human Environment: Past, Present and Future Projections.” *Proceedings of the Estonian Academy of Sciences: Engineering* 13 (11): 605–30.

Arrigo, Kevin R., and David N. Thomas. 2004. “Large Scale Importance of Sea Ice Biology in the Southern Ocean.” *Antarctic Science / Blackwell Scientific Publications* 16 (4): 471–86.

Ashcroft, Michael B. 2010. “Identifying Refugia from Climate Change.” *Journal of Biogeography*, April, no – no.

Aubin-Horth, Nadia, and Susan C. P. Renn. 2009. “Genomic Reaction Norms: Using Integrative Biology to Understand Molecular Mechanisms of Phenotypic Plasticity.” *Molecular Ecology* 18 (18): 3763–80.

Banks, Jonathan, Amy Van Buren, Yves Cherel, and James B. Whitfield. 2006. “Genetic Evidence for Three Species of Rockhopper Penguins, Eudyptes Chrysocome.” *Polar Biology* 30 (1): 61–67.

Barrat, A., 1976. “Quelques aspects de la biologie et de l'écologie du Manchot royal (Aptenodytes patagonicus) des îles Crozet.”

Barrett, Rowan D. H., and Dolph Schluter. 2008. “Adaptation from Standing Genetic Variation.” *Trends in Ecology & Evolution* 23 (1): 38–44.

Bauer, A., 1967. Denombrement des manchotieres de lárchipel des Crozet et des iles Kerguelen a láide de photographies aeriennes verticales. La Documentation Française.

Beaumont, Hubertus J. E., Jenna Gallie, Christian Kost, Gayle C. Ferguson, and Paul B. Rainey. 2009. “Experimental Evolution of Bet Hedging.” *Nature* 462 (7269): 90–93.

Benning, Tracy L., Dennis LaPointe, Carter T. Atkinson, and Peter M. Vitousek. 2002. “Interactions of Climate Change with Biological Invasions and Land Use in the Hawaiian Islands: Modeling the Fate of Endemic Birds Using a Geographic Information System.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (22): 14246–49.

Bertorelle, Giorgio, Francesca Raffini, Mirte Bosse, Chiara Bortoluzzi, Alessio Iannucci, Emiliano Trucchi, Hernán E. Morales, and Cock van Oosterhout. 2022. “Genetic Load: Genomic Estimates and Applications in Non-Model Animals.” *Nature Reviews. Genetics* 23 (8): 492–503.

Bitter, M. C., L. Kapsenberg, J-P Gattuso, and C. A. Pfister. 2019. “Standing Genetic Variation Fuels Rapid Adaptation to Ocean Acidification.” *Nature Communications* 10 (1): 5821.

Bond, David P. G., and Stephen E. Grasby. 2017. “On the Causes of Mass Extinctions.” *Palaeogeography, Palaeoclimatology, Palaeoecology* 478 (July): 3–29.

Boonstra, Rudy, Korryn Bodner, Curtis Bosson, Brendan Delehanty, Evan S. Richardson, Nicholas J. Lunn, Andrew E. Derocher, and Péter K. Molnár. 2020. “The Stress of Arctic Warming on Polar Bears.” *Global Change Biology* 26 (8): 4197–4214.

Borboroglu, Pablo Garcia, and P. Dee Boersma. 2015. *Penguins: Natural History and Conservation*. University of Washington Press.

Bost, C. A., J. B. Charrassin, Y. Clerquin, Y. Ropert-Coudert, and Y. Le Maho. 2004. “Exploitation of Distant Marginal Ice Zones by King Penguins during Winter.” *Marine Ecology Progress Series* 283: 293–97.

Bost, C. A., J. Y. Georges, C. Guinet, Y. Cherel, K. Pütz, J. B. Charrassin, Y. Handrich, T. Zorn, J. Lage, and Y. Le Maho. 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.

Bost, Charles A., Cedric Cotté, Pascal Terray, Christophe Barbraud, Cécile Bon, Karine Delord, Olivier Gimenez, et al. 2015. “Large-Scale Climatic Anomalies Affect Marine Predator Foraging

Behaviour and Demography." *Nature Communications* 6 (October): 8220.

Bost, Charles-André, Karine Delord, Christophe Barbraud, Cédric Cotte, Clara Péron, and Henri Weimerskirch. 2013. "The King Penguin: Life History, Current Status and Priority Conservation Actions. In: Penguins Book," 360.

Both, Christiaan, Margriet van Asch, Rob G. Bijlsma, Arnold B. van den Burg, and Marcel E. Visser. 2009. "Climate Change and Unequal Phenological Changes across Four Trophic Levels: Constraints or Adaptations?" *The Journal of Animal Ecology* 78 (1): 73–83.

Both, C., and M. E. Visser. 2001. "Adjustment to Climate Change Is Constrained by Arrival Date in a Long-Distance Migrant Bird." *Nature* 411 (6835): 296–98.

Bried, J. and Jouventin, P., 2001. The King Penguin *Aptenodytes patagonicus*, a non-nesting bird which selects its breeding habitat. Ibis, 143(3), pp.670-673.

Briga, M., Koetsier, E., Boonekamp, J.J., Jimeno, B. and Verhulst, S., 2017. Food availability affects adult survival trajectories depending on early developmental conditions. Proceedings of the Royal Society B: Biological Sciences, 284(1846), p.20162287.

Buse, A., S. J. Dury, R. J. W. Woodburn, C. M. Perrins, and J. E. G. Good. 1999. "Effects of Elevated Temperature on Multi-Species Interactions: The Case of Pedunculate Oak, Winter Moth and Tits." *Functional Ecology* 13 (s1): 74–82.

Butterfield, Nicholas J. 2007. "Macroevolution and Macroecology through Deep Time." *Palaeontology* 50 (1): 41–55.

Cahill, Abigail E., Matthew E. Aiello-Lammens, M. Caitlin Fisher-Reid, Xia Hua, Caitlin J. Karanewsky, Hae Yeong Ryu, Gena C. Sbeglia, et al. 2013. "How Does Climate Change Cause Extinction?" *Proceedings. Biological Sciences / The Royal Society* 280 (1750): 20121890.

Capilla-Lasheras, Pablo, Xavier Harrison, Emma M. Wood, Alastair J. Wilson, and Andrew J. Young. 2021. "Altruistic Bet-Hedging and the Evolution of Cooperation in a Kalahari Bird." *Science Advances* 7 (39): eabe8980.

Cardoso-Moreira, Margarida, Jean Halbert, Delphine Valloton, Britta Velten, Chunyan Chen, Yi Shao, Angélica Liechti, et al. 2019. "Gene Expression across Mammalian Organ Development." *Nature* 571 (7766): 505–9.

Catling, David C., and Kevin J. Zahnle. 2020. "The Archean Atmosphere." *Science Advances* 6 (9): eaax1420.

Charmantier, Anne, Robin H. McCleery, Lionel R. Cole, Chris Perrins, Loeske E. B. Kruuk, and Ben C. Sheldon. 2008. "Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population." *Science* 320 (5877): 800–803.

Chen, Song-Can, Guo-Xin Sun, Yu Yan, Konstantinos T. Konstantinidis, Si-Yu Zhang, Ye Deng, Xiao-Min Li, et al. 2020. "The Great Oxidation Event Expanded the Genetic Repertoire of Arsenic Metabolism and Cycling." *Proceedings of the National Academy of Sciences of the United States of America* 117 (19): 10414–21.

Cherel, Yves, Jean-Claude Stahl, and Yvon Le Maho. 1987. "Ecology and Physiology of Fasting in King Penguin Chicks." *The Auk* 104 (2): 254–62.

Chiari, Ylenia, and Nicolas Galtier. 2011. "RNA Extraction from Sauropsids Blood: Evaluation and Improvement of Methods." *Amphibia-Reptilia: Publication of the Societas Europaea Herpetologica* 32 (1): 136–39.

Childs, Dylan Z., C. J. E. Metcalf, and Mark Rees. 2010. "Evolutionary Bet-Hedging in the Real World: Empirical Evidence and Challenges Revealed by Plants." *Proceedings. Biological Sciences / The Royal Society* 277 (1697): 3055–64.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. "A Program for Annotating and Predicting the

Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain w1118; Iso-2; Iso-3." *Fly* 6 (2): 80–92.

Clucas, Gemma V., Jane L. Younger, Damian Kao, Louise Emmerson, Colin Southwell, Barbara Wienecke, Alex D. Rogers, et al. 2018. "Comparative Population Genomics Reveals Key Barriers to Dispersal in Southern Ocean Penguins." *Molecular Ecology* 27 (23): 4680–97.

Cole, Ella F., Charlotte E. Regan, and Ben C. Sheldon. 2021. "Spatial Variation in Avian Phenological Response to Climate Change Linked to Tree Health." *Nature Climate Change* 11 (10): 872–78.

Cole, Theresa L., Ludovic Dutoit, Nicolas Dussex, Tom Hart, Alana Alexander, Jane L. Younger, Gemma V. Clucas, et al. 2019. "Receding Ice Drove Parallel Expansions in Southern Ocean Penguins." *Proceedings of the National Academy of Sciences of the United States of America* 116 (52): 26690–96.

Cole, Theresa L., Chengran Zhou, Miaoquan Fang, Hailin Pan, Daniel T. Ksepka, Steven R. Fiddaman, Christopher A. Emerling, et al. 2022. "Genomic Insights into the Secondary Aquatic Transition of Penguins." *Nature Communications* 13 (1): 3912.

Convey, P., A. Morton, and J. Poncet. 1999. "Survey of Marine Birds and Mammals of the South Sandwich Islands." *The Polar Record* 35 (193): 107–24.

Coop, Graham. 2020. "Population and Quantitative Genetics" Davis: University of California.

Cooper, Gregory M., Eric A. Stone, George Asimenos, NISC Comparative Sequencing Program, Eric D. Green, Serafim Batzoglou, and Arend Sidow. 2005. "Distribution and Intensity of Constraint in Mammalian Genomic Sequence." *Genome Research* 15 (7): 901–13.

Cooper, John, Robert J. M. Crawford, Marienne S. de Villiers, B. M. Dyer, G. J. G. Hofmeyr, and A. Jonker. 2009. "Disease Outbreaks among Penguins at Sub-Antarctic Marion Island: A Conservation Concern." *Marine Ornithology* 37: 193–96.

Cresswell, Will, and Robin Mccleery. 2003. "How Great Tits Maintain Synchronization of Their Hatch Date with Food Supply in Response to Long-Term Variability in Temperature." *The Journal of Animal Ecology* 72 (2): 356–66.

Cristofari, Robin, Xiaoming Liu, Francesco Bonadonna, Yves Cherel, Pierre Pistorius, Yvon Le Maho, Virginie Raybaud, Nils Christian Stenseth, Céline Le Bohec, and Emiliano Trucchi. 2018. "Climate-Driven Range Shifts of the King Penguin in a Fragmented Ecosystem." *Nature Climate Change* 8 (3): 245–51.

Cristofari, Robin, Emiliano Trucchi, Jason D. Whittington, Stéphanie Vigetta, Hélène Gachot-Neveu, Nils Christian Stenseth, Yvon Le Maho, and Céline Le Bohec. 2015. "Spatial Heterogeneity as a Genetic Mixing Mechanism in Highly Philopatric Colonial Seabirds." *PloS One* 10 (2): e0117981.

Crowley, Thomas J. 1983. "The Geologic Record of Climatic Change." *Reviews of Geophysics* 21 (4): 828.

Croxall, John P., Stuart H. M. Butchart, Ben Lascelles, Alison J. Stattersfield, Ben Sullivan, Andy Symes, and Phil Taylor. 2012. "Seabird Conservation Status, Threats and Priority Actions: A Global Assessment." *Bird Conservation International* 22 (1): 1–34.

Cushing, D. H. 1969. "The Regularity of the Spawning Season of Some Fishes." *ICES Journal of Marine Science: Journal Du Conseil* 33 (1): 81–92.

Cushing, D.H., 1974. "The possible density-dependence of larval mortality and adult mortality in fishes." The Early Life History of Fish" JHS, 103-111.

Cushing, D. H. 1990. "Plankton Production and Year-Class Strength in Fish Populations: An Update of the Match/Mismatch Hypothesis." In *Advances in Marine Biology*, edited by J. H. S. Blaxter and A. J. Southward, 26:249–93. Academic Press.

Cushing, J. M., and M. Saleem. 1982. "A Predator Prey Model with Age Structure." *Journal of Mathematical Biology* 14 (2): 231–50.

Darnis, Gérald, Dominique Robert, Corinne Pomerleau, Heike Link, Philippe Archambault, R. John Nelson, Maxime Geoffroy, et al. 2012. “Current State and Trends in Canadian Arctic Marine Ecosystems: II. Heterotrophic Food Web, Pelagic-Benthic Coupling, and Biodiversity.” *Climatic Change* 115 (1): 179–205.

Darwin, Charles. 1859. "On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life." London: Murray.

De Lisle, Stephen P., Maarit I. Mäenpää, and Erik I. Svensson. 2022. “Phenotypic Plasticity Is Aligned with Phenological Adaptation on Both Micro- and Macroevolutionary Timescales.” *Ecology Letters* 25 (4): 790–801.

Descamps, Sébastien, Céline Le Bohec, Yvon Le Maho, Jean-Paul Gendner, and Michel Gauthier-Clerc. 2009. “Relating Demographic Performance to Breeding-Site Location in the King Penguin.” *The Condor* 111 (1): 81–87.

Descamps, Sebastien, M. Gauthier-Clerc, J. P. Gendner, and Y. Le Maho. 2002. “The Annual Cycle of Unbanded King Penguins Aptenodytes Patagonicus on Possession Island (Crozet)” 2 (January): 87–98.

Descamps, Sébastien, Michel Gauthier-Clerc, Céline Le Bohec, Jean-Paul Gendner, and Yvon Le Maho. 2005. “Impact of Predation on King Penguin Aptenodytes Patagonicus in Crozet Archipelago.” *Polar Biology* 28 (4): 303–10.

Deutsch, Curtis A., Joshua J. Tewksbury, Raymond B. Huey, Kimberly S. Sheldon, Cameron K. Ghalambor, David C. Haak, and Paul R. Martin. 2008. “Impacts of Climate Warming on Terrestrial Ectotherms across Latitude.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (18): 6668–72.

Diamond, Sarah E., and Ryan A. Martin. 2016. “The Interplay between Plasticity and Evolution in Response to Human-Induced Environmental Change.” *F1000Research* 5 (December): 2835.

Dias, Maria P., Rob Martin, Elizabeth J. Pearmain, Ian J. Burfield, Cleo Small, Richard A. Phillips, Oliver Yates, Ben Lascelles, Pablo Garcia Borboroglu, and John P. Croxall. 2019. “Threats to Seabirds: A Global Assessment.” *Biological Conservation* 237 (September): 525–37.

Díaz, Sandra, Josef Settele, Eduardo S. Brondízio, Hien T. Ngo, John Agard, Almut Arneth, Patricia Balvanera, et al. 2019. “Pervasive Human-Driven Decline of Life on Earth Points to the Need for Transformative Change.” *Science* 366 (6471). https://doi.org/10.1126/science.aax3100.

Doiron, Madeleine, Gilles Gauthier, and Esther Lévesque. 2015. “Trophic Mismatch and Its Effects on the Growth of Young in an Arctic Herbivore.” *Global Change Biology* 21 (12): 4364–76.

Dunkley Jones, Tom, Daniel J. Lunt, Daniela N. Schmidt, Andy Ridgwell, Appy Sluijs, Paul J. Valdes, and Mark Maslin. 2013. “Climate Model and Proxy Data Constraints on Ocean Warming across the Paleocene–Eocene Thermal Maximum.” *Earth-Science Reviews* 125 (October): 123–45.

Dunn, Peter. 2004. “Breeding Dates and Reproductive Performance.” In *Advances in Ecological Research*, 35:69–87. Academic Press.

Durance, Isabelle, and S. J. Ormerod. 2010. “Evidence for the Role of Climate in the Local Extinction of a Cool-Water Triclad.” *Journal of the North American Benthological Society* 29 (4): 1367–78.

Durant, J. M., D. Ø. Hjermann, G. Ottersen, and N. C. Stenseth. 2007. “Climate and the Match or Mismatch between Predator Requirements and Resource Availability.” *Climate Research* 33 (April): 271–83.

Durant, Joël M., Dag Ø. Hjermann, Tycho Anker-Nilssen, Grégory Beaugrand, Atle Mysterud, Nathalie Pettorelli, and Nils Chr Stenseth. 2005. “Timing and Abundance as Key Mechanisms Affecting Trophic Interactions in Variable Environments.” *Ecology Letters* 8 (9): 952–58.

Edwards, Martin, and Anthony J. Richardson. 2004. “Impact of Climate Change on Marine Pelagic Phenology and Trophic Mismatch.” *Nature* 430 (7002): 881–84.

Eichhorn, Götz, René Groscolas, Gaële Le Glaunec, Camille Parisel, Laurent Arnold, Patrice Medina, and Yves Handrich. 2011. “Heterothermy in Growing King Penguins.” *Nature Communications* 2 (August): 435.

Evans, Tyler G., and Gretchen E. Hofmann. 2012. “Defining the Limits of Physiological Plasticity: How Gene Expression Can Assess and Predict the Consequences of Ocean Change.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367 (1596): 1733–45.

Exposito-Alonso, Moises, Tom R. Booker, Lucas Czech, Lauren Gillespie, Shannon Hateley, Christopher C. Kyriazis, Patricia L. M. Lang, et al. 2022. “Genetic Diversity Loss in the Anthropocene.” *Science* 377 (6613): 1431–35.

Frankham, Richard. 2010. “Challenges and Opportunities of Genetic Approaches to Biological Conservation.” *Biological Conservation* 143 (9): 1919–27.

Frankham, Richard. 2015. “Genetic Rescue of Small Inbred Populations: Meta-Analysis Reveals Large and Consistent Benefits of Gene Flow.” *Molecular Ecology* 24 (11): 2610–18.

Frugone, M. J., A. Lowther, D. Noll, B. Ramos, P. Pistorius, G. P. M. Dantas, M. V. Petry, et al. 2018. “Contrasting Phylogeographic Pattern among Eudyptes Penguins around the Southern Ocean.” *Scientific Reports* 8 (1): 17481.

Gauthier-Clerc, M., Y. Le Maho, Y. Clerquin, C. A. Bost, and Y. Handrich. 2002. “Seabird Reproduction in an Unpredictable Environment: How King Penguins Provide Their Young Chicks with Food.” *Marine Ecology Progress Series* 237: 291–300.

Gendner, Jean-Paul, Michel Gauthier-Clerc, Celine Le Bohec, Sebastien Descamps, and Yvon Le Maho. 2005. “A New Application for Transponders in Studying Penguins.” *Journal of Field Ornithology* 76 (2): 138–42.

Ghalambor, Cameron K., Kim L. Hoke, Emily W. Ruell, Eva K. Fischer, David N. Reznick, and Kimberly A. Hughes. 2015. “Non-Adaptive Plasticity Potentiates Rapid Adaptive Evolution of Gene Expression in Nature.” *Nature* 525 (7569): 372–75.

Ghalambor, C. K., J. K. McKAY, S. P. Carroll, and D. N. Reznick. 2007. “Adaptive versus Non-Adaptive Phenotypic Plasticity and the Potential for Contemporary Adaptation in New Environments.” *Functional Ecology* 21 (3): 394–407.

Gienapp, P., Teplitsky, C., Alho, J.S., Mills, J.A. and Merilä, J., 2008. Climate change and evolution: disentangling environmental and genetic responses. Molecular ecology, 17(1), pp.167-178.

Gluckman, Peter D., Mark A. Hanson, and Hamish G. Spencer. 2005a. “Predictive Adaptive Responses and Human Evolution.” *Trends in Ecology & Evolution* 20 (10): 527–33.

Gluckman, Peter D., Mark A. Hanson, Hamish G. Spencer, and Patrick Bateson. 2005b. “Environmental Influences during Development and Their Later Consequences for Health and Disease: Implications for the Interpretation of Empirical Studies.” *Proceedings. Biological Sciences / The Royal Society* 272 (1564): 671–77.

Gradinger, R., 1995. Climate change and biological oceanography of the Arctic Ocean. Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences, 352(1699), pp.277-286.

Grafen, Alan. 1988. "On the uses of data on lifetime reproductive success." Reproductive success 454-485.

Groscolas, R., and J. P. Robin. 2001. “Long-Term Fasting and Re-Feeding in Penguins.” *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 128 (3): 645–55.

Grosser, Stefanie, Christopher P. Burridge, Amanda J. Peucker, and Jonathan M. Waters. 2015. “Coalescent Modelling Suggests Recent Secondary-Contact of Cryptic Penguin Species.” *PloS One* 10 (12): e0144966.

Gryz, Piotr, Alina Gerlée, and Małgorzata Korczak-Abshire. 2018. “New Breeding Site and Records of

King Penguins (Aptenodytes Patagonicus) on King George Island (South Shetlands, Western Antarctic)." *The Polar Record* 54 (4): 275–83.

Hallam, A. 1987. "End-Cretaceous Mass Extinction Event: Argument for Terrestrial Causation." *Science* 238 (4831): 1237–42.

Hallam, A., and P. B. Wignall. 1999. "Mass Extinctions and Sea-Level Changes." *Earth-Science Reviews* 48 (4): 217–50.

Hamel, Sandra, Jean-Michel Gaillard, Marco Festa-Bianchet, and Steeve D. Côté. 2009. "Individual Quality, Early-Life Conditions, and Reproductive Success in Contrasted Populations of Large Herbivores." *Ecology* 90 (7): 1981–95.

Harley, Christopher D. G. 2011. "Climate Change, Keystone Predation, and Biodiversity Loss." *Science* 334 (6059): 1124–27.

Hayward, Adam D., Ian J. Rickard, and Virpi Lummaa. 2013. "Influence of Early-Life Nutrition on Mortality and Reproductive Success during a Subsequent Famine in a Preindustrial Population." *Proceedings of the National Academy of Sciences of the United States of America* 110 (34): 13886–91.

Hearing, Thomas W., Thomas H. P. Harvey, Mark Williams, Melanie J. Leng, Angela L. Lamb, Philip R. Wilby, Sarah E. Gabbott, Alexandre Pohl, and Yannick Donnadieu. 2018. "An Early Cambrian Greenhouse Climate." *Science Advances* 4 (5): eaar5690.

Hibbets, Eric M., Katelyn I. Schumacher, Hannah B. Scheppler, P. Dee Boersma, and Juan L. Bouzat. 2020. "Genetic Evidence of Hybridization between Magellanic (Sphensicus Magellanicus) and Humboldt (Spheniscus Humboldti) Penguins in the Wild." *Genetica* 148 (5-6): 215–28.

Hodgskiss, Malcolm S. W., Peter W. Crockford, Yongbo Peng, Boswell A. Wing, and Tristan J. Horner. 2019. "A Productivity Collapse to End Earth's Great Oxidation." *Proceedings of the National Academy of Sciences* 116 (35): 17207–12.

Del Hoyo, J., J., Elliott, A. and Sargatal, J., 1992. Handbook of the birds of the world (Vol. 1, No. 8). Barcelona: Lynx edicions.

Hughes, I., I. 2000. "Biological Consequences of Global Warming: Is the Signal Already Apparent?" *Trends in Ecology & Evolution* 15 (2): 56–61.

IUCN Red List of Threatened Species; link <https://www.iucnredlist.org/species/22697748/184637776#bibliography> accessed the 10/02/2023

Jiang, Guangshun, Haiyi Sun, Jianmin Lang, Lijuan Yang, Cheng Li, Arnaud Lyet, Barney Long, et al. 2014. "Effects of Environmental and Anthropogenic Drivers on Amur Tiger Distribution in Northeastern China." *Ecological Research* 29 (5): 801–13.

Ji, Rubao, Meibing Jin, and Øystein Varpe. 2013. "Sea Ice Phenology and Timing of Primary Production Pulses in the Arctic Ocean." *Global Change Biology* 19 (3): 734–41.

Jouventin, P. and Lagarde, F., 1995. Evolutionary ecology of the King Penguin Aptenodytes patagonicus: the self-regulation of the breeding cycle. The penguins: ecology and management. Chipping Norton: Surrey Beatty & Sons, pp.80-95.

Jun, Sang-Yoon, Joo-Hong Kim, Jung Choi, Seong-Joong Kim, Baek-Min Kim, and Soon-Il An. 2020. "The Internal Origin of the West-East Asymmetry of Antarctic Climate Change." *Science Advances* 6 (24): eaaz1490.

Kardos, Marty, Ellie E. Armstrong, Sarah W. Fitzpatrick, Samantha Hauser, Philip W. Hedrick, Joshua M. Miller, David A. Tallmon, and W. Chris Funk. 2021. "The Crucial Role of Genome-Wide Genetic Variation in Conservation." *Proceedings of the National Academy of Sciences of the United States of America* 118 (48). https://doi.org/10.1073/pnas.2104642118.

Keogan, Katharine, Francis Daunt, Sarah Wanless, Richard A. Phillips, Craig A. Walling, Philippa

Agnew, David G. Ainley, et al. 2018. “Global Phenological Insensitivity to Shifting Ocean Temperatures among Seabirds.” *Nature Climate Change* 8 (4): 313–18.

Kharouba, Heather M., Johan Ehrlén, Andrew Gelman, Kjell Bolmgren, Jenica M. Allen, Steve E. Travers, and Elizabeth M. Wolkovich. 2018. “Global Shifts in the Phenological Synchrony of Species Interactions over Recent Decades.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (20): 5211–16.

Kharouba, Heather M., and Elizabeth M. Wolkovich. 2020. “Disconnects between Ecological Theory and Data in Phenological Mismatch Research.” *Nature Climate Change* 10 (5): 406–15.

Kozlov, A. N., K. V. Shust, and A. V. Zemsky. 1991. "Seasonal and inter-annual variability in the distribution of Electrona carlsbergi in the Southern Polar Front area (the area to the north of South Georgia is used as an example)." CCAMLR Sel Sci Pap (7) : 337-368.

Kramer-Schadt, Stephanie, Eloy Revilla, Thorsten Wiegand, and Urs Breitenmoser. 2004. “Fragmented Landscapes, Road Mortality and Patch Connectivity: Modelling Influences on the Dispersal of Eurasian Lynx.” *The Journal of Applied Ecology* 41 (4): 711–23.

Kriesell, Hannah J., Thierry Aubin, Víctor Planas-Bielsa, Marine Benoiste, Francesco Bonadonna, Hélène Gachot-Neveu, Yvon Le Maho, et al. 2018. “Sex Identification in King Penguins*Aptenodytes Patagonicus*through Morphological and Acoustic Cues.” *The Ibis* 160 (4): 755–68.

Kriesell, Hannah Joy, Thierry Aubin, Víctor Planas-Bielsa, Quentin Schull, Francesco Bonadonna, Clément Cornec, Yvon Le Maho, Laura Troudet, and Céline Le Bohec. 2021. “How King Penguins Advertise Their Sexual Maturity.” *Animal Behaviour* 177 (July): 253–67.

Kroeker, Kristy J., Cassandra Powell, and Emily M. Donham. 2020. “Windows of Vulnerability: Seasonal Mismatches in Exposure and Resource Identity Determine Ocean Acidification’s Effect on a Primary Consumer at High Latitude.” *Global Change Biology*, December. https://doi.org/10.1111/gcb.15449.

Ksepka, Daniel T., Daniel J. Field, Tracy A. Heath, Walker Pett, Daniel B. Thomas, Simone Giovanardi, and Alan J. D. Tennyson. n.d. “Largest-Known Fossil Penguin Provides Insight into the Early Evolution of Sphenisciform Body Size and Flipper Anatomy.” *Journal of Paleontology*, 1–20. Accessed March 28, 2023.

Kucheravy, Caila E., Jane M. Waterman, Elaine A. C. Dos Anjos, James F. Hare, Chris Enright, and Charlene N. Berkvens. 2021. “Extreme Climate Event Promotes Phenological Mismatch between Sexes in Hibernating Ground Squirrels.” *Scientific Reports* 11 (1): 21684.

Kusch, Alejandro, and Manuel Marín. 2012. “Sobre La Distribución Del Pingüino Rey Aptenodytes Patagonicus (Aves: Spheniscidae) En Chile.” *Anales Del Instituto de La Patagonia* 40 (1): 157–63.

Laidre, Kristin L., Harry Stern, Kit M. Kovacs, Lloyd Lowry, Sue E. Moore, Eric V. Regehr, Steven H. Ferguson, et al. 2015. “Arctic Marine Mammal Population Status, Sea Ice Habitat Loss, and Conservation Recommendations for the 21st Century.” *Conservation Biology: The Journal of the Society for Conservation Biology* 29 (3): 724–37.

Lansch-Justen, Lucy, Davide Cusseddu, Mark A. Schmitz, and Claudia Bank. 2022. “The Extinction Time under Mutational Meltdown.” *bioRxiv*. https://doi.org/10.1101/2022.02.01.478601.

Le Bohec, Céline. 2007. “Stratégies D’histoire De Vie D’un Oiseau Longévif: Le Manchot Royal (*Aptenodytes patagonicus*)” PhD thesis. University of Strasbourg.

Le Bohec, Céline, Michel Gauthier-Clerc, David Grémillet, Roger Pradel, Arnaud Béchet, Jean-Paul Gendner, and Yvon Le Maho. 2007. “Population Dynamics in a Long-Lived Seabird: I. Impact of Breeding Activity on Survival and Breeding Probability in Unbanded King Penguins.” *The Journal of Animal Ecology* 76 (6): 1149–60.

Le Bohec, Céline, Michel Gauthier-Clerc, and Yvon Le Maho. 2005. “The Adaptive Significance of Crèches in the King Penguin.” *Animal Behaviour* 70 (3): 527–38.

Leonard, Aoife M., and Lesley T. Lancaster. 2020. “Maladaptive Plasticity Facilitates Evolution of Thermal Tolerance during an Experimental Range Shift.” *BMC Evolutionary Biology* 20 (1): 47.

Lepot, Kevin. 2020. “Signatures of Early Microbial Life from the Archean (4 to 2.5 Ga) Eon.” *Earth-Science Reviews* 209 (October): 103296.

Liew, Choong-Chin, Jun Ma, Hong-Chang Tang, Run Zheng, and Adam A. Dempsey. 2006. “The Peripheral Blood Transcriptome Dynamically Reflects System Wide Biology: A Potential Diagnostic Tool.” *The Journal of Laboratory and Clinical Medicine* 147 (3): 126–32.

Ligrone, Roberto. 2019. “The Great Oxygenation Event.” In *Biological Innovations That Built the World: A Four-Billion-Year Journey through Life and Earth History*, edited by Roberto Ligrone, 129–54. Cham: Springer International Publishing.

Lindström, J. 1999. “Early Development and Fitness in Birds and Mammals.” *Trends in Ecology & Evolution* 14 (9): 343–48.

Lof, Marjolein E., Thomas E. Reed, John M. McNamara, and Marcel E. Visser. 2012. “Timing in a Fluctuating Environment: Environmental Variability and Asymmetric Fitness Curves Can Lead to Adaptively Mismatched Avian Reproduction.” *Proceedings. Biological Sciences / The Royal Society* 279 (1741): 3161–69.

Lyons, Timothy W., Christopher T. Reinhard, and Noah J. Planavsky. 2014. “The Rise of Oxygen in Earth’s Early Ocean and Atmosphere.” *Nature* 506 (7488): 307–15.

Marcil-Ferland, Dominique, Marco Festa-Bianchet, Alexandre M. Martin, and Fanie Pelletier. 2013. “Despite Catch-Up, Prolonged Growth Has Detrimental Fitness Consequences in a Long-Lived Vertebrate.” *The American Naturalist* 182 (6): 775–85.

Marvelde, Luc, Simone L. Webber, Harro A. J. Meijer, and Marcel E. Visser. 2011. “Mismatched Reproduction Is Energetically Costly for Chick Feeding Female Great Tits.” *Functional Ecology* 25 (6): 1302–8.

Massa, Piergiorgio. 2021. “STRATEGIA RIPRODUTTIVA BET-HEDGING NEL PINGUINO REALE: UNO STUDIO SIMULATIVO FORWARD-IN-TIME”. Master dissertation. Marche Polytechnic University

Mayor, Stephen J., Robert P. Guralnick, Morgan W. Tingley, Javier Otegui, John C. Withey, Sarah C. Elmendorf, Margaret E. Andrew, Stefan Leyk, Ian S. Pearse, and David C. Schneider. 2017. “Increasing Phenological Asynchrony between Spring Green-up and Arrival of Migratory Birds.” *Scientific Reports* 7 (1): 1902.

McLean, M. A., M. J. Angilletta, and K. S. Williams. 2005. “If You Can’t Stand the Heat, Stay out of the City: Thermal Reaction Norms of Chitinolytic Fungi in an Urban Heat Island.” *Journal of Thermal Biology* 30 (5): 384–91.

Millon, Alexandre, Steve J. Petty, Brian Little, and Xavier Lambin. 2011. “Natal Conditions Alter Age-Specific Reproduction but Not Survival or Senescence in a Long-Lived Bird of Prey.” *The Journal of Animal Ecology* 80 (5): 968–75.

Moline, Mark A., Nina J. Karnovsky, Zachary Brown, George J. Divoky, Thomas K. Frazer, Charles A. Jacoby, Joseph J. Torres, and William R. Fraser. 2008. “High Latitude Changes in Ice Dynamics and Their Impact on Polar Marine Ecosystems.” *Annals of the New York Academy of Sciences* 1134: 267–319.

Monaghan, Pat. 2008. “Early Growth Conditions, Phenotypic Development and Environmental Change.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1497): 1635–45.

Montes-Hugo, Martin, Scott C. Doney, Hugh W. Ducklow, William Fraser, Douglas Martinson, Sharon E. Stammerjohn, and Oscar Schofield. 2009. “Recent Changes in Phytoplankton Communities

Associated with Rapid Regional Climate Change along the Western Antarctic Peninsula." *Science* 323 (5920): 1470–73.

Moody, Allison T., Keith A. Hobson, and Anthony J. Gaston. 2012. "High-Arctic Seabird Trophic Variation Revealed through Long-Term Isotopic Monitoring." *Journal of Ornithology / DO-G* 153 (4): 1067–78.

Moritz, Craig, James L. Patton, Chris J. Conroy, Juan L. Parra, Gary C. White, and Steven R. Beissinger. 2008. "Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA." *Science* 322 (5899): 261–64.

Mugabo, M., O. Marquis, S. Perret, and J. F. LE Galliard. 2010. "Immediate and Delayed Life History Effects Caused by Food Deprivation Early in Life in a Short-Lived Lizard." *Journal of Evolutionary Biology* 23 (9): 1886–98.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, and J. Kent. 2000. "Biodiversity Hotspots for Conservation Priorities." *Nature* 403 (6772): 853–58.

National Research Council, Division on Earth and Life Studies, Board on Earth Sciences and Resources, and Committee on Grand Research Questions in the Solid-Earth Sciences. 2008. *Origin and Evolution of Earth: Research Questions for a Changing Planet*. National Academies Press.

Nevoux, Marie, Jaume Forcada, Christophe Barbraud, John Croxall, and Henri Weimerskirchi. 2010. "Bet-Hedging Response to Environmental Variability, an Intraspecific Comparison." *Ecology* 91 (8): 2416–27.

Olsson, Olof. 1996. "Seasonal Effects of Timing and Reproduction in the King Penguin: A Unique Breeding Cycle." *Journal of Avian Biology* 27 (1): 7–14.

Olsson, Olof. 1997. "Clutch Abandonment: A State-Dependent Decision in King Penguins." *Journal of Avian Biology* 28 (3): 264–67.

Oostra, Vicencio, Marjo Saastamoinen, Bas J. Zwaan, and Christopher W. Wheat. 2018. "Strong Phenotypic Plasticity Limits Potential for Evolutionary Responses to Climate Change." *Nature Communications* 9 (1): 1005.

Oremland, Ronald S., Chad W. Saltikov, Felisa Wolfe-Simon, and John F. Stolz. 2009. "Arsenic in the Evolution of Earth and Extraterrestrial Ecosystems." *Geomicrobiology Journal* 26 (7): 522–36.

Pantalacci, Sophie, and Marie Sémon. 2015. "Transcriptomics of Developing Embryos and Organs: A Raising Tool for Evo-Devo." *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution* 324 (4): 363–71.

Park, Eunmi, Jens Hefter, Gerhard Fischer, Morten Hvitfeldt Iversen, Simon Ramondenc, Eva-Maria Nöthig, and Gesine Mollenhauer. 2019. "Seasonality of Archaeal Lipid Flux and GDGT-Based Thermometry in Sinking Particles of High-Latitude Oceans: Fram Strait (79° N) and Antarctic Polar Front (50° S)." *Biogeosciences* 16 (11): 2247–68.

Parmesan, Camille. 2006. "Ecological and Evolutionary Responses to Recent Climate Change." *Annual Review of Ecology, Evolution, and Systematics* 37: 637–69.

Parmesan, Camille, and Gary Yohe. 2003. "A Globally Coherent Fingerprint of Climate Change Impacts across Natural Systems." *Nature* 421 (6918): 37–42.

Pelletier, F., D. Garant, and A. P. Hendry. 2009. "Eco-Evolutionary Dynamics." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1523): 1483–89.

Péron, Clara, Henri Weimerskirch, and Charles-André Bost. 2012. "Projected Poleward Shift of King Penguins' (Aptenodytes Patagonicus) Foraging Range at the Crozet Islands, Southern Indian Ocean." *Proceedings. Biological Sciences / The Royal Society* 279 (1738): 2515–23.

Perrins, C. M. 1970. "The Timing of Birds' Breeding Seasons." *The Ibis* 112 (2): 242–55.

Perrins, C. M., and R. H. McCleery. 1989. "Laying Dates and Clutch Size in the Great Tit." *The Wilson*

Bulletin 101 (2): 236–53.

Peterson, Brant K., Jesse N. Weber, Emily H. Kay, Heidi S. Fisher, and Hopi E. Hoekstra. 2012. "Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species." *PloS One* 7 (5): e37135.

Petrullo, Lauren, Stan Boutin, Jeffrey E. Lane, Andrew G. McAdam, and Ben Dantzer. 2023. "Phenotype-Environment Mismatch Errors Enhance Lifetime Fitness in Wild Red Squirrels." *Science* 379 (6629): 269–72.

Philipp, Eva E. R., Wiebke Wessels, Heike Gruber, Julia Strahl, Anika E. Wagner, Insa M. A. Ernst, Gerald Rimbach, et al. 2012. "Gene Expression and Physiological Changes of Different Populations of the Long-Lived Bivalve Arctica Islandica under Low Oxygen Conditions." *PloS One* 7 (9): e44621.

Philippi, T., and J. Seger. 1989. "Hedging One's Evolutionary Bets, Revisited." *Trends in Ecology & Evolution* 4 (2): 41–44.

Pigeon, Gabriel, Leif Egil Loe, Richard Bischof, Christophe Bonenfant, Mads Forchhammer, R. Justin Irvine, Erik Ropstad, Audun Stien, Vebjørn Veiberg, and Steve Albon. 2019. "Silver Spoon Effects Are Constrained under Extreme Adult Environmental Conditions." *Ecology* 100 (12): e02886.

Pigliucci, Massimo, and Professor of Ecology and Evolution Department of Botany Massimo Pigliucci. 2001. *Phenotypic Plasticity: Beyond Nature and Nurture*. JHU Press.

Pimm, S. L., C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, P. H. Raven, C. M. Roberts, and J. O. Sexton. 2014. "The Biodiversity of Species and Their Rates of Extinction, Distribution, and Protection." *Science* 344 (6187): 1246752.

Pinsky, Malin L., Anne Maria Eikeset, Douglas J. McCauley, Jonathan L. Payne, and Jennifer M. Sunday. 2019. "Greater Vulnerability to Warming of Marine versus Terrestrial Ectotherms." *Nature* 569 (7754): 108–11.

Pirri, Federica, Lino Ometto, Silvia Fuselli, Flávia A. N. Fernandes, Lorena Ancona, Nunzio Perta, Daniele Di Marino, Céline Le Bohec, Lorenzo Zane, and Emiliano Trucchi. 2022. "Selection-Driven Adaptation to the Extreme Antarctic Environment in the Emperor Penguin." *Heredity*, October. https://doi.org/10.1038/s41437-022-00564-8.

Planton, Serge, Michel Déqué, Fabrice Chauvin, and Laurent Terray. 2008. "Expected Impacts of Climate Change on Extreme Climate Events." *Comptes Rendus: Geoscience* 340 (9): 564–74.

Pounds, J. Alan, Martín R. Bustamante, Luis A. Coloma, Jamie A. Consuegra, Michael P. L. Fogden, Pru N. Foster, Enrique La Marca, et al. 2006. "Widespread Amphibian Extinctions from Epidemic Disease Driven by Global Warming." *Nature* 439 (7073): 161–67.

Pütz, K., and Y. Cherel. 2005. "The Diving Behaviour of Brooding King Penguins (Aptenodytes Patagonicus) from the Falkland Islands: Variation in Dive Profiles and Synchronous Underwater Swimming Provide New Insights into Their Foraging Strategies." *Marine Biology* 147 (2): 281–90.

Pütz, K., R. P. Wilson, J-B Charrassin, T. Raclot, J. Lage, Y. Le Maho, M. A. M. Kierspel, B. M. Culik, and D. Adelung. 1998. "Foraging Strategy of King Penguins (Aptenodytes Patagonicus) during Summer at the Crozet Islands." *Ecology* 79 (6): 1905–21.

Pyron, R. Alexander, and Matt Pennell. 2022. "Macroevolutionary Perspectives on Anthropocene Extinction." *Biological Conservation* 274 (October): 109733.

Ramp, Christian, Julien Delarue, Per J. Palsbøll, Richard Sears, and Philip S. Hammond. 2015. "Adapting to a Warmer Ocean--Seasonal Shift of Baleen Whale Movements over Three Decades." *PloS One* 10 (3): e0121374.

Reed, Thomas E., Stephanie Jenouvrier, and Marcel E. Visser. 2013. "Phenological Mismatch Strongly Affects Individual Fitness but Not Population Demography in a Woodland Passerine." *The*

Journal of Animal Ecology 82 (1): 131–44.

Robinson, Jacqueline A., Jannikke Räikkönen, Leah M. Vucetich, John A. Vucetich, Rolf O. Peterson, Kirk E. Lohmueller, and Robert K. Wayne. 2019. “Genomic Signatures of Extensive Inbreeding in Isle Royale Wolves, a Population on the Threshold of Extinction.” *Science Advances* 5 (5): eaau0757.

Robinson, Sharon A. 2022. “Climate Change and Extreme Events Are Changing the Biology of Polar Regions.” *Global Change Biology* 28 (20): 5861–64.

Román-Palacios, Cristian, and John J. Wiens. 2020. “Recent Responses to Climate Change Reveal the Drivers of Species Extinction and Survival.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (8): 4211–17.

Saastamoinen, Marjo, Dominique van der Sterren, Nienke Vastenhout, Bas J. Zwaan, and Paul M. Brakefield. 2010. “Predictive Adaptive Responses: Condition-Dependent Impact of Adult Nutrition and Flight in the Tropical Butterfly Bicyclus Anynana.” *The American Naturalist* 176 (6): 686–98.

Sæther, Bernt-Erik, and Steinar Engen. 2015. “The Concept of Fitness in Fluctuating Environments.” *Trends in Ecology & Evolution* 30 (5): 273–81.

Saino, Nicola, Roberto Ambrosini, Diego Rubolini, Jost von Hardenberg, Antonello Provenzale, Kathrin Hüppop, Ommo Hüppop, et al. 2011. “Climate Warming, Ecological Mismatch at Arrival and Population Decline in Migratory Birds.” *Proceedings. Biological Sciences / The Royal Society* 278 (1707): 835–42.

Sanghvi, Krish, Felix Zajitschek, Maider Iglesias-Carrasco, and Megan L. Head. 2021. “Sex- and Trait-Specific Silver-Spoon Effects of Developmental Environments, on Ageing.” *Evolutionary Ecology* 35 (3): 367–85.

Sanz, Juan José, Jaime Potti, Juan Moreno, Santiago Merino, and Oscar FRÍAs. 2003. “Climate Change and Fitness Components of a Migratory Bird Breeding in the Mediterranean Region.” *Global Change Biology* 9 (3): 461–72.

Saraux, Claire, Vincent A. Viblanc, Nicolas Hanuise, Yvon Le Maho, and Céline Le Bohec. 2011. “Effects of Individual Pre-Fledging Traits and Environmental Conditions on Return Patterns in Juvenile King Penguins.” *PloS One* 6 (6): e20407.

Schneider, David P., Clara Deser, and Yuko Okumura. 2012. “An Assessment and Interpretation of the Observed Warming of West Antarctica in the Austral Spring.” *Climate Dynamics* 38 (1): 323–47.

Schweiger, Oliver, Risto K. Heikkinen, Alexander Harpke, Thomas Hickler, Stefan Klotz, Otakar Kudrna, Ingolf Kühn, Juha Pöyry, and Josef Settele. 2012. “Increasing Range Mismatching of Interacting Species under Global Change Is Related to Their Ecological Characteristics.” *Global Ecology and Biogeography: A Journal of Macroecology* 21 (1): 88–99.

Scott, M. and Lindsey, R., 2020. What’s the Hottest Earth’s Ever Been?.

Seabloom, Eric W., Andy P. Dobson, and David M. Stoms. 2002. “Extinction Rates under Nonrandom Patterns of Habitat Loss.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (17): 11229–34.

Seger, J., and H. J. Brockmann. 1987. "What is bet-hedging, in``Oxford Surveys in Evolutionary Biology'' (PH Harvey and L. Partridge, Eds.), Vol. 4." (1987): 182.

Sheriff, Michael J., G. Jim Kenagy, Melanie Richter, Trixie Lee, Øivind Tøien, Franziska Kohl, C. Loren Buck, and Brian M. Barnes. 2011. “Phenological Variation in Annual Timing of Hibernation and Breeding in Nearby Populations of Arctic Ground Squirrels.” *Proceedings. Biological Sciences / The Royal Society* 278 (1716): 2369–75.

Siepel, Adam, Gill Bejerano, Jakob S. Pedersen, Angie S. Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, et al. 2005. “Evolutionarily Conserved Elements in Vertebrate, Insect, Worm,

and Yeast Genomes." *Genome Research* 15 (8): 1034–50.

Silbereis, John C., Sirisha Pochareddy, Ying Zhu, Mingfeng Li, and Nenad Sestan. 2016. "The Cellular and Molecular Landscapes of the Developing Human Central Nervous System." *Neuron* 89 (2): 248–68.

Simons, Andrew M. 2011. "Modes of Response to Environmental Change and the Elusive Empirical Evidence for Bet Hedging." *Proceedings. Biological Sciences / The Royal Society* 278 (1712): 1601–9.

Slatkin, Montgomery. 1974. "Hedging One's Evolutionary Bets." Nature Publishing Group UK. August 1974. https://doi.org/10.1038/250704b0.

Somero, G. N. 2010. "The Physiology of Climate Change: How Potentials for Acclimatization and Genetic Adaptation Will Determine 'Winners' and 'Losers.'" *The Journal of Experimental Biology* 213 (6): 912–20.

Stanley, Steven M. 2016. "Estimates of the Magnitudes of Major Marine Mass Extinctions in Earth History." *Proceedings of the National Academy of Sciences of the United States of America* 113 (42): E6325–34.

Steig, Eric J., David P. Schneider, Scott D. Rutherford, Michael E. Mann, Josefino C. Comiso, and Drew T. Shindell. 2009. "Warming of the Antarctic Ice-Sheet Surface since the 1957 International Geophysical Year." *Nature* 457 (7228): 459–62.

Stevenson, I. R., and D. M. Bryant. 2000. "Climate Change and Constraints on Breeding." *Nature* 406 (6794): 366–67.

Stier, Antoine, Vincent A. Viblanc, Sylvie Massemin-Challet, Yves Handrich, Sandrine Zahn, Emilio R. Rojas, Claire Saraux, et al. 2014. "Starting with a Handicap: Phenotypic Differences between Early- and Late-born King Penguin Chicks and Their Survival Correlates." *Functional Ecology* 28 (3): 601–11.

Stonehouse, Bernard. 1960. "The King Penguin... of South Georgia. I. Breeding Behaviour and Development." HM Stationery Office.

Stott, Peter. 2016. "CLIMATE CHANGE. How Climate Change Affects Extreme Weather Events." *Science* 352 (6293): 1517–18.

Suttle, K. B., Meredith A. Thomsen, and Mary E. Power. 2007. "Species Interactions Reverse Grassland Responses to Changing Climate." *Science* 315 (5812): 640–42.

Taborsky, Barbara. 2006. "The Influence of Juvenile and Adult Environments on Life-History Trajectories." *Proceedings. Biological Sciences / The Royal Society* 273 (1587): 741–50.

Teixeira, João C., and Christian D. Huber. 2021. "The Inflated Significance of Neutral Genetic Diversity in Conservation Genetics." *Proceedings of the National Academy of Sciences of the United States of America* 118 (10). https://doi.org/10.1073/pnas.2015096118.

Therkildsen, Nina Overgaard, and Stephen R. Palumbi. 2017. "Practical Low-Coverage Genomewide Sequencing of Hundreds of Individually Barcoded Samples for Population and Evolutionary Genomics in Nonmodel Species." *Molecular Ecology Resources* 17 (2): 194–208.

Trathan, Phil N., Pablo García-Borboroglu, Dee Boersma, Charles-André Bost, Robert J. M. Crawford, Glenn T. Crossin, Richard J. Cuthbert, et al. 2015. "Pollution, Habitat Loss, Fishing, and Climate Change as Critical Threats to Penguins." *Conservation Biology: The Journal of the Society for Conservation Biology* 29 (1): 31–41.

Twining, Cornelia W., J. Ryan Shipley, and Blake Matthews. 2022. "Climate Change Creates Nutritional Phenological Mismatches." *Trends in Ecology & Evolution* 37 (9): 736–39.

Van Noordwijk, A. J., R. H. McCleery, and C. M. Perrins. 1995. "Selection for the Timing of Great Tit Breeding in Relation to Caterpillar Growth and Temperature." *The Journal of Animal Ecology* 64 (4): 451–58.

Veening, Jan-Willem, Wiep Klaas Smits, and Oscar P. Kuipers. 2008. “Bistability, Epigenetics, and Bet-Hedging in Bacteria.” *Annual Review of Microbiology* 62: 193–210.

Verboven, Nanette, Joost M. Tinbergen, and Simon Verhulst. 2001. “Food, Reproductive Success and Multiple Breeding in the Great Tit Parus Major.” *Ardea* 89 (2): 387–406.

Vianna, Juliana A., Flávia A. N. Fernandes, María José Frugone, Henrique V. Figueiró, Luis R. Pertierra, Daly Noll, Ke Bi, et al. 2020. “Genome-Wide Analyses Reveal Drivers of Penguin Diversification.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (36): 22303–10.

Villa Martín, Paula, Miguel A. Muñoz, and Simone Pigolotti. 2019. “Bet-Hedging Strategies in Expanding Populations.” *PLoS Computational Biology* 15 (4): e1006529.

Vincenzi, Simone, Scott Hatch, Marc Mangel, and Alexander Kitaysky. 2013. “Food Availability Affects Onset of Reproduction in a Long-Lived Seabird.” *Proceedings. Biological Sciences / The Royal Society* 280 (1760): 20130554.

Visser, Marcel E., Frank Adriaensen, Johan H. Van Balen, Jacques Blondel, André A. Dhondt, Stefan Van Dongen, Chris Du Feu, et al. 2003. “Variable Responses to Large-Scale Climate Change in European Parus Populations.” *Proceedings. Biological Sciences / The Royal Society* 270 (1513): 367–72.

Visser, Marcel E., and Christiaan Both. 2005. “Shifts in Phenology due to Global Climate Change: The Need for a Yardstick.” *Proceedings. Biological Sciences / The Royal Society* 272 (1581): 2561–69.

Visser, Marcel E., and Phillip Gienapp. 2019. “Evolutionary and Demographic Consequences of Phenological Mismatches.” *Nature Ecology & Evolution* 3 (6): 879–85.

Visser, Marcel E., Luc te Marvelde, and Marjolein E. Lof. 2012. “Adaptive Phenological Mismatches of Birds and Their Food in a Warming World.” *Journal of Ornithology / DO-G* 153 (1): 75–84.

Visser, M. E., and L. J. Holleman. 2001. “Warmer Springs Disrupt the Synchrony of Oak and Winter Moth Phenology.” *Proceedings. Biological Sciences / The Royal Society* 268 (1464): 289–94.

Visser, M. E., A. J. van Noordwijk, J. M. Tinbergen, and C. M. Lessells. 1998. “Warmer Springs Lead to Mistimed Reproduction in Great Tits (Parus Major).” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265 (1408): 1867–70.

Vitasse, Yann, Sylvain Ursenbacher, Geoffrey Klein, Thierry Bohnenstengel, Yannick Chittaro, Anne Delestrade, Christian Monnerat, et al. 2021. “Phenological and Elevational Shifts of Plants, Animals and Fungi under Climate Change in the European Alps.” *Biological Reviews of the Cambridge Philosophical Society* 96 (5): 1816–35.

Voosen, P.A., 2019. 500-million-year survey of Earth’s climate reveals dire warning for humanity. Science.

Vurusaner, Beyza, Giuseppe Poli, and Huveyda Basaga. 2012. “Tumor Suppressor Genes and ROS: Complex Networks of Interactions.” *Free Radical Biology and Medicine* 52 (1): 7–18.

Waldvogel, Ann-Marie, Barbara Feldmeyer, Gregor Rolshausen, Moises Exposito-Alonso, Christian Rellstab, Robert Kofler, Thomas Mock, et al. 2020. “Evolutionary Genomics Can Improve Prediction of Species’ Responses to Climate Change.” *Evolution Letters* 4 (1): 4–18.

Walther, Gian-Reto, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor J. C. Beebee, Jean-Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. 2002. “Ecological Responses to Recent Climate Change.” *Nature* 416 (6879): 389–95.

Wan, Xinru, Guangshun Jiang, Chuan Yan, Fangliang He, Rongsheng Wen, Jiayin Gu, Xinhai Li, Jianzhang Ma, Nils Chr Stenseth, and Zhibin Zhang. 2019. “Historical Records Reveal the Distinctive Associations of Human Disturbance and Extreme Climate Change with Local Extinction of Mammals.” *Proceedings of the National Academy of Sciences of the United States*

of America 116 (38): 19001–8.

Weelden, Celine van, Jared R. Towers, and Thijs Bosker. 2021. “Impacts of Climate Change on Cetacean Distribution, Habitat and Migration.” *Climate Change Ecology* 1 (July): 100009.

Weimerskirch, Henri, Fabrice Le Bouard, Peter G. Ryan, and C. A. Bost. 2018. “Massive Decline of the World’s Largest King Penguin Colony at Ile Aux Cochons, Crozet.” *Antarctic Science / Blackwell Scientific Publications* 30 (4): 236–42.

Weimerskirch, H., J. C. Stahl, and P. Jouventin. 1992. “The Breeding Biology and Population Dynamics of King Penguins Aptenodytes Patagonica on the Crozet Islands.” *The Ibis* 134 (2): 107–17.

Wienecke, Barbara, Graham Robertson, Roger Kirkwood, and Kieran Lawton. 2007. “Extreme Dives by Free-Ranging Emperor Penguins.” *Polar Biology* 30 (2): 133–42.

Willard, Debra A., Timme H. Donders, Tammo Reichgelt, David R. Greenwood, Francesca Sangiorgi, Francien Peterse, Klaas G. J. Nierop, Joost Frieling, Stefan Schouten, and Appy Sluijs. 2019. “Arctic Vegetation, Temperature, and Hydrology during Early Eocene Transient Global Warming Events.” *Global and Planetary Change* 178 (July): 139–52.

Wint, William. 1983. “The Role of Alternative Host-Plant Species in the Life of a Polyphagous Moth, Operophtera Brumata (Lepidoptera: Geometridae).” *The Journal of Animal Ecology* 52 (2): 439–50.

Yin, Ying, Mengyu Wu, Lejla Zubcevic, William F. Borschel, Gabriel C. Lander, and Seok-Yong Lee. 2018. “Structure of the Cold- and Menthol-Sensing Ion Channel TRPM8.” *Science* 359 (6372): 237–41.

Zahnle, K., M. Claire, and D. Catling. 2006. “The Loss of Mass-Independent Fractionation in Sulfur due to a Palaeoproterozoic Collapse of Atmospheric Methane.” *Geobiology* 4 (4): 271–83.

Zhu, Yong-Guan, Masafumi Yoshinaga, Fang-Jie Zhao, and Barry P. Rosen. 2014. “Earth Abides Arsenic Biotransformations.” *Annual Review of Earth and Planetary Sciences* 42 (May): 443–67.

Zverkov, Nikolay G., Mikhail A. Rogov, Viktor A. Zakharov, Igor G. Danilov, Dmitry V. Grigoriev, and Martin Košťák. 2023. “Northernmost Occurrences of Plesiosaurs and Turtles in the Upper Cretaceous of Eurasia.” *Cretaceous Research* 148 (August): 105537.

Appendix

Appendix General Material and Methods

RFIDeep: unfolding the potential of deep learning for radio-frequency identification (Accepted in *Methods in Ecology and Evolution* with minor revisions)

Gaël Bardon[1,2], Robin Cristofari[3], Alexander Winterl[4], Téo Barracho[2,5], Marine Benoiste[2], Claire Ceresa[1], Nicolas Chatelain[2], Julien Courtecuisse[2], Flavia A.N. Fernandes[6], Benjamin Friess[2], Michel Gauthier-Clerc[7], Jean-Paul Gendner[2], Yves Handrich[2], Aymeric Houstin[1,2,8], Adélie Krellenstein[2,4], Nicolas Lecomte[5], Charles-Edouard Salmon[9], Emiliano Trucchi[6], Benoit Vallas[2,10], Emily M. Wong[8,11], Daniel P. Zitterbart[4,8*] & Céline Le Bohec[1,2*]

1 - Centre Scientifique de Monaco, Département de Biologie Polaire, Monaco, Principality of Monaco
2 - Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France
3 - University of Turku, Turku, Turun Yliopisto, Finland
4 - Friedrich-Alexander-University Erlangen-Nürnberg, Department of Physics, Erlangen, Germany
5 - University of Moncton, Canada Research Chair in Polar and Boreal Ecology, Department of Biology, Moncton, NB, Canada
6 - Marche Polytechnic University, Department of Life and Environmental Sciences, Ancona, Italy
7 - Université de Genève, Faculté des Sciences, Genève, Switzerland
8 - Woods Hole Oceanographic Institution, Applied Ocean Physics and Engineering Department, Woods Hole, MA, USA
9 - Beefutures, Nantes, France
10 - Réserve Nationale des Terres Australes Françaises, TAAF, Saint-Pierre, La Réunion, France
11 - Stanford University, Stanford, CA, USA
*co-last authors

Correspondence:
Gaël Bardon : gael.bardon2@gmail.com
Céline Le Bohec : celine.le-bohec@cnrs.fr

Abstract

1. Automatic monitoring of wildlife is becoming a critical tool in the field of ecology. In particular, radio-frequency identification (RFID) is now a widespread technology to assess the phenology, breeding, and survival of many species. While RFID produces massive datasets, no major fast and accurate methods are yet available for this type of data processing. Deep learning approaches have been used to overcome similar problems in other scientific fields and hence might bear the potential to easily overcome analytical challenges and unlock the full potential of RFID studies.

2. Here, we present a deep learning workflow, coined "RFIDeep", to derive ecological features, such as breeding status and outcome, from RFID mark-recapture data. To demonstrate the performance of RFIDeep with complex datasets, we used long-term automatic monitoring of a long-lived seabird breeding in densely packed colonies, i.e., king penguins (*Aptenodytes patagonicus*).

3. To determine individual breeding status and phenology at the individual level and for each breeding season, we first developed a one-dimensional convolution neural network (1D-CNN) architecture. Second, to account for variance in breeding phenology and technological limitations of field data acquisition, we added a new data augmentation step mimicking a shift in breeding dates and missing RFID detections (i.e., missing recaptures). Third, to identify segments of the breeding activity used during classification, we also included a visualisation tool, allowing users to understand what is usually considered a "black box" step of deep learning. With these three steps, we achieved a high accuracy for all breeding parameters (breeding status accuracy = 95.1%; phenological accuracy = 87.2%; breeding success accuracy = 97.3%).

4. RFIDeep has unfolded the potential of artificial intelligence for tracking changes in animal populations, multiplying the benefit of automated mark-recapture monitoring of undisturbed wildlife populations. RFIDeep's code is open source to facilitate their use, adaptation, or enhancement in a wide variety of species. In addition to a tremendous time saving, our study shows the capacities of CNN models to blindly detect ecologically meaningful patterns in data through visualisation techniques seldom used in ecology.

KEYWORDS

Artificial intelligence, behaviour, machine learning, RFID, wildlife monitoring

1 | INTRODUCTION

Electronic monitoring systems have been widely used over the past two decades to better understand animal populations without human disturbance (Fagerstone & Johns, 1987; Schooley et al., 1993). Radio-frequency identification (RFID) technology allows the monitoring of uniquely identified individuals and automated recording of the presence of tagged individuals at chosen locations (Whitfield et al., 2004). By placing RFID antennas along animal paths at perches or narrow entries of the breeding site (Gendner et al., 2005; Bonter et al., 2011), individual survival and breeding rates as well as behaviour and movements, can be precisely estimated, e.g., in the classical capture-mark-recapture framework (Descamps et al., 2002, 2009; Gauthier-Clerc et al., 2004; Le Bohec et al., 2003, 2007). While RFID technology allows the recording of vast amounts of data, it also creates new challenges for data treatment, even if the data structure itself is rather simple (i.e., id, date and time, and location for each detection). Because RFID data are not directly linked with biological parameters, one of the classic approaches is human expert interpretation (Descamps et al., 2002; Afanasyev et al., 2015). Still, most of the information extraction from such detection data and the ecological interpretation is done manually, although it remains extremely time-consuming and potentially biassed by human interpretation. In addition to this difficulty in manually processing potentially large numbers of detection data, RFID data also suffer from possible missing detections (Hughes et al., 2021).

A natural solution to these challenges is the search for accurate and robust methods in the automated data processing that can mimic the behaviour of an expert analyst. Artificial intelligence has been the focus of intense methodological effort in ecology; being used to process various sources of data, including imagery, passive and active acoustic data, to detect, classify, localise, identify, estimate, and predict, at every biological scale, i.e., from individual to ecosystem (Christin et al., 2019; Pichler & Hartig, 2022). Among artificial intelligence methods, deep learning has a wide and promising scope but often lacks approachable workflows for ecologists. Convolutional Neural Networks (CNN) have been initially developed for image content classification (Krizhevsky et al., 2012), but have also been used for classifying signals (Hinton et al., 2012) such as human activity classification (Mutegeki & Han, 2020), birds vocalisation classification (Kahl et al., 2021) or marine mammal detection (Shiu et al., 2020). Yet, CNN capacities remain unexplored in numerous fields such as RFID data processing.

Recent efforts have been made to automatically infer biological patterns such as behaviour from diverse types of biologgers (Fannjiang et al., 2019; Wang, 2019). For instance, accelerometers have shown promising capacities to detect food-catching events (Brisson-Curadeau et al., 2021) or activity classification (Sakamoto et al., 2009; Jeantet et al., 2021). Yet, biologgers record time is limited because of the required trade-off between miniaturisation, storage capacity and power consumption while their impact on wildlife is still existent (Bodey et al., 2018). In contrast, passive RFID tags do not need batteries to run and can be small enough to be attached or implanted in animals for life. Although the tag moves with the animal, with RFID based mark-recapture technology (as opposed to biologging), detections occur at one or more fixed points (the antenna): it is a rather unique observation situation creating a specific challenge for data interpretation. RFID technology is also exposed to two major constraints because of the impossibility to detect multiple tags at the same time with a single antenna and the impossibility to install several antennas at the same place. By increasing probability to miss detections, this tag collision problem and reader collision problem create a trade-off between the number of deployed tags and the quality of the dataset. This leads to challenges in inferring missing detections to correct the locations and movement patterns of

individuals. Like in other automated data processes, such data imperfections need to be considered and if possible repaired with suitable algorithms.

Only a few ecological studies using RFID technology described fully automatic data treatment despite the vast amount of generated data. Recent studies have shown advances in facilitating RFID data processing, as with the R package *feedr* (LaZerte et al., 2017) allowing RFID data visualisation and pre-processing. Automatic pre-processing of RFID data has been tested with blue tits (*Cyanistes caeruleus*) by Iserbyt et al., (2018) with a video recording system and showed high accuracy in three behavioural estimates. However, such RFID data processing remained coupled with additional instrumentation and a small number of monitored individuals.

Here, we demonstrate that non-explicit detection data from fixed observation points contain comprehensive information to infer the general behavioural patterns of individuals. Taking advantage of the recent developments in deep learning methods, we developed a deep learning workflow, called "RFIDeep", to automatically extract breeding status from detection data acquired by RFID antennas using convolutional neural networks. We illustrate how deep learning methods detect biological features in RFID data with very high classification accuracy and a visualisation method not yet commonly used in ecology.

We use an "archetypal" RFID dataset to illustrate the application of RFIDeep to real-life biological data with a 20 years-long RFID detection time series collected on king penguins (*Aptenodytes patagonicus*) at Crozet Archipelago. Since 1998, RFID tags have been implanted subcutaneously in ca. 15,000 penguins, mostly of known age (ca. 11-month-old chicks), to record every transit between the colony and the sea throughout their life (Gendner et al., 2005). Unlike flipper bands used until then, which are detrimental to the individuals (Gauthier-Clerc et al., 2004; Dugger et al., 2006; Saraux et al., 2011), these RFID tags allowed a more accurate and unbiased description of the reproductive patterns of the species (Descamps et al., 2002), and of the demographic parameters (Le Bohec et al., 2007, 2008). In these previous studies, all RFID detections were manually analysed by human experts and none of them used the entire dataset of RFID-tagged penguins. Since king penguins express stereotyped movement patterns during their breeding (Descamps et al., 2002), they were good candidates for artificial intelligence classification of their detection data.

Based on field observations made between 2011 and 2019, we trained several CNN to infer RFID-tagged penguins' sex, breeding status (Breeding vs. Non-Breeding; Success vs. Failure), and breeding dates. We developed RFID-specific data augmentation steps to account for biological variance and data acquisition imperfections. We trained our classification process with field observation data and tested it using manually annotated data to compare the performance of automatic classification with the human experts' one.

We provide all source codes used in RFIDeep workflow that could be applicable for any study using RFID data acquisition and that could inspire ecologists to develop their deep learning process. Finally, a software named *Sphenotron,* developed to represent movements and locations (in or outside the breeding site) based on RFID detections, is provided with a sample dataset as an example of an RFID data visualisation method used for penguins.

2 | MATERIALS AND METHODS

2.1 | Overall structure of RFIDeep workflow

Figure 1 summarises the steps needed to classify RFID data with a deep learning framework. To illustrate a comprehensive view of the use of the RFIDeep workflow on a typical dataset of RFID data, we present an application where each step is detailed in the following sections.

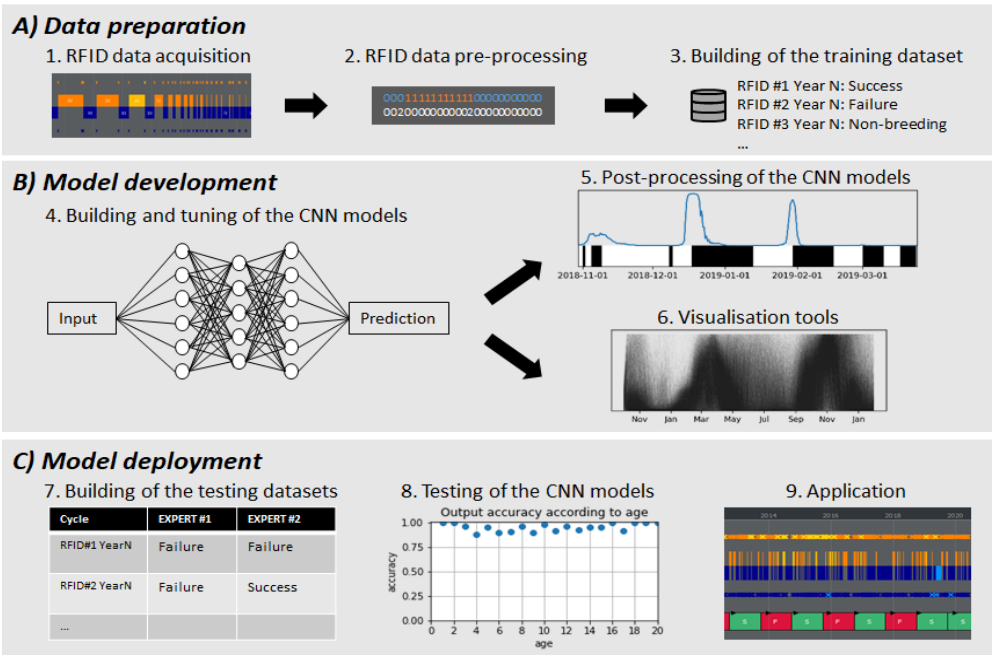


FIGURE 1 Overall structure of the RFIDeep workflow classifying RFID data with deep learning. The workflow is divided into three phases: data preparation, model development, and model deployment. **A) Data preparation.** 1) RFID data acquisition: many individuals are equipped with RFID tags and antenna systems are installed at key locations to register the detections. A software called *Sphenotron* (Supplement A) has been developed to represent detections and transitions (in or out) of RFID-tagged individuals. 2) RFID data pre-processing: a correction of missing detections is applied and data are formatted to have a unique and readable format for deep learning. 3) Building of the training dataset: direct observations of RFID-tagged individuals are used to build a ground truth dataset of labelled vectors giving the true classification. **B) Model development.** 4) Building and tuning of the CNN models: the architecture of deep learning models and hyperparameters are tuned with the training dataset. Data augmentation is implemented to cover more biological and technical variance. An individual network is built for each classification problem (e.g., breeding status, sex). 5) Post-processing of the CNN models: classification networks are derived to get other biological information needing a post-processing step such as location of stereotyped patterns in RFID data (e.g., determination of the breeding dates with a probability curve (in blue) over presence/absence pattern in black and white, respectively). 6) Visualisation tools: models are validated and interpreted with visualisation tools (e.g., with black curves representing the focus of the model during the season). **C) Model deployment.** 7) Building of the testing datasets: a testing step is used to remove biases induced during parameterisation with comparison of model classifications and manual techniques of data processing (i.e., human expert classifications). Multiple manual classifications are used to assess variability between human and automatic classification. 8) Testing of the CNN models: model tests using expert-labelled datasets assess performance but also ensure that model performances are consistent according to classes and individual characteristics (e.g., age, sex, life stage). 9) Application: classifications are applied to all detection data after pre-processing and formatting (i.e., after correction of missing detections and building of vectors), and results are represented in the *Sphenotron* for each individual (successful breeding cycles in green, failed breeding cycles in red).

2.2| Application on a seabird species long-term monitored by RFID

2.2.1 | RFID data acquisition

Here, we used data collected from the colony of king penguins (*Aptenodytes patagonicus*) named 'La Grande Manchotière' and located at Possession Island, Crozet Archipelago (46°25S, 51°45E). A sub-area of the colony of ca. 10,000 breeding pairs has been electronically monitored since 1998 with RFID technology. In 2022, four pathways between the sea and the colony (the only ways in or out of the colony) have been equipped with permanent automatic identification systems (the detailed information of the field site and systems are described in Gendner et al., 2005). In short, these automatic systems are composed of paired antennas to record the direction of each commuting bird that has been implanted with RFID tags. Patterns of presence and absence of ca. 15,000 RFID-tagged birds throughout their breeding seasons and life have then been recorded since 1998. This has generated a large (and increasing) number of detection data, with, for instance, 7 million individual detections as of 2022. To manage, visualise and use information in the field (e.g., select specific groups of birds of known age or history), we developed a python software, called *Sphenotron*, that displays the location (in or out of the colony) of the individuals during their life, based on the latest known location transition (entrance or exit) for each bird (see Figure 2). More details on the *Sphenotron* are given in Supplement A.

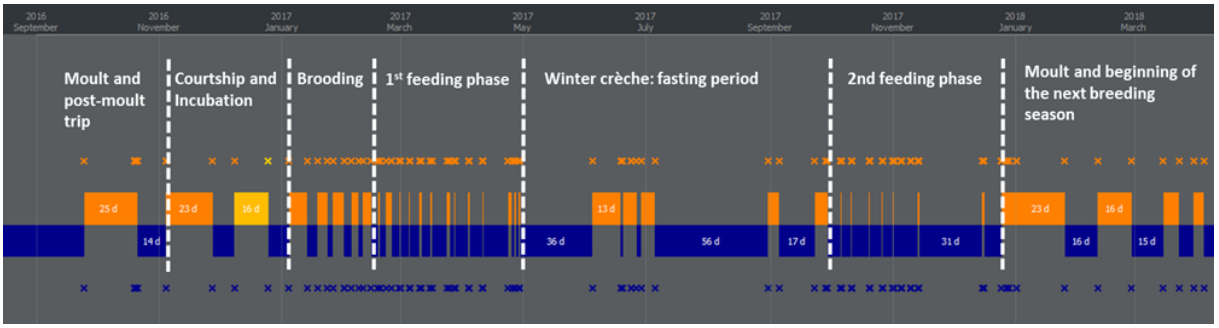


FIGURE 2 Representation of the presence/absence patterns at the breeding site of a given RFID-tagged individual for one successful breeding cycle. Each cross corresponds to one RFID detection (outside antenna in blue, and inside antenna in orange). The periods outside the colony (in blue) and inside (in orange, or in yellow after the correction of missing detections) are interpreted from the sequence of detections. The presence/absence patterns presented here correspond to the annual activities of a male king penguin but can be applied to any individual or any targeted species to understand, for instance, how long an individual stays or leaves a specific study area where RFID antennas are installed. Phases of the breeding cycle, identified from the presence/absence patterns of a bird, are indicated by the white dashed lines. The duration in or out of the colony is given in days (d).

Thanks to stereotyped presence/absence patterns at the breeding site of the targeted species and a strong knowledge of the ecology of the species (Figure 2), we can classify the breeding status of any RFID-tagged individual. For king penguins, for which a breeding cycle starts between November and March, and a successful one lasts e.g. 12 to 14 months (Barrat, 1976), three status can be listed: (1) Successful breeding with regular presence/absence patterns during the first austral summer and after the austral winter, (2) Failed breeding when at least one major pattern of (1) is lacking, and (3) Non-breeding when no regular presence/absence pattern is identified. Start of a breeding cycle (breeding date) is defined as the beginning of the stereotyped pattern characteristic of the courtship and incubation period, i.e., the first long sojourn at the colony following the annual moult (Descamps et al., 2002). Additionally, the sex of an individual can also be derived from presence/absence

patterns at the colony. An automatic sex detection has great potential application for many species where sex determination is challenging (e.g. monomorphic species like king penguins; Kriesell et al., 2018).

2.2.2 | RFID data pre-processing

Input data

To prepare the detection data in an appropriate format, we chose to represent absence and presence time-series for each breeding cycle with two vectors providing the location of the individual at the end of 12-hour periods (states 0 and 1) and the number of detections occurring during the 12 hours. For one individual and one given year *n*, we built vectors encompassing the breeding cycle. For the King penguin, vectors start October 1st of the year Y and end January 31th of the year Y+2 to cover the entire >1-year breeding cycle of the species (Figure S1). We obtained two vectors of 974 elements for each individual and each year.

Missing detection correction

To tackle missing detections that can occur when individuals exit or enter their breeding site, an algorithm has been developed to repair the simple missing detections (i.e., those when the detection of one antenna in a pair was missing, resulting in uncertainty in the individual's walking direction). These corrections are usually trivial: for example, when an individual is detected only on the inside antenna followed by an entrance (i.e., outside-inside transition), an outside detection is inferred to restore a valid pattern in detections corresponding to the missed exit from the colony. We simply built the algorithm to detect all unrealistic successions of detections and to add the corresponding missing detection in all possible cases (See Supplement A for more details).

2.2.3 | Building of the training dataset

To build a training/ground truth dataset, we visually monitored 295 RFID-tagged individuals over 9 years (2011-2019), assessing their breeding status and behaviour directly through field observations.

Birds were monitored from the beginning of the breeding season (November-January), thereby we were able to detect early breeding failures that may have been difficult to distinguish from non-breeding behaviour using RFID detections alone. Breeding outcome (S: Success; F: Failure) from these study birds was determined according to the survival of their chicks until they fledged.

The sex of individuals was determined with the observation of their first period in the colony as females leave the breeding site right after hatching, while males care for the egg (Barrat 1976). A ground truth database with breeding status, timing of breeding, and sex for 463 breeding cycles was then compiled over the years.

2.2.4 | Building and tuning of the CNN models

Overall classification workflow

Several models were built to describe breeding activities from regular ecological patterns with a classification workflow (all classification steps are shown in Figure S2):

1) two models to determine if an individual in a given year was a breeder (Breeding vs. Non Breeding) and if the breeding cycle was successful (Success vs. Failure),

2) a model to distinguish the sex of an individual through classification of male and female breeding cycles and a prediction compiling all the sexes identified over the lifetime breeding seasons,

3) a model to determine the most likely breeding date of males and females separately, through post-processing of a CNN model.

Convolution Neural Network (CNN) architecture

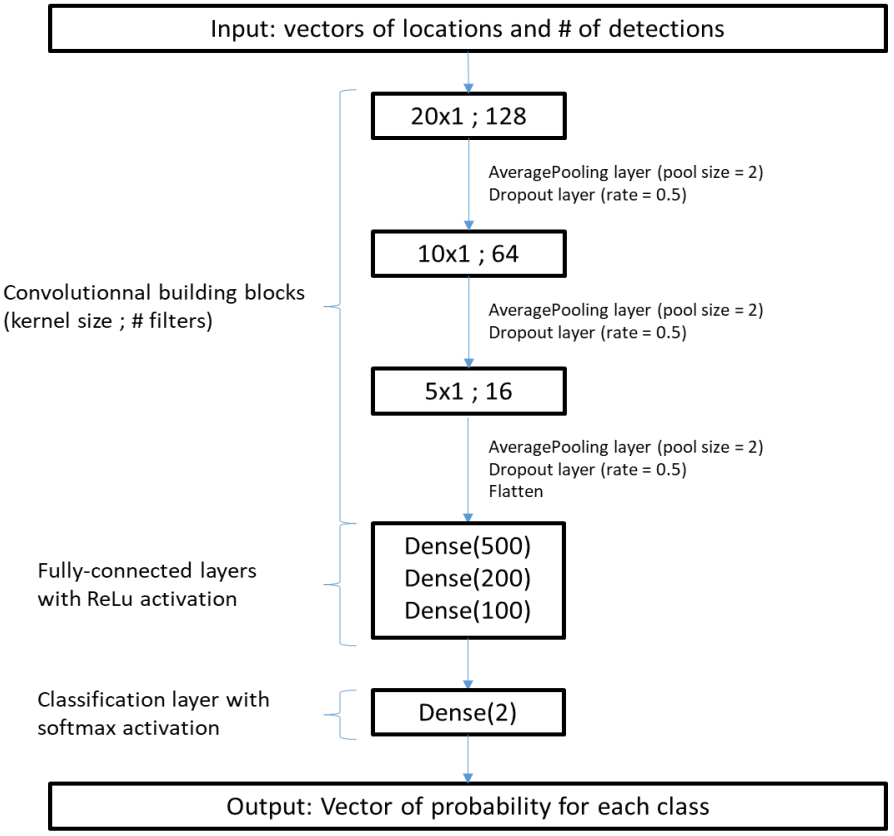


FIGURE 3 CNN architecture used for the classification model of Success vs. Failure. The CNN architecture consisted of three 1-D convolutional layers (with the number of kernels and their sizes shown in the three black boxes on the top). Here, we kept the standard 2-D nomenclature of kernel sizes, but they correspond to 1-D windows. Each convolutional layer was interleaved with a dropout layer (to avoid overfitting) and an average pooling layer (with pool size of 2) to keep the most essential elements (see LeCun et al. (2015) for details on CNN architectures). A flatten function was used at the end of convolution blocks to obtain a single 1-D vector from the previous layers. Three fully connected layers (Dense) followed the building blocks with the Rectified Linear Unit (ReLU) and were included just before the prediction layer in order to interpret the learned features. The final classification uses softmax activation functions with a fully connected layer (Dense) producing 2 output probabilities.

Several CNN models were trained to classify the detection vectors into different classes (Breeding vs. Non-Breeding; Success vs. Failure; Male vs. Female), but the same CNN architecture was used for each classification (Figure 3), meaning that the same layers were used in the same order. Only kernel sizes changed between classification with 20x1, 10x1, and 5x1 for Success vs. Failure and sex classification, and 50x1, 20x1, and 10x1 for Breeding vs. Non-Breeding classification.

Each model was trained on a training set of 80% of the dataset, and the remaining 20% was used as a validation set to measure model performances and avoid overfitting (shown by low validation accuracy and high training accuracy), as suggested by Christin et al. (2019). Multiple training of the models with randomly splitting of training/validation sets was performed to cross-validate the hyperparameters. Once the final hyperparameters were chosen, the validation accuracies with the 20% validation set were recorded and the final models were trained using 100% of the training datasets.

When the models were applied to detection vectors to generate the classifications, the most probable class was chosen for the classification.

The CNN was implemented using the Keras tensorflow framework (Abadi, 2015) in Python 3.9.7.

Data augmentation process

To extend the generalisation capacities of our models, we used a data augmentation process during the training of the models (LeCun et al., 1998). We used two types of augmentation: the first one consisted in shifting the breeding cycles by a random number of days, as usually done with imagery data to make the models translational invariant. At each iteration of the training, we shifted each training vector by a zero padding at the end or at the beginning of the vector, while trimming the same number of elements on the opposite side. We used a random offset between -30 days and 30 days to cover a large biological variability in the phenology of the birds.

The second augmentation process consisted in simulating missing RFID detections. In the actual dataset, the most frequent problem is the loss of a single detection due to a RFID-tag collision problem (two or more RFID-tags passing over an antenna at the same time), which is solved by our correction algorithm. Therefore, we chose to remove 10% of the detections at each iteration, before applying our correction algorithm, allowing a complete recovery of the original detections for at least 50% of penguins (see Supplement C) and leaving uncorrected detections and erroneous locations to improve training generality.

Models for determining the breeding status were trained with and without the data augmentation processes to assess the benefits of this step.

2.2.5 | Post-processing of the CNN models

Sex determination: With RFID detections, males and females can only be distinguished based on a few features at the beginning of the breeding cycles, therefore prediction over a single breeding season may be less reliable than prediction over lifetime breeding seasons. To increase the accuracy of sex classification of the individuals with lifetime data, we averaged the classification probabilities of each sex for the classified breeding cycle and took the maximum. Then, we obtained the most probable sex over the lifetime of the individuals, not only over separated breeding cycles.

We also used the sex classification for each breeding cycle to measure the benefit of this pooling in classification performance.

Breeding date: We used CNN models to determine the breeding date, defined as the beginning of the first long sojourn at the colony following moult, by scanning all possible breeding cycles in a year and determining the most probable one.

We trained a new CNN model (with the same architecture and hyperparameters as before) that classifies whether the detection vector is aligned to the breeding date (positive class) or not (negative class). To build the positive class, a dataset where each vector was aligned to a known breeding date was designed. We used ground truth breeding cycles with a known breeding date and truncated the detection vectors around the breeding date (30 days before and 75 days after). These vectors constituted the positive class of our training dataset with vectors aligned on the breeding date. The dataset was completed with a negative class corresponding to breeding cycles that were not aligned to the breeding date (e.g., starting at an unrealistic date): we simply took a random breeding date for each correct breeding cycle and truncated the vectors around this random date, giving us the second half part of our training dataset.

This data generation was processed at each iteration of the training to cover the maximum number of unrealistic breeding cycles possible while keeping a 50/50 ratio of positive to negative classes at each iteration. We removed the data augmentation process with shifting breeding cycles that would make our classification irrelevant.

To apply this model and obtain the most probable breeding date of a given breeding cycle, we classified detection vectors that were aligned to each 12-hour period between November 1st to April 1st, and assessed the probability of having a correctly aligned vector with the previous trained model. We obtained a certainty curve along the year, with the maximum corresponding to the most probable breeding date (as shown for example in Figure 4 with two true breeding cycles).

In our king penguin study case, we trained two different models for males and females separately to account for the difference in patterns at the beginning of their breeding cycles.

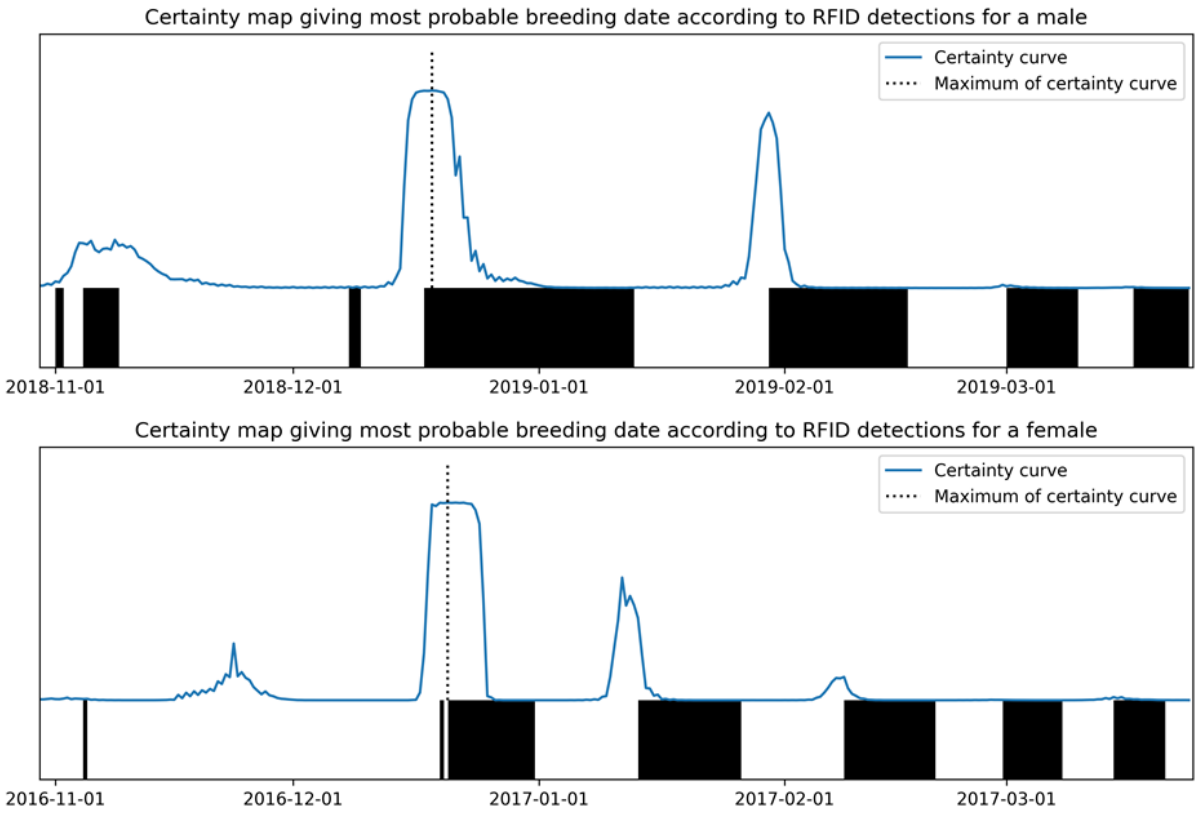


FIGURE 4 Examples of certainty maps produced by the scanning algorithm to detect the beginning of a stereotyped pattern. Here, the most probable breeding date of a male and a female was determined. The blue curve represents the probability (between 0 and 1) that the breeding cycle starts on a relevant date. The black and white bars in the lower part of the figures represent the location of the RFID-tagged individual (inside and outside, respectively). The most probable breeding date corresponds to the maximum of the blue curve (dashed line).

2.2.6 | Visualisation tools: identify important regions for decision making

We used visual explanation techniques to show parts of the input data that are identified by the convolutional layers and used to perform the classification. We leveraged techniques recently developed since CNN was first used for image classification, notably with saliency maps (Zeiler & Fergus, 2013; Simonyan et al., 2013; Springenberg et al., 2014) and class activation mapping (Zhou et al., 2016).

Standard visualisation techniques have been developed and used to produce heat maps on images classified by a 2-D CNN algorithm to show which pixels contribute most to the classification. Thus, we produced this type of visualisation on our breeding cycles to interpret the CNN decision process and to identify the critical parts of the breeding cycle from an ecological point of view. To produce heat maps on our breeding cycles, we used the GRADient-weighted Class Activation Mapping (Grad-CAM) algorithm (Selvaraju et al., 2017) that was directly applicable to the 1-D CNN layers. In short, the Grad-CAM uses the gradients of the final convolutional layer to produce a coarse localization map from an input image (or vector) by searching for pixels whose intensity should be increased to increase the probability of a given class.

We used the implementation of the Grad-CAM algorithm in *tf-keras-vis* package (https://github.com/keisen/tf-keras-vis) that worked directly for 1-D structure, and we obtained a

graph of importance value for each element of the vector (each 12-hour period in our example) for a particular class of interest (e.g., classification as a successful breeding).

We ran this algorithm on all breeding cycles in our dataset to identify the regions important to the detection vector in the decision making: we computed our activation maps 1) for the Breeding vs. Non-Breeding model with all breeding cycles classified as Breeding to identify where the algorithm was able to detect a breeding cycle, and 2) for the Success vs. Failure model with all successful breeding cycles to identify which regions of the breeding cycle indicating a success. Activation maps were then generated for both classifications and compared to the raw input detection data to highlight the critical parts of the detection vectors and then the critical biological phases of the breeding cycle.

2.2.7 | Building of the testing datasets and testing of the CNN models

Since we are interested in the overall classification performance and not the perfect classification of a specific class, the global accuracy metric was effective for comparing the performance of different models. We chose a global accuracy metric (Powers, 2020) given by: $\frac{number\ of\ correct\ predictions}{number\ of\ predictions}$

Since our ground truth datasets were well balanced across classes (168 Non-Breeding; 131 Failure; 164 Success), the global accuracy metric did not suffer from its limits with unbalanced classes and it provided a simple and effective metric of overall classification performance. To provide a measure of classification accuracy for all possible classification thresholds, we also used the AUC-ROC score (Area Under the Receiver Operating Characteristic Curve) (Fawcett, 2006).

To assess the accuracy of breeding date determination, we used a threshold of 5 days between the true breeding date and the predicted date to define whether a breeding date was correctly predicted. The breeding date is defined as the beginning of a first sojourn on the breeding site exceeding 10 days, so any date that is no more than 5 days away from this true breeding date can be easily corrected and then defined as a good prediction (see Supplement C).

To quantify an unbiased estimate of model performance, the accuracy of the classification models was then tested using a dataset not included in the training (Kuhn & Johnson, 2013). This testing dataset encompassed 917 breeding cycles of penguin individuals that were never used in model training. These breeding cycles were blind-labelled, i.e., breeding status and breeding date were not determined through field observations, but by human experts who examined the RFID detections of individuals using our custom-designed *Sphenotron* software (see Supplement A). Human experts, with a strong knowledge and experience of the species in the field, were trained using the ground truth dataset, blindly examining detection data to infer breeding cycles, and cross-checking previously assigned breeding cycles. Two human experts were chosen to label the same dataset, and we tested our models with both classifications. We also computed the global accuracy metric between the datasets labelled by the two human experts to assess human variability in classification.

The performance of the lifetime sexing method was compared to a molecular sexing dataset of 6,196 birds with the same metric (molecular sexing method adapted from Griffiths et al. (1998), see Kriesell et al. (2018)). However, the accuracy of the molecular method itself was not measured despite the known imperfections of the method (98% of accuracy). Because sex was estimated with a variable number of breeding cycles between individuals (we did indeed use all available breeding

cycles for each bird), we also tested whether the accuracy of pooled sexing increased with additional breeding cycles used for sexing.

Finally, we computed the accuracy of the models for each age class and for males and females separately to test whether the performance of our models was consistent over the whole dataset.

3 | RESULTS

3.1 | Training of the models

We chose 200 epochs (i.e., training iterations) for training of each model, which yielded the best results for validating model accuracy while avoiding overfitting. Each model took approximately 1 hour to train using a CPU Intel Core i7-10750H (2.60GHz) and a GPU Nvidia GeForce GTX 1660 Ti, a non-prohibitive technology for many.

The performance of models, according to the validation datasets used to select the CNN architecture and hyperparameters, reached near perfection for the three models, i.e., Breeding vs. Non Breeding, Success vs. Failure, and Male vs. Female, with global accuracy of 99.1%, 99.7%, and 100%, respectively.

As expected, the three models without a data augmentation step achieved lower performances with global accuracy of 94.6%, 91.5%, and 96.6% for Breeding vs. Non Breeding, Success vs. Failure, and Male vs. Female, respectively.

3.2 | Visualisation of the models

The activation maps $L^{c}_{grad-CAM}$ of visualisation techniques are given in Figure 5 with an example for the breeding class (Figure 5A) and the breeding outcome class (Figure 5B). The blue curve shows the weights applied to the elements of the breeding cycle during the last convolutional layer of the CNN: it detects parts of the breeding cycle that can be used by the CNN procedure to produce the classifications. It also displays where the differences in the breeding cycles lie for various classes. The median maps (Figures 5C and 5D) illustrate the maps for all the breeding cycles classified as Breeding and as Success, respectively.

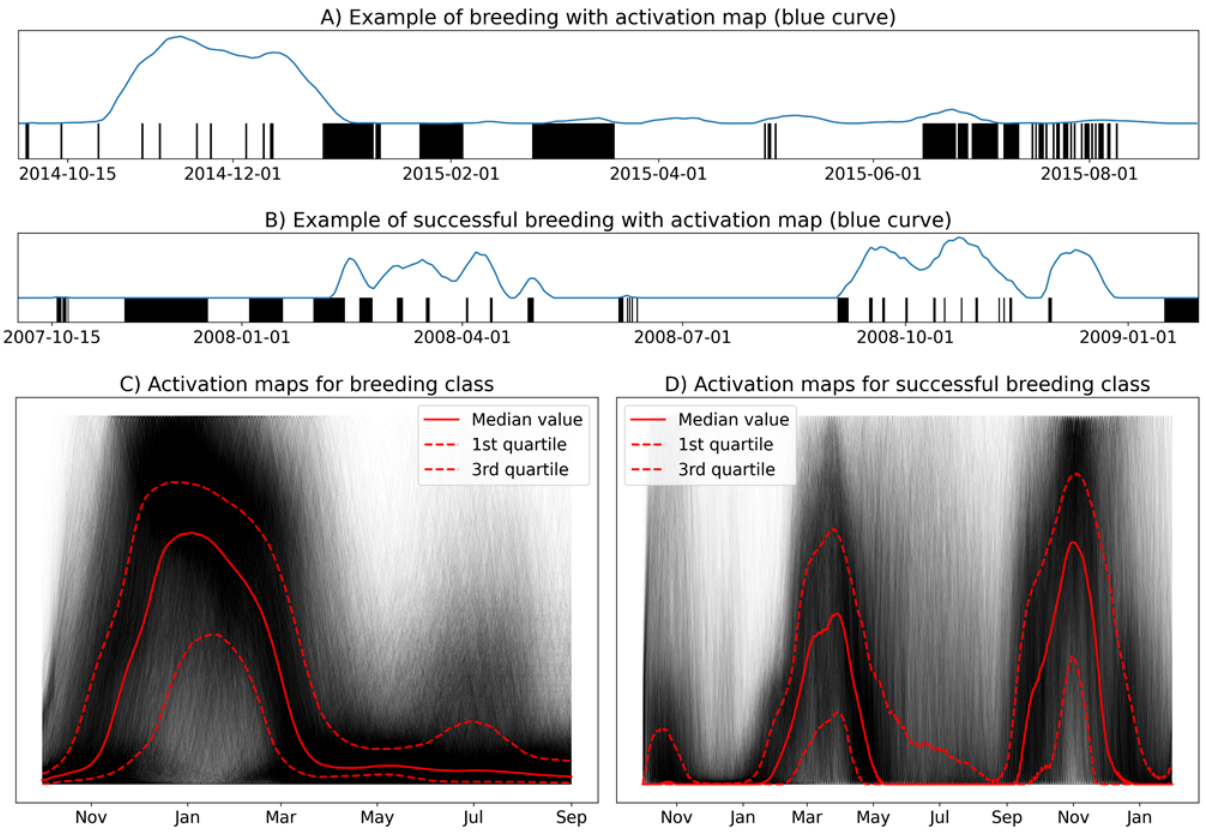


FIGURE 5 Activation map (blue curve) and simplified presence/absence pattern (black: inside; white: outside) for two true breeding cycles (A and B), and median maps (C and D) illustrating all maps with the median curve, the first quartile, and the third quartile. A and C correspond to the Breeding class (all breeding cycles classified as Breeding) and to the Success class (all breeding cycles classified as Success), respectively.

For the vectors classified in the Breeding class (i.e., individuals that attempt to breed) (Figures 5A and 5C), the model focused on the beginning of breeding, when long periods in the colony occur (approximately in January and February). This indeed corresponds to the incubation phase, which occurs for both successful and failed breeding, but not for non-breeding birds that do not have long fasting periods on the breeding site.

For the Success class (Figures 5B and 5D), the model focused on two regions: the pre-winter period, when chicks are fed alternately by both parents, and the post-winter feeding period. As expected, these are the parts of the breeding cycle that can be missing if the breeding fails during incubation, brooding, or even during the winter fasting period.

As expected, the visualisation maps relied on the same regions that human experts have used as criteria for determining whether individuals actually attempted to breed and succeeded to breed.

3.3 | Model deployment

The trained models were used to predict the breeding status and dates of all RFID-tagged individuals since 1998 (i.e., 79,805 breeding cycles from 14,111 different individuals). On the laptop computer used here, prediction (from raw RFID data to classification) of breeding status and sex of all birds required 140 seconds, but it took 1.1 hours for the determination of the breeding date due to

the number of predictions needed (320 for each breeding cycle, i.e., 160 days). In comparison, it took a human about 1 minute to make the same decision as RFIDeep for one bird and one breeding cycle, which corresponded to 1320 hours or 165 work days (8h per day) to classify all breeding cycles.

We tested our model's classifications against two human expert classification datasets. These datasets were well balanced across classes, with about 50% of Failure and 25% of Success and Non breeding. Testing accuracy using these datasets that were never used in model training was carried out with the global accuracy metric (Table 1) and AUC-ROC (Figure S3).

TABLE 1 Results of the comparison between the predictions yielded by RFIDeep and two datasets labelled by human experts. For each model (with and without data augmentation 'DA' procedure), global accuracy metrics between CNN predictions and the two datasets labelled by human experts are given, as well as the global accuracy between the two datasets. These accuracy metrics are given for the classifications of Breeding vs. Non-Breeding (B vs. NB), Success vs. Failure (S vs. F), compiled Non-Breeding vs. Failure vs. Success (NB vs. F vs. S), and Breeding date.

	B vs. NB	**B vs. NB (without DA)**	**S vs. F**	**S vs. F (without DA)**	**Compiled NB vs. F vs. S**	**Compiled NB vs. F vs. S (without DA)**	**Breeding date**
Prediction vs. Dataset 1	0.931	0.904	0.973	0.945	0.910	0.828	0.814
Prediction vs. Dataset 2	0.951	0.938	0.967	0.898	0.938	0.863	0.872
Dataset 1 vs Dataset 2	0.951	0.951	0.975	0.975	0.932	0.932	0.858

As expected, models with data augmentation consistently performed better than models without any transformation of the input data, highlighting the benefits of the augmentation process to cover more biological variability than the one present in the ground truth data, and to compete with human performance.

Furthermore, the similarities between the expert-labelled datasets were globally equivalent to the accuracy of our CNN models, indicating the high performance of the automatic classification procedure. The AUC scores of Breeding vs. Non-Breeding and Success vs. Failure computed with the human expert classification were even higher (e.g., $AUC_{B\ vs.NB} = 0.993$ and $AUC_{S\ vs.F} = 0.992$ for Prediction vs. Dataset 2, see Figure S3).

The lifetime classification sexing procedure yielded an accuracy of 88.5% compared to the molecular data. Before pooling lifetime sex probabilities, the global accuracy of sexing was only 81.7%. The AUC-ROC score also yielded a high accuracy of 93.0%. As expected, we also found that sex classification accuracy from the pooling of lifetime sex probabilities increased with the number of breeding cycles used to determine the sex an individual (see Supplement C).

Age- and sex-specific performances were also computed with global accuracy metrics (Figure S4). Predictions were slightly better for males than for females for the breeding status (males: 94.5% vs. females: 93.2%) and the breeding dates (males: 89.4% vs. females: 83.7%). The breeding dates also appeared to be less predictable for young individuals (see Figure S4).

4 | DISCUSSION

In this study, we developed, tested, and provided a complete workflow based on CNN models to automatically infer behavioural and fitness traits from RFID-tagged animal detection data. Based on a train-test split approach (i.e., a classical 20% validation dataset during the training, and a testing dataset never used in training), we showcased the potential of deep learning to adequately replace human expertise in RFID data processing in a much shorter time span. Remarkably, human-like performance to translate patterns from detection data into biological parameters was reached with a rather simple CNN architecture and a standard desktop computing power. To improve results, we used time-shift data augmentation to mimic the variability that could occur due to biological mechanisms (e.g., a shift in breeding dates) and simulated data dropouts to mimic technical constraints (e.g., missing detections). We also developed a post-processing step of CNN models to extract dates of breeding and, with a visualisation technique, we identified the regions of the dataset used by the models to classify the breeding cycles. We argue that such a framework can be used beyond our example dataset, and help to quickly classify the breeding activities of many individuals, even more so for long-term projects for which pre-processing analysis is very time- and labor-consuming (in our example, we worked on ca. 15,000 individuals over 20 years).

While it is still challenging to successfully transfer pre-trained deep learning from a study case to another (Marcus, 2018), RFIDeep workflow is tailored for any study classifying behaviours based on RFID-tagged animal detections. RFIDeep was successfully tested and used on another species, the Adélie penguin (*Pygoscelis adeliae),* for which breeding is markedly different from our first dataset example, yet monitored with a similar automatic RFID setup. Given that our model performed well for these contrasting datasets (see Supplement E for details), we argue that any RFID-monitored species with stereotyped movements during a given life stage could certainly benefit from the RFIDeep workflow, such as bumblebees (Molet et al., 2008), Leach's Storm-petrels (Zangmeister et al., 2009), hummingbirds (Bandivadekar et al., 2018), as well as other penguin species (Kerry et al., 1993; Ballard et al., 2001; Ballerini et al., 2009; Chiaradia et al., 1999; Horswill et al., 2014). Furthermore, the missing detection correction and data augmentation algorithms implemented in RFIDeep have great potential to tackle uncompleted and/or low-quality datasets, such as those produced by mobile RFID antennas temporarily deployed (Cristofari et al., 2018). We are confident that the RFIDeep workflow will help biologists to more easily adopt deep learning applications, either by using the software directly or by adapting it for their requirements.

Both validation and testing steps showed the high performances of the RFIDeep models, on the one hand, in reference to the ground truth data and, on the other hand, from a human-machine comparison point of view. Even though we developed a software to efficiently display detections and locations (inside or outside the colony) of RFID-tagged individuals during their life (*Sphenotron*, see Supplement A), the distinction between specific breeding status can be challenging, if not impossible, like in our case between non-breeding and failed breeding when the failure occurs very early in the season. By using automatic classification, we standardised the bias among all breeding classifications throughout the years of monitoring by removing variability related to potential differences in human expert interpretation. It allows for remarkably fast extraction of life history parameters of the monitored individuals, necessary to estimate population vital rates (e.g., survival, breeding success) and viability, in addition to other breeding and/or phenological traits. For example, breeding success with a very good classification accuracy (97.3% of accuracy in the

classification of successful vs. non- or failed breeding) can then be used to estimate fecundity rates of the monitored population with high confidence in estimates for all monitored years.

Our analysis highlighted the benefits of data augmentation to cope with more biological variance than that contained in our ground truth data. This part of the process, commonly used to improve deep learning application (Taylor, 2018) and sometimes developed in the application of deep learning in ecology (e.g., with image data (Kalin et al., 2018) or audio data (Kahl et al., 2021)), has significantly enhanced our classification process. While data augmentation is usually done by adding random noise to the dataset (e.g., in pictures for 2-D CNN classification with image rotations for instance, Pawara et al., 2017), here we aimed to mimic biological variance and technical limitations of the RFID data acquisition systems. Doing so, we covered a large variance in breeding dates, enabling us to anticipate breeding seasons that could begin earlier or later than those existing in our ground truth data, as a consequence of environmental shifts already observed or expected in the coming years/decades (Visser et al. 2021). It also highlighted the fact that it is not critical to have a ground truth dataset that does not fully cover the entire biological variability and that enlarging the training dataset by simulating the missing variability improves predictive power. This applies not only for the Sphenisciformes species used in this study but likely to other species with a high variance in breeding phenology (de Villemereuil et al., 2020).

Another interesting aspect of the automatic classification of breeding cycles is the independence among predictions. Indeed, each breeding cycle was analysed without supplement information about the year (e.g., average breeding success, phenological data), the individual (e.g., age, body condition), and/or previous and future breeding cycles. The breeding classification of lifetime datasets by human experts can induce bias for quantifying the inter-individual and intra-individual heterogeneity in breeding cycles since they are not classified independently. However, while there may be an advantage to having independent classifications, the lifetime information may also be beneficial, for instance to better determine the breeding date of the very first breeding seasons that tend to be less predictable for numerous species (see Figure S4). It would also be useful to train CNN models with mixed data (e.g., RFID detections and weights at the detection) to increase the classification accuracy and/or complexity to refine further some of the analyses (e.g., the stage of breeding failure), as it has also been done in other fields (Ahsan, 2020).

With visualisation techniques (e.g., see Figure 5), we showed which parts of the datasets are mostly used to perform classification by providing a peek into the deep learning ‘black box’, making the process more transparent for the user, a shortcoming that often prevents its use by ecologists (Borowiec et al., 2021). We argue that such a step can help expand the potential of deep learning to describe and analyse ecological big data. In our example, while activation maps are primarily used by CNN for classification, their visualisation allows the detection of the particular breeding activities or features, such as seasonal phenology. In our application in king penguins, the CNN models showed that the presence or absence of pre- and post-winter chick feeding patterns were the most important criteria for predicting breeding outcome. Although it is clear that these regions can be used to distinguish between failure and success, it reinforces our interest in using these visualisation techniques not only to understand how our deep learning models work, but also to detect regions of interest in our datasets. It also highlights the use of CNN models that are not frequently found in ecological studies but have great potential, for instance, to detect hidden patterns in large datasets. Moreover, to cope with the recent explosion of big data acquisition due to increasingly sophisticated, miniaturised, autonomous, and powerful data collection instruments (Williams et al., 2020), visualisation tools are critical and they could be powerful in detecting similar patterns in

given classes or differences between similar classes. For instance, identifying parts of the vocalisation essential to distinguish between species or even individuals is key in bioacoustic studies (Stowell et al., 2016; Kobayashi et al., 2021). Visualisation techniques have also been used to select the most informative variables to infer animal behaviours from multi-sensor data (in green turtles (*Chelonia mydas*), Jeantet et al., 2021) or to select the most relevant morphological characters to identify species (among midges (Milošević et al., 2020) and mosquitoes (Park et al., 2020)).

By developing tools to help users unleash the vast potential of machine learning in ecology and to increase numerous benefits of RFID technology, we also aim with RFIDeep to foster low-impact monitoring of sensitive species by reducing human presence and intervention in wild habitats (Hughes et al., 2021; Rafiq et al., 2021; Harrison & Kelly, 2022). We are convinced that combining automatic data collection and real-time data analysis and storage will help secure key ecological information over time necessary to continuously monitor the health of wild populations and their ecosystems.

ACKNOWLEDGEMENTS

This study was supported by the Institut Polaire Français Paul-Emile Victor (IPEV) within the framework of the Program 137-ANTAVIA, by the Centre Scientifique de Monaco with additional support from the LIA-647 and RTPI-NUTRESS (CSM/CNRS-UNISTRA), by the Centre National de la Recherche Scientifique (CNRS) through the Programme Zone Atelier de Recherches sur l'Environnement Antarctique et Subantarctique (ZATA), and by the Deutsche Forschungsgemeinschaft (DFG) grants FA336/5-1 and ZI1525/3-1 in the framework of the priority program ‘‘Antarctic research with comparative investigations in Arctic ice areas”. This study was approved by the French ethics committee (last: APAFIS#29338-2020070210516365) and the French Polar Environmental Committee, and permits to handle animals and access breeding sites were delivered by the “Terres Australes et Antarctiques Françaises” (TAAF). We are deeply grateful to all the wintering and summering members of Program 137 since the beginning of the data collection (1997/1998), all the members of the missions in Crozet and Dumont d'Urville since then, Denis Allemand, Yvon Le Maho, Victor Planas-Bielsa, Claire Saraux, and all the other colleagues and students within the team who have contributed over the last 20 years to the improvement of hardware, software and databases. We also sincerely thank the IPEV logistics team in Crozet, Kerguelen, and Dumont d'Urville for their important and continued support in the field.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR' CONTRIBUTIONS

G.B., R.C., N.C., M.G.-C., J.-P.G., Y.H., C.-E.S, D.P.Z., C.L.B. conceived the ideas and designed methodology; G.B., R.C., T.B., M.B., N.C., F.A.N.F., M.G.-C., A.H., A.K., N.L., C.-E.S, E.T., B.V., C.L.B. collected the data; G.B., R.C., T.B., M.B., C.C., F.A.N.F., M.G.-C., A.H., C.-E.S, B.V., C.L.B. cured and/or analysed the data; G.B., R.C., A.W., C.C., J.C., B.F., A.H., C.-E.S, E.M.W., D.P.Z., C.L.B. wrote codes for

RFIDeep and software; R.C., N.C., J.C., B.F., M.G.-C., J.-P.G., Y.H., D.P.Z., C.L.B. developed the hardware; Project administration and supervision: C.L.B.; G.B. and C.L.B. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The codes used to build the models and use them on a sample dataset are accessible at the following location: https://github.com/g-bardon/RFIDeep

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems.

Ahsan, M. M., E. Alam, T., Trafalis, T., & Huebner, P. (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. Symmetry, 12(9), 1526.

Ballard, G., Ainley, D. G., Ribic, C. A., & Barton, K. R. (2001). Effect of instrument attachment and other factors on foraging trip duration and nesting success of Adélie penguins. The Condor, 103(3), 481-490.

Ballerini, T., Tavecchia, G., Olmastroni, S., Pezzo, F., & Focardi, S. (2009). Nonlinear effects of winter sea ice on the survival probabilities of Adélie penguins. Oecologia, 161(2), 253-265.

Bandivadekar, R. R., Pandit, P. S., Sollmann, R., Thomas, M. J., Logan, S. M., Brown, J. C., ... & Tell, L. A. (2018). Use of RFID technology to characterize feeder visitations and contact network of hummingbirds in urban habitats. PloS one, 13(12), e0208057.

Barrat, A. (1976). Quelques aspects de la biologie et de l'écologie du Manchot Royal (*Aptenodytes patagonicus*) des îles Crozet. Com. Natl Fr. Rech. Antarct. 40, 9–51

Bodey, T. W., Cleasby, I. R., Bell, F., Parr, N., Schultz, A., Votier, S. C., & Bearhop, S. (2018). A phylogenetically controlled meta-analysis of biologging device effects on birds: Deleterious effects and a call for more standardized reporting of study data. Methods in Ecology and Evolution, 9(4), 946-955.

Bonter, D.N. and Bridge, E.S. (2011), Applications of radio frequency identification (RFID) in ornithological research: a review. Journal of Field Ornithology, 82: 1-10.

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. Methods in Ecology and Evolution, 13(8), 1640-1660.

Brisson-Curadeau, É., Handrich, Y., Elliott, K. H., & Bost, C.-A. (2021). Accelerometry predicts prey-capture rates in the deep-diving king penguin *Aptenodytes patagonicus*. Marine Biology, 168(10), 156.

Chiaradia, A. F., & Kerry, K. R. (1999). Daily nest attendance and breeding performance in the little penguin Eudyptula minor at Phillip Island, Australia. Marine Ornithology, 27, 13-20.

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. Methods in Ecology and Evolution, 10(10), 1632-1644.

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., ... & Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. Nature Climate Change, 8(3), 245-251.

de Villemereuil, P., Charmantier, A., Arlt, D., Bize, P., Brekke, P., Brouwer, L., Cockburn, A., Côté, S. D., Dobson, F. S., Evans, S. R., Festa-Bianchet, M., Gamelon, M., Hamel, S., Hegelbach, J., Jerstad, K., Kempenaers, B., Kruuk, L. E. B., Kumpula, J., Kvalnes, T., McAdam, A. G., ... Chevin, L. M. (2020). Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. Proceedings of the National Academy of Sciences of the United States of America, 117(50), 31969–31978.

Descamps, S & Gauthier-Clerc, M. & Gendner, JP & Le Maho, Y. (2002). The annual cycle of unbanded king penguins *Aptenodytes patagonicus* on Possession Island (Crozet). Avian Science. 2. 87-98.

Dugger, K., Ballard, G., Ainley, D. & Barton, K. Effects of flipper-bands on foraging behavior and survival of Adélie penguins (*Pygoscelis Adeliae*). Auk 123, 858–869 (2006).

Fagerstone, K. A., & Johns, B. E. (1987). Transponders as Permanent Identification Markers for Domestic Ferrets, Black-Footed Ferrets, and Other Wildlife. The Journal of Wildlife Management, 51(2), 294.

Fannjiang, C., Mooney, T. A., Cones, S., Mann, D., Shorter, K. A., & Katija, K. (2019). Augmenting biologging with supervised machine learning to study in situ behavior of the medusa Chrysaora fuscescens. Journal of Experimental Biology, 222(16), jeb207654.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

Gauthier-Clerc, M., Gendner, J. P., Ribic, C. A., Fraser, W. R., Woehler, E. J., Descamps, S., Gilly, C., Le Bohec, C., & Le Maho, Y. (2004). Long-term effects of flipper bands on penguins. Proceedings. Biological sciences, 271 Suppl 6(Suppl 6), S423–S426.

Gendner, J. P., Gauthier-Clerc, M., Le Bohec, C., Descamps, S., & Le Maho, Y. (2005). A new application for transponders in studying penguins. Journal of Field Ornithology, 76(2), 138-142.

Hinton, G., Deng, L. I., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., ...Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82–97.

Gibbons, W. J., & Andrews, K. M. (2004). PIT tagging: simple technology at its best. Bioscience, 54(5), 447-454.

Goldwater, M., Zitterbart D. P., Wright, D., & Bonnel, J. (2023). Machine-learning-based simultaneous detection and ranging of impulsive baleen whale vocalizations using a single hydrophone. The Journal of the Acoustical Society of America 153, 1094-1107

Harrison, N. D., & Kelly, E. L. (2022). Affordable RFID loggers for monitoring animal movement, activity, and behaviour. Plos one, 17(10), e0276388.

Horswill, C., Matthiopoulos, J., Green, J. A., Meredith, M. P., Forcada, J., Peat, H., ... & Ratcliffe, N. (2014). Survival in macaroni penguins and the relative importance of different drivers: individual traits, predation pressure and environmental variability. Journal of Animal Ecology, 83(5), 1057-1067.

Hughes, E. J., Mady, R. P., & Bonter, D. N. (2021). Evaluating the accuracy and biological meaning of visits to RFID-enabled bird feeders using video. Ecology and Evolution, 11(23), 17132-17141.

Jeantet, L., Vigon, V., Geiger, S., & Chevallier, D. (2021). Fully Convolutional Neural Network: A solution to infer animal behaviours from multi-sensor data. Ecological Modelling, 450, 109555.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, 61, 101236.

Kälin, U., Lang, N., Hug, C., Gessler, A., & Wegner, J. D. (2019). Defoliation estimation of forest trees from ground-level images. Remote Sensing of Environment, 223, 143-153.

Kerry, K., Clarke, J., & Grant, E. L. S. E. (1993). The use of an automated weighing and recording system for the study of the biology of adelie penguins (*Pygoscelis adeliae*) (14th Symposium on Polar Biology).

Kobayashi, K., Masuda, K., Haga, C., Matsui, T., Fukui, D., & Machimura, T. (2021). Development of a species identification system of Japanese bats from echolocation calls using convolutional neural networks. Ecological Informatics, 62, 101253.

Kriesell, H. J., Aubin, T., Planas-Bielsa, V., Benoiste, M., Bonadonna, F., Gachot-Neveu, H., ... & Le Bohec, C. (2018). Sex identification in king penguins Aptenodytes patagonicus through morphological and acoustic cues. Ibis, 160(4), 755-768.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Le Bohec, C., Gauthier-Clerc, M., Grémillet, D., Pradel, R., Béchet, A., Gendner, J. P., & Le Maho, Y. (2007). Population dynamics in a long-lived seabird: I. Impact of breeding activity on survival and breeding probability in unbanded king penguins. The Journal of animal ecology, 76(6), 1149–1160.

Le Bohec, C., Durant, J. M., Gauthier-Clerc, M., Stenseth, N. C., Park, Y. H., Pradel, R., Grémillet, D., Gendner, J. P., & Le Maho, Y. (2008). King penguin population threatened by Southern Ocean warming. Proceedings of the National Academy of Sciences of the United States of America, 105(7), 2493–2497.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, 2.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278{2324.

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).

Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.

Milošević, D., Milosavljević, A., Predić, B., Medeiros, A. S., Savić-Zdravković, D., Piperac, M. S., ... & Leese, F. (2020). Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. Science of the Total Environment, 711, 135160.

Molet, M., Chittka, L., Stelzer, R. J., Streit, S., & Raine, N. E. (2008). Colony nutritional status modulates worker responses to foraging recruitment pheromone in the bumblebee *Bombus terrestris*. Behavioral Ecology and Sociobiology, 62(12), 1919-1926.

Mutegeki, R., & Han, D. S. (2020, February). A CNN-LSTM approach to human activity recognition. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 362-366). IEEE.

Park, J., Kim, D. I., Choi, B., Kang, W., & Kwon, H. W. (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. Scientific reports, 10(1), 1012.

Pawara, P., Okafor, E., Schomaker, L., & Wiering, M. (2017). Data augmentation for plant classification. In Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18 (pp. 615-626). Springer International Publishing.

Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. Methods in Ecology and Evolution, 00, 1– 23.

Powers, David MW. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." arXiv preprint arXiv:2010.16061 (2020).

Rafiq, K., Appleby, R. G., Edgar, J. P., Radford, C., Smith, B. P., Jordan, N. R., ... & Cochrane, M. (2021). WildWID: An open‐source active RFID system for wildlife research. Methods in Ecology and Evolution, 12(9), 1580-1587.

Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell **1,** 206–215 (2019).

Sakamoto, K. Q., Sato, K., Ishizuka, M., Watanuki, Y., Takahashi, A., Daunt, F., & Wanless, S. (2009). Can Ethograms Be Automatically Generated Using Body Acceleration Data from Free-Ranging Birds? PLOS ONE, 4(4), e5379.

Saraux, C., Le Bohec, C., Durant, J. *et al.* Reliability of flipper-banded penguins as indicators of climate change. Nature **469,** 203–206 (2011).

Schooley, R. L., Van Horne, B., & Burnham, K. P. (1993). Passive integrated transponders for marking free-ranging Townsend's ground squirrels. Journal of Mammalogy, 74(2), 480-484.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Shiu, Y., Palmer, K. J., Roch, M. A., Fleishman, E., Liu, X., Nosal, E. M., ... & Klinck, H. (2020). Deep neural networks for automated detection of marine mammal species. Scientific reports, 10(1), 1-12.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806.

Stonehouse, B. (1960). The king penguin Aptenodytes patagonica of South Georgia. I.Breeding behaviour and development. Falkl. Isl. Depend. Surv. Sci. Rep. 23, 1–81.

Stowell, D., Morfi, V., Gill, L.F. (2016) Individual Identity in Songbirds: Signal Representations and Metric Learning for Locating the Information in Complex Corvid Calls. Proc. Interspeech 2016, 2607-2611

Taylor, L., & Nitschke, G. (2018, November). Improving deep learning with generic data augmentation. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1542-1547). IEEE.

Visser, M. E., Lindner, M., Gienapp, P., Long, M. C., & Jenouvrier, S. (2021). Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (*Parus major*). Proceedings of the Royal Society B, 288(1963), 20211337.

Wang, G. (2019). Machine learning for inferring animal behavior from location and movement data. Ecological informatics, 49, 69-76.

Weimerskirch, H., Stahl, J. C., & Jouventin, P. (1992). The breeding biology and population dynamics of king penguins Aptenodytes patagonica on the Crozet Islands. Ibis, 134(2), 107-117.

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., ... & Börger, L. (2020). Optimizing the use of biologgers for movement ecology research. Journal of Animal Ecology, 89(1), 186-206.

Zangmeister, J. L., Haussmann, M. F., Cerchiara, J., & Mauck, R. A. (2009). Incubation failure and nest abandonment by Leach's Storm-Petrels detected using PIT tags and temperature loggers. Journal of Field Ornithology, 80(4), 373-379.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

SUPPORTING INFORMATION

Supplement A: *Sphenotron* software

Sphenotron is open-source software written in Python. The current version of *Sphenotron* has been developed to interact with MySQL databases composed of several tables compiling all known information on each RFID-tagged individual, such as biometric or phenological data, sampling and recapture events if any, past breeding territories/coordinates, etc., in addition to detections (see Figure SI A1). *Sphenotron* was initially developed to interact with databases related to specific species (order Sphenisciformes). Yet, the code can be reused and adjusted to manage and interact with databases of similar or different formats of other species. The complete Sphenotron software and codes, with examples of database, are downloadable at the following link (https://github.com/g-bardon/RFIDeep), and can be fully modified for a wide range of species (e.g., sea or terrestrial birds or mammals) and monitoring scheme (e.g., with and without mass tracking).

With automatic data pre-processing and analysis, *Sphenotron* allows the organisation, aggregation, management, and storage of biological time series in near real time. The development, improvement and tests have been implemented to long-term monitored penguin populations since 2002. With the development of this novel interface, we aim to optimise data reuse following the FAIR principles (Wilkinson, 2016). In addition to automating the time-consuming pre-processing of detection data into biologically meaningful data (e.g., breeding outcome, sexing), the other advantage of this automatisation is to assist scientists in the field by accessing information from RFID-tagged individuals in order to conduct specific experiments or observations on selected individuals with the desired characteristics.

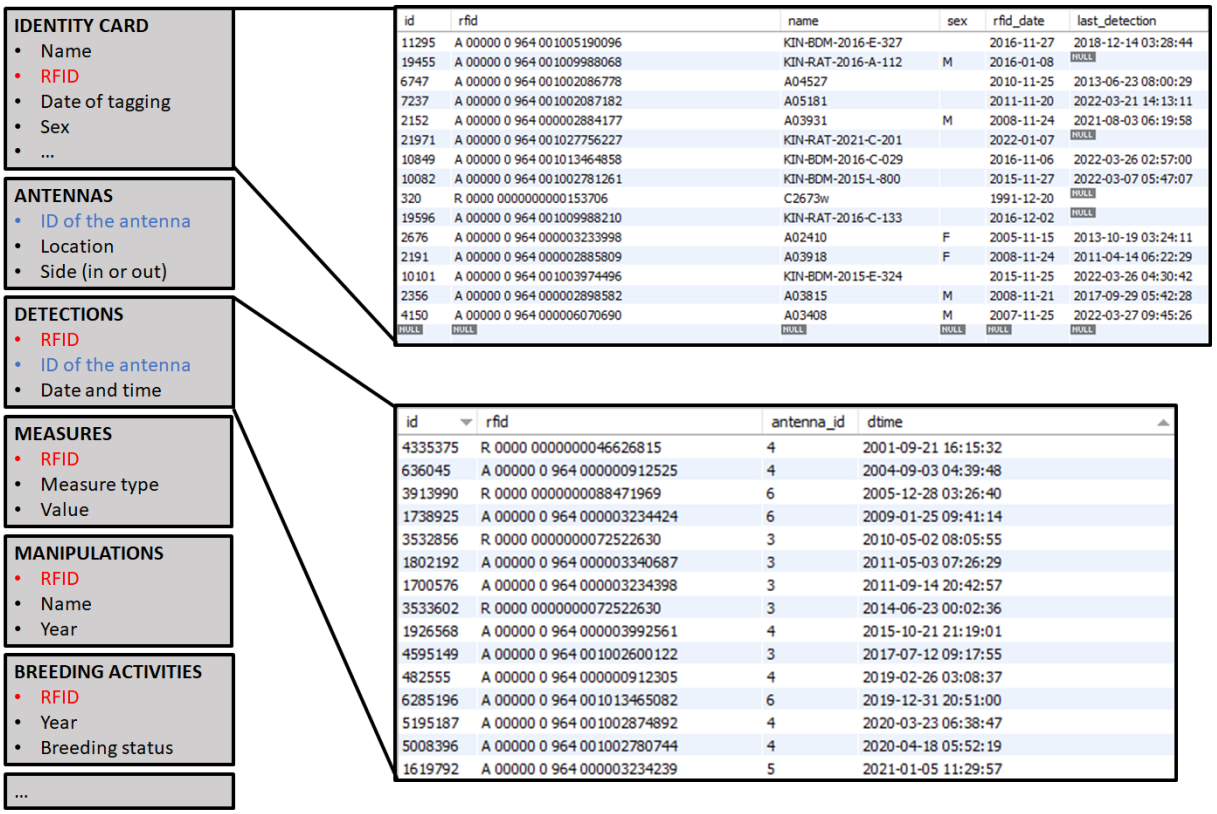


FIGURE SI A1 Schematic illustration of the main tables of the database. Each light grey box corresponds to one table. The database is built around the main table (IDENTITY CARD) containing

all the key information of the individuals. The other tables refer to this main IDENTITY CARD table with the identity of the individuals given by the RFID numbers. For example, the DETECTIONS table records all RFID detections with the RFID identity of the detected individual and the identity of the antenna, also linked to ANTENNAS table compiling all information about the different antennas. Additional information on the individuals is given in other tables for better clarity and flexibility in data storage. Extract of the IDENTITY CARD and DETECTIONS table are given on the top and bottom right side, respectively. The 'id' column stands for the unique identification number assigned to each row of the table. The 'rfid' column corresponds to the RFID-tag number (and by extension the identity of the associated individual). The 'name' column gives the name of the individuals used on the field for simplicity and clarity. The 'sex' column gives the molecular sex if known. The 'rfid date' column gives the date of RFID-tagging. The 'last_detection' column gives the date of the last detection on the antennas and is continuously updated. The column 'antenna_id' corresponds to the identification number of the antenna. The 'dtime' column gives the date and time of the detection.

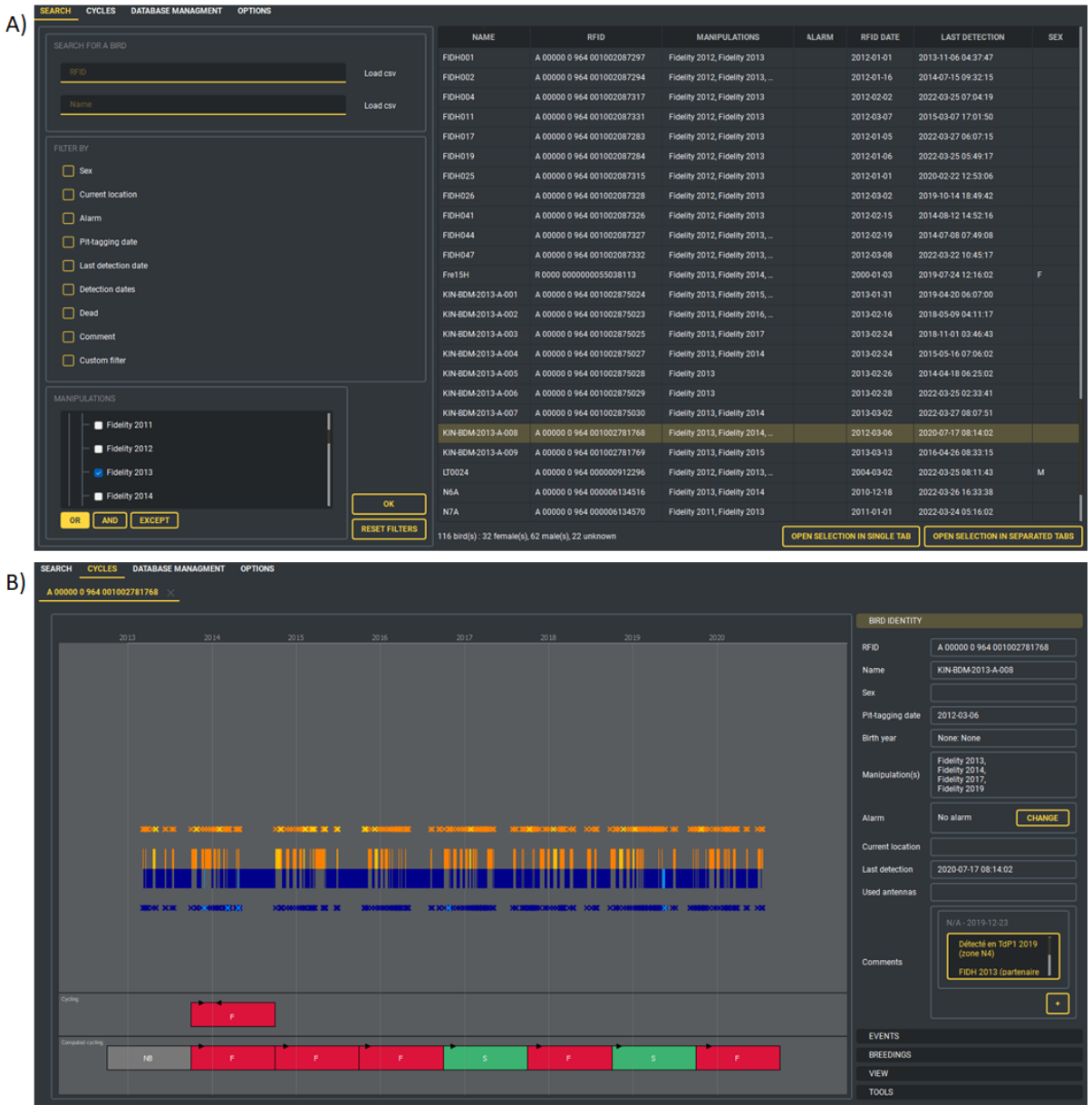


FIGURE SI A2 (A) Main window of the *Sphenotron* showing a search feature to easily find one or more individuals in the database . The left side of the window contains common filters used to

search for specific birds according to chosen filters (e.g., filter by Sex, Pit-tagging date, or Manipulations). The right side gives the result of the search in the database and the main information known on the individual. These individuals can be selected and the detections can be displayed in (B) Individual window showing the presence/absence pattern of a selected individual (here KIN-BDM-2013-A-008) during its electronically monitored lifetime. The right panel displays the individual's information. The central chart displays the lifetime locations of the individual: inside the colony in orange, and outside the colony in blue. Each cross corresponds to one detection. The bottom panel displays the breeding cycles analysed by human experts (top) and by RFIDeep function (bottom): Success in green, Failure in red, and Non-Breeding in grey. Black triangles correspond to the Breeding dates.

Supplement B: Missing detection correction algorithm

To build the missing detection correction algorithm, the detections have been converted into short or long transitions (duration between two following detections and the side of both detections), which were coded in 3 bits: the 1st bit gives the side of the first detection (0 for outside and 1 for inside), the 2nd bit gives the duration between the two detections (0 for less than 10 minutes and 1 for more), and the 3rd bit gives the side of the second detection. These encoded transitions have then been converted into numerical and the vectors of all transitions have been built. A correct schema of transition is therefore 1-7-4-2 (corresponding to 001 – 111 – 100 – 010 in 3-bit code) as shown in Figure SI B1. The algorithm has then been built to detect the incorrect successions of transitions and to correct each possible error by adding one or several detections to recover the 1-7-4-2 successive transitions. Some missing detections remain impossible to correct in this way, e.g., an individual leaving the colony two successive times without entering in between, because no information is known on the date of the entering.

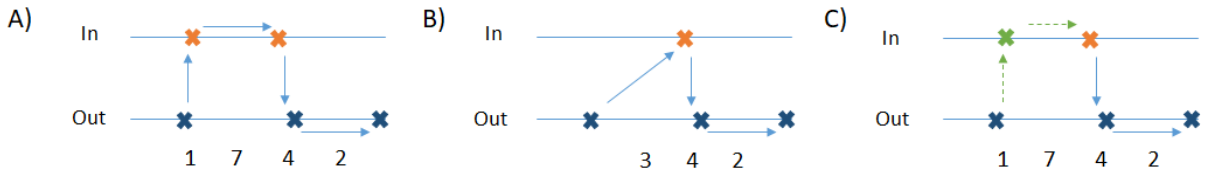


FIGURE SI B1 The blue crosses correspond to true detections at the outside antenna and the orange crosses to true detections at the inside antenna. The green cross corresponds to a logically added detection. Arrows represent transitions. A) Correct schema of transition. B) Incorrect schema of transition leading to an impossible long duration between detection time recorded by the outside and inside antennas. C) Schema that gives the correction of B) with the addition of a logical detection (green cross) just after the first one to recover the correct schema.

Supplement C: Missing detection correction algorithm performance

The algorithm to solve the missing detections was tested based on the detections from our ground truth dataset in order to assess its performances according to various degrees of missing detections. The first step was to get a cleaned detection dataset with corrected detections (because we used detections from the field). Thus, detections have been corrected with the algorithm to remove original missing detections and we manually validated the new corrected detections.

Then, this cleaned dataset was used to test the algorithm performance by removing a various number of detections and by comparing the cleaned dataset with the newly corrected one.

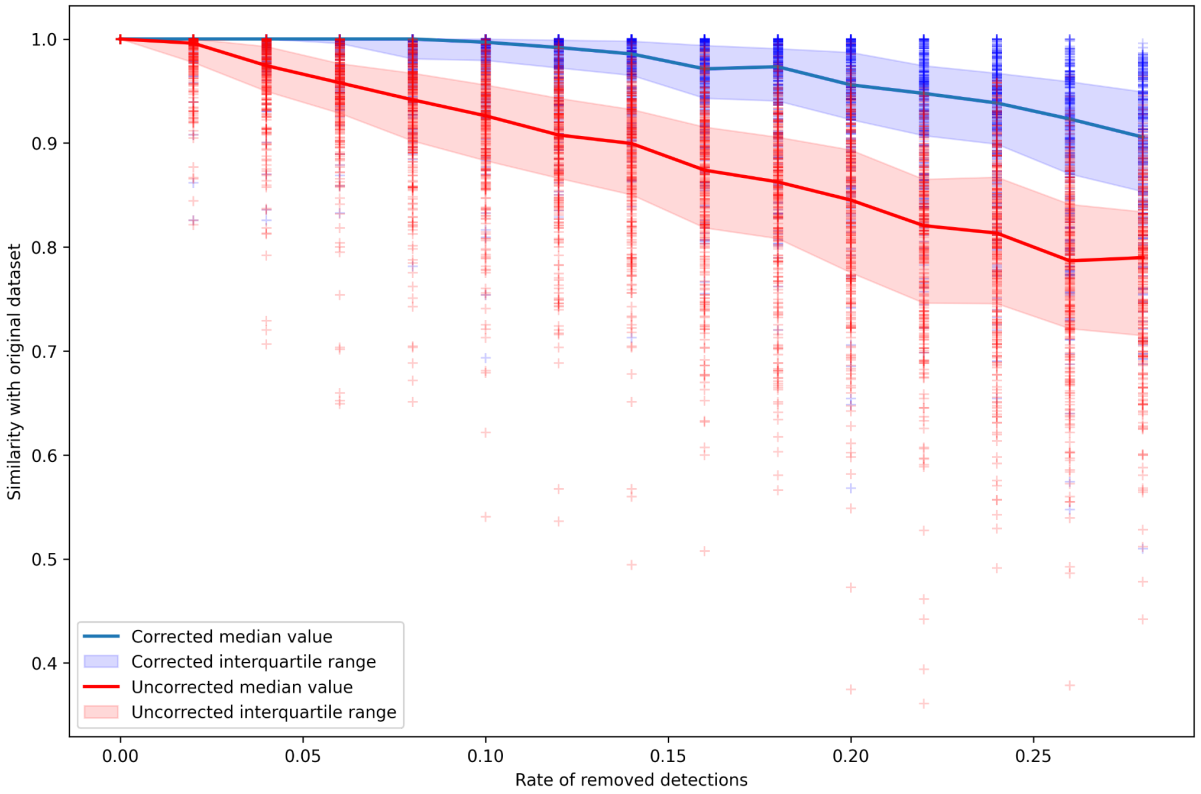


FIGURE SI C1 Performance of missing detection correction algorithm with various rates of removed detections. Similarity of location vectors giving the location (inside or outside) of the individuals in 12-hour periods during the breeding season are compared to the cleaned location vector where raw RFID data was corrected by the algorithm. The blue curve gives the median of the similarity of the vectors for the corrected one while the red curve gives the uncorrected vectors similarity.

Supplement D:

To compute the accuracy of breeding date determination, we used a score giving the proportion of breeding dates that were correctly determined by taking a threshold of 5 days between the true breeding date and the prediction. Indeed, most predictions are either perfectly accurate or completely false (Figure D1). This threshold were chosen to take into account uncertainties in the definition of the breeding date (the beginning of the first long period on the breeding site is not always well clearly recognisable) and because the period of 5 days before and after the predicted date leads to only one possible breeding date given that the first period on the breeding site lasts more than 10 days for our species.

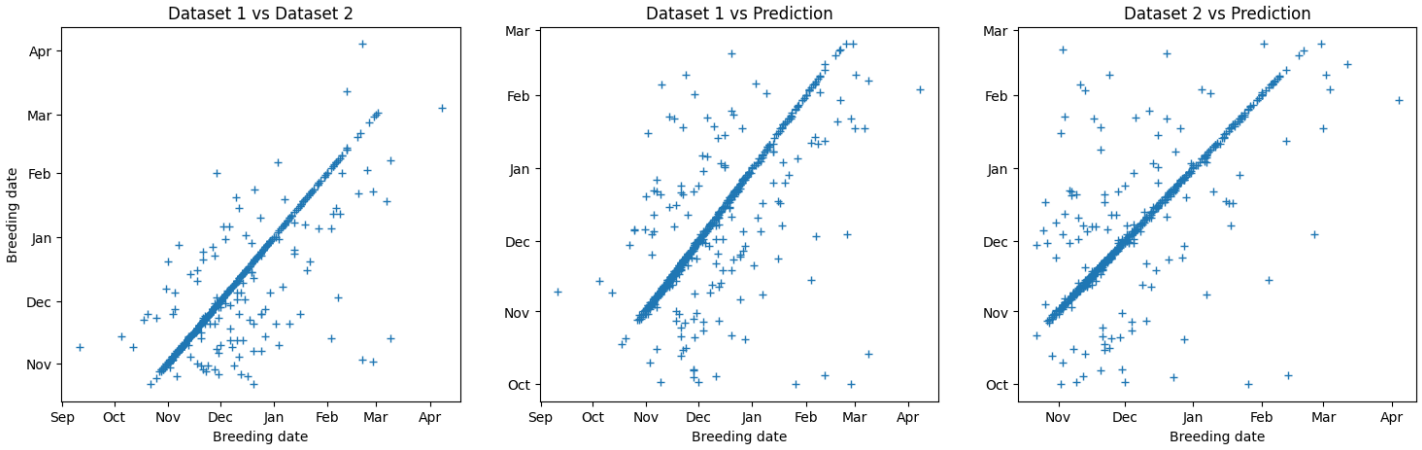


FIGURE D1 Regression of the Breeding date datasets obtained by the two human experts (Datasets 1 and 2) and the one predicted by RFIDeep (Prediction).

Supplement E: Applying RFIDeep to another species, the Adélie penguin

Similar to king penguins, Adélie penguins (*Pygoscelis adeliae*) routinely perform foraging trips between their breeding colonies and the sea during reproduction. However, their breeding cycle is much more constrained temporally than that of the king penguin. All land-based breeding activities take place during a 6-month window between October and March (Ainley, 2002).

As part of the long-term monitoring program 137 of the French polar Institute Paul-Emile Victor (IPEV), an Adélie penguin colony of ca. 300 breeding pairs are electronically monitored with RFID tags off the coast of Adélie Land, Antarctica (ANTAVIA colony, Île des Pétrels, Pointe Géologie archipelago). Every breeding season since 2009-2010, two-access pathways equipped with RFID antennas record the colony attendance patterns of known RFID-tagged individuals. As for king penguins, these patterns are informative on both the breeding outcome (Success or Failure) and the sex of individuals. These biological parameters are not easily determined by direct observations, yet are critical for understanding population processes from individual-based data. There is thus a genuine interest in an automated assessment of individual sex and breeding outcome, especially given the potential for comparisons with other locations around Antarctica, where similar electronic monitoring setups exist (Kerry et al., 1993; Olmastroni et al., 2000; Ballard et al., 2001; Lescroël et al., 2014; Afanasyev et al., 2015).

Here, the RFIDeep approach was applied to 3,959 breeding cycles collected between 2009-2010 and 2021-2022 (n=907 unique individuals). One breeding cycle corresponds to all the detections of an individual in a given breeding season (Figure SI D1). The algorithm was trained using ground truth data from 319 breeding cycles, where breeding outcome and sex were determined by (labor-intensive) field observations. The resulting model's accuracy was then tested on a separate dataset of 1,164 breeding cycles, where breeding outcome and sex were determined by human observation of breeding cycles (e.g., as in Figure SI E1). In this new application, no hyperparameter in the architecture of the model was changed (i.e., we kept the same number of layers, filters, and kernel), but the length of the input vector (364,2) was modified to fit the shorter breeding season of

the Adélie penguins (here it corresponds to 182 days with a 12 hours timesteps, meaning 6 months, encompassing all possible breeding season lengths).

The accuracy of RFIDeep reached 95.2% for breeding outcome determination and 93.5% for sex. These values are comparable to those obtained for king penguins.

These results further demonstrate the broad applicability and effectiveness of RFIDeep for extracting biologically meaningful parameters from RFID data. This electronic monitoring is also coupled with weighbridges, meaning that weights are recorded when an individual crosses the bridge and is detected by the antenna. In future development of the RFIDeep, mass information could be coupled with the detection vectors to help the model to determine individual status and sex.

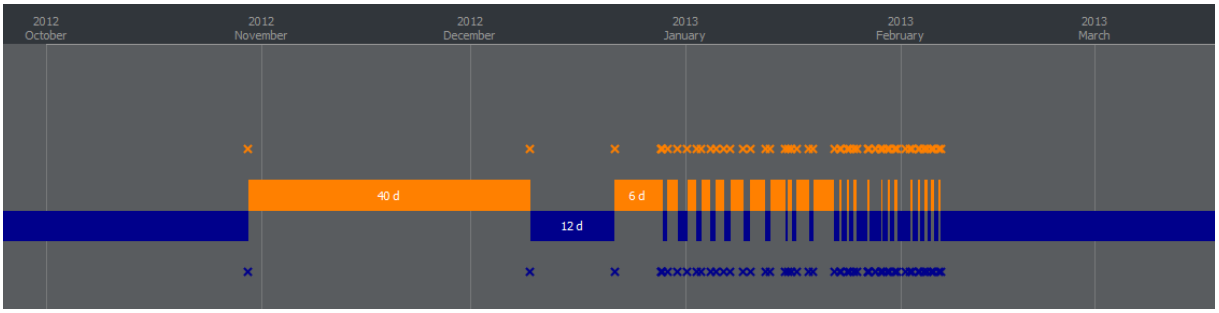


FIGURE SI E1: Typical pattern of presence/absence of a successfully breeding male Adélie penguin . Each cross corresponds to one detection by an antenna (blue and orange indicating outside and inside, respectively).

References (Supplementary Material):

Ainley, D. G. (2002). The Adélie penguin: Bellwether of climate change. Columbia University Press.

Afanasyev, V., Buldyrev, S. V., Dunn, M. J., Robst, J., Preston, M., Bremner, S. F., ... & Peat, H. J. (2015). Increasing Accuracy: A New Design and Algorithm for Automatically Measuring Weights, Travel Direction and Radio Frequency Identification (RFID) of Penguins. Plos one, *10*(4), e0126292.

Ballard, G., Ainley, D. G., Ribic, C. A., & Barton, K. R. (2001). Effect of instrument attachment and other factors on foraging trip duration and nesting success of Adélie penguins. The Condor, 103(3), 481-490.

Kerry, K., Clarke, J., & Grant, E. L. S. E. (1993). The use of an automated weighing and recording system for the study of the biology of adelie penguins (*Pygoscelis adeliae*)(14th Symposium on Polar Biology).

Lescroël, A., Ballard, G., Grémillet, D., Authier, M., & Ainley, D. G. (2014). Antarctic climate change: extreme events disrupt plastic phenotypic response in Adélie penguins. PloS one, 9(1), e85291.

Olmastroni S, Corsolini S, Pezzo F, Focardi S, Kerry K (2000) The first 5 years of the Italian–Australian joint programme on the Ade´lie penguin: an overview. Ital J Zool 67:141–145

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.

FIGURES (SUPPORTING INFORMATION):

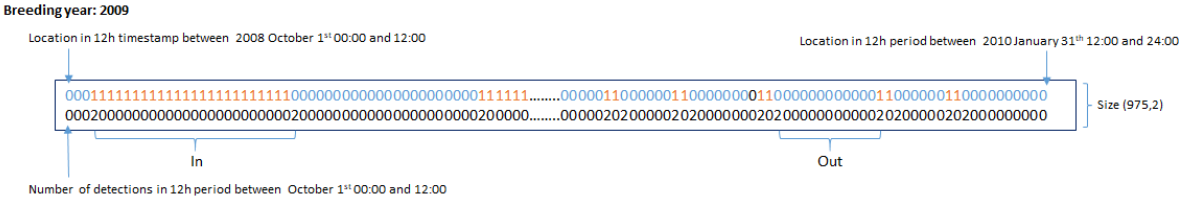


FIGURE S1: Example of one breeding cycle converted into 2 vectors of length 975. The first row corresponds to the location of the individual at each 12-hour period (0 for outside and 1 for inside). The second row gives the number of detections per 12-hour period.

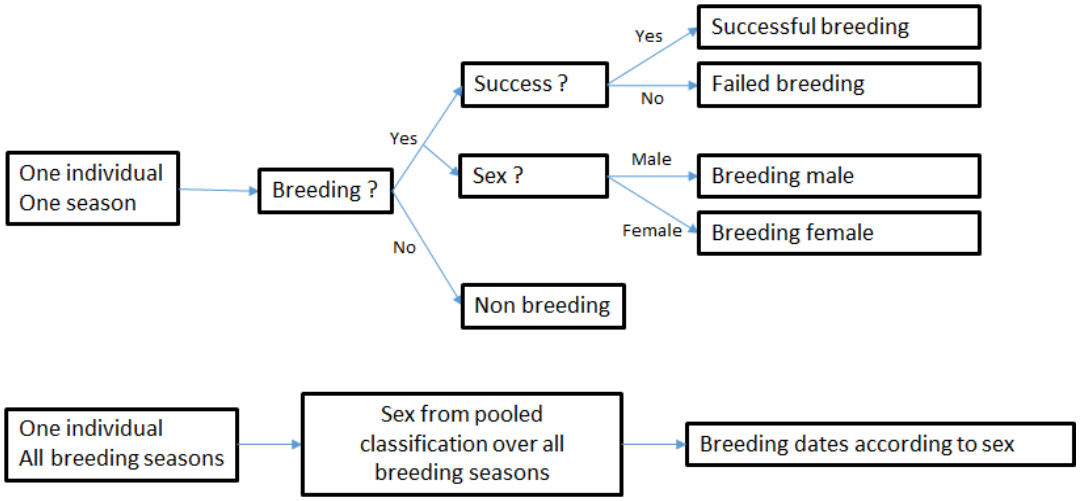


FIGURE S2 Classification scheme used to determine breeding specifications. The top part of the scheme gives the classification procedure performed on each individual and each season: it leads to classification of the breeding status (Breeding vs. Non Breeding), the breeding outcome (Success vs. Failure) and the sex (Male vs. Female). The lower part of the scheme corresponds to the determination of sex from the pooled classification of breeding cycles and to the determination of breeding dates of each breeding season according to the sex.

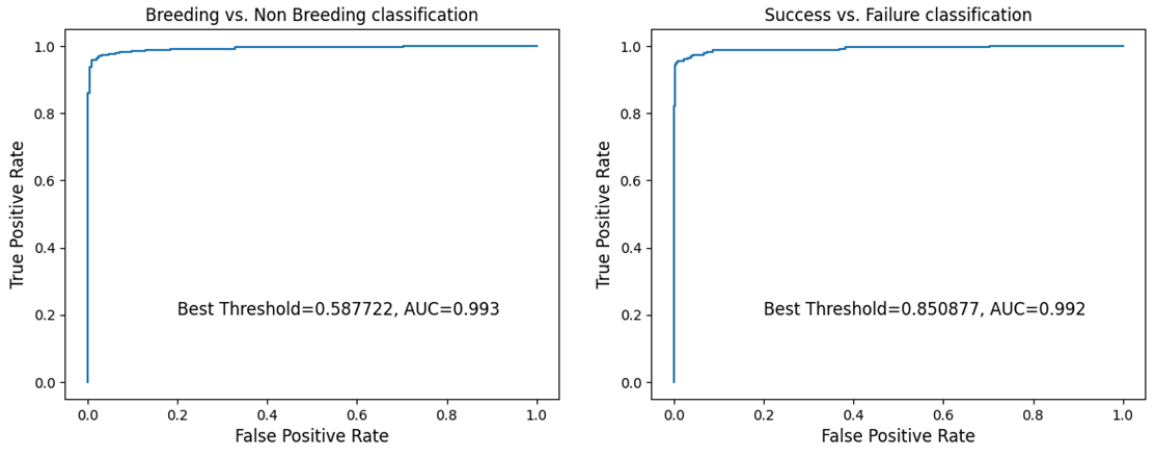


FIGURE S3 Receiver Operating Characteristic (ROC) curves and associated Area Under the Curve (AUC)-ROC scores, and best classification threshold for the Breeding vs. Non-Breeding model and

for the Failure vs. Success model. These metrics were computed using human expert (2) labelled dataset. AUC-ROC scores of 0.993 and 0.992 reveal high accuracies in comparison to the human performances. The values of the threshold reveal that the models tend to be less restrictive to classify a breeding cycle as a success than the human expert.

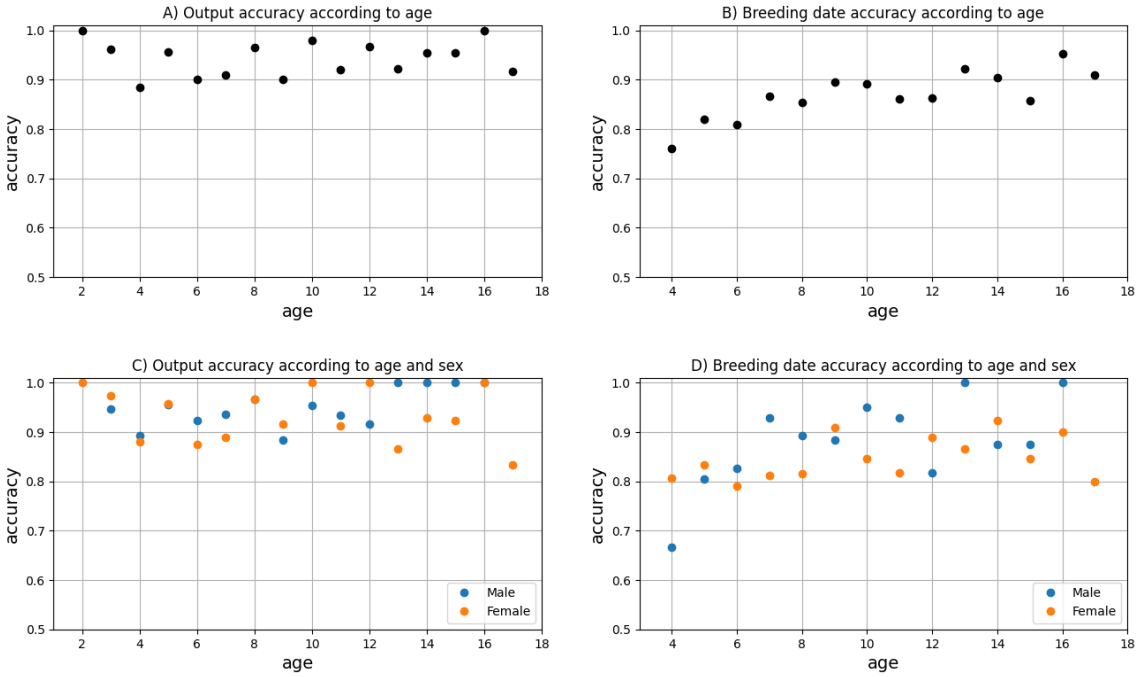


FIGURE S4 Age-specific (A, B, C, D) and sex-specific (C, D) accuracy metrics of output classification and breeding date determination. Output classification (Non-Breeder vs. Failure vs. Success) accuracy appeared stable according to age and sex. Age-specific accuracy metrics are given for age classes with sample sizes greater than 10 individuals. Accuracy of breeding date determination is slightly lower for young birds (breakpoint at 7 years old, from segmented regression analysis, 'segmented' R package).

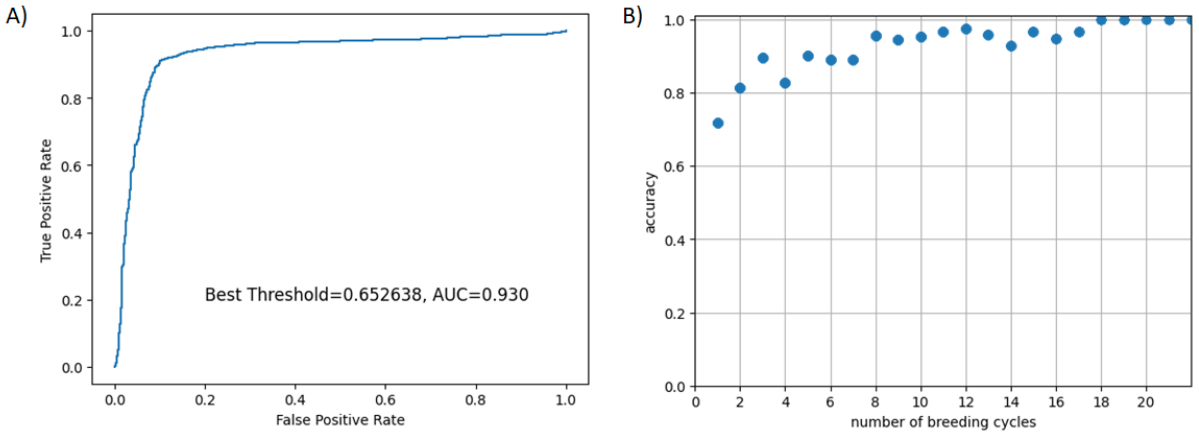


FIGURE S5 (A) Receiver Operating Characteristic (ROC) curve of the sex probability for each individual given by the pooling of its lifetime sex probabilities against the molecular sexing. The Area Under the Curve (AUC)-ROC score and the best threshold are indicated in the figure. (B) Accuracy of sex classification according to the number of breeding cycles pooled together to

determine the most probable sex of an individual. Higher number of breeding cycles leads to a better classification of the sex supporting the benefit of the pooling method.

Appendix Chapter 1

Identification of sex-linked scaffolds

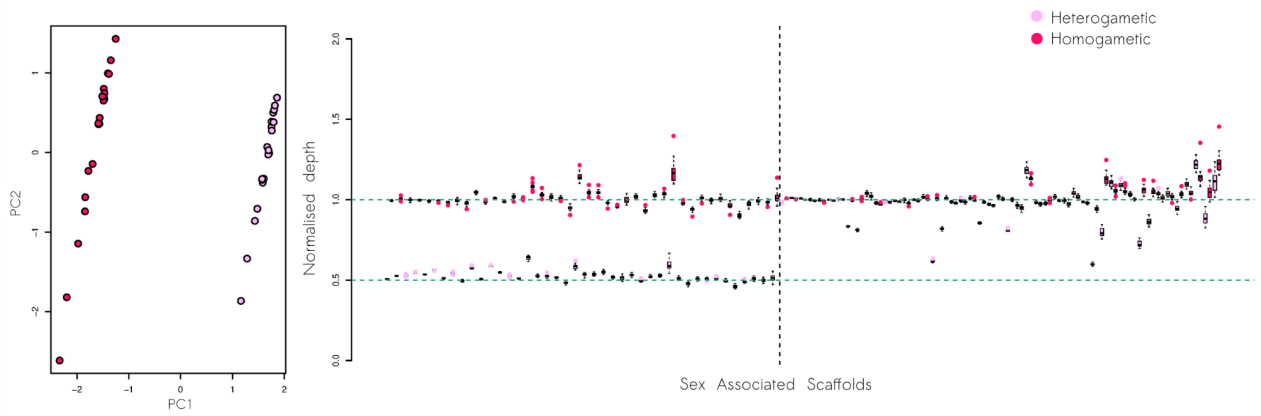


Figure S1. Identification of sex-linked scaffolds and sex identification of individuals with SATC. On the left, the PCA projection of the 40 individuals' sequencing depth; on the right, the normalised sequencing depth per sex associated scaffold (both sex- and XZ-linked).

Appendix Chapter 2

Control for outliers and overrepresented genes

We started with a routine control on our data to check for sources of noise that could affect the main analyses (e.g., overrepresented genes, sex bias, batch effects). First, we checked for overrepresented sequences that could affect the differential expression analysis. In the data generated by the genome alignment approach, we found that more than half of each individual's reads mapped to two haemoglobin genes, HBA and HBB (**Figure S2**), which code for the alpha and beta haemoglobin subunits, respectively.

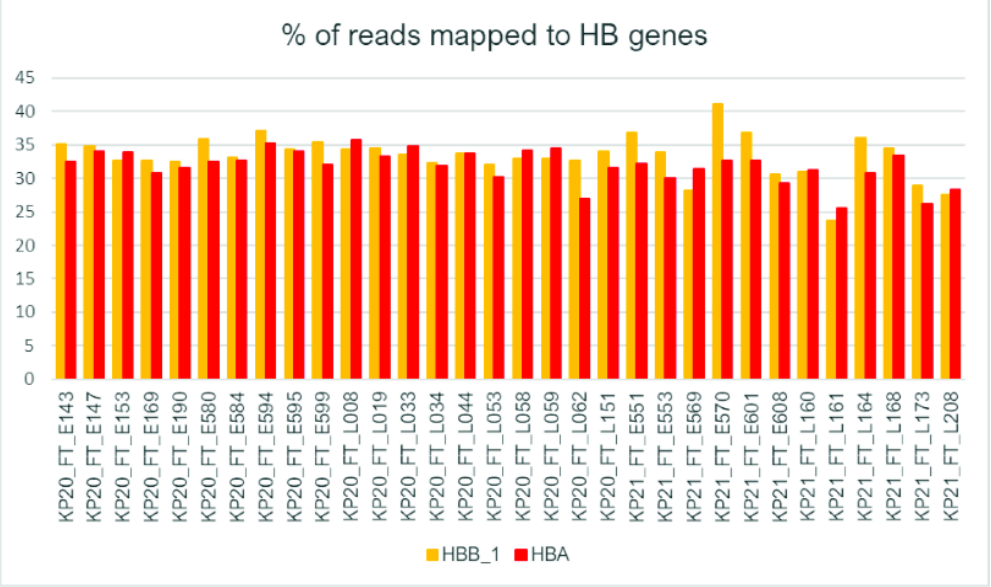


Figure S2. Histogram of the percentage of reads mapped to haemoglobin (HB) genes. On the x-axis, hatching individuals from 2020 and 2021. Yellow bars represent the percentage of reads mapped to the HBB_1 gene (i.e., beta haemoglobin), and red bars represent the percentage of reads mapped to the HBA gene (i.e., alpha haemoglobin).

As it can be seen in **Figure S2** the two overrepresented haemoglobin genes are highly expressed in all samples (around 30-35% of reads aligning to each gene, around 60-75% of reads considering the two genes). The abundance of haemoglobin subunits' transcripts in the whole blood is well known, (mostly (Melé et al. 2015; Désert et al. 2016; Meitern, Andreson, and Hõrak 2014), and was, therefore, already expected. Generally, blood transcriptome studies in human samples deal with the overrepresentation of haemoglobin genes by performing a haemoglobin depletion step before sequencing (Field et al. 2007). Although this depletion step is usually indicated for blood transcriptome analyses, the available kits for globin depletion are designed for humans (e.g., GLOBINclear or Ribo-Zero), and the attempt of using them in non-model species can compromise RNA integrity and gene discovery (Choi et al. 2014).

Moreover, even if the production of a custom-made species-specific depletion method is possible, as it has been shown in polar bears (Byrne et al. 2019), it was not in the scope of this thesis in both terms of time and methodology. For all the reasons mentioned

above, we chose not to perform a haemoglobin depletion library preparation, proceeding with a bioinformatic removal of these gene's read counts before the differential expression analysis step. In any case, haemoglobin reads did not seem to affect the differential gene expression analyses in our study (i.e., the same differentially expressed genes were detected with or without the removal of these sequences).

We performed a second quality to check for outlier samples in our data. We clustered the log transformed counts per million (CPM) of all individuals, and observed a clear separation of one 2021 sample: an early individual at hatching (KP21_E570) in the genome-aligned dataset (**Figure S3**, bottom figure). The isolation of this individual was also detected in PCA analysis (**Figure S3**, top figure). It can also be seen that the outlier sample also has a slightly higher percentage of HBB_1 reads in **Figure S2** (yellow bar). However, even after the removal of haemoglobin gene's this outlier remained (plots in **Figure S3** were done after haemoglobin removal). For these reasons, we removed this sample from all further analysis.

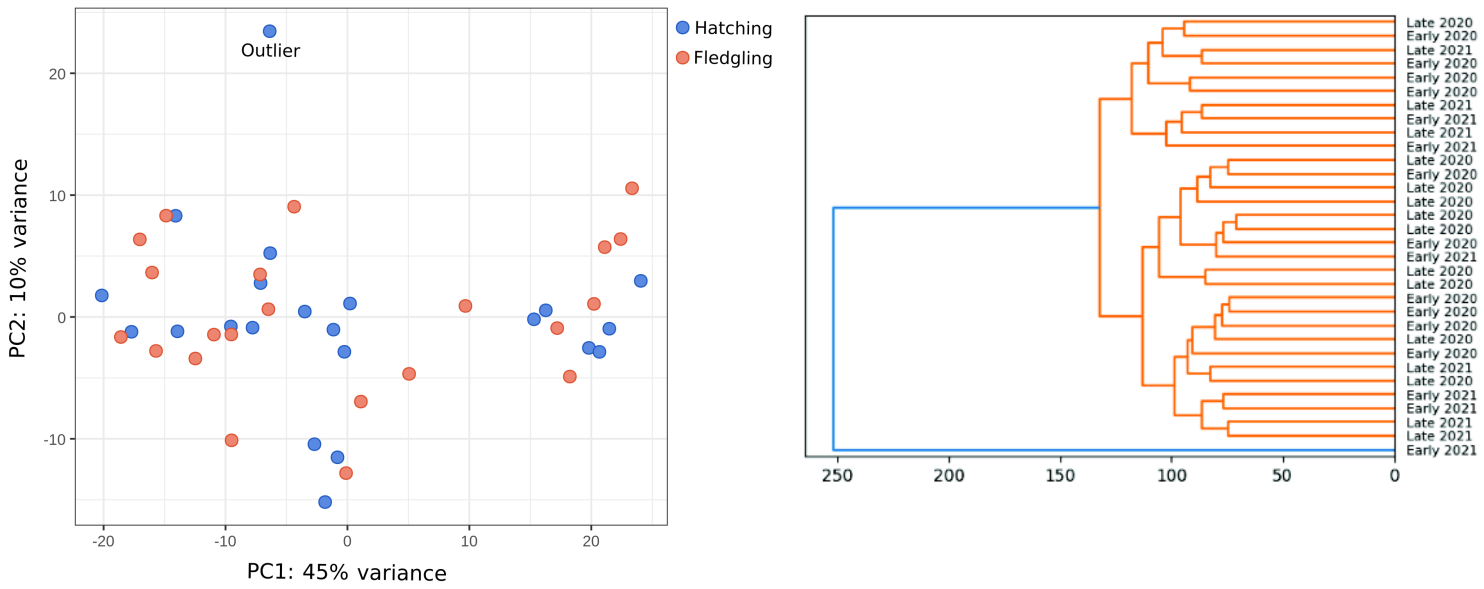


Figure S3. Detection of outlier sample. Top figure shows the principal component analysis (PCA) of normalised genome-mapped counts for all samples, indicating the outlier sample, coloured by developmental stage; at the bottom, clustering of the log transformed raw CPM of hatching individuals with the genome-mapped data.

Finally, we confirmed that the use of the RUVSeq normalisation allowed the removal of excessive noise in both genome-aligned and transcriptome-mapped counts (**Figure S4**). This can be visualised in the boxplots of the relative log expression (RLE) distributions (i.e., the log-ratio of gene counts) in Figure 4. After normalisation, RLE distributions are centred around zero, which is the expected pattern of unwanted noise elimination (Risso et al. 2014). We are not showing the transcriptome-mapped normalisation plots, as the visualisation of this dataset is much larger the initial non-normalised transcript counts dataset was much

larger (i.e., 80 Kb of transcripts, versus 14 Kb of genes) and no clear visualisation of the normalisation effects is perceptible.

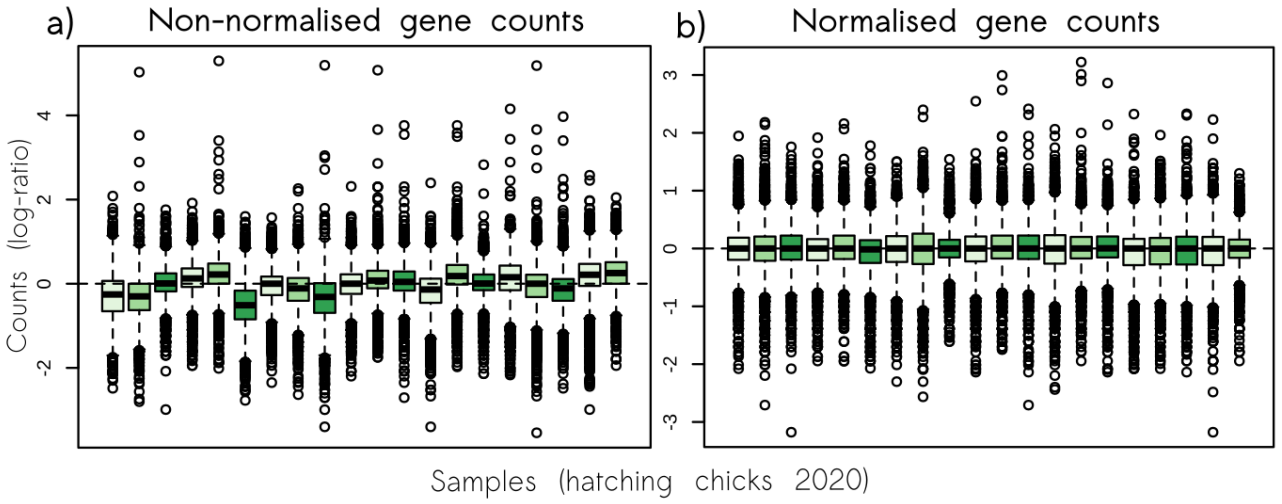


Figure S4. Relative log expression (RLE) plots of the log-ratio of gene counts of the genome-aligned data of 2020 hatching samples. a) RLE of gene counts before normalisation with RUVSeq; b) RLE of gene counts after normalisation with RUVSeq.

Gene Ontology terms of blood characterisation

Table S1. Biological process gene ontology (GO) terms of all genes detected in the blood of hatching and fledging King penguins from PANTHER. The second column represents the GO description and annotation ID; the third column contains the number of genes in the reference list (*Gallus gallus*) that map to this category; the fourth column contains the number of genes in our uploaded list; the fifth column contains the expected number of genes expected in our list for this category based on the reference list; the sixth column shows the fold enrichment of genes observed in the uploaded list over the expected; the seventh column shows the *P-value* of the Fisher exact test; the eighth column shows the false discovery rate (FDR) using the Benjamini-Hochberg procedure.

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
1	cytoplasmic translation (GO:0002181)	52	49	23.65	02.07	3.34E-04	1.34E-02
2	translational initiation (GO:0006413)	53	49	24.1	02.03	3.84E-04	1.50E-02
3	ribosomal large subunit biogenesis (GO:0042273)	65	60	29.56	02.03	9.51E-05	4.30E-03
4	ribosome biogenesis (GO:0042254)	226	204	102.79	1.98	1.84E-12	2.71E-10
5	rRNA processing (GO:0006364)	160	144	72.77	1.98	5.00E-09	5.67E-07
6	rRNA metabolic process (GO:0016072)	169	152	76.86	1.98	1.43E-09	1.73E-07
7	ribosomal small subunit biogenesis (GO:0042274)	58	52	26.38	1.97	4.27E-04	1.64E-02
8	translation (GO:0006412)	290	254	131.89	1.93	3.16E-14	5.81E-12
9	mitochondrial gene expression (GO:0140053)	55	48	25.01	1.92	1.27E-03	4.27E-02
10	ribonucleoprotein complex biogenesis (GO:0022613)	321	280	145.99	1.92	2.62E-15	5.54E-13
11	peptide biosynthetic process (GO:0043043)	298	257	135.53	1.9	7.99E-14	1.36E-11
12	ncRNA processing (GO:0034470)	308	264	140.08	1.88	6.44E-14	1.13E-11
13	tRNA modification (GO:0006400)	75	64	34.11	1.88	3.09E-04	1.25E-02
14	regulation of protein stability (GO:0031647)	74	63	33.66	1.87	2.93E-04	1.19E-02
15	nuclear-transcribed mRNA catabolic process (GO:0000956)	70	59	31.84	1.85	5.68E-04	2.11E-02
16	mRNA catabolic process (GO:0006402)	82	69	37.29	1.85	2.00E-04	8.47E-03
17	endoplasmic reticulum to Golgi vesicle-mediated transport (GO:0006888)	87	73	39.57	1.84	1.50E-04	6.56E-03
18	vacuolar transport (GO:0007034)	93	78	42.3	1.84	8.91E-05	4.09E-03
19	tRNA metabolic process (GO:0006399)	164	137	74.59	1.84	2.35E-07	1.94E-05
20	ncRNA metabolic process (GO:0034660)	398	331	181.01	1.83	6.16E-16	1.35E-13
21	ribonucleoprotein complex subunit organisation (GO:0071826)	136	113	61.85	1.83	3.52E-06	2.24E-04
22	tRNA processing (GO:0008033)	105	87	47.75	1.82	4.36E-05	2.16E-03
23	ribonucleoprotein complex assembly (GO:0022618)	130	107	59.12	1.81	7.92E-06	4.72E-04
24	amide biosynthetic process (GO:0043604)	374	307	170.1	1.8	2.22E-14	4.15E-12
25	protein targeting (GO:0006605)	133	109	60.49	1.8	7.12E-06	4.32E-04
26	negative regulation of cellular amide metabolic process (GO:0034249)	99	81	45.03	1.8	1.34E-04	5.94E-03
27	RNA methylation (GO:0001510)	71	58	32.29	1.8	1.13E-03	3.85E-02
28	Golgi vesicle transport (GO:0048193)	188	153	85.5	1.79	1.36E-07	1.17E-05

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
29	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	102	83	46.39	1.79	1.19E-04	5.37E-03
30	negative regulation of translation (GO:0017148)	95	77	43.21	1.78	1.92E-04	8.16E-03
31	cytosolic transport (GO:0016482)	82	66	37.29	1.77	6.80E-04	2.48E-02
32	mitochondrion organisation (GO:0007005)	234	186	106.42	1.75	2.04E-08	2.06E-06
33	peptide metabolic process (GO:0006518)	377	299	171.46	1.74	8.68E-13	1.37E-10
34	RNA localization (GO:0006403)	98	77	44.57	1.73	4.06E-04	1.58E-02
35	RNA catabolic process (GO:0006401)	98	77	44.57	1.73	4.06E-04	1.57E-02
36	RNA modification (GO:0009451)	139	109	63.22	1.72	2.47E-05	1.34E-03
37	protein folding (GO:0006457)	157	123	71.4	1.72	9.22E-06	5.38E-04
38	RNA processing (GO:0006396)	650	508	295.62	1.72	3.60E-20	1.35E-17
39	establishment of RNA localization (GO:0051236)	91	71	41.39	1.72	8.54E-04	3.03E-02
40	RNA transport (GO:0050658)	91	71	41.39	1.72	8.54E-04	3.02E-02
41	nucleic acid transport (GO:0050657)	91	71	41.39	1.72	8.54E-04	3.01E-02
42	establishment of protein localization to membrane (GO:0090150)	99	77	45.03	1.71	5.67E-04	2.11E-02
43	regulation of protein modification by small protein conjugation or removal (GO:1903320)	88	68	40.02	1.7	1.28E-03	4.27E-02
44	endosomal transport (GO:0016197)	140	108	63.67	1.7	4.51E-05	2.23E-03
45	regulation of cellular amide metabolic process (GO:0034248)	213	164	96.87	1.69	5.49E-07	4.27E-05
46	proteasomal protein catabolic process (GO:0010498)	250	192	113.7	1.69	6.27E-08	5.76E-06
47	double-strand break repair via homologous recombination (GO:0000724)	86	66	39.11	1.69	1.54E-03	4.99E-02
48	cellular response to DNA damage stimulus (GO:0006974)	434	332	197.38	1.68	1.08E-12	1.66E-10
49	protein catabolic process (GO:0030163)	472	361	214.67	1.68	1.26E-13	2.07E-11
50	modification-dependent macromolecule catabolic process (GO:0043632)	412	315	187.38	1.68	4.14E-12	5.95E-10
51	regulation of translation (GO:0006417)	198	151	90.05	1.68	2.22E-06	1.48E-04
52	DNA repair (GO:0006281)	324	247	147.36	1.68	1.37E-09	1.69E-07
53	proteolysis involved in protein catabolic process (GO:0051603)	440	335	200.11	1.67	1.44E-12	2.18E-10
54	proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	226	172	102.79	1.67	4.69E-07	3.70E-05
55	double-strand break repair (GO:0006302)	146	111	66.4	1.67	6.05E-05	2.90E-03
56	modification-dependent protein catabolic process (GO:0019941)	403	306	183.28	1.67	1.68E-11	2.33E-09
57	histone modification (GO:0016570)	203	154	92.32	1.67	2.17E-06	1.46E-04
58	regulation of proteolysis involved in protein catabolic process (GO:1903050)	91	69	41.39	1.67	1.51E-03	4.91E-02
59	cellular respiration (GO:0045333)	103	78	46.84	1.67	9.12E-04	3.19E-02
60	cellular macromolecule catabolic process (GO:0044265)	523	396	237.86	1.66	2.03E-14	3.86E-12
61	ubiquitin-dependent protein catabolic process (GO:0006511)	395	299	179.65	1.66	3.33E-11	4.55E-09

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
62	nucleobase-containing compound transport (GO:0015931)	127	96	57.76	1.66	2.04E-04	8.59E-03
63	protein polyubiquitination (GO:0000209)	138	104	62.76	1.66	1.20E-04	5.37E-03
64	regulation of cellular macromolecule biosynthetic process (GO:2000112)	223	168	101.42	1.66	1.16E-06	8.21E-05
65	macromolecule catabolic process (GO:0009057)	605	455	275.15	1.65	4.42E-16	1.01E-13
66	DNA-templated transcription elongation (GO:0006354)	137	103	62.31	1.65	1.47E-04	6.48E-03
67	macromolecule methylation (GO:0043414)	169	127	76.86	1.65	2.23E-05	1.21E-03
68	negative regulation of cell cycle phase transition (GO:1901988)	104	78	47.3	1.65	9.64E-04	3.34E-02
69	mRNA metabolic process (GO:0016071)	407	305	185.1	1.65	4.49E-11	6.06E-09
70	chromatin organisation (GO:0006325)	306	229	139.17	1.65	1.74E-08	1.77E-06
71	cellular amide metabolic process (GO:0043603)	533	398	242.41	1.64	6.92E-14	1.19E-11
72	energy derivation by oxidation of organic compounds (GO:0015980)	130	97	59.12	1.64	3.02E-04	1.23E-02
73	regulation of protein catabolic process (GO:0042176)	128	95	58.21	1.63	3.59E-04	1.42E-02
74	autophagy (GO:0006914)	147	109	66.86	1.63	1.38E-04	6.13E-03
75	process utilising autophagic mechanism (GO:0061919)	147	109	66.86	1.63	1.38E-04	6.11E-03
76	DNA replication (GO:0006260)	139	103	63.22	1.63	2.10E-04	8.80E-03
77	post-transcriptional regulation of gene expression (GO:0010608)	243	180	110.52	1.63	1.01E-06	7.24E-05
78	peptidyl-lysine modification (GO:0018205)	161	119	73.22	1.63	7.20E-05	3.40E-03
79	cellular response to stress (GO:0033554)	758	559	344.74	1.62	1.91E-18	5.79E-16
80	protein modification by small protein removal (GO:0070646)	117	86	53.21	1.62	8.03E-04	2.88E-02
81	response to endoplasmic reticulum stress (GO:0034976)	112	82	50.94	1.61	1.37E-03	4.49E-02
82	intracellular transport (GO:0046907)	853	624	387.95	1.61	6.10E-20	2.07E-17
83	mRNA processing (GO:0006397)	308	225	140.08	1.61	8.39E-08	7.40E-06
84	protein localization to organelle (GO:0033365)	341	249	155.09	1.61	1.70E-08	1.76E-06
85	cell division (GO:0051301)	211	154	95.96	1.6	1.09E-05	6.34E-04
86	generation of precursor metabolites and energy (GO:0006091)	188	137	85.5	1.6	3.09E-05	1.62E-03
87	cellular macromolecule biosynthetic process (GO:0034645)	603	439	274.25	1.6	4.91E-14	8.73E-12
88	RNA splicing (GO:0008380)	258	187	117.34	1.59	1.47E-06	1.03E-04
89	establishment of protein localization to organelle (GO:0072594)	194	140	88.23	1.59	3.87E-05	1.96E-03
90	phosphatidylinositol metabolic process (GO:0046488)	122	88	55.49	1.59	1.25E-03	4.22E-02
91	regulation of catabolic process (GO:0009894)	344	248	156.45	1.59	3.93E-08	3.83E-06
92	negative regulation of cell cycle process (GO:0010948)	118	85	53.67	1.58	1.37E-03	4.49E-02
93	intracellular protein transport (GO:0006886)	435	313	197.84	1.58	6.18E-10	7.78E-08

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
94	chromatin remodelling (GO:0006338)	196	141	89.14	1.58	4.18E-05	2.08E-03
95	gene expression (GO:0010467)	1525	1095	693.57	1.58	4.72E-33	3.52E-30
96	methylation (GO:0032259)	207	148	94.14	1.57	3.12E-05	1.63E-03
97	protein transport (GO:0015031)	668	477	303.81	1.57	3.99E-14	7.21E-12
98	mitotic cell cycle process (GO:1903047)	296	211	134.62	1.57	7.31E-07	5.42E-05
99	establishment of protein localization (GO:0045184)	695	494	316.09	1.56	2.03E-14	3.92E-12
100	protein modification by small protein conjugation or removal (GO:0070647)	609	431	276.97	1.56	1.69E-12	2.53E-10
101	cellular component disassembly (GO:0022411)	136	96	61.85	1.55	1.31E-03	4.35E-02
102	regulation of cellular catabolic process (GO:0031329)	288	203	130.98	1.55	2.18E-06	1.46E-04
103	mRNA splicing, via spliceosome (GO:0000398)	199	140	90.51	1.55	9.39E-05	4.30E-03
104	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	199	140	90.51	1.55	9.39E-05	4.28E-03
105	RNA splicing, via transesterification reactions (GO:0000375)	199	140	90.51	1.55	9.39E-05	4.26E-03
106	regulation of mRNA metabolic process (GO:1903311)	174	122	79.14	1.54	3.09E-04	1.25E-02
107	protein modification by small protein conjugation (GO:0032446)	494	346	224.67	1.54	7.68E-10	9.57E-08
108	chromosome organisation (GO:0051276)	280	196	127.34	1.54	4.14E-06	2.61E-04
109	RNA biosynthetic process (GO:0032774)	345	241	156.91	1.54	4.05E-07	3.22E-05
110	establishment of organelle localization (GO:0051656)	179	125	81.41	1.54	2.92E-04	1.20E-02
111	cellular macromolecule metabolic process (GO:0044260)	1850	1288	841.38	1.53	5.88E-35	5.99E-32
112	phospholipid biosynthetic process (GO:0008654)	144	100	65.49	1.53	1.37E-03	4.48E-02
113	DNA-templated transcription (GO:0006351)	340	236	154.63	1.53	6.72E-07	5.06E-05
114	nucleic acid-templated transcription (GO:0097659)	340	236	154.63	1.53	6.72E-07	5.02E-05
115	organelle localization (GO:0051640)	254	176	115.52	1.52	2.03E-05	1.13E-03
116	nucleic acid metabolic process (GO:0090304)	1708	1181	776.8	1.52	5.92E-31	3.49E-28
117	nucleobase-containing compound catabolic process (GO:0034655)	168	116	76.41	1.52	6.26E-04	2.31E-02
118	organonitrogen compound catabolic process (GO:1901565)	678	468	308.36	1.52	2.65E-12	3.86E-10
119	protein localization (GO:0008104)	991	684	450.71	1.52	1.20E-17	3.45E-15
120	cellular macromolecule localization (GO:0070727)	993	685	451.62	1.52	1.27E-17	3.56E-15
121	protein ubiquitination (GO:0016567)	453	312	206.03	1.51	1.72E-08	1.77E-06
122	cellular catabolic process (GO:0044248)	994	684	452.07	1.51	2.03E-17	5.30E-15
123	regulation of cell cycle phase transition (GO:1901987)	173	119	78.68	1.51	5.88E-04	2.18E-02
124	establishment of localization in cell (GO:0051649)	1020	700	463.9	1.51	1.45E-17	3.97E-15
125	RNA metabolic process (GO:0016070)	1245	854	566.23	1.51	2.00E-21	8.95E-19

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
126	vesicle organisation (GO:0016050)	156	107	70.95	1.51	1.29E-03	4.29E-02
127	cellular nitrogen compound metabolic process (GO:0034641)	2557	1744	1162.93	1.5	3.47E-45	5.56E-42
128	macromolecule biosynthetic process (GO:0009059)	1045	712	475.27	1.5	2.40E-17	6.12E-15
129	regulation of cellular response to stress (GO:0080135)	216	147	98.24	1.5	2.01E-04	8.51E-03
130	positive regulation of catabolic process (GO:0009896)	197	134	89.6	1.5	4.13E-04	1.59E-02
131	positive regulation of cellular catabolic process (GO:0031331)	165	112	75.04	1.49	1.36E-03	4.49E-02
132	cell cycle (GO:0007049)	638	432	290.16	1.49	1.39E-10	1.83E-08
133	cellular nitrogen compound biosynthetic process (GO:0044271)	1062	719	483	1.49	3.90E-17	9.71E-15
134	protein localization to membrane (GO:0072657)	207	140	94.14	1.49	3.61E-04	1.41E-02
135	nitrogen compound transport (GO:0071705)	919	621	417.96	1.49	1.08E-14	2.19E-12
136	mitotic cell cycle (GO:0000278)	365	246	166	1.48	2.15E-06	1.46E-04
137	glycerophospholipid metabolic process (GO:0006650)	190	128	86.41	1.48	6.40E-04	2.35E-02
138	nucleobase-containing compound metabolic process (GO:0006139)	2067	1392	940.07	1.48	3.55E-33	3.06E-30
139	cellular localization (GO:0051641)	1457	980	662.65	1.48	1.11E-22	5.40E-20
140	organonitrogen compound biosynthetic process (GO:1901566)	969	651	440.7	1.48	5.06E-15	1.05E-12
141	cell cycle process (GO:0022402)	466	313	211.94	1.48	1.08E-07	9.38E-06
142	heterocycle metabolic process (GO:0046483)	2156	1446	980.55	1.47	6.03E-34	5.63E-31
143	non-membrane-bounded organelle assembly (GO:0140694)	191	128	86.87	1.47	8.05E-04	2.88E-02
144	localization within membrane (GO:0051668)	266	178	120.98	1.47	8.30E-05	3.83E-03
145	negative regulation of gene expression (GO:0010629)	324	216	147.36	1.47	1.51E-05	8.53E-04
146	cellular nitrogen compound catabolic process (GO:0044270)	194	129	88.23	1.46	8.81E-04	3.10E-02
147	heterocycle catabolic process (GO:0046700)	191	127	86.87	1.46	9.94E-04	3.42E-02
148	cellular aromatic compound metabolic process (GO:0006725)	2182	1450	992.38	1.46	1.07E-32	7.52E-30
149	catabolic process (GO:0009056)	1188	787	540.3	1.46	5.42E-17	1.29E-14
150	organic substance catabolic process (GO:1901575)	1034	683	470.26	1.45	1.16E-14	2.31E-12
151	macromolecule modification (GO:0043412)	2126	1392	966.91	1.44	3.15E-29	1.68E-26
152	organic cyclic compound metabolic process (GO:1901360)	2299	1493	1045.59	1.43	2.50E-30	1.40E-27
153	regulation of cell cycle process (GO:0010564)	288	187	130.98	1.43	1.73E-04	7.44E-03
154	macromolecule localization (GO:0033036)	1268	822	576.69	1.43	4.12E-16	9.62E-14
155	protein modification process (GO:0036211)	1972	1273	896.87	1.42	8.96E-25	4.56E-22
156	cellular metabolic process (GO:0044237)	4993	3191	2270.82	1.41	2.80E-72	1.57E-68
157	cellular biosynthetic process (GO:0044249)	1696	1083	771.34	1.4	8.47E-20	2.79E-17
158	organic substance biosynthetic process (GO:1901576)	1740	1106	791.35	1.4	9.25E-20	2.96E-17

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
159	regulation of cell cycle (GO:0051726)	452	287	205.57	1.4	1.10E-05	6.35E-04
160	biosynthetic process (GO:0009058)	1787	1132	812.73	1.39	6.05E-20	2.12E-17
161	macromolecule metabolic process (GO:0043170)	4503	2851	2047.97	1.39	8.36E-59	1.56E-55
162	protein-containing complex organisation (GO:0043933)	730	462	332.01	1.39	2.06E-08	2.06E-06
163	regulation of cellular localization (GO:0060341)	266	168	120.98	1.39	9.81E-04	3.38E-02
164	nitrogen compound metabolic process (GO:0006807)	5103	3213	2320.85	1.38	1.45E-67	4.06E-64
165	protein-containing complex assembly (GO:0065003)	641	403	291.53	1.38	3.05E-07	2.46E-05
166	protein metabolic process (GO:0019538)	2941	1834	1337.57	1.37	4.21E-31	2.62E-28
167	cell death (GO:0008219)	320	199	145.54	1.37	5.64E-04	2.11E-02
168	programmed cell death (GO:0012501)	312	194	141.9	1.37	6.85E-04	2.49E-02
169	peptidyl-amino acid modification (GO:0018193)	556	345	252.87	1.36	5.56E-06	3.40E-04
170	primary metabolic process (GO:0044238)	5565	3450	2530.97	1.36	5.16E-69	1.93E-65
171	organophosphate biosynthetic process (GO:0090407)	352	218	160.09	1.36	3.59E-04	1.41E-02
172	organelle assembly (GO:0070925)	456	282	207.39	1.36	5.46E-05	2.64E-03
173	apoptotic process (GO:0006915)	293	181	133.26	1.36	1.35E-03	4.46E-02
174	regulation of response to stress (GO:0080134)	353	218	160.54	1.36	4.31E-04	1.65E-02
175	organonitrogen compound metabolic process (GO:1901564)	3655	2243	1662.3	1.35	3.14E-36	3.52E-33
176	cellular component biogenesis (GO:0044085)	1451	889	659.92	1.35	4.80E-13	7.68E-11
177	organic substance metabolic process (GO:0071704)	5871	3588	2670.14	1.34	1.60E-67	3.58E-64
178	organophosphate metabolic process (GO:0019637)	609	372	276.97	1.34	6.79E-06	4.14E-04
179	metabolic process (GO:0008152)	6180	3772	2810.67	1.34	1.78E-72	2.00E-68
180	organic substance transport (GO:0071702)	1205	734	548.04	1.34	1.57E-10	2.04E-08
181	DNA metabolic process (GO:0006259)	594	360	270.15	1.33	1.62E-05	9.10E-04
182	heterocycle biosynthetic process (GO:0018130)	702	425	319.27	1.33	2.57E-06	1.70E-04
183	nucleobase-containing compound biosynthetic process (GO:0034654)	638	386	290.16	1.33	8.89E-06	5.22E-04
184	negative regulation of protein metabolic process (GO:0051248)	403	243	183.28	1.33	4.96E-04	1.87E-02
185	regulation of organelle organisation (GO:0033043)	487	293	221.49	1.32	1.63E-04	7.08E-03
186	negative regulation of cellular metabolic process (GO:0031324)	872	521	396.59	1.31	5.86E-07	4.50E-05
187	negative regulation of biosynthetic process (GO:0009890)	640	381	291.07	1.31	3.02E-05	1.60E-03
188	negative regulation of cellular biosynthetic process (GO:0031327)	635	378	288.8	1.31	3.28E-05	1.70E-03
189	negative regulation of macromolecule biosynthetic process (GO:0010558)	617	367	280.61	1.31	4.05E-05	2.04E-03

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
190	aromatic compound biosynthetic process (GO:0019438)	699	415	317.91	1.31	1.48E-05	8.41E-04
191	carboxylic acid metabolic process (GO:0019752)	539	320	245.14	1.31	1.54E-04	6.73E-03
192	cellular lipid metabolic process (GO:0044255)	607	359	276.06	1.3	7.34E-05	3.44E-03
193	oxoacid metabolic process (GO:0043436)	544	321	247.41	1.3	2.18E-04	9.13E-03
194	phosphorus metabolic process (GO:0006793)	1373	810	624.44	1.3	1.37E-09	1.67E-07
195	regulation of protein metabolic process (GO:0051246)	1021	601	464.35	1.29	3.05E-07	2.47E-05
196	phosphate-containing compound metabolic process (GO:0006796)	1360	800	618.53	1.29	2.45E-09	2.89E-07
197	organelle organisation (GO:0006996)	2032	1189	924.16	1.29	3.08E-13	5.00E-11
198	phosphorylation (GO:0016310)	719	419	327	1.28	4.94E-05	2.41E-03
199	organic cyclic compound biosynthetic process (GO:1901362)	769	448	349.74	1.28	2.50E-05	1.35E-03
200	small molecule metabolic process (GO:0044281)	1069	617	486.18	1.27	1.50E-06	1.05E-04
201	carbohydrate derivative metabolic process (GO:1901135)	673	388	306.08	1.27	1.73E-04	7.47E-03
202	organic acid metabolic process (GO:0006082)	560	322	254.69	1.26	7.65E-04	2.77E-02
203	negative regulation of nitrogen compound metabolic process (GO:0051172)	977	561	444.34	1.26	7.37E-06	4.44E-04
204	negative regulation of metabolic process (GO:0009892)	1186	681	539.39	1.26	6.46E-07	4.89E-05
205	cellular component assembly (GO:0022607)	1269	725	577.14	1.26	4.90E-07	3.84E-05
206	response to stress (GO:0006950)	1366	780	621.26	1.26	1.71E-07	1.44E-05
207	negative regulation of macromolecule metabolic process (GO:0010605)	1130	644	513.93	1.25	2.99E-06	1.94E-04
208	positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	844	477	383.85	1.24	1.23E-04	5.49E-03
209	positive regulation of RNA metabolic process (GO:0051254)	778	437	353.84	1.24	3.41E-04	1.35E-02
210	positive regulation of nitrogen compound metabolic process (GO:0051173)	1276	707	580.33	1.22	1.46E-05	8.32E-04
211	positive regulation of macromolecule biosynthetic process (GO:0010557)	775	429	352.47	1.22	9.30E-04	3.23E-02
212	proteolysis (GO:0006508)	922	509	419.33	1.21	3.45E-04	1.37E-02
213	regulation of cellular component organisation (GO:0051128)	855	470	388.86	1.21	8.17E-04	2.92E-02
214	intracellular signal transduction (GO:0035556)	861	473	391.58	1.21	8.49E-04	3.02E-02
215	positive regulation of macromolecule metabolic process (GO:0010604)	1374	753	624.9	1.2	2.17E-05	1.18E-03
216	negative regulation of cellular process (GO:0048523)	1849	1012	840.93	1.2	6.40E-07	4.88E-05
217	cellular component organisation or biogenesis (GO:0071840)	3660	2001	1664.57	1.2	1.22E-13	2.04E-11
218	vesicle-mediated transport (GO:0016192)	909	494	413.41	1.19	1.25E-03	4.22E-02
219	positive regulation of metabolic process (GO:0009893)	1531	821	696.3	1.18	7.40E-05	3.45E-03

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
220	negative regulation of biological process (GO:0048519)	2170	1162	986.92	1.18	1.93E-06	1.32E-04
221	positive regulation of cellular metabolic process (GO:0031325)	1304	696	593.06	1.17	4.41E-04	1.68E-02
222	cellular component organisation (GO:0016043)	3489	1848	1586.8	1.16	3.90E-09	4.50E-07
223	regulation of nitrogen compound metabolic process (GO:0051171)	3100	1637	1409.88	1.16	8.05E-08	7.22E-06
224	regulation of cellular metabolic process (GO:0031323)	3079	1621	1400.33	1.16	1.73E-07	1.45E-05
225	regulation of primary metabolic process (GO:0080090)	3152	1657	1433.53	1.16	1.53E-07	1.31E-05
226	regulation of macromolecule metabolic process (GO:0060255)	3330	1750	1514.49	1.16	6.34E-08	5.77E-06
227	localization (GO:0051179)	3015	1578	1371.23	1.15	7.77E-07	5.73E-05
228	regulation of macromolecule biosynthetic process (GO:0010556)	2210	1156	1005.11	1.15	4.10E-05	2.05E-03
229	regulation of metabolic process (GO:0019222)	3609	1880	1641.38	1.15	9.52E-08	8.34E-06
230	regulation of cellular biosynthetic process (GO:0031326)	2264	1178	1029.67	1.14	6.77E-05	3.21E-03
231	regulation of biosynthetic process (GO:0009889)	2283	1186	1038.31	1.14	7.77E-05	3.61E-03
232	regulation of nucleobase-containing compound metabolic process (GO:0019219)	2313	1201	1051.96	1.14	7.22E-05	3.40E-03
233	regulation of gene expression (GO:0010468)	2622	1355	1192.49	1.14	3.76E-05	1.91E-03
234	establishment of localization (GO:0051234)	2730	1401	1241.61	1.13	6.76E-05	3.22E-03
235	regulation of RNA metabolic process (GO:0051252)	2182	1115	992.38	1.12	7.94E-04	2.86E-02
236	transport (GO:0006810)	2665	1357	1212.05	1.12	2.53E-04	1.04E-02
237	cellular process (GO:0009987)	11844	5991	5386.67	1.11	1.30E-32	8.57E-30
238	biological_process (GO:0008150)	14632	7165	6654.66	01.08	2.54E-36	3.17E-33
239	response to stimulus (GO:0050896)	4343	1779	1975.2	0.9	2.15E-05	1.18E-03
240	response to chemical (GO:0042221)	1562	600	710.4	0.84	2.31E-04	9.64E-03
241	developmental process (GO:0032502)	2350	870	1068.78	0.81	2.36E-08	2.34E-06
242	system development (GO:0048731)	1527	564	694.48	0.81	8.51E-06	5.02E-04
243	multicellular organism development (GO:0007275)	1691	621	769.07	0.81	1.45E-06	1.02E-04
244	anatomical structure development (GO:0048856)	2109	765	959.18	0.8	9.52E-09	1.02E-06
245	cellular developmental process (GO:0048869)	1500	543	682.2	0.8	1.59E-06	1.10E-04
246	signal transduction (GO:0007165)	2919	1051	1327.57	0.79	9.39E-13	1.46E-10
247	cell differentiation (GO:0030154)	1491	536	678.11	0.79	8.79E-07	6.36E-05
248	anatomical structure morphogenesis (GO:0009653)	963	345	437.97	0.79	7.84E-05	3.63E-03
249	cell communication (GO:0007154)	3157	1129	1435.81	0.79	1.62E-14	3.19E-12
250	response to external stimulus (GO:0009605)	837	298	380.67	0.78	1.88E-04	8.03E-03
251	animal organ development (GO:0048513)	1124	398	511.2	0.78	7.60E-06	4.55E-04
252	signalling (GO:0023052)	3113	1097	1415.8	0.77	8.79E-16	1.89E-13

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
253	nervous system development (GO:0007399)	911	313	414.32	0.76	8.33E-06	4.94E-04
254	biological process involved in interspecies interaction between organisms (GO:0044419)	489	167	222.4	0.75	1.02E-03	3.46E-02
255	multicellular organismal process (GO:0032501)	2555	858	1162.02	0.74	4.36E-17	1.06E-14
256	neurogenesis (GO:0022008)	591	194	268.79	0.72	4.68E-05	2.30E-03
257	defence response (GO:0006952)	437	142	198.75	0.71	3.37E-04	1.34E-02
258	cell surface receptor signalling pathway (GO:0007166)	1188	382	540.3	0.71	5.10E-10	6.49E-08
259	positive regulation of immune system process (GO:0002684)	334	107	151.9	0.7	1.29E-03	4.30E-02
260	neuron differentiation (GO:0030182)	519	165	236.04	0.7	3.40E-05	1.76E-03
261	generation of neurons (GO:0048699)	552	175	251.05	0.7	1.64E-05	9.17E-04
262	defence response to other organism (GO:0098542)	336	104	152.81	0.68	4.12E-04	1.59E-02
263	Unclassified (UNCLASSIFIED)	3477	1071	1581.34	0.68	2.54E-36	3.56E-33
264	animal organ morphogenesis (GO:0009887)	368	113	167.37	0.68	1.58E-04	6.86E-03
265	cell junction organisation (GO:0034330)	235	70	106.88	0.65	1.49E-03	4.86E-02
266	regulation of ion transport (GO:0043269)	266	78	120.98	0.64	4.32E-04	1.65E-02
267	axon development (GO:0061564)	220	64	100.06	0.64	1.25E-03	4.22E-02
268	immune response (GO:0006955)	502	143	228.31	0.63	2.28E-07	1.89E-05
269	positive regulation of immune response (GO:0050778)	227	64	103.24	0.62	4.58E-04	1.73E-02
270	cell-cell signalling (GO:0007267)	479	131	217.85	0.6	6.21E-08	5.80E-06
271	locomotion (GO:0040011)	270	70	122.8	0.57	1.20E-05	6.90E-04
272	immune response-regulating signalling pathway (GO:0002764)	182	47	82.77	0.57	3.25E-04	1.31E-02
273	chemotaxis (GO:0006935)	262	65	119.16	0.55	4.49E-06	2.80E-04
274	taxis (GO:0042330)	263	65	119.61	0.54	3.51E-06	2.25E-04
275	response to bacterium (GO:0009617)	250	61	113.7	0.54	4.65E-06	2.88E-04
276	cell adhesion (GO:0007155)	541	132	246.05	0.54	6.85E-12	9.60E-10
277	pattern specification process (GO:0007389)	201	49	91.42	0.54	3.64E-05	1.85E-03
278	phagocytosis (GO:0006909)	132	32	60.03	0.53	9.25E-04	3.22E-02
279	axon guidance (GO:0007411)	157	38	71.4	0.53	2.53E-04	1.04E-02
280	neuron projection guidance (GO:0097485)	158	38	71.86	0.53	2.55E-04	1.05E-02
281	synaptic signalling (GO:0099536)	250	60	113.7	0.53	2.50E-06	1.66E-04
282	trans-synaptic signalling (GO:0099537)	230	54	104.6	0.52	4.23E-06	2.65E-04
283	regulation of membrane potential (GO:0042391)	192	44	87.32	0.5	1.49E-05	8.41E-04
284	regionalization (GO:0003002)	166	38	75.5	0.5	5.35E-05	2.59E-03
285	chemical synaptic transmission (GO:0007268)	223	51	101.42	0.5	2.82E-06	1.85E-04
286	anterograde trans-synaptic signalling (GO:0098916)	223	51	101.42	0.5	2.82E-06	1.84E-04
287	positive regulation of cell activation (GO:0050867)	171	39	77.77	0.5	3.63E-05	1.86E-03
288	anterior/posterior pattern specification (GO:0009952)	115	26	52.3	0.5	7.09E-04	2.57E-02

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
289	positive regulation of leukocyte activation (GO:0002696)	167	37	75.95	0.49	2.70E-05	1.45E-03
290	activation of immune response (GO:0002253)	169	37	76.86	0.48	2.06E-05	1.14E-03
291	positive regulation of lymphocyte activation (GO:0051251)	160	35	72.77	0.48	3.46E-05	1.78E-03
292	system process (GO:0003008)	714	156	324.73	0.48	1.66E-19	5.15E-17
293	cell-cell adhesion (GO:0098609)	307	65	139.62	0.47	1.71E-09	2.04E-07
294	regulation of B cell activation (GO:0050864)	111	23	50.48	0.46	2.49E-04	1.03E-02
295	cell fate commitment (GO:0045165)	129	26	58.67	0.44	5.51E-05	2.65E-03
296	immune response-regulating cell surface receptor signalling pathway (GO:0002768)	142	26	64.58	0.4	3.49E-06	2.25E-04
297	adaptive immune response (GO:0002250)	173	30	78.68	0.38	7.44E-08	6.73E-06
298	immune effector process (GO:0002252)	194	33	88.23	0.37	8.58E-09	9.33E-07
299	viral process (GO:0016032)	71	12	32.29	0.37	5.36E-04	2.02E-02
300	nervous system process (GO:0050877)	487	82	221.49	0.37	8.15E-21	3.51E-18
301	external encapsulating structure organisation (GO:0045229)	167	28	75.95	0.37	6.12E-08	5.82E-06
302	extracellular matrix organisation (GO:0030198)	167	28	75.95	0.37	6.12E-08	5.77E-06
303	immune response-activating cell surface receptor signalling pathway (GO:0002429)	138	23	62.76	0.37	8.62E-07	6.31E-05
304	immune response-activating signal transduction (GO:0002757)	138	23	62.76	0.37	8.62E-07	6.27E-05
305	extracellular structure organisation (GO:0043062)	169	28	76.86	0.36	4.41E-08	4.26E-06
306	positive regulation of B cell activation (GO:0050871)	99	16	45.03	0.36	2.78E-05	1.48E-03
307	antigen receptor-mediated signalling pathway (GO:0050851)	131	21	59.58	0.35	9.80E-07	7.04E-05
308	defence response to bacterium (GO:0042742)	154	24	70.04	0.34	5.46E-08	5.23E-06
309	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (GO:0002460)	142	22	64.58	0.34	1.67E-07	1.42E-05
310	leukocyte mediated immunity (GO:0002443)	149	23	67.77	0.34	8.15E-08	7.25E-06
311	adenylate cyclase-modulating G protein-coupled receptor signalling pathway (GO:0007188)	165	25	75.04	0.33	9.23E-09	9.94E-07
312	G protein-coupled receptor signalling pathway (GO:0007186)	672	100	305.63	0.33	4.36E-33	3.49E-30
313	adenylate cyclase-activating G protein-coupled receptor signalling pathway (GO:0007189)	102	15	46.39	0.32	4.99E-06	3.07E-04
314	membrane invagination (GO:0010324)	104	15	47.3	0.32	3.56E-06	2.26E-04
315	regulation of postsynaptic membrane potential (GO:0060078)	78	11	35.47	0.31	4.74E-05	2.32E-03
316	humoral immune response (GO:0006959)	131	18	59.58	0.3	6.26E-08	5.80E-06
317	sensory perception (GO:0007600)	313	43	142.35	0.3	1.92E-17	5.13E-15
318	plasma membrane invagination (GO:0099024)	103	14	46.84	0.3	1.73E-06	1.19E-04

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (8236 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
319	cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	174	23	79.14	0.29	1.24E-10	1.65E-08
320	B cell mediated immunity (GO:0019724)	107	14	48.66	0.29	5.61E-07	4.33E-05
321	lymphocyte mediated immunity (GO:0002449)	139	18	63.22	0.28	6.81E-09	7.55E-07
322	immunoglobulin mediated immune response (GO:0016064)	106	13	48.21	0.27	2.55E-07	2.09E-05
323	homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)	135	16	61.4	0.26	2.93E-09	3.42E-07
324	chemical synaptic transmission, postsynaptic (GO:0099565)	68	8	30.93	0.26	3.12E-05	1.63E-03
325	phagocytosis, engulfment (GO:0006911)	97	11	44.12	0.25	3.38E-07	2.71E-05
326	cell recognition (GO:0008037)	124	14	56.4	0.25	5.56E-09	6.23E-07
327	zymogen activation (GO:0031638)	45	5	20.47	0.24	6.31E-04	2.32E-02
328	excitatory postsynaptic potential (GO:0060079)	65	7	29.56	0.24	3.07E-05	1.62E-03
329	positive regulation of pathway-restricted SMAD protein phosphorylation (GO:0010862)	40	4	18.19	0.22	8.87E-04	3.11E-02
330	neuropeptide signalling pathway (GO:0007218)	91	7	41.39	0.17	1.02E-08	1.08E-06
331	B cell receptor signalling pathway (GO:0050853)	91	7	41.39	0.17	1.02E-08	1.07E-06
332	complement activation (GO:0006956)	94	7	42.75	0.16	4.24E-09	4.85E-07
333	humoral immune response mediated by circulating immunoglobulin (GO:0002455)	83	6	37.75	0.16	2.87E-08	2.82E-06
334	entry into host (GO:0044409)	31	2	14.1	0.14	1.00E-03	3.43E-02
335	viral entry into host cell (GO:0046718)	31	2	14.1	0.14	1.00E-03	3.42E-02
336	detection of stimulus (GO:0051606)	209	13	95.05	0.14	1.50E-20	5.78E-18
337	complement activation, classical pathway (GO:0006958)	82	5	37.29	0.13	8.50E-09	9.34E-07
338	sensory perception of chemical stimulus (GO:0007606)	143	7	65.04	0.11	5.81E-16	1.30E-13
339	phagocytosis, recognition (GO:0006910)	75	2	34.11	0.06	2.43E-10	3.13E-08
340	detection of chemical stimulus (GO:0009593)	152	4	69.13	0.06	5.64E-20	2.04E-17
341	DNA integration (GO:0015074)	86	2	39.11	0.05	4.87E-12	6.91E-10
342	sensory perception of smell (GO:0007608)	131	3	59.58	0.05	7.97E-18	2.35E-15
343	detection of stimulus involved in sensory perception (GO:0050906)	149	3	67.77	0.04	1.30E-20	5.22E-18
344	detection of chemical stimulus involved in sensory perception (GO:0050907)	135	1	61.4	0.02	8.31E-21	3.45E-18
345	cytotoxic T cell differentiation (GO:0045065)	21	0	9.55	< 0.01	5.61E-04	2.10E-02
346	detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	128	0	58.21	< 0.01	1.92E-21	8.96E-19

Gene Ontology terms at hatching and fledging DEGs

Table S2. Biological process gene ontology (GO) terms of differentially expressed genes between hatching and fledging from PANTHER. The second column represents the GO description and annotation ID; the third column contains the number of genes in the reference list (*Gallus gallus*) that map to this category; the fourth column contains the number of genes in our uploaded list; the fifth column contains the expected number of genes expected in our list for this category based on the reference list; the sixth column shows the fold enrichment of genes observed in the uploaded list over the expected; the seventh column shows the *P-value* of the Fisher exact test; the eighth column shows the false discovery rate (FDR) using the Benjamini-Hochberg procedure.

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (2014 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
1	attachment of mitotic spindle microtubules to kinetochore (GO:0051315)	8	7	0.89	7.87	3.10E-04	1.58E-02
2	mitotic chromosome condensation (GO:0007076)	10	8	1.11	7.19	1.73E-04	9.74E-03
3	kinetochore organization (GO:0051383)	17	13	1.89	6.88	2.27E-06	2.17E-04
4	kinetochore assembly (GO:0051382)	14	10	1.56	6.42	5.17E-05	3.41E-03
5	mitotic metaphase plate congression (GO:0007080)	12	8	1.33	5.99	4.14E-04	1.99E-02
6	centromere complex assembly (GO:0034508)	26	17	2.89	5.88	3.14E-07	3.70E-05
7	attachment of spindle microtubules to kinetochore (GO:0008608)	17	10	1.89	5.29	1.68E-04	9.55E-03
8	metaphase plate congression (GO:0051310)	19	11	2.11	5.21	8.83E-05	5.44E-03
9	formation of cytoplasmic translation initiation complex (GO:0001732)	14	8	1.56	5.14	8.77E-04	3.60E-02
10	nuclear DNA replication (GO:0033260)	16	9	1.78	05.06	4.56E-04	2.14E-02
11	cell cycle DNA replication (GO:0044786)	16	9	1.78	05.06	4.56E-04	2.13E-02
12	mitotic sister chromatid segregation (GO:0000070)	76	42	8.45	4.97	5.87E-14	1.43E-11
13	chromosome localization (GO:0050000)	24	13	2.67	4.87	3.45E-05	2.41E-03
14	cytoplasmic translation (GO:0002181)	52	28	5.78	4.84	1.38E-09	2.39E-07
15	DNA unwinding involved in DNA replication (GO:0006268)	15	8	1.67	4.8	1.23E-03	4.73E-02
16	mitotic spindle assembly checkpoint signaling (GO:0007094)	17	9	1.89	4.76	6.35E-04	2.79E-02
17	mitotic spindle checkpoint signaling (GO:0071174)	17	9	1.89	4.76	6.35E-04	2.78E-02
18	spindle assembly checkpoint signaling (GO:0071173)	17	9	1.89	4.76	6.35E-04	2.77E-02
19	spindle checkpoint signaling (GO:0031577)	17	9	1.89	4.76	6.35E-04	2.76E-02
20	mitotic G2 DNA damage checkpoint signaling (GO:0007095)	21	11	2.34	4.71	1.73E-04	9.78E-03
21	establishment of chromosome localization (GO:0051303)	23	12	2.56	4.69	9.06E-05	5.55E-03
22	sister chromatid segregation (GO:0000819)	94	48	10.45	4.59	9.56E-15	2.82E-12
23	negative regulation of mitotic metaphase/anaphase transition (GO:0045841)	18	9	2	4.5	8.69E-04	3.58E-02
24	cytoplasmic translational initiation (GO:0002183)	24	12	2.67	4.5	1.24E-04	7.29E-03

25	chromosome condensation (GO:0030261)	29	14	3.23	4.34	4.60E-05	3.13E-03
26	negative regulation of mitotic nuclear division (GO:0045839)	21	10	2.34	4.28	6.08E-04	2.69E-02
27	negative regulation of mitotic sister chromatid separation (GO:2000816)	19	9	2.11	4.26	1.17E-03	4.58E-02
28	negative regulation of metaphase/anaphase transition of cell cycle (GO:1902100)	19	9	2.11	4.26	1.17E-03	4.56E-02
29	negative regulation of mitotic sister chromatid segregation (GO:0033048)	19	9	2.11	4.26	1.17E-03	4.55E-02
30	negative regulation of sister chromatid segregation (GO:0033046)	19	9	2.11	4.26	1.17E-03	4.53E-02
31	mitotic G2/M transition checkpoint (GO:0044818)	28	13	3.11	4.17	1.16E-04	6.90E-03
32	mitotic DNA damage checkpoint signaling (GO:0044773)	38	17	4.23	04.02	1.58E-05	1.28E-03
33	mitotic nuclear division (GO:0140014)	106	47	11.79	3.99	9.88E-13	2.26E-10
34	negative regulation of chromosome organization (GO:2001251)	25	11	2.78	3.96	5.52E-04	2.51E-02
35	mitotic cytokinesis (GO:0000281)	37	16	4.11	3.89	3.86E-05	2.67E-03
36	negative regulation of cell cycle G2/M phase transition (GO:1902750)	35	15	3.89	3.85	7.26E-05	4.65E-03
37	negative regulation of G2/M transition of mitotic cell cycle (GO:0010972)	35	15	3.89	3.85	7.26E-05	4.63E-03
38	chromosome segregation (GO:0007059)	172	73	19.13	3.82	3.57E-18	1.43E-15
39	nuclear chromosome segregation (GO:0098813)	141	59	15.68	3.76	9.74E-15	2.80E-12
40	mitotic cell cycle checkpoint signaling (GO:0007093)	65	27	7.23	3.73	1.84E-07	2.34E-05
41	mitotic DNA integrity checkpoint signaling (GO:0044774)	41	17	4.56	3.73	3.41E-05	2.41E-03
42	signal transduction by p53 class mediator (GO:0072331)	27	11	3	3.66	9.17E-04	3.71E-02
43	translation (GO:0006412)	290	113	32.25	3.5	5.57E-25	6.24E-22
44	DNA-templated DNA replication (GO:0006261)	90	35	10.01	3.5	1.15E-08	1.78E-06
45	execution phase of apoptosis (GO:0097194)	31	12	3.45	3.48	7.81E-04	3.29E-02
46	DNA conformation change (GO:0071103)	81	31	09.01	3.44	1.03E-07	1.41E-05
47	heterochromatin organization (GO:0070828)	34	13	3.78	3.44	5.25E-04	2.39E-02
48	cell cycle checkpoint signaling (GO:0000075)	84	32	9.34	3.43	7.04E-08	9.99E-06
49	ATP synthesis coupled electron transport (GO:0042773)	50	19	5.56	3.42	3.18E-05	2.29E-03
50	peptide biosynthetic process (GO:0043043)	298	113	33.14	3.41	3.36E-24	3.14E-21
51	electron transport chain (GO:0022900)	74	28	8.23	3.4	5.07E-07	5.75E-05
52	DNA damage checkpoint signaling (GO:0000077)	53	20	5.89	3.39	2.16E-05	1.69E-03
53	oxidative phosphorylation (GO:0006119)	61	23	6.78	3.39	5.41E-06	4.89E-04
54	nuclear division (GO:0000280)	176	66	19.57	3.37	1.76E-14	4.92E-12
55	DNA integrity checkpoint signaling (GO:0031570)	59	22	6.56	3.35	9.91E-06	8.35E-04
56	double-strand break repair via homologous recombination (GO:0000724)	86	32	9.56	3.35	1.09E-07	1.47E-05

57	signal transduction in response to DNA damage (GO:0042770)	62	23	6.9	3.34	6.72E-06	5.93E-04
58	cytoskeleton-dependent cytokinesis (GO:0061640)	54	20	06.01	3.33	2.68E-05	2.06E-03
59	protein-DNA complex assembly (GO:0065004)	92	34	10.23	3.32	5.04E-08	7.25E-06
60	recombinational repair (GO:0000725)	87	32	9.68	3.31	1.35E-07	1.78E-05
61	organelle fission (GO:0048285)	192	70	21.35	3.28	8.80E-15	2.66E-12
62	chromosome organization (GO:0051276)	280	102	31.14	3.28	6.34E-21	3.38E-18
63	regulation of G2/M transition of mitotic cell cycle (GO:0010389)	44	16	4.89	3.27	1.98E-04	1.09E-02
64	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	47	17	5.23	3.25	1.33E-04	7.70E-03
65	mitotic cell cycle process (GO:1903047)	296	106	32.92	3.22	3.02E-21	1.69E-18
66	negative regulation of mitotic cell cycle phase transition (GO:1901991)	73	26	8.12	3.2	3.18E-06	2.97E-04
67	double-strand break repair (GO:0006302)	146	52	16.24	3.2	4.85E-11	9.71E-09
68	aerobic respiration (GO:0009060)	90	32	10.01	3.2	2.51E-07	3.13E-05
69	aerobic electron transport chain (GO:0019646)	45	16	5	3.2	2.44E-04	1.31E-02
70	meiotic cell cycle process (GO:1903046)	94	33	10.45	3.16	2.08E-07	2.62E-05
71	cytokinesis (GO:0000910)	57	20	6.34	3.15	5.00E-05	3.34E-03
72	cell division (GO:0051301)	211	74	23.47	3.15	7.76E-15	2.42E-12
73	meiotic cell cycle (GO:0051321)	120	42	13.35	3.15	5.20E-09	8.45E-07
74	respiratory electron transport chain (GO:0022904)	60	21	6.67	3.15	3.37E-05	2.39E-03
75	regulation of cell cycle G2/M phase transition (GO:1902749)	49	17	5.45	3.12	2.00E-04	1.10E-02
76	mitotic cell cycle (GO:0000278)	365	125	40.59	03.08	1.76E-23	1.32E-20
77	DNA geometric change (GO:0032392)	73	25	8.12	03.08	8.53E-06	7.30E-04
78	protein-DNA complex subunit organization (GO:0071824)	106	36	11.79	03.05	1.17E-07	1.56E-05
79	translational initiation (GO:0006413)	53	18	5.89	03.05	1.64E-04	9.41E-03
80	amide biosynthetic process (GO:0043604)	374	127	41.59	03.05	1.98E-23	1.39E-20
81	mitochondrial transmembrane transport (GO:1990542)	62	21	6.9	03.05	4.99E-05	3.35E-03
82	DNA replication (GO:0006260)	139	47	15.46	03.04	1.62E-09	2.76E-07
83	microtubule cytoskeleton organization involved in mitosis (GO:1902850)	81	27	09.01	3	5.70E-06	5.07E-04
84	negative regulation of cell cycle phase transition (GO:1901988)	104	34	11.57	2.94	5.36E-07	6.00E-05
85	negative regulation of mitotic cell cycle (GO:0045930)	99	32	11.01	2.91	1.40E-06	1.47E-04
86	ribosomal large subunit biogenesis (GO:0042273)	65	21	7.23	2.9	8.71E-05	5.43E-03
87	meiosis I cell cycle process (GO:0061982)	62	20	6.9	2.9	1.29E-04	7.56E-03
88	peptide metabolic process (GO:0006518)	377	121	41.93	2.89	7.98E-21	4.07E-18
89	spindle assembly (GO:0051225)	53	17	5.89	2.88	4.26E-04	2.03E-02
90	cellular respiration (GO:0045333)	103	33	11.46	2.88	1.69E-06	1.68E-04
91	regulation of chromosome segregation (GO:0051983)	50	16	5.56	2.88	6.36E-04	2.75E-02

92	meiotic chromosome segregation (GO:0045132)	50	16	5.56	2.88	6.36E-04	2.74E-02
93	mitotic spindle organization (GO:0007052)	57	18	6.34	2.84	3.42E-04	1.72E-02
94	meiotic nuclear division (GO:0140013)	83	26	9.23	2.82	3.06E-05	2.27E-03
95	meiosis I (GO:0007127)	61	19	6.78	2.8	2.74E-04	1.43E-02
96	intrinsic apoptotic signaling pathway (GO:0097193)	55	17	6.12	2.78	6.05E-04	2.69E-02
97	spindle organization (GO:0007051)	91	28	10.12	2.77	1.67E-05	1.34E-03
98	DNA duplex unwinding (GO:0032508)	69	21	7.67	2.74	2.48E-04	1.32E-02
99	chromatin remodeling (GO:0006338)	196	59	21.8	2.71	6.55E-10	1.15E-07
100	energy derivation by oxidation of organic compounds (GO:0015980)	130	38	14.46	2.63	1.56E-06	1.57E-04
101	DNA repair (GO:0006281)	324	94	36.03	2.61	4.15E-14	1.03E-11
102	cell cycle process (GO:0022402)	466	135	51.83	2.6	8.79E-20	3.94E-17
103	cell cycle (GO:0007049)	638	184	70.96	2.59	1.88E-26	2.34E-23
104	negative regulation of cell cycle process (GO:0010948)	118	34	13.12	2.59	7.37E-06	6.46E-04
105	cellular response to DNA damage stimulus (GO:0006974)	434	124	48.27	2.57	7.19E-18	2.78E-15
106	DNA recombination (GO:0006310)	186	53	20.69	2.56	2.46E-08	3.68E-06
107	generation of precursor metabolites and energy (GO:0006091)	188	53	20.91	2.53	3.11E-08	4.59E-06
108	nucleoside monophosphate metabolic process (GO:0009123)	71	20	7.9	2.53	6.34E-04	2.80E-02
109	regulation of chromosome organization (GO:0033044)	83	23	9.23	2.49	4.52E-04	2.13E-02
110	rRNA processing (GO:0006364)	160	44	17.79	2.47	8.40E-07	9.32E-05
111	ribosome biogenesis (GO:0042254)	226	62	25.13	2.47	7.25E-09	1.16E-06
112	cellular amide metabolic process (GO:0043603)	533	146	59.28	2.46	2.04E-19	8.81E-17
113	nucleoside triphosphate metabolic process (GO:0009141)	99	27	11.01	2.45	1.32E-04	7.64E-03
114	negative regulation of cell cycle (GO:0045786)	147	40	16.35	2.45	3.42E-06	3.17E-04
115	chromatin organization (GO:0006325)	306	82	34.03	2.41	5.35E-11	1.05E-08
116	regulation of mitotic cell cycle phase transition (GO:1901990)	120	32	13.35	2.4	5.41E-05	3.54E-03
117	rRNA metabolic process (GO:0016072)	169	45	18.8	2.39	1.51E-06	1.54E-04
118	regulation of cell cycle phase transition (GO:1901987)	173	46	19.24	2.39	1.13E-06	1.22E-04
119	purine nucleoside triphosphate metabolic process (GO:0009144)	79	21	8.79	2.39	1.10E-03	4.38E-02
120	ribonucleoside triphosphate metabolic process (GO:0009199)	84	22	9.34	2.35	8.84E-04	3.60E-02
121	ribonucleoprotein complex biogenesis (GO:0022613)	321	84	35.7	2.35	8.22E-11	1.56E-08
122	cellular macromolecule biosynthetic process (GO:0034645)	603	149	67.06	2.22	2.35E-16	7.74E-14
123	negative regulation of organelle organization (GO:0010639)	130	32	14.46	2.21	1.76E-04	9.85E-03
124	regulation of mitotic cell cycle (GO:0007346)	188	46	20.91	2.2	8.05E-06	6.99E-04
125	regulation of cell cycle process (GO:0010564)	288	69	32.03	2.15	9.26E-08	1.28E-05

126	ribonucleoprotein complex assembly (GO:0022618)	130	31	14.46	2.14	3.19E-04	1.61E-02
127	non-membrane-bounded organelle assembly (GO:0140694)	191	45	21.24	2.12	2.48E-05	1.92E-03
128	ribonucleoprotein complex subunit organization (GO:0071826)	136	32	15.13	2.12	3.89E-04	1.89E-02
129	purine-containing compound biosynthetic process (GO:0072522)	136	32	15.13	2.12	3.89E-04	1.88E-02
130	nucleoside phosphate biosynthetic process (GO:1901293)	182	42	20.24	02.07	6.87E-05	4.42E-03
131	cellular nitrogen compound biosynthetic process (GO:0044271)	1062	242	118.11	02.05	2.31E-22	1.44E-19
132	organonitrogen compound biosynthetic process (GO:1901566)	969	220	107.77	02.04	3.74E-20	1.82E-17
133	ncRNA metabolic process (GO:0034660)	398	89	44.26	02.01	1.69E-08	2.56E-06
134	ncRNA processing (GO:0034470)	308	68	34.25	1.99	1.38E-06	1.45E-04
135	DNA metabolic process (GO:0006259)	594	131	66.06	1.98	1.41E-11	2.93E-09
136	nucleotide metabolic process (GO:0009117)	306	67	34.03	1.97	2.01E-06	1.98E-04
137	nucleoside phosphate metabolic process (GO:0006753)	321	70	35.7	1.96	1.43E-06	1.48E-04
138	establishment of organelle localization (GO:0051656)	179	39	19.91	1.96	3.64E-04	1.77E-02
139	regulation of cell cycle (GO:0051726)	452	98	50.27	1.95	1.33E-08	2.04E-06
140	nucleotide biosynthetic process (GO:0009165)	176	38	19.57	1.94	5.01E-04	2.29E-02
141	purine-containing compound metabolic process (GO:0072521)	252	54	28.03	1.93	4.50E-05	3.10E-03
142	mitochondrion organization (GO:0007005)	234	50	26.02	1.92	8.45E-05	5.32E-03
143	nucleobase-containing small molecule metabolic process (GO:0055086)	372	78	41.37	1.89	1.45E-06	1.49E-04
144	cellular response to stress (GO:0033554)	758	158	84.3	1.87	5.58E-12	1.20E-09
145	purine nucleotide metabolic process (GO:0006163)	226	47	25.13	1.87	2.21E-04	1.20E-02
146	regulation of translation (GO:0006417)	198	41	22.02	1.86	6.75E-04	2.89E-02
147	cellular nitrogen compound metabolic process (GO:0034641)	2557	527	284.38	1.85	2.21E-40	2.47E-36
148	cellular macromolecule metabolic process (GO:0044260)	1850	381	205.75	1.85	5.30E-28	7.43E-25
149	ribose phosphate metabolic process (GO:0019693)	240	49	26.69	1.84	2.29E-04	1.23E-02
150	protein-containing complex assembly (GO:0065003)	641	130	71.29	1.82	2.11E-09	3.48E-07
151	gene expression (GO:0010467)	1525	309	169.6	1.82	1.71E-21	1.01E-18
152	regulation of cellular amide metabolic process (GO:0034248)	213	43	23.69	1.82	7.12E-04	3.01E-02
153	macromolecule biosynthetic process (GO:0009059)	1045	210	116.22	1.81	2.98E-14	7.59E-12
154	protein-containing complex organization (GO:0043933)	730	146	81.19	1.8	5.14E-10	9.14E-08
155	ribonucleotide metabolic process (GO:0009259)	231	46	25.69	1.79	5.60E-04	2.53E-02
156	regulation of cellular macromolecule biosynthetic process (GO:2000112)	223	44	24.8	1.77	9.28E-04	3.73E-02
157	nucleobase-containing compound metabolic process (GO:0006139)	2067	396	229.88	1.72	9.35E-24	7.49E-21

158	heterocycle metabolic process (GO:0046483)	2156	413	239.78	1.72	7.40E-25	7.54E-22
159	nucleic acid metabolic process (GO:0090304)	1708	324	189.96	1.71	1.01E-18	4.17E-16
160	cellular aromatic compound metabolic process (GO:0006725)	2182	412	242.67	1.7	7.86E-24	6.78E-21
161	cellular biosynthetic process (GO:0044249)	1696	316	188.62	1.68	3.21E-17	1.20E-14
162	organic substance biosynthetic process (GO:1901576)	1740	322	193.51	1.66	3.63E-17	1.31E-14
163	organic cyclic compound metabolic process (GO:1901360)	2299	423	255.68	1.65	1.74E-22	1.15E-19
164	RNA processing (GO:0006396)	650	119	72.29	1.65	1.52E-06	1.53E-04
165	biosynthetic process (GO:0009058)	1787	327	198.74	1.65	8.12E-17	2.84E-14
166	regulation of organelle organization (GO:0033043)	487	89	54.16	1.64	3.00E-05	2.24E-03
167	organophosphate biosynthetic process (GO:0090407)	352	64	39.15	1.63	4.92E-04	2.26E-02
168	cellular component biogenesis (GO:0044085)	1451	258	161.37	1.6	4.96E-12	1.09E-09
169	heterocycle biosynthetic process (GO:0018130)	702	121	78.07	1.55	1.38E-05	1.15E-03
170	modification-dependent macromolecule catabolic process (GO:0043632)	412	71	45.82	1.55	9.27E-04	3.74E-02
171	microtubule-based process (GO:0007017)	538	92	59.83	1.54	1.98E-04	1.09E-02
172	nucleobase-containing compound biosynthetic process (GO:0034654)	638	109	70.96	1.54	5.09E-05	3.38E-03
173	macromolecule catabolic process (GO:0009057)	605	103	67.29	1.53	9.54E-05	5.81E-03
174	cellular macromolecule catabolic process (GO:0044265)	523	89	58.17	1.53	3.53E-04	1.74E-02
175	cellular metabolic process (GO:0044237)	4993	844	555.3	1.52	7.96E-39	4.46E-35
176	aromatic compound biosynthetic process (GO:0019438)	699	118	77.74	1.52	3.83E-05	2.67E-03
177	cellular catabolic process (GO:0044248)	994	167	110.55	1.51	1.20E-06	1.28E-04
178	protein catabolic process (GO:0030163)	472	79	52.49	1.5	1.14E-03	4.50E-02
179	nitrogen compound metabolic process (GO:0006807)	5103	850	567.53	1.5	6.43E-37	1.80E-33
180	macromolecule metabolic process (GO:0043170)	4503	740	500.8	1.48	8.06E-29	1.29E-25
181	organophosphate metabolic process (GO:0019637)	609	100	67.73	1.48	4.37E-04	2.07E-02
182	cellular component assembly (GO:0022607)	1269	208	141.13	1.47	2.56E-07	3.15E-05
183	organelle organization (GO:0006996)	2032	332	225.99	1.47	2.69E-11	5.49E-09
184	organic cyclic compound biosynthetic process (GO:1901362)	769	125	85.52	1.46	1.02E-04	6.16E-03
185	organonitrogen compound metabolic process (GO:1901564)	3655	589	406.49	1.45	8.36E-20	3.91E-17
186	primary metabolic process (GO:0044238)	5565	892	618.91	1.44	1.53E-33	3.43E-30
187	RNA metabolic process (GO:0016070)	1245	199	138.46	1.44	2.07E-06	2.01E-04
188	organic substance catabolic process (GO:1901575)	1034	165	115	1.43	1.88E-05	1.49E-03
189	organonitrogen compound catabolic process (GO:1901565)	678	108	75.4	1.43	6.62E-04	2.84E-02

190	catabolic process (GO:0009056)	1188	189	132.12	1.43	4.87E-06	4.44E-04
191	metabolic process (GO:0008152)	6180	982	687.31	1.43	2.77E-37	1.04E-33
192	protein metabolic process (GO:0019538)	2941	465	327.08	1.42	7.40E-14	1.76E-11
193	organic substance metabolic process (GO:0071704)	5871	925	652.95	1.42	1.06E-32	1.98E-29
194	small molecule metabolic process (GO:0044281)	1069	168	118.89	1.41	3.13E-05	2.26E-03
195	cellular component organization or biogenesis (GO:0071840)	3660	558	407.05	1.37	2.75E-14	7.17E-12
196	regulation of cellular component organization (GO:0051128)	855	129	95.09	1.36	1.29E-03	4.95E-02
197	response to stress (GO:0006950)	1366	201	151.92	1.32	1.89E-04	1.06E-02
198	cellular component organization (GO:0016043)	3489	511	388.03	1.32	2.30E-10	4.22E-08
199	macromolecule modification (GO:0043412)	2126	309	236.44	1.31	5.47E-06	4.91E-04
200	protein modification process (GO:0036211)	1972	286	219.32	1.3	1.53E-05	1.26E-03
201	negative regulation of cellular process (GO:0048523)	1849	255	205.64	1.24	8.34E-04	3.45E-02
202	cellular process (GO:0009987)	11844	1497	1317.24	1.14	2.54E-16	8.14E-14
203	biological_process (GO:0008150)	14632	1762	1627.3	01.08	2.69E-14	7.35E-12
204	response to stimulus (GO:0050896)	4343	411	483.01	0.85	2.91E-04	1.50E-02
205	system development (GO:0048731)	1527	123	169.83	0.72	2.26E-04	1.23E-02
206	multicellular organism development (GO:0007275)	1691	133	188.07	0.71	2.92E-05	2.21E-03
207	developmental process (GO:0032502)	2350	180	261.36	0.69	8.26E-08	1.16E-05
208	anatomical structure development (GO:0048856)	2109	161	234.55	0.69	3.47E-07	4.05E-05
209	signal transduction (GO:0007165)	2919	221	324.64	0.68	3.83E-10	6.92E-08
210	cell communication (GO:0007154)	3157	235	351.11	0.67	1.08E-11	2.28E-09
211	cellular response to chemical stimulus (GO:0070887)	1039	77	115.55	0.67	2.59E-04	1.38E-02
212	animal organ development (GO:0048513)	1124	83	125.01	0.66	1.10E-04	6.57E-03
213	anatomical structure morphogenesis (GO:0009653)	963	71	107.1	0.66	3.55E-04	1.74E-02
214	signaling (GO:0023052)	3113	228	346.21	0.66	3.27E-12	7.34E-10
215	Unclassified (UNCLASSIFIED)	3477	252	386.7	0.65	2.69E-14	7.18E-12
216	regulation of multicellular organismal process (GO:0051239)	816	57	90.75	0.63	2.72E-04	1.43E-02
217	cell differentiation (GO:0030154)	1491	104	165.82	0.63	4.23E-07	4.89E-05
218	cell development (GO:0048468)	703	49	78.18	0.63	7.96E-04	3.32E-02
219	cellular developmental process (GO:0048869)	1500	104	166.82	0.62	2.86E-07	3.45E-05
220	response to external stimulus (GO:0009605)	837	57	93.09	0.61	1.27E-04	7.44E-03
221	response to chemical (GO:0042221)	1562	105	173.72	0.6	3.44E-08	5.01E-06
222	multicellular organismal process (GO:0032501)	2555	170	284.16	0.6	1.26E-13	2.94E-11
223	response to organic substance (GO:0010033)	995	66	110.66	0.6	9.45E-06	8.03E-04
224	cellular response to organic substance (GO:0071310)	787	49	87.53	0.56	1.62E-05	1.30E-03
225	cell motility (GO:0048870)	457	28	50.83	0.55	9.59E-04	3.84E-02

226	biological process involved in interspecies interaction between organisms (GO:0044419)	489	29	54.38	0.53	3.53E-04	1.74E-02
227	response to other organism (GO:0051707)	432	24	48.05	0.5	3.52E-04	1.75E-02
228	response to external biotic stimulus (GO:0043207)	433	24	48.16	0.5	2.71E-04	1.43E-02
229	regulation of immune system process (GO:0002682)	508	28	56.5	0.5	7.44E-05	4.71E-03
230	response to biotic stimulus (GO:0009607)	444	24	49.38	0.49	1.68E-04	9.59E-03
231	cell surface receptor signaling pathway (GO:0007166)	1188	64	132.12	0.48	1.44E-10	2.69E-08
232	defense response to other organism (GO:0098542)	336	18	37.37	0.48	1.20E-03	4.65E-02
233	defense response (GO:0006952)	437	23	48.6	0.47	1.09E-04	6.53E-03
234	regulation of immune response (GO:0050776)	337	17	37.48	0.45	4.57E-04	2.12E-02
235	cell-cell adhesion (GO:0098609)	307	15	34.14	0.44	6.83E-04	2.91E-02
236	positive regulation of immune system process (GO:0002684)	334	16	37.15	0.43	2.91E-04	1.51E-02
237	regulation of ion transport (GO:0043269)	266	12	29.58	0.41	8.03E-04	3.33E-02
238	immune response (GO:0006955)	502	22	55.83	0.39	9.18E-07	1.01E-04
239	cell adhesion (GO:0007155)	541	23	60.17	0.38	1.62E-07	2.11E-05
240	chemotaxis (GO:0006935)	262	11	29.14	0.38	3.42E-04	1.71E-02
241	taxis (GO:0042330)	263	11	29.25	0.38	3.44E-04	1.72E-02
242	locomotion (GO:0040011)	270	11	30.03	0.37	1.91E-04	1.06E-02
243	cell-cell signaling (GO:0007267)	479	19	53.27	0.36	3.05E-07	3.63E-05
244	cell surface receptor signaling pathway involved in cell-cell signaling (GO:1905114)	203	8	22.58	0.35	1.13E-03	4.49E-02
245	regulation of hormone levels (GO:0010817)	204	8	22.69	0.35	1.14E-03	4.51E-02
246	system process (GO:0003008)	714	27	79.41	0.34	5.40E-11	1.04E-08
247	muscle structure development (GO:0061061)	201	7	22.35	0.31	4.41E-04	2.09E-02
248	cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	174	6	19.35	0.31	1.08E-03	4.32E-02
249	immune response-regulating signaling pathway (GO:0002764)	182	6	20.24	0.3	5.74E-04	2.56E-02
250	regulation of lymphocyte activation (GO:0051249)	194	6	21.58	0.28	3.13E-04	1.59E-02
251	nervous system process (GO:0050877)	487	15	54.16	0.28	1.67E-09	2.79E-07
252	regulation of leukocyte activation (GO:0002694)	211	6	23.47	0.26	8.53E-05	5.34E-03
253	synaptic signaling (GO:0099536)	250	7	27.8	0.25	1.31E-05	1.10E-03
254	regulation of cell activation (GO:0050865)	220	6	24.47	0.25	2.98E-05	2.24E-03
255	chemical synaptic transmission (GO:0007268)	223	6	24.8	0.24	3.10E-05	2.27E-03
256	anterograde trans-synaptic signaling (GO:0098916)	223	6	24.8	0.24	3.10E-05	2.26E-03
257	trans-synaptic signaling (GO:0099537)	230	6	25.58	0.23	1.54E-05	1.26E-03
258	defense response to bacterium (GO:0042742)	154	4	17.13	0.23	4.66E-04	2.15E-02
259	sensory perception (GO:0007600)	313	8	34.81	0.23	2.63E-07	3.21E-05
260	response to bacterium (GO:0009617)	250	6	27.8	0.22	2.76E-06	2.60E-04

261	antigen receptor-mediated signaling pathway (GO:0050851)	131	3	14.57	0.21	7.82E-04	3.28E-02
262	phagocytosis (GO:0006909)	132	3	14.68	0.2	7.85E-04	3.28E-02
263	immune response-activating cell surface receptor signaling pathway (GO:0002429)	138	3	15.35	0.2	5.71E-04	2.57E-02
264	immune response-activating signal transduction (GO:0002757)	138	3	15.35	0.2	5.71E-04	2.56E-02
265	immune response-regulating cell surface receptor signaling pathway (GO:0002768)	142	3	15.79	0.19	4.00E-04	1.93E-02
266	G protein-coupled receptor signaling pathway (GO:0007186)	672	14	74.74	0.19	8.50E-17	2.89E-14
267	regulation of membrane potential (GO:0042391)	192	4	21.35	0.19	2.03E-05	1.61E-03
268	positive regulation of lymphocyte activation (GO:0051251)	160	3	17.79	0.17	6.43E-05	4.19E-03
269	adenylate cyclase-modulating G protein-coupled receptor signaling pathway (GO:0007188)	165	3	18.35	0.16	4.51E-05	3.08E-03
270	positive regulation of leukocyte activation (GO:0002696)	167	3	18.57	0.16	4.75E-05	3.21E-03
271	developmental growth (GO:0048589)	112	2	12.46	0.16	8.81E-04	3.60E-02
272	positive regulation of cell activation (GO:0050867)	171	3	19.02	0.16	3.30E-05	2.35E-03
273	adaptive immune response (GO:0002250)	173	3	19.24	0.16	2.18E-05	1.69E-03
274	lymphocyte mediated immunity (GO:0002449)	139	2	15.46	0.13	9.89E-05	5.99E-03
275	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (GO:0002460)	142	2	15.79	0.13	6.64E-05	4.30E-03
276	leukocyte mediated immunity (GO:0002443)	149	2	16.57	0.12	3.09E-05	2.28E-03
277	skeletal system morphogenesis (GO:0048705)	94	1	10.45	0.1	8.90E-04	3.62E-02
278	immune effector process (GO:0002252)	194	2	21.58	0.09	4.74E-07	5.42E-05
279	immunoglobulin mediated immune response (GO:0016064)	106	1	11.79	0.08	2.78E-04	1.45E-02
280	sensory perception of smell (GO:0007608)	131	1	14.57	0.07	2.74E-05	2.09E-03
281	sensory perception of chemical stimulus (GO:0007606)	143	1	15.9	0.06	8.46E-06	7.29E-04
282	detection of stimulus involved in sensory perception (GO:0050906)	149	1	16.57	0.06	3.71E-06	3.41E-04
283	detection of stimulus (GO:0051606)	209	1	23.24	0.04	9.64E-09	1.52E-06
284	humoral immune response (GO:0006959)	131	0	14.57	< 0.01	2.35E-06	2.24E-04
285	complement activation, classical pathway (GO:0006958)	82	0	9.12	< 0.01	3.02E-04	1.55E-02
286	complement activation (GO:0006956)	94	0	10.45	< 0.01	8.80E-05	5.45E-03
287	phagocytosis, recognition (GO:0006910)	75	0	8.34	< 0.01	7.04E-04	2.99E-02
288	humoral immune response mediated by circulating immunoglobulin (GO:0002455)	83	0	9.23	< 0.01	3.10E-04	1.59E-02
289	regulation of postsynaptic membrane potential (GO:0060078)	78	0	8.67	< 0.01	4.56E-04	2.12E-02
290	detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	128	0	14.24	< 0.01	2.12E-06	2.05E-04
291	detection of chemical stimulus involved in sensory perception (GO:0050907)	135	0	15.01	< 0.01	9.31E-07	1.01E-04

292	detection of chemical stimulus (GO:0009593)	152	0	16.9	< 0.01	1.81E-07	2.33E-05

Gene Ontology terms enriched for males and females DEGs

Table S3. Biological process gene ontology (GO) terms of differentially expressed genes between males and females of 2020 from PANTHER. The second column represents the GO description and annotation ID; the third column contains the number of genes in the reference list (*Gallus gallus*) that map to this category; the fourth column contains the number of genes in our uploaded list; the fifth column contains the expected number of genes expected in our list for this category based on the reference list; the sixth column shows the fold enrichment of genes observed in the uploaded list over the expected; the seventh column shows the *P-value* of the Fisher exact test; the eighth column shows the false discovery rate (FDR) using the Benjamini-Hochberg procedure.

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (297 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
1	alcohol biosynthetic process (GO:0046165)	66	8	01.08	7.39	2.50E-05	2.81E-02
2	small molecule biosynthetic process (GO:0044283)	293	16	4.81	3.33	4.52E-05	4.22E-02
3	cellular nitrogen compound metabolic process (GO:0034641)	2557	76	41.94	1.81	2.42E-07	6.78E-04
4	heterocycle metabolic process (GO:0046483)	2156	63	35.36	1.78	5.88E-06	9.42E-03
5	nucleobase-containing compound metabolic process (GO:0006139)	2067	59	33.9	1.74	3.12E-05	3.18E-02
6	cellular aromatic compound metabolic process (GO:0006725)	2182	62	35.79	1.73	2.15E-05	2.67E-02
7	organic cyclic compound metabolic process (GO:1901360)	2299	63	37.71	1.67	4.93E-05	4.25E-02
8	cellular metabolic process (GO:0044237)	4993	129	81.89	1.58	6.63E-09	7.43E-05
9	nitrogen compound metabolic process (GO:0006807)	5103	128	83.69	1.53	5.62E-08	3.15E-04
10	primary metabolic process (GO:0044238)	5565	136	91.27	1.49	8.29E-08	3.10E-04
11	macromolecule metabolic process (GO:0043170)	4503	108	73.85	1.46	1.38E-05	1.93E-02
12	organic substance metabolic process (GO:0071704)	5871	139	96.29	1.44	3.61E-07	6.74E-04
13	metabolic process (GO:0008152)	6180	145	101.36	1.43	2.62E-07	5.87E-04

Gene Ontology terms enriched for early and late DEGs

Table S4. Biological process gene ontology (GO) terms of differentially expressed genes between early and late chicks of 2020 from PANTHER. The second column represents the GO description and annotation ID; the third column contains the number of genes in the reference list (*Gallus gallus*) that map to this category; the fourth column contains the number of genes in our uploaded list; the fifth column contains the expected number of genes expected in our list for this category based on the reference list; the sixth column shows the fold enrichment of genes observed in the uploaded list over the expected; the seventh column shows the *P-value* of the Fisher exact test; the eighth column shows the false discovery rate (FDR) using the Benjamini-Hochberg procedure. GOs with one asterisk (*) represent offspring GO terms of Protein catabolic process GOslim, and GOs with two asterisks (**) represent offspring GOterms of Protein modification process GOslim.

	GO biological process complete	Gallus gallus (18109 genes)	upload_1 (106 genes)	upload_1 (expected)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
1	peptidyl-cysteine S-nitrosylation (GO:0018119)**	3	3	0.02	> 100	3.78E-06	2.23E-03
2	protein nitrosylation (GO:0017014)**	3	3	0.02	> 100	3.78E-06	2.12E-03
3	response to interferon-gamma (GO:0034341)	34	4	0.2	20.1	6.87E-05	2.57E-02
4	cytoplasmic translation (GO:0002181)	52	6	0.3	19.71	1.07E-06	7.47E-04
5	reactive oxygen species metabolic process (GO:0072593)	40	4	0.23	17.08	1.23E-04	4.05E-02
6	translation (GO:0006412)	290	13	1.7	7.66	2.30E-08	5.16E-05
7	peptide biosynthetic process (GO:0043043)	298	13	1.74	7.45	3.13E-08	5.85E-05
8	amide biosynthetic process (GO:0043604)	374	14	2.19	6.4	5.61E-08	6.98E-05
9	peptide metabolic process (GO:0006518)	377	14	2.21	6.34	6.17E-08	6.91E-05
10	regulation of proteolysis (GO:0030162)	303	9	1.77	05.07	8.67E-05	3.13E-02
11	cellular amide metabolic process (GO:0043603)	533	15	3.12	4.81	6.44E-07	4.81E-04
12	ubiquitin-dependent protein catabolic process (GO:0006511)*	395	11	2.31	4.76	2.46E-05	1.25E-02
13	modification-dependent protein catabolic process (GO:0019941)*	403	11	2.36	4.66	2.94E-05	1.37E-02
14	modification-dependent macromolecule catabolic process (GO:0043632)	412	11	2.41	4.56	3.57E-05	1.54E-02
15	proteolysis involved in protein catabolic process (GO:0051603)*	440	11	2.58	4.27	6.37E-05	2.55E-02
16	protein catabolic process (GO:0030163)	472	11	2.76	3.98	1.17E-04	3.98E-02
17	cellular macromolecule biosynthetic process (GO:0034645)	603	14	3.53	3.97	1.35E-05	7.22E-03
18	cellular macromolecule catabolic process (GO:0044265)	523	12	03.06	3.92	6.53E-05	2.52E-02
19	RNA processing (GO:0006396)	650	14	3.8	3.68	3.05E-05	1.37E-02
20	gene expression (GO:0010467)	1525	29	8.93	3.25	1.06E-08	5.95E-05
21	cellular nitrogen compound biosynthetic process (GO:0044271)	1062	20	6.22	3.22	3.68E-06	2.29E-03
22	regulation of protein metabolic process (GO:0051246)	1021	18	5.98	03.01	2.89E-05	1.41E-02
23	organonitrogen compound biosynthetic process (GO:1901566)	969	17	5.67	3	5.25E-05	2.18E-02

24	macromolecule biosynthetic process (GO:0009059)	1045	17	6.12	2.78	1.30E-04	4.17E-02
25	cellular macromolecule metabolic process (GO:0044260)	1850	30	10.83	2.77	1.87E-07	1.75E-04
26	cellular nitrogen compound metabolic process (GO:0034641)	2557	37	14.97	2.47	7.43E-08	7.57E-05
27	protein metabolic process (GO:0019538)	2941	41	17.21	2.38	3.40E-08	5.45E-05
28	cellular biosynthetic process (GO:0044249)	1696	23	9.93	2.32	1.42E-04	4.31E-02
29	biosynthetic process (GO:0009058)	1787	24	10.46	2.29	1.07E-04	3.75E-02
30	organonitrogen compound metabolic process (GO:1901564)	3655	45	21.39	2.1	2.95E-07	2.54E-04
31	cellular metabolic process (GO:0044237)	4993	61	29.23	02.09	1.22E-10	1.36E-06
32	macromolecule metabolic process (GO:0043170)	4503	54	26.36	02.05	1.16E-08	4.35E-05
33	nitrogen compound metabolic process (GO:0006807)	5103	58	29.87	1.94	1.33E-08	3.73E-05
34	regulation of macromolecule metabolic process (GO:0060255)	3330	36	19.49	1.85	1.32E-04	4.11E-02
35	primary metabolic process (GO:0044238)	5565	58	32.57	1.78	4.02E-07	3.22E-04
36	metabolic process (GO:0008152)	6180	64	36.17	1.77	4.58E-08	6.41E-05
37	organic substance metabolic process (GO:0071704)	5871	59	34.37	1.72	1.08E-06	7.14E-04

Differentially expressed genes between early and late groups

Table S5. Differentially expressed genes (DEGs) between the early and late chicks of 2020 from DESeq2. The second column contains gene symbols; the third column contains the average of the normalised count values, divided by size factors, taken over all samples; the fourth column contains the log2 fold changes in expression between early and late chicks; the fifth column contains the log fold change standard error; the sixth column contains the Wald statistic; the seventh column contains the *P-value* of the Wald test; the eighth column contains the adjusted *P-value* for false discovery rate using the Benjamini-Hochberg test; the ninth column indicates the methodology in which the gene was detected, HTSeq corresponds to the genome-aligned approach, Salmon corresponds to the transcriptome-mapped approach. Genes are ordered by log2 fold change level.

	Gene code	baseMean	log2FC	lfcSE	stat	*P-value*	*padj*	Method
1	APC	52.405213025 3108	1.28287814036 756	0.2796947677 9304	4.58670768312 988	4.50290453574 814E-06	0.002529506 622957	HTSeq
2	BFAR	12.638628061 7075	0.9709228097 1226	0.266890965 911925	3.6379006175 6032	0.00027486944 2082	0.0189279860 99251	HTSeq
3	MXD4	31.944882586 541	0.9598639217 46805	0.19735318946 8154	4.8636858838 387	1.1521971666278 3E-06	0.0017823487 39104	HTSeq
4	ZBTB24	15.306867950 7831	0.9466379506 09543	0.2818591665 5391	3.3585494563 9487	0.00078352699 2643	0.0279019079 95983	HTSeq
5	CSNK1A1	18.590056682 5816	0.9383985143 90872	0.24171642078 4787	3.8822290655 4776	0.000103503307 864	0.0127577685 06694	HTSeq
6	KIF11	20.238104327 6004	0.9358635175 70107	0.2282548178 21684	4.1000822085 6576	4.130034015253 92E-05	0.0084369781 51847	HTSeq
7	MINDY3	19.3282707120 966	0.9136703326 99737	0.2046462110 42855	4.46463351578 204	8.020592031363 88E-06	0.00300371171 5746	HTSeq
8	DDX59	57.643866708 9479	0.9076103219 24952	0.2098101927 6418	4.3258638199 0067	1.519352704609 33E-05	0.0166694696 73428	SalmonCDS
9	CCNL1	39.8381411942 259	0.8253613204 50506	0.2013042021 73063	4.10007000122 599	4.130251876738 85E-05	0.0084369781 51847	HTSeq
10	THUMPD3	47.939305677 9282	0.8104570748 55295	0.2458224616 9505	3.2969203435 1479	0.000977512205 379	0.029575242 208395	HTSeq
11	PPP1R21	20.519519969 6551	0.7888097899 34993	0.2037132381 85925	3.8721577299 5106	0.000107876102 193	0.0127577685 06694	HTSeq
12	FQA23_000 4894	45.928688516 1317	0.7563904687 53231	0.252560898 821883	2.9948835005 0049	0.00274549680 948	0.049835556 086759	HTSeq
13	PLAT	820.3994634 44244	0.7487980261 00747	0.1856494584 43492	4.0333973089 863	5.497621147458 44E-05	0.0364197456 81228	SalmonCDS
14	UBR3	18.262736244 3969	0.7478612909 60235	0.2187100636 44009	3.41941874324 319	0.000627550721 714	0.0254987499 56827	HTSeq
15	CPSF2_1	17.4183457285 984	0.7475197630 38389	0.2287914871 37219	3.2672533947 4742	0.001085964319 172	0.0312841259 6384	HTSeq
16	FQA23_0010 887	221.014676987 154	0.7466409842 70329	0.1583036907 91707	4.71651027551 055	2.39924255438 012E-06	0.0017970326 73231	HTSeq
17	BNIP3L_1	25.581223484 402	0.7446048517 36776	0.2165357665 98762	3.4387152913 9351	0.00058448158 0603	0.025256348 300274	HTSeq
18	WASHC2D	20.261794365 5115	0.7442358832 268	0.2315962152 50113	3.2135062415 552	0.001311249480 518	0.033866408 996837	HTSeq
19	DDX60	34.843070713 7958	0.7394675697 23897	0.2107038439 73143	3.5095115294 5341	0.00044893059 5043	0.0231140313 38794	HTSeq
20	CHMP1B	27.6738148193 436	0.7318812360 85087	0.2408780166 32781	3.0383894982 0332	0.002378463164 028	0.0466241459 7435	HTSeq
21	ITSN2	47.2051613490 126	0.7180536422 17883	0.2096758894 5145	3.4245885118 0477	0.000615731237 65	0.0254987499 56827	HTSeq
22	SAMD9L	70.7908111882 568	0.7161632071 61809	0.2247762327 29754	3.1861162475 4315	0.0014419664511 42	0.0361817218 77918	HTSeq

23	ST3GAL1	68.578142181797737562	0.714517179823756	0.175876007051955	4.062618840400900	4.85252292408882E-05	0.008961566578613	HTSeq
24	DROSHA	24.628193984436401070	0.706607126110107	0.197039368696234	3.586121549137980	0.000335632564815	0.021170971002915	HTSeq
25	ADSS	41.152653224149942417	0.704368437042417	0.201123289888977	3.502172410905080	0.000461480888556	0.023114031338794	HTSeq
26	CLINT1	78.711389557929968949	0.694623362768949	0.167821555360296	4.139059260162750	3.48732869768694E-05	0.008436978151847	HTSeq
27	FQA23_0000462	35.630655914339150153	0.691592990350153	0.187795001850466	3.682701794698680	0.000230775059312	0.017881088216366	HTSeq
28	MARCH6	103.6393037296270411	0.685463583470411	0.152611935038302	4.491546374131060	7.070790040532 29E-06	0.009050611251881	SalmonCDS
29	NOS2	27.929932151134329041	0.683809468129041	0.185169322606303	3.692887452976860	0.000221722165 38	0.017881088216366	HTSeq
30	USP7	29.088189104428869595	0.681895366069595	0.211977755385921	3.216825108975020	0.001296175853856	0.033866362135049	HTSeq
31	CEP89	42.788498457591558710	0.673024151615871	0.192385038334935	3.498318566715850	0.000468201524531	0.023114031338794	HTSeq
32	SMG1	25.655912869663344363	0.666563772044363	0.218010618207525	3.057483059884080	0.002232042570346	0.045594542 32335	HTSeq
33	FQA23_0002226	210.83754012296	0.637821562977996	0.183401094009034	3.477741321142 72	0.000505657653524	0.024115789862584	HTSeq
34	HNRNPH1	503.097665772997	0.622227952740846	0.185483368784465	3.354629349350091	0.000794713890406	0.027901907995983	HTSeq
35	DDX17	185.113150609345	0.621675374695389	0.134417114051865	4.624971894988888	3.746491979691 23E-06	0.008202786802969	SalmonCDS
36	EPB41	367.820688482241	0.621663085703062	0.179297752609104	3.467210696490089	0.000525889498561	0.024115789862584	HTSeq
37	FKBP3	270.038607383876	0.619396602285104	0.193447247629068	3.201888938077777	0.001365296257 66	0.034861598760929	HTSeq
38	FOXN3	110.634506061213	0.607131159626611	0.132051732597418	4.597676589958853	4.272284793212 86E-06	0.008202786802969	SalmonCDS
39	FQA23_0013493	32.1464372138066	0.604797319728662	0.180987515527524	3.341652146369 66	0.000832813595 28	0.028048731516336	HTSeq
40	ATM	34.885758996424907617	0.601759979907617	0.179755155932763	3.347664642 9697	0.000814955633417	0.028048731516336	HTSeq
41	TMEM131L	226.531279695061	0.589913325969411	0.122746367835376	4.805953417380003	1.540157594053 5E-06	0.008202786802969	SalmonCDS
42	ZMYND8	144.638755372176	0.583818473727288	0.166984135504308	3.496251137655 08	0.000471844391839	0.023114031338794	HTSeq
43	FQA23_0013572	63.4529382564448	0.583672216623614	0.161345638440939	3.617527082 00581	0.000297431217473	0.019656704284144	HTSeq
44	UBR4	141.504734295808	0.572276545807159	0.188637847510152	3.033731318 29423	0.002415494510248	0.046789794521793	HTSeq
45	RBFA	26.787201200 0847	0.568991352998012	0.187327705330523	3.037411641775 44	0.002386193496685	0.046624145 97435	HTSeq
46	NFE2L2	402.699550926629	0.558622214182136	0.141345524156 29	3.952174768 29653	7.7444145267314 8E-05	0.045751618127152	SalmonCDS
47	NARF	40.480625813 6797	0.554921318418 2	0.182523190545263	3.040278425774	0.002363595347239	0.046624145 97435	HTSeq
48	TRANK1	86.5290684665527	0.547508647144579	0.176123520507455	3.108662860 96867	0.001879360356697	0.042229227214971	HTSeq
49	FQA23_0013833	585.3566185 65614	0.534044583524951	0.111258612124081	4.800029169 24181	1.586425223946 52E-06	0.001782348739104	HTSeq
50	UBE3A	42.107014136555 4	0.514320741305583	0.150782191586504	3.411017812474 99	0.000647208632602	0.025498749956827	HTSeq
51	FQA23_0004721	148.623858502673	0.510985867029387	0.153875670911327	3.320771009 45605	0.000897691575345	0.028815899568561	HTSeq

52	SKP1	455.705123514204	0.504118863821513	0.163326626191859	3.086568770662282	0.00202481103278	0.043674269792324	HTSeq
53	CCDC66	82.275259692636	0.496381895927346	0.165758184072669	2.994614707589953	0.002747917102967	0.049835556086759	HTSeq
54	DOT1L	88.1582022223959	0.494117613018766	0.165091436946079	2.992993592878787	0.00276255548 7538	0.049835556086759	HTSeq
55	NCOA4	103.883036508295	0.489945109134651	0.159630454209829	3.069245850109974	0.002145999107725	0.044239082523465	HTSeq
56	NUB1	158.963760250866	0.488212330282241	0.122576286840984	3.982926411497421	6.8071838321112 1E-05	0.009899913494661	HTSeq
57	FXR1	335.974219560437	0.484359354904482	0.139580877147072	3.470098231250 74	0.00052026805 2271	0.024115789862584	HTSeq
58	BCLAF1	169.90771579377	0.479445716566401	0.136996926939127	3.499682272 2851	0.000465813017476	0.023114031338794	HTSeq
59	ST3GAL1	215.159357890846	0.475601085298786	0.114401459169541	4.157299117959 27	3.22032228803 488E-05	0.027480083524564	SalmonCDS
60	STK17A	99.2339767110918	0.469826834059553	0.152332375078209	3.084221813113 21	0.00204085372 8613	0.043674269792324	HTSeq
61	BRD1	296.711184326107	0.450980888223132	0.139894927015676	3.223711522953 2	0.001265408047055	0.033866362135049	HTSeq
62	ITPK1	229.323115610619	0.450185077305 09	0.149575774544146	3.009745920936 03	0.002614663144908	0.048818248819783	HTSeq
63	CIR1	165.94791207044	0.440489218099075	0.123577497013763	3.564477584863335	0.000364581604525	0.022140942307225	HTSeq
64	MARCH6	136.870479411106	0.427828967208073	0.125594442441016	3.406432314144 78	0.000658178681574	0.025498749956827	HTSeq
65	PDE3B	342.744332686354	0.423354415124271	0.139247347318794	3.040305063442 53	0.002363386291	0.046624145 97435	HTSeq
66	FQA23_0003308	312.43641098058	0.390684155159161	0.125350259923614	3.116739888670 65	0.001828628372348	0.041585540665053	HTSeq
67	THRAP3	246.600276423727	0.387635699508195	0.128005115041105	3.028282888412 06	0.002459477347466	0.047234577775701	HTSeq
68	RPS13	1615.26295278566	-0.349208969450977	0.115586486963553	-3.021191997651 85	0.0025178165134 1	0.047542300047332	HTSeq
69	ATP5F1D	201.145420312994	-0.372686494736343	0.115789450220823	-3.218656743128 06	0.001287925587805	0.033866362135049	HTSeq
70	RPS3	1513.028529835567	-0.374742475805347	0.120856248844678	-3.100729001501 26	0.001930448772812	0.042947706856528	HTSeq
71	RPL30	2510.71463820668	-0.386693085683625	0.127813935096685	-3.025437604992 84	0.00248273624 8192	0.047277189404131	HTSeq
72	COX7B	523.363851241169	-0.407450860720324	0.132683944635957	-3.070837710155 83	0.002134591394528	0.044239082523465	HTSeq
73	NACA	1257.686440550 21	-0.411204873294368	0.127787896014466	-3.217870284426 76	0.001291462093253	0.033866362135049	HTSeq
74	CDC42	305.443777668824	-0.433580436 68615	0.104091889851411	-4.165362328468 41	3.108584670461 21E-05	0.027480083524564	SalmonCDS
75	FTH	10231.41342391 48	-0.477773829141069	0.138276129765101	-3.455215516609 38	0.000549853111687	0.024225881214905	HTSeq
76	IBA57	75.7036317002356	-0.479473586138431	0.153866634558419	-3.116163471791 46	0.001832206731571	0.041585540665053	HTSeq
77	PSMB3	232.785203901655	-0.488683298 66791	0.146835276533505	-3.328105549325 56	0.000874387314812	0.028474612991043	HTSeq
78	EEF1D	1082.44087003222	-0.489929059751027	0.149338038322058	-3.280671590813 72	0.00103560249 9873	0.030618405489668	HTSeq
79	ADRM1	101.044924446747	-0.498728202876282	0.156824956866952	-3.180158393407 97	0.001471945722987	0.036345736698374	HTSeq
80	MRPL54	386.22287881234	-0.518899334007845	0.168739029117 18	-3.075158940540 65	0.00210390402 0063	0.044181984421319	HTSeq

81	RPL7	1593.974640569957	-0.523223970549794	0.161613472103366	-3.237502194217774	0.001205810123408	0.033042138381684	HTSeq
82	NDK_COLLI	51.775308778442	-0.524752127175573	0.150122435402327	-3.495494366111121	0.000473184442183	0.023114031338794	HTSeq
83	FQA23_0003811	50.8851239504379	-0.530389356621533	0.170173676323856	-3.11675323751103	0.001828545579732	0.041585540665053	HTSeq
84	RAP1B	442.966970122936	-0.541653939646548	0.114374729185572	-4.735783363139135	2.18210582851297E-06	0.008202786802969	SalmonCDS
85	RPL9	454.184614770024	-0.546033994620701	0.174820810991291	-3.123392412633389	0.001787792130527	0.041585540665053	HTSeq
86	FQA23_0009141	121.477863785842	-0.547049108678712	0.148427275253145	-3.685637344922276	0.000228131012474	0.017881088216366	HTSeq
87	SSU72	158.256974119526	-0.547415673376262	0.179913180610551	-3.042665754218581	0.002344926343704	0.046624145957435	HTSeq
88	NMRK2_1	891.097933377794	-0.572575574426227	0.162511881720853	-3.523284379967633	0.000426233753779	0.023114031338794	HTSeq
89	MCTS1	74.33202796852	-0.573274760245681	0.180010833822646	-3.184668100646077	0.001449201143308	0.036181721877918	HTSeq
90	DNPEP	35.6520051891813	-0.593712295093969	0.180231690050542	-3.294161503603911	0.000987157617103	0.029575242208395	HTSeq
91	UQCC2	301.385446950915	-0.606206641018077	0.177797746222864	-3.409529388849597	0.000650750660736	0.025498749956827	HTSeq
92	RPLP2	400.443789943896	-0.618167464292431	0.200387838103433	-3.084855199512431	0.002036512746775	0.043674269792324	HTSeq
93	HIGD2A	66.3160552679402	-0.628736944234255	0.190036352146641	-3.308508804040544	0.000937942356394	0.029271617705796	HTSeq
94	ATP5F1E	591.408584764556	-0.637162443404394	0.194569916183537	-3.274722299838851	0.001057658875654	0.030864409007715	HTSeq
95	EIF3M	157.46167479791	-0.638071873421017	0.202727082936973	-3.147442680953344	0.001647053948864	0.038957160243126	HTSeq
96	FAM96B	79.0266782579119	-0.645598497673482	0.174861012079114	-3.692066573311588	0.000222439216803	0.017881088216366	HTSeq
97	RPL37	4608.13432500499	-0.658946568144329	0.186668812291386	-3.530030325122302	0.000415512038195	0.023114031338794	HTSeq
98	RPS21	1213.930482107732	-0.659413209021383	0.196187267098309	-3.361141723285	0.000776209861335	0.027901907995983	HTSeq
99	CPSF3	50.8317166295884	-0.659970731817112	0.202367508882968	-3.261248485293308	0.001109227891327	0.031549811035585	HTSeq
100	RPS29	5685.150840339377	-0.661717948365618	0.168902202933006	-3.917757950309977	8.937638371214 97E-05	0.011813454953012	HTSeq
101	FQA23_0009255	106.2894647455321	-0.661995085368642	0.151694561432209	-4.364000127087749	1.27705502309062E-05	0.003586928296106	HTSeq
102	CSTB	75.0077023728553	-0.666744918154435	0.165250623613117	-4.034749785364436	5.46605144585525E-05	0.036419745681228	SalmonCDS
103	TXD17_EPICO	87.645070661726	-0.682269622036207	0.215139220541229	-3.171293548056605	0.001517616993612	0.037066145485281	HTSeq
104	TRAPPC6B	54.6629332643803	-0.695140478047752	0.206927716637595	-3.359339625174433	0.000781289845923	0.027901907995983	HTSeq
105	COX7C	1577.31969761161	-0.705775618370135	0.193189021183167	-3.653290513341 2	0.000258901024436	0.018927986099251	HTSeq
106	FQA23_0007151	36.418853550 6183	-0.721808186978042	0.241252941417561	-2.991914555473368	0.002772338456095	0.049835556086759	HTSeq
107	URM1	32.477905623 4986	-0.735835191784082	0.218816950703592	-3.362788803 23508	0.000771593698208	0.027901907995983	HTSeq
108	IER3IP1	40.667025148 6324	-0.738791908242249	0.223935138446485	-3.299133460555247	0.000969837904759	0.029575242208395	HTSeq
109	EIF3H	129.250194609103	-0.751232264444455	0.167512961104485	-4.484621724 14158	7.30433993688873E-06	0.003003711715746	HTSeq

110	GPX1	1690.957715186261	-0.75308388261185	0.243116061774798	-3.09763113598576	0.001950740680061	0.042973669687219	HTSeq
111	ARHGDIB	77.2912266497695	-0.758865824235935	0.188524647175649	-4.025287067790 66	5.6905852626919E-05	0.036419745681228	SalmonCDS
112	GAPDH	23.3563396533257	-0.759518047189944	0.240840138541143	-3.15361904286646	0.001612594433451	0.038957160243126	HTSeq
113	RPS15	2839.73147464179	-0.761748924206153	0.209559564035747	-3.634999565 45153	0.000277981104261	0.018927986099251	HTSeq
114	FQA23_0005809	32.7370946025632	-0.781960575142752	0.226035170174031	-3.459464182 23633	0.000541251258122	0.024225881214905	HTSeq
115	MTREX	30.5342779054336	-0.790621315436186	0.237265663884452	-3.332219683 59059	0.000861562166656	0.028469561595224	HTSeq
116	ASNSD1	50.2763145244215	-0.793530019682884	0.206966073049961	-3.834106759 57179	0.000126021402755	0.014153820817216	HTSeq
117	CYT_CHICK	87.5213519182504	-0.793893268078682	0.245121570975983	-3.238773580 46333	0.001200448301028	0.033042138381684	HTSeq
118	COMMD5	17.4191371126908	-0.809326474159626	0.236259391020104	-3.425584357 36744	0.000613478395185	0.025498749956827	HTSeq
119	MRPS22	79.1390053683258	-0.827962302450415	0.243421362474909	-3.401354318 42289	0.000670528570721	0.025536910142538	HTSeq
120	FQA23_0012943	114.110795465009	-0.835570505094911	0.210226810332496	-3.974614388 01914	7.0493376018945 9E-05	0.009899913494661	HTSeq
121	MBD4	22.1811810174559	-0.854287740317459	0.262812594209659	-3.250558607 68967	0.001151785257695	0.032350768425504	HTSeq
122	PRDX4	54.3080290822148	-0.863951496551648	0.241100181630772	-3.583371404 81598	0.000339187786429	0.021170971002915	HTSeq
123	OSGEPL1	15.2266240568057	-0.876625471442037	0.284973881926363	-3.076160753 80745	0.002096847626056	0.044181984421319	HTSeq
124	CCDC173	557.2424550600073	-0.919883018514477	0.203865524958638	-4.512204889 47854	6.4157166467792 8E-06	0.009050611251881	SalmonCDS
125	FQA23_0014950	44.7168837712483	-0.942190247258839	0.299126987518821	-3.149800207 17642	0.001633821522628	0.038957160243126	HTSeq
126	TPM2	13.6828520069546	-0.955828196980428	0.317751175395407	-3.008102789 20606	0.002628842059276	0.048818248819783	HTSeq
127	MIOS	12.9464730202 69	-0.976702693148323	0.268451764484412	-3.638280027 78889	0.000274464909849	0.018927986099251	HTSeq
128	NOX4	22.0118528697825	-0.979434245491621	0.293201852633295	-3.340477683 53153	0.000836344019401	0.028048731516336	HTSeq
129	MRPS27	15.9305478038829	-0.980620791926069	0.295697161456366	-3.316300998 95556	0.000912175254129	0.028868419662373	HTSeq
130	EIF2AK1	15.2743491754692	-0.983380849931737	0.280799970732972	-3.502068918 89916	0.000461660183777	0.023114031338794	HTSeq
131	RPL22L1	419.177537513913	-1.0523213250 2717	0.261137871336026	-4.029753783 48347	5.5835305785 7514E-05	0.008961566578613	HTSeq
132	NDFIP1	32.6709380529 59	-1.1697698526698	0.289840996409569	-4.035901984 74484	5.4392921993035 7E-05	0.008961566578613	HTSeq
133	SUB1_0	21.5153807008816	-1.3655817549 6401	0.309717347947156	-4.409122588 75795	1.0379026451894 E-05	0.003331667491058	HTSeq

Other publications during the Ph.D.

Other publications during the Ph.D. in which I participated and that contributed to the implementation in my project

1. Bardon et al. *Accepted in Methods in Ecology and Evolution with minor revisions*

This publication is currently available in Bioarchive (link in the QRcode), but has been accepted with minor revisions in *Methods in Ecology and Evolution*.

2. Pirri et al. 2022. *Heredity*

the geneticssociety

www.nature.com/hdy

ARTICLE

Check for updates

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

Federica Pirri[1,2], Lino Ometto[3], Silvia Fuselli[4], Flávia A. N. Fernandes[1,5], Lorena Ancona[1], Nunzio Perta[1], Daniele Di Marino[1], Céline Le Bohec[5,6], Lorenzo Zane[2] and Emiliano Trucchi[1]✉

3. Vianna et al. 2020. *Proceedings of the National Academy of Sciences (PNAS)*

RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Genome-wide analyses reveal drivers of penguin diversification

Juliana A. Vianna, Flávia A. N. Fernandes, María José Frugone, +19, and Rauri C. K. Bowie Authors Info & Affiliations

Edited by Scott V. Edwards, Harvard University, Cambridge, MA, and approved July 8, 2020 (received for review April 8, 2020)

August 17, 2020 | 117 (36) 22303-22310 | https://doi.org/10.1073/pnas.2006659117