Prey differences drive local genetic adaptation in Antarctic fur seals

Alison C Cleary[1,8]*, Marthán Bester[2], Jaume Forcada[3], Michael Goebel[4], Simon D Goldsworthy[5], Christophe Guinet[6], Joseph I Hoffman[7], Kit M Kovacs[1], Christian Lydersen[1], Andrew D Lowther[1]

1 – Norwegian Polar Institute, Fram Centre, 9296 Tromsø, Norway
2 – Mammal Research Institute, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa
3 - British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 OET, UK
4 - Antarctic Ecosystem Research Division, Southwest Fisheries Science Center, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA
5 - South Australian Research and Development Institute, 2 Hamra Avenue, West Beach, South Australia 5024, Australia
6 - Centre d'Etudes Biologiques de Chize´ (CEBC), CNRS and Universite´ de La Rochelle – UMR 7372, 79360 Villiers en Bois, France
7 - Department of Animal Behaviour, Bielefeld University, 33501 Bielefeld, Germany
8 – Department of Natural Sciences, University of Agder, 4630 Kristiansand, Norway
ORCID: ACC - 0000-0001-9880-5816, MB - 0000-0002-2265-764X, JF - 0000-0002-2115-0150, JIH – 0000-0001-5895-8949
* alison_cleary@my.uri.edu, +47 96 88 53 94

Abstract

Antarctic fur seal, *Arctocephalus gazella*, colonies are found on sub-Antarctic islands around the continent. These islands experience a range of conditions in terms of physical and biological habitat, creating a natural laboratory to investigate local genetic adaptation. One striking habitat difference is in the availability of *Euphausia superba* krill as prey, which has led to *A. gazella* exhibiting a range of diets. *A. gazella* in some colonies consuming exclusively krill, while their conspecifics in other colonies feed mainly on fish and consume few to no krill. To investigate potential adaptations to these different prey fields, reduced representation genome sequencing was conducted on *A. gazella* from all eight of the major colonies. Twenty seven genomic regions exhibiting signatures of natural selection were identified. Two of these genomic regions were clearly associated with seals living in krill-dominated areas or those in fish-dominated areas. Twenty-two additional genomic regions under selection showed a pattern consistent with prey differences as the driver of selection, after historical migrations from krill-dominated habitats where lineages evolved to present krill-poor habitat areas were taken into account. Only one of the genomic regions identified appeared to be explained by any other environmental variable analysed (depth). Genomic regions under prey-driven selection included genes associated with regulation of gene expression, skeletal development, and lipid metabolism. Adaptation to local prey has implications for spatial management of this species, and for the potential impacts of climate or harvest driven reductions in krill abundance on these seals.

Key words

Arctocephalus gazella, ddRAD sequencing, diet, *Euphausia superba*, natural selection

Running Head

Seal genetic adaptation to prey type

1. Introduction

The Antarctic fur seal, *Arctocephalus gazella*, is an abundant pinniped which breeds in colonies on sub-Antarctic islands surrounding the Antarctic continent. The large circumpolar biomass of this species makes it a potentially important component of Southern Ocean ecosystems, and its distribution on isolated islands makes it an ideal case study to investigate local adaptation to variations in the physical and biological habitat. One particularly interesting habitat variation is in the availability of the common *A. gazella* prey item, Antarctic krill, *Euphausia superba*.

It has been recognized since the 1800s that differences in available prey can drive natural selection and local genetic adaptation, such as the classic case of Darwin's finches, in which selection is driven by variation in the size and type of available seeds and other food sources (Grant & Grant 2003). Starvation is the main cause of mortality in *A. gazella* pups, (Reid & Forcada, 2005), and interannual variations in available prey biomass have been correlated with variations in *A. gazella* recruitment (Forcada & Hoffman 2014). Thus, genetic adaptations which increase the probability that lactating females are able to consume sufficient prey, and efficiently use that prey to fuel metabolism, will be subject to strong positive selection.

A. gazella is one of the many megafauna species in the Southern Ocean which typically relies on *E. superba* (hereafter also "krill") as prey (Quetin & Ross 1991). However, not all *A. gazella* feed exclusively, or even primarily, on krill. E. superba are generally restricted to waters south of the Polar Front (Siegel 2005), whereas *A. gazella* colonies are located on islands located south of, on, and north of this front (Wynen et al. 2000). Thus, at least during the breeding season when adult female seals are restricted in their travels by the need to return to the colony at least every four to five days to provision their offspring (Boyd et al. 1991), seals at certain colonies feed almost exclusively on krill, while their conspecifics at other colonies consume no krill at all. *A. gazella* along the West Antarctic Peninsula and at South Georgia feed predominantly on krill (Casaux et al. 2003). By contrast, *A. gazella* at Marion and Macquarie Islands consume mainly myctophid fish (90% of the diet), those at Iles Kerguelen consume mainly myctophids and icefish (87% of the diet), and those at Heard Island feed on various fish groups, mainly myctophids and icefish (*Champsocephalus* spp.) (present in over 95% of scats) (Cherel et al. 1997, Green et al. 1989, Goldsworthy et al. 1997, 2010, Klages & Bester 1998, Robinson et al. 2002,

Jeanniard du Dot et al. 2017). The South Shetland Islands, South Georgia, and Bouvetøya are all located in krill-dominated areas, while Marion Island, Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island are located in krill-poor (or krill-absent) areas. This creates an ideal natural laboratory of replicated island systems across a sharp prey gradient, that lends itself to investigating the impacts of prey differences on natural selection and genetic adaptation in a megafaunal predator.

Krill and fish differ in many aspects of their biology that are potentially relevant to their predators, such as size, behaviour, and nutritional composition. Krill are smaller than most adult fish; *E. superba* adults are typically 30 to 60 mm in length, while the fish consumed by *A. gazella* typically have lengths ranging from 60 to 390 mm (Casaux et al. 1998, Makhado et al. 2008). Krill are typically found in dense schools, whereas fish are found both in less dense schools (myctophids) and in relatively dispersed distributions (icefish) (Frolkina 2002). In terms of nutritional composition, krill are relatively low in lipid (around 1.5%), as compared to fish (4-8%), and krill contain particularly high levels of fluoride (Soevik & Braekkan 1979, Tou et al. 2007). These differences between krill and fish likely make different genetic adaptations advantageous to seals feeding under different prey regimes.

In addition to their interest as a case study of local genetic adaptation to prey differences, regional-scale genomic adaptation to diet in *A. gazella* has important implications for management of the species, and for the use of this species as a proxy for krill abundance. Although unregulated commercial exploitation in the 1830s-1930s decimated most *A. gazella* colonies, populations have subsequently recovered across much of their historical range (Wynen et al. 2000). Currently, population sizes range from approximately 150 individuals at Macquarie Island to over 1 million individuals at South Georgia, where it is estimated they have significantly exceeded their pre-harvest population size (Boyd 1993, Hodgson et al. 1998). The strong predator-prey relationship between *A. gazella* and *E. superba* has been used to justify monitoring populations of *A. gazella* as an ecosystem indicator for krill abundance, including by the international CCAMLR Ecosystem Monitoring Program (Reid et al. 2005). If diet is a strong driver of natural selection, seals that have adapted genetically to a particular prey regime will have reduced fitness under an alternative prey regime. As such, the diet of various seal stocks should be taken into account when considering management units for this species. Alternatively, a lack of prey-driven genetic adaptation would suggest a relatively low threshold to prey-switching, indicating *A. gazella*

are potentially a poor proxy for krill abundance, as they could relatively easily switch to fish prey in low-krill years.

To explore the adaptations of *A. gazella* to different prey environments, this study investigated genomic signatures of natural selection in *A. gazella* across their circumpolar range. Over 60,000 single nucleotide polymorphisms (SNPs) from 104 individual seals, across eight colonies, were analysed to determine overall population structure, and detect genomic regions under selection. The genomic regions showing signatures of selection were investigated further by comparing their sequences with gene databases to identify their biological function, and comparing the distribution of genotypes across the islands with that of environmental variables to identify the drivers of selection.

2. Materials and Methods

Arcotcephalus gazella samples were collected from eight islands, encompassing the entire circumpolar breeding range of this species (Table 1, Figure 1): Livingston Island in the South Shetlands (13 individuals), Bird Island on South Georgia (13), Bouvetøya (13), Marion Island (8), Iles Crozet (8), Iles Kerguelen (13), Heard Island (13) and Macquarie Island (13). Additionally, ten samples were collected from the congeneric species *Arctocephalus tropicalis*, in order to detect any hybridization of *A. tropicalis* into the study individuals of *A. gazella* (Marion Island – 5, Iles Crozet – 5). Either blood or skin samples were collected from individuals in the breeding areas, under ethics permits from the relevant national authorities. Samples were stored in sodium chloride saturated dimethyl sulphide, or ethanol, and/or frozen until processing.

Genomic DNA was extracted from all samples using a chloroform-isoamyl alcohol protocol, adapted from Sambrook et al. (1989). Double digest restriction-site associated DNA (ddRAD) library preparation and sequencing was performed by a commercial facility (IGA technology). In brief, genomic DNA from each individual was doubly digested with the enzymes SphI and EcoRI. Individual identification tags (dual, variable length tags) and sequencing adaptors were attached to these DNA fragments with ligation, and paired-end sequencing was conducted on pooled samples on an Illumina HiSeq 2500.

Sequence data was analysed to calculate the likelihood of each base occurring at each position in each individual. This approach allows for uncertainty in the sequence data to be taken into account in all stages of analyses, and maximizes the information which can be used, by allowing low read depth data to be included without introducing errors. Sequence

data was quality controlled and de-multiplexed in Stacks using default parameters of process_radtags (Catchen et al. 2013). DNA sequences were mapped against a reference genome for *A. gazella* (NCBI accession #SRP148937) with Bowtie2 (Langmead and Salzberg 2012). Only sequences that mapped to a single unique location in the genome were retained for further analyses. Subsequently, mapped sequences were placed into contig order, and re-formatted for further analysis with SamTools (Li et al. 2009). All base positions within each read were included in analyses. The genotype likelihoods were calculated in Analysis of Next Generation Sequencing Data (ANGSD), using the Genome Analysis ToolKit (GATK) model. Quality control filtering restricted analysis to SNP positions that had a minimum p-value of being genuinely variable of 10^{-6}, and that were present in a minimum of 47 individuals (half of the *A. gazella* samples in this study) at a minimum sequencing depth in each individual of 2 (Korneliussen et al. 2014). These relatively lenient filtering thresholds allow for retaining the maximum amount of data, while the use of genotype likelihoods (rather than called SNPs) down-weights base positions with lower certainty, such as those with low read depth. Likelihoods were calculated for all samples (*A. gazella* and *A. tropicalis* together), and for *A. gazella* alone.

In order to account for genetic connectivity in the analysis of selection, population structure was analysed to set a backdrop. Initial investigations of overall population structure included both *A. gazella* and *A. tropicalis*. *A. gazella* and *A. tropicalis* are known to form hybrids under certain conditions (Wynen et al. 2000, Lancaster et al. 2006), so both species were analysed together in order to detect any hybrid individuals. NGSadmix, an analytical approach similar to the more commonly known STRUCTURE, but which takes SNP calling uncertainty into account, was run for from 2 to 9 distinct groupings (K), with a minimum minor allele frequency of 0.05, and all other parameters at program defaults (Skotte et al. 2013). Additional detail on the patterns of similarity across individuals were explored with a Principle Components Analysis (PCA), calculated in PCangsd (Meisner & Albrechtsen 2018).

A two-step process was used to detect regions of the genome under selection. Selection analyses were restricted to *A. gazella* individuals. The detection of genomic regions under selection did not include any information on the sampling location of each individual seal, and thus identified genomic regions under selection as driven by any factor, within or between islands. In the first stage of this analysis, the probability that each single base position was under selection was calculated in PCangsd, using an extended model of PCadapt

(Meisner & Albrechtsen 2018). This analysis is based on the extent to which the pattern of allele frequency across seals observed for each SNP position alone deviates from the overall pattern derived from all SNP positions combined. In essence, a PCA is generated for each base position alone, and then subtracted from the PCA generated from all base positions, and this residual provides an indication of the likelihood that each base position is experiencing selection. Base positions which have large distances to the overall PCA, such as those which are either much more variable between individual seals (as would be the result of directional selection operating differently across colonies), or much less variable (as could result from stabilizing selection) will have a higher probability of being under selection. This approach therefore takes into account any differences in genotype frequency driven by overall genetic structure, such that the overall population structure, and the presence of admixed individuals, do not impact the detection of SNPs under selection. In the second stage of the section analysis, Fariello et al.'s (2017) Local Scores approach, as implemented in R, was used to take into account the combined effects of selection on SNP positions which are physically adjacent in the genome. This approach allows for more sensitive detection of selection, particularly for genomic regions characterized by multiple SNPs each with small adaptive advantages. The Local Scores approach takes the physical location along genome scaffolds of each base position, and the p-values that each individual base position is under selection (as calculated with PCangsd in the first stage), and identifies and delimits regions of the genome showing significant selection.

Fst measures were calculated for each of the genomic regions under selection, for every possible pair of islands using realSFS in ANGSD with a minimum of four individuals for each SNP for each island (Nielsen et al. 2012). These Fst values were used both to compare with environmental distances (described later) and to test for the type of selection influencing each of the identified genomic regions – directional selection or stabilizing selection. Fst deviations were calculated as the Fst for each gene region minus the overall genome-wide Fst, and the mean was taken of all possible pairs of islands to calculate an overall Fst deviation for each genomic region. Positive deviations indicate directional selection (the region of interest is more differentiated across islands than the genome as a whole), while negative deviations indicate stabilizing selection (the region of interest is more homogenous across islands than the genome as a whole).

Gene regions identified as being under selection were further explored using clustering analysis, to compare the distributions of alleles with those of potential

environmental drivers. The average likelihood of the major allele (ranging from 0 for minor allele homozygotes, through 2 for major allele homozygotes) at each SNP position was calculated for each island, from the overall genotype likelihoods described above. These values were then used to cluster the islands, for each gene region under selection, with Ward clustering as implemented in MatLab. Dendrograms were used to visualize the clustering patterns for each gene region under selection, and were cut using an automatic threshold (70% of the total dendrogram length), to separate the islands into a natural number of groups (one to eight groups). It is conceivable that overall population structure could influence clustering patterns for genomic regions only weakly selected for by environmental drivers which differ between islands, potentially giving a false indication of the environmental drivers of selection. In such a situation, most or all of the genome would be expected to cluster similarly. In order to test for this possibility, a control set of random gene regions of the same length as the gene regions under selection was generated using a random number generator to select a contig and start position. These random gene regions were clustered in the same manner as the gene regions identified as being under selection. These random gene regions provide a control to detect any potential artefacts or biases of the clustering approach.

Repetitive gene regions, such as multicopy genes and gene families, can show false signals of selection, due to the difficulties of aligning sequence reads to the correct version of the gene in the genome, leading to non-homologous mappings. Such non-homologous mappings can sometimes be detected as Hardy-Weinberg equilibrium outliers (Hosking et al. 2004). However, genes under selection can also display deviations from Hardy-Weinberg equilibrium. As a compromise, after the selection analyses described above, Hardy-Weinberg equilibrium was analysed for each SNP position in each island separately. SNP positions that were significantly ($p < 0.01$) out of equilibrium in half or more of the islands indicated potentially unreliable data. None of the genomic regions identified as being under selection contained such unreliable SNP positions, so all regions were retained for further analyses.

The biological functions of the gene regions identified as being under selection were investigated by searching for homologous sequences in annotated databases. The complete sequence for each of the gene regions was retrieved from the reference *A. gazella* genome. These sequences were BLAST searched against the KEGG and NCBI GenBank databases (Altschul et al. 1990; Kanehisa et al. 2016). While GenBank is a larger database, KEGG is curated, so using these two different databases in parallel increases the probability that

functions identified are true, and not simply a result of random chance when searching very large databases. The highest scoring match with functional annotation was recorded, unless the top five matches contained results from a pinniped, in which case the pinniped match was recorded as these were considered more accurate annotations than matches to more distantly related model organisms which dominate the databases such as human and mouse. Matches without meaningful functional annotations, such as database sequences identified as "uncharacterized" or "hypothetical" or those with only positional notations, such as "chromosome", "BAC", or "contig" were ignored.

Data on the physical and biological habitat at each island were retrieved from Quantarctica (Matsuoka et al. 2018). The physical habitat was characterized by latitude (a proxy for light regime), longitude, sea surface temperature during the summer breeding season calculated from satellite observations and interpolated across the Southern Ocean (Locarnini et al. 2013), proportion of the year with sea ice (Spreen et al. 2008), mean ocean depth (Amante and Eakins 2009), and mean land elevation (Amante and Eakins 2009) . The biological habitat was characterized by chlorophyll *a* averaged over the austral summer seasons (days 355 to 80) (a proxy for primary production) (Johnson et al. 2017), and krill abundance calculated from standardized KrillBase data using all net types except CPR and all seasons (Atkinson et al. 2017). All variables were calculated as arithmetic means (initial explorations with minimum/maximum/median values for ice and temperature provided similar results) of all available data within 160 km of the island, as this is the maximum foraging distance for *A. gazella* during the breeding season (Guinet et al. 2001, Staniland et al. 2004).

Two approaches were used to compare the environmental variability with the distribution of genes – a linear distance correlation approach and a clustering approach. In the distance correlation approach, the correlation was calculated between the environmental distance (i.e. the absolute difference in the values of each environmental parameter between each pair of islands), and the Fst of each of the genomic regions identified as being under selection. Any correlations above 0.5 were further investigated with scatter plots to differentiate true correlations from correlations driven by many invariant points. The clustering approach is more appropriate for data which contains many zeros, such as krill and sea ice, which are in essence presence/absence data. In this approach all environmental data were clustered using the same calculations applied to the genomic sequence data, and the clustering patterns were compared between genomic regions and environmental variables

3. Results

Close to 0.3 billion sequences of 200 bp in length were obtained. After quality control, genomic sequence data were analyzed from a mean of 22,423,483 base positions from each of the 104 seals. Of these base positions, 76,816 SNPs were identified within *Arctocephalus gazella*, which were sufficiently variable, and present in a sufficient number of individuals and colonies, for comparative analyses. Most SNPs were in Hardy-Weinberg Equilibrium (HWE) within colonies, only 6.6% of SNPs were out of HWE at more than one colony, and only 1% were out of HWE in all eight colonies. None of the SNP positions which deviated significantly from HWE for more than four colonies fell within the genomic regions exhibiting signatures of selection.

3.1 Population structure

Four population groups within *A. gazella* provided the most plausible clustering of individuals, which is to say four population groups most clearly clustered individuals by geographic location of sampling (Figure 2), while greater numbers of population groups did not provide any additional geographic resolution. There was very little indication of hybridization – with only minor contributions from *A. tropicalis* to two individuals from the *A. gazella* population at Iles Crozet (2% and 14% *A. tropicalis*-type), and to a single individual from the *A. gazella* population at Bouvetøya (2% *A. tropicalis*-type). There were no indications of hybridized individuals at any of the other islands. South Georgia and the South Shetland Islands were each composed of a single population group, although each also contained a single individual of ancestry from the other, and a single individual of mixed ancestry between the two islands. Bouvetøya was composed of another population group, and was quite homogenous, with little indication of immigration from other colonies. Similarly, Iles Kerguelen was a relatively homogenous population. Marion Island, Iles Crozet, and Heard Island showed a mixing gradient between the Bouvetøya-type and Iles Kerguelen-type. Macquarie Island showed a mixture between the Iles Kerguelen-type and South Georgia-type. When the analysis was restricted to a smaller number of population groups, the overall east vs west patterns were similar. With three groups, the South Shetland Islands and South Georgia populations merged into one, and with only two groups, South Shetland Islands, South Georgia and Bouvetøya all merged into a single western group. These broad clustering patterns are also reflected in the PCA (Figure 3).

3.2 Genomic regions under selection

A total of 37 regions of the *A. gazella* genome were identified as being under selection (table 2). Selective gene regions were distributed across 34 contigs, and ranged in length from 2 base pairs to 585,781 base pairs. Most of the selective gene regions showed positive mean Fst deviations (table 2), indicating the identified selection was mainly directional.

The genomic regions identified as being under selection showed homologies with a variety of annotated genes from other organisms (table 2). Ten of the gene regions had highest annotated matches in the NCBI database to MHC genes from domesticated dogs, but to KCNQ1 genes from humans in the KEGG database, and had sequence similarity with a mink retrotransposon of 80-89%. (Anistoroaei et al. 2011). These ten regions were thus identified as likely retrotransposons (small repetitive gene fragments which can be present within other genes, such as MHC and KCNQ1), and were removed from all analyses, as the repetitive nature of retrotransposons may cause false signals of selection. The remaining 27 genomic regions showed homology with a wide range of annotated genes, from a variety of different vertebrate groups. Seven of these genomic regions identified the same functional annotation when compared with both KEGG and NCBI, and only these gene regions were considered further herein in terms of biological function. Two of these genomic regions were associated with development (#33 – cell-type differentiation, #37 – skeletal development), two gene regions were associated with regulatory processes (#16 – mRNA regulation, #28 – protein degradation), one gene region was associated with the mitochondrial genome (#12), one was associated with metabolism (#15), and one gene region was associated with an extracellular kinase (#34) (Table 2).

Selective gene regions showed seven different clustering patterns, out of a possible over 300 ways in which eight items (colonies) could cluster (Figure 4). Eleven gene regions clustered the South Shetland Islands, South Georgia, Bouvetøya, Marion Island and Iles Crozet together, with Iles Kerguelen, Heard Island, and Macquarie Island as a second group. Five gene regions clustered the South Shetland Islands, South Georgia, Bouvetøya, Marion Island, Iles Crozet and Macquarie Island together, with Iles Kerguelen and Heard Island as a second group. Six gene regions clustered the South Shetland Islands, South Georgia, Bouvetøya, and Marion Island together, with Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island as a second group. Two gene regions clustered the South Shetland Islands, South Georgia, and Bouvetøya together, with Marion Island, Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island as a second group. The remaining three selective gene

regions each showed unique clustering patterns. The 27 matched random control genomic regions clustered in 25 different ways, including 18 regions clustering into three groups, and four regions clustering into four groups. Only two of the random control regions matched the top four clustering patterns observed in the selective gene regions, confirming that clustering patterns observed for the selective genomic regions are unlikely to be an artefact of the analytical method or overall population structure.

3.3 Environmental data

Each environmental variable clustered into different groupings. Only longitude clustered islands in the same general pattern as krill abundance. Krill abundance as inferred from Krill Base reflected the same overall pattern which has been observed across many studies, with highest abundances in the West Antarctic Peninsula and Scotia Sea area, moderate abundances in the downstream areas around the prime meridian, and zero or near zero abundances north of 60 degrees around the rest of the continent (Siegel 2005).

In the linear correlational approach, only one genomic region under selection showed a meaningful correlation with an environmental factor (with a correlation coefficient above 0.5). Selective region 27 had a correlational coefficient of 0.615 with mean depth.

4. Discussion

4.1 Population structure and historical processes

Arctocephalus gazella individuals showed clear clustering by location. Four of the islands were each composed of a single independent population (the South Shetlands, South Georgia, Bouvetøya and Iles Kerguelen), while the other four islands were composed of seals showing admixed ancestry (Marion Island, Iles Crozet, Heard Island and Macquarie Island). This is largely consistent with previous results from mitochondrial amplicon sequencing, which found a western population cluster lumping together seals from South Shetland Islands, South Georgia, Bouvetøya, Marion Island, and Heard Island, and an eastern cluster lumping together those from Iles Kerguelen and Macquarie Island (Wynen et al. 2000). There are two plausible explanations for the mixed-ancestry islands; firstly, there could be ongoing migration between specific islands, or secondly mixed ancestry could result from complete (or near complete) extirpation of the presently admixed islands during harvest, followed by re-colonization by seals from multiple other islands.

The observed geographic pattern of mixed ancestry islands argues against significant ongoing migration as the cause of this mixed ancestry. In cases where ongoing migration drives mixed ancestry, a pattern of isolation by distance is frequently observed, which is to say there tends to be the most migration between the most physically proximate habitats. This is not the case in the current study; there is relatively little indication of mixing between South Georgia and the South Shetland Islands, the two most physically proximate colonies, while by contrast there is mixing between much more geographically distant colonies. *A. gazella* are additionally generally thought to have very high fidelity to breeding sites – consistently returning to the same beaches, even the same areas on a beach, to breed year after year (Lunn & Boyd 1991, Boyd et al. 1998). Such high site fidelity also argues against significant ongoing migration as an explanation for admixed ancestry islands. Both of these factors suggest that post-harvest founder events are a more parsimonious explanation for the observed genetic structure than ongoing migration.

Commercial harvest of *A. gazella* was extensive, and while some colonies were relatively little exploited, such as Bouvetøya (Hofmeyr et al. 2005), it has been suggested that colonies on some islands were completely extirpated (Bonner 1968, Hucke-Gaete et al. 2004, Lancaster et al. 2006, Goldsworthy et al. 2009). Unfortunately, systematic surveys were not conducted during the period immediately following the cessation of harvest, so historical records alone cannot be used to determine which colonies were extirpated or nearly extirpated, and which had more sizeable populations persisting. The extirpation, or near extirpation, of certain colonies would have resulted in empty habitat, where a vagrant seal or two would find no competition for beach space needed for breeding, and likely reduced competition for prey in the nearshore foraging areas. A completely, or nearly completely, extirpated colony would be strongly influenced by the genetic signature of a few founders. This is unlike the case with migrants to a large and established colony, where the genetic signature of rare migrants would be diluted among that of the many long-term residents. The observed population structure is most consistent with populations persisting through the period of commercial harvesting in four areas: the South Shetland Islands, South Georgia, Bouvetøya, and Iles Kerguelen. The remaining colonies would then have been re-populated following the cessation of harvest, with founder individuals from Bouvetøya going east to Marion Island and Iles Crozet, founders from Iles Kerguelen going west to Iles Crozet, and east to Heard Island, and Macquarie Island, and finally founders from South Georgia going west to Macquarie Island.

This is broadly similar to, though slightly more complex than, previous mitochondrial DNA amplicon sequencing results, which were interpreted to suggest Macquarie Island was repopulated by founders from Iles Kerguelen, and that the South Shetlands, Marion Island, and Heard Island were repopulated by founders from South Georgia or Bouvetøya (Wynen et al. 2000). The persistence of remnant colonies at the South Shetlands, South Georgia, Bouvetøya, and Iles Kerguelen in the face of fairly intense harvest pressure across the Southern Ocean is at first glance somewhat puzzling, particularly as South Georgia had the most intense human activity during this period of all the sub-Antarctic islands due to the whaling and elephant sealing station of Grytviken. The South Shetlands, South Georgia, and Iles Kerguelen are the most geographically complex of the sub-Antarctic islands, with many small coves and rocky outcroppings. Bouvetøya is one of the most geographically simple of the sub-Antarctic islands, being nearly circular in form, but it is also the most remote of these island groups, and is notably lacking in suitable landing sites. It seems the most parsimonious explanation then for these four surviving colonies is two-fold; the South Shetlands, South Georgia, and Iles Kerguelen populations likely survived due to the presence of inaccessible areas of coastline which provided refuges for seals, while the Bouvetøya population likely survived due to the combination of extreme remoteness (and concomitantly higher costs to harvest) and lack of landing sites, of this island. Future work with more complex modelling has the potential to shed light on the details of this recovery process, although the recency of this recovery makes such modelling mathematically challenging.

Two individuals were observed which appear to be non-admixed seals present in the "wrong" colony – one South-Georgia-type seal found in the South Shetlands samples, and one South Shetlands-type seal found in the South Georgia samples. *A. gazella* are strong swimmers, which are able to transit long distances even against the prevailing oceanic currents. Tracking studies have documented individuals covering distances in excess of 2000 km in a single winter (Staniland et al. 2012). It is thus not surprising that a few individuals were found away from what would be their likely natal ground. These individuals may represent recent migrants, or vagrant individuals who may not necessarily interbreed with the population where they were sampled.

Post-harvest founder events have occurred very recently in terms of evolutionally time. Commercial exploitation of *A. gazella* declined in the beginning of the 1900s, ceasing nearly completely by 1920; founder events occurring after this time have thus taken place within the past 100 years (Bonner 1968). By contrast, *A. gazella* are thought to have

colonized the Southern Ocean in the early Pliocene, around 4-5 million years ago (Yonezawa et al. 2009). Thus, *A. gazella* had hundreds of thousands of generations to adapt to their local environmental conditions prior to harvest, but have only had 10-20 generations to adapt to new conditions since post-harvest founder events. It would therefore be expected that genomic signatures of selection to particular types of prey may be found at "expatriate" colonies; colonies which were founded by seals adapted to a particular prey, but which are located in an area with another prey type currently. Specifically, legacy genomic adaptations to a krill-dominated diet could be expected at Marion Island, Iles Crozet, and Macquarie Island (Figure 5).

4.2 Patterns of Genomic Regions under selection

The 27 genomic regions identified as being under selection clustered into seven patterns, all of which divided the islands into two groups. Four of the patterns contained 24 of the genomic regions. These four patterns all clustered the islands where krill is abundant together (South Shetland Islands, South Georgia and Bouvetøya), with various mixed-ancestry islands also clustering with the krill-islands (Marion Island, Iles Crozet and Macquarie Island) (Figure 4). These clustering patterns are thus consistent with a prey-driven selective pressure. The presence of four different patterns may reflect the different degrees of admixture between krill and fish feeding ancestors, or may reflect the ongoing process of adapting to new habitats. These patterns are similar to the overall genetic structure (Figures 2,3). However, they are unlikely to be artefacts of this structure. Genomic regions under selection were initially identified based on the difference between individual SNPs and the genome as a whole, which means that overall population structure is already accounted for. Random control regions of the genome subjected to the same analysis did not show any consistent clustering patterns, confirming overall population structure cannot explain the patterns observed in the selective gene regions. Additionally, the Fst deviations for most of the identified genomic regions were positive, showing greater differentiation between islands than the genome as a whole, again indicating these genomic regions have adaptive value, rather than simply reflecting genome-wide population structure.

Linear distance correlation analysis failed to identify this link between selective genomic regions and available prey type, likely due to krill being absent at five of the eight islands, making correlation analysis challenging. The one genomic region under selection which showed a clear association with the environment in the distance correlation analysis

was associated with depth. Although depth is not correlated with krill abundance, it does influence the availability of different prey items, and the effort required to capture them. For example, demersal fish such as icefish are less accessible in deeper waters. It is notable that Iles Kerguelen, the fish-dominated island which persisted through harvest, is on a plateau, and thus the adjacent foraging areas are much shallower at this colony than at other colonies. Further research will be necessary to definitively disentangle the influence of depth and prey field in the adaptations identified here.

Of the other environmental factors explored (temperature, bathymetry, land topography, sea ice, light regime, location), only longitude clustered together islands in patterns matching the clustering patterns of the gene regions under selection, even when founder events were taken into account. Longitude is correlated with krill abundance, and is not in and of itself likely to drive natural selection. Krill abundance is also typically influenced by depth over small spatial scales, however when considering depth over the entirely of the likely foraging range (160 km radius circle around the colony), it was not correlated with krill abundance. Other environmental factors known to correlate with krill abundance, such as water mass and position relative to the fronts were not explored, as these factors are strongly linked to variables in the analysis (specifically temperature and chlorophyll *a*) which are more plausible proximate drivers of selection. It thus appears that differences in the available prey field were the likely driver behind most of the observed selection. It is perhaps not surprising that no evidence was observed for selection related to temperature; the three warmest islands, which cluster out from the other five, Marion Island, Macquarie Island, and Iles Crozet, are all among the islands that were likely extirpated, removing any adaptations to a warmer habitat.

4.3 Biological Function of Genomic Regions under Selection

Biological functions were identified for many of the gene regions under selection. Ten of the gene regions identified as being under selection were affiliated with retrotransposons, and were removed from further analyses. Retrotansposons are small pieces of DNA that have the ability to insert copies of themselves throughout genomes, using a form of molecular cut and paste, and are found roughly 500,000 times in the human genome, mainly in introns and pseudogenes (Goodier 2016). Retrotransposons can be targets of natural selection, if their insertion site causes them to impact functional genes. However, due to their repetitive nature, retrotransposons are also particularly prone to errors in assembly of reference genomes, and

mapping of ddRAD data back onto the genome, either or both of which can lead to erroneous signals of selection (Catchen et al. 2013). Hardy-Weinberg equilibrium filtering was implemented to address this problem herein, but due to compromises necessary to avoid removing genomic regions under selection, this filtering may not have been sufficient. Because of this potential error, the genomic regions associated with retrotransposon sequences cannot be confidently considered as being under selection. These retrotransposons highlight the importance of considering gene functions when investigating signals of selection, and the utility of applying multiple reference databases in the identification of gene functions.

One of the genomic regions was associated with the mitochondrial genome. Mitochondrial DNA is inherited only maternally, and thus may show a different pattern across populations as compared with bulk genomic DNA, if there are sexual differences in movement. Given that male and female *A. gazella* exhibit different movement patterns across large spatial scales (Boyd et al. 1998), it cannot be excluded that the signature of selection observed in mitochondrial genes is an artefact of the method, which "blanks" against the genome as a whole.

Two of the gene regions were associated with post-transcriptional regulation of gene expression, one with mRNA silencing, and one with protein degradation. Regulatory mechanisms such as these are often strong targets of selection, more so than the protein-coding genes, which they regulate (Carroll 2000). Mutations that change when, where, or how much a gene is expressed are more likely to be beneficial by chance than mutations which change the amino acid building blocks of the expressed proteins themselves (Carroll 2000).

One of the gene regions under selection was associated with Homeobox gene NKX3-2, which is involved in skeletal development, specifically in the ossification and longitudinal growth of bone (Jeong et al. 2017). Selection for genes involved in skeletal development between krill-feeding and fish feeding seal colonies could be a result of the very high levels of fluoride which are found in krill, but not in fish. Fluoride is an essential component of bone – and a lack of fluoride can lead to reduced bone strength; alternately, an excess of fluoride can cause skeletal deformities or the ossification of non-bone tissues (Ranjan & Ranjan 2015). Although previous studies of genetic adaptation to fluoride are few, studies with penguins have suggested they digest krill at lower efficiencies than fish, potentially as

an adaptation to reduce fluoride intake (Culik 1987, Kirkwood & Robertson 1997), and work with moths has shown adaptation to different levels of fluoride in the locally available trees they feed on as caterpillars (Chen 2003). Skeletal genes could also be under selection for their role in body morphology, as both different prey and different depth regimes may favour different body shapes.

Lastly, one of the genomic regions under selection was associated with the nicotinamide N-methyltransferase (NNMT) gene, which functions in metabolism, particularly with lipids (Kraus et al. 2014). In mice, changing NNMT expression influences the balance between metabolizing lipids for heat and storing lipids in adipose tissue (Trammell and Brenner 2015). Selection for differences in lipid metabolism in krill feeding versus fish feeding seals could be caused by a couple of mechanisms. Fish are roughly five times higher in lipid content than krill, which likely alters the optimal strategy for lipid metabolism. Krill-dominated areas also have a higher total availability of prey for seals, which may impact the optimal strategy for using versus storing energy.

Previous studies of *A. gazella* have generally been limited to a few adjacent islands. The only study to date investigating seals across their circumpolar range focused on a single one-base-pair loss-of-function genetic polymorphism. This relatively rare polymorphism, which causes a very light "blond" fur colour, was only present in the krill-dominated habitat area of the South Shetlands, South Georgia, and Bouvetøya (Hoffman et al. 2018). These nearly white "blond" seals are highly visible under water. One might speculate that selection against this polymorphism would be much stronger in areas where the predominant prey item is fish, as fish have higher visual acuity and a higher capacity to escape predators than krill. Thus, this may potentially indicate another genomic region under different natural selection pressures in krill-dominated as opposed to fish-dominated habitats. Given that this allele is determined by a single base position, it is perhaps not surprising that our analysis, which is based on a representative sub-set of the genome, did not detect this adaptation. This undetected adaptation suggests there may be other additional adaptations that could be detected with more extensive sequencing, offering exciting future possibilities as the costs of DNA sequencing continue to fall.

4.4 Conclusions and implications

A. gazella genomic DNA indicated local scale prey-driven natural selection. Of 27 gene regions identified as under selection, 24 displayed patterns consistent with different prey

regimes, krill-dominated or fish-dominated, as the driving factor for selection. These genomic regions under prey-driven selection included sequences associated with gene regulation, skeletal development, and lipid metabolism. Future investigations and controlled experiments, incorporating measures of physiological performance, will be needed to understand the precise mechanisms of these genetic adaptations and the impacts of these adaptations on growth and reproductive success. One promising initial avenue for such research would be comparative studies between islands that are currently inhabited by seal lineages that evolved in the present prey regime, compared with islands which are currently inhabited by seal lineages that migrated to the area after harvest-driven extirpation, having evolved under a different prey regime, a form of "natural" common garden experiment.

Local genetic adaptation for distinct prey regimes has implications for management of *A. gazella*. Our results suggest that seals which have evolved with an abundance of krill would be less fit for eating fish, as compared to their conspecifics that evolved in the absence of krill. The differences between seals endemic to different areas should be taken into account when considering appropriate management units, and when considering the potential impacts on *A. gazella* of climate or harvest driven changes in krill abundance.

Data availability

Sequence data is available in the NCBI short read archive under bioproject RJNA521705

Ethics Statement

Samples of *A. gazella* skin or blood were collected following standard protocols under the following permits: South Shetland Islands – US National Marine Fisheries Service, Marine Mammal Protection Act permit # 774-1847-04, South Georgia – British Antarctic Survey Animal Welfare and Ethics Review Body permit #PEA6, Bouvetøya – Norwegian Department of Plants, Fish, Animals and Food permit #7001, Marion Island – University of Pretoria, South Africa, Animal Ethics permit # PN859, Heard Island – Territory of Heard Island and McDonald Islands permit #00/18, Macquarie Island – Tasmanian Parks and Wildlife Service permit #FA99167.

Acknowledgements

We thank Debbie Baird-Bower, Caroline Bonin, and Iain Staniland for assistance with sample logistics. We thank the Abel computing cluster at the University of Oslo for computational resources and technical advice. Additional technical advice from Thorfinn

Korneliussen (ANGSD) and George Roth (Quantarctica) was also much appreciated. We additionally thank the anonymous manuscript reviewers for their comments. This research was funded by the Norwegian Antarctic Research Expeditions (NARE) programme.

Author contributions

KMK & ADL conceived of the study. MB, JF, MG, SG, CG & ADL provided samples. JIH provided access to laboratory facilities. ACC conducted lab work, analysed the data, and wrote the initial manuscript. MB, JF, SG, CG, JIH, KMK, ADL & CL contributed comments to the manuscript.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3): 403-410

Amante C, Eakins BW (2009) ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M

Anistoroaei R., ten Hallers B, Nefedov M., Christensen K, de Jong P (2011) Construction of an American mink Bacterial Artificial Chromosome (BAC) library and sequencing candidate genes important for the fur industry. BMC genomics 12(1): 354

Atkinson A, Hill SL, Pakhomov EA, Siegel V, Anadon R, Chiba S, Daly KL, Downie R., Fielding S, Fretwell P, Gerrish L, Hosie GW, Jessopp MJ, Kawaguchi S, Krafft BA, Loeb V, Nishikawa J, Peat HJ, Reiss CS, Ross RM, Quetin LB, Schmidt K, Steinberg DK, Subramaniam RC, Tarling GA, Ward P (2017) KRILLBASE: a circumpolar database of Antarctic krill and salp numerical densities, 1926–2016, Earth Syst. Sci. Data, 9, 193-210, https://doi.org/10.5194/essd-9-193-2017

Bonner WN (1968) The fur seal of South Georgia (Vol. 56). British Antarctic Survey, London

Boyd IL (1993) Pup production and distribution of breeding Antarctic fur seals (*Arctocephalus gazella*) at South Georgia. Antarctic Science, 5(01): 17-24

Boyd IL, McCafferty DJ, Reid K, Taylor R, Walker TR (1998) Dispersal of male and female Antarctic fur seals (*Arctocephalus gazella*). Canadian Journal of Fisheries and Aquatic Sciences 55(4): 845-852

Boyd IL, Lunn NJ, Barton T (1991) Time budgets and foraging characteristics of lactating Antarctic fur seals. Journal of Animal Ecology 60(2):577-592

Carroll SB (2000) Endless forms: the evolution of gene regulation and morphological diversity. Cell 101(6): 577-580

Casaux R, Baroni A, Arrighetti F, Ramón A, Carlini A (2003). Geographical variation in the diet of the Antarctic fur seal *Arctocephalus gazella*. Polar Biology 26(12):753-758

Casaux R, Baroni A, Carlini A (1998) The diet of the Antarctic fur seal *Arctocephalus gazella* at Harmony Point, Nelson Island, South Shetland Islands. Polar Biology 20 (6): 424-428.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013). Stacks: an analysis tool set for population genomics. Molecular ecology 22(11): 3124-3140

Cherel Y, Guinet C, Tremblay Y (1997). Fish prey of Antarctic fur seals *Arctocephalus gazella* at Ile de Croy, Kerguelen. Polar Biology 17(1): 87-90

Chen Y (2003) Variable tolerance of the silkworm *Bombyx mori* to atmospheric fluoride pollution. Fluoride 36(3): 157-162

Culik B (1987). Fluoride turnover in adélie penguins (*Pygoscelis adeliae*) and other bird species. Polar Biology 7(3): 179-187

Fariello MI, Boitard S, Mercier S, Robelin D, Faraut T, Arnould C, Recoquillay J, Bouchez O, Salin G, Dehais P, Gourichon D (2017) Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: The local score approach. Molecular Ecology 26(14): 3700-3714

Forcada J, Hoffman, JI (2014). Climate change selects for heterozygosity in a declining fur seal population. Nature 511(7510):462

Frolkina ZA (2002) Distribution of mackerel icefish (*Champsocephalus gunnari*) (Channichthyidae) around South Georgia at various stages of its life cycle. CCAMLR Science 9: 49-69

Goldsworthy SD, Hindell MA, Crowley HM (1997) Diet and diving behaviour of sympatric fur seals Arctocephalus gazella and A. tropicalis at Macquarie Island. In: Hindell MA,

Kemper C (eds) Marine mammal research in the southern hemisphere, Vol 1. Status, ecology and medicine. Surrey Beatty and Sons, Chipping Norton

Goldsworthy SD, McKenzie J, Page B, Lancaster ML, Shaughnessy PD, Wynen LP, Robinson SA, Peters KJ, Baylis AM, McIntosh RR (2009) Fur seals at Macquarie Island: post-sealing colonisation, trends in abundance and hybridisation of three species. Polar Biology 32(10): 1473-1486

Goldsworthy SD, Page B, Welling A, Chambellant M, Bradshaw CJA (2010) Selection of diving strategy by Antarctic fur seals depends on where and when foraging takes place. Marine Ecology Progress Series 409: 255-266

Goodier JL (2016) Restricting retrotransposons: a review. Mobile DNA 7:16 doi: 10.1186/s13100-016-0070-zGrant RB, Grant PR (2003) What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. BioScience 53(10): 965-975

Grant RB, Grant PR, (2003) What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. BioScience 53(10): 965-75.

Green K, Burton HR, Williams R (1989) The diet of Antarctic fur seals *Arctocephalus gazella* (Peters) during the breeding season at Heard Island. Antarctic Science 1(4): 317-324

Guinet C, Dubroca L, Lea MA, Goldsworthy S, Cherel Y, Duhamel G, Bonadonna F Donnay JP (2001). Spatial distribution of foraging in female Antarctic fur seals *Arctocephalus gazella* in relation to oceanographic variables: a scale-dependent approach using geographic information systems. Marine Ecology Progress Series 219: 251-264

Hodgson DA, Johnston NM, Caulkett AP, Jones VJ (1998) Palaeolimnology of Antarctic fur seal *Arctocephalus gazella* populations and implications for Antarctic management. Biological Conservation 83(2)145-154

Hoffman JI, Bauer E, Paijmans AP, Humble E, Beckmann LM, Kubetschek C, Christaller F, Kröcker N, Fuchs B, Moreras A, Shihlomule Y, Bester MN, Cleary AC, De Bruyn PJN, Forcada J, Goebel ME, Goldsworthy SD, Guinet C, Hoelzel AR, Lydersen C,

Kovaks KM, Lowther A (2018) A global cline in a colour polymorphism suggests a limited contribution of gene flow towards the recovery of a heavily exploited marine mammal. Royal Society Open Science 5: 181227

Hofmeyr GJG, Krafft BA, Kirkman SP, Bester MN, Lydersen C, Kovacs KM (2005) Population changes of Antarctic fur seals at Nyrøysa, Bouvetøya. Polar Biology 28(9): 725-731

Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, Riley J, Purvis I, Xu CF (2004) Detection of genotyping errors by Hardy–Weinberg equilibrium testing. European Journal of Human Genetics 12(5): 395

Hucke-Gaete R, Osman LP, Moreno CA, Torres D (2004) Examining natural population growth from near extinction: the case of the Antarctic fur seal at the South Shetlands, Antarctica. Polar Biology, 27(5): 304-311

Jeanniard du Dot T., Thomas A, Cherel Y, Trites A, Guinet C (2017). Combining hard-part and DNA analyses in scats with biologging and stable isotopes can reveal different diet compositions and feeding strategies within a population. Marine Ecology Progress Series 584: 1-6

Jeong DU, Choi JY, Kim DW, (2017) Cartilage-Specific and Cre-Dependent Nkx3. 2 Overexpression In Vivo Causes Skeletal Dwarfism by Delaying Cartilage Hypertrophy. Journal of Cellular Physiology 232(1): 78-90

Johnson R, Sumner M, Raymond B, (2017) Southern Ocean summer chlorophyll-*a* climatology Australian Antarctic Data Centre - doi:10.4225/15/5906b48f70bf9.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44: D457-D462

Kirkwood R, Robertson G (1997) The energy assimilation efficiency of emperor penguins, *Aptenodytes forsteri*, fed a diet of Antarctic krill, *Euphausia superba*. Physiological Zoology 70(1): 27-32

Klages NTW, Bester MN (1998) Fish prey of fur seals Arctocephalus spp. at subantarctic Marion Island. Marine Biology 131(3): 559-566

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. BMC bioinformatics 15(1): 356

Kraus D, Yang Q, Kong D, Banks AS, Zhang L, Rodgers JT, Pirinen E, Pulinilkunnil TC, Gong F, Wang YC, Cen Y, (2014) Nicotinamide N-methyltransferase knockdown protects against diet-induced obesity. Nature 508(7495): 258

Lancaster ML, Gemmell NJ, Negro S, Goldsworthy S, Sunnucks P (2006) Ménage à trois on Macquarie Island: hybridization among three species of fur seal (*Arctocephalus* spp.) following historical population extinction. Molecular Ecology 15: 3681-3692

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359

Li H*, Handsaker B*, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078-9

Locarnini RA, Mishonov AV, Antonov JI, Boyer T, Garcia HE, Baranova OK, Zweng MM, Paver CR, Reagan JR, Johnson DR, Hamilton M, Seidov D, (2013). World Ocean Atlas 2013, Volume 1: Temperature. S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 73

Lunn NJ, Boyd IL (1991) Pupping-site fidelity of Antarctic fur seals at Bird Island, South Georgia. Journal of Mammalogy 72(1): 202-206

Makhado AB, Bester MN, Kirkman SP, Pistorius PA, Ferguson JWH, Klages NTW (2008) Prey of the Antarctic fur seal *Arctocephalus gazella* at Marion Island. Polar Biology, 31(5): 575

Matsuoka K, Skoglund A, Roth G (2018) Quantarctica [Data set]. Norwegian Polar Institute. https://doi.org/10.21334/npolar.2018.8516e961

Meisner J, Albrechtsen A (2018) Inferring Population Structure and Admixture Proportions in Low Depth NGS Data, GENETICS doi: 10.1534/genetics.118.301336

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One 7:e37558

Quetin LB, Ross RM (1991) Behavioral and physiological characteristics of the Antarctic krill, *Euphausia superba*. American Zoologist 31(1): 49-63

Ranjan R, Ranjan A (2015) Toxic Effects. In: Fluoride Toxicity in Animals. Springer Briefs in Animal Sciences. Springer, Cham

Reid K, Croxall JP, Briggs DR, Murphy EJ (2005). Antarctic ecosystem monitoring: quantifying the response of ecosystem indicators to variability in Antarctic krill. ICES Journal of Marine Science 62(3): 366-373

Reid K, Forcada J (2005). Causes of offspring mortality in the Antarctic fur seal, *Arctocephalus gazella*: the interaction of density dependence and ecosystem variability. Canadian Journal of Zoology 83(4): 604-609

Robinson SA, Goldsworthy SD, van den Hoff J, Hindell MA (2002) The foraging ecology of two sympatric fur seal species, Arctocephalus gazella and Arctocephalus tropicalis, at Macquarie Island during the austral summer. Mar Freshwater Research 53: 1071–1082

Sambrook J, Fritsch EF, Maniatis T, (1989) Molecular Cloning, Cold Spring Harbor Laboratory Press, New York

Siegel V (2005) Distribution and population dynamics of *Euphausia superba*: summary of recent findings. Polar Biology 29: 1-22

Soevik T, Braekkan OR (1979) Fluoride in Antarctic krill (*Euphausia superba*) and Atlantic krill (*Meganyctiphanes norvegica*). Journal of the Fisheries Board of Canada 36(11): 1414-1416

Spreen G, Kaleschke L, Heygster G (2008) Sea ice remote sensing using AMSR-E 89 GHz channels, J. Geophys. Res doi:10.1029/2005JC003384

Staniland IJ, Reid K, Boyd IL (2004) Comparing individual and spatial influences on foraging behaviour in Antarctic fur seals *Arctocephalus gazella*. Marine Ecology Progress Series 275: 263-274

Staniland IJ, Robinson SL, Silk JRD, Warren N, Trathan PN (2012) Winter distribution and haul-out behaviour of female Antarctic fur seals from South Georgia. Marine Biology 159(2): 291-301

Skotte L1, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. Genetics 195(3):693-702. doi: 10.1534/genetics.113.154138

Tou JC, Jaczynski J, Chen YC, (2007) Krill for human consumption: nutritional value and potential health benefits. Nutrition reviews 65(2): 63-77

Trammell SAJ, Brenner C (2015) NNMT: a bad actor in fat makes good in liver. Cell metabolism 22(2): 200-201

Wynen LP, Goldsworthy SD, Guinet C, Bester MN, Boyd IL, Gjertz I, Hofmeyr GJ, White RW, Slade R (2000) Postsealing genetic variation and population structure of two species of fur seal (*Arctocephalus gazella* and *A. tropicalis*). Molecular Ecology 9(3): 299-314

Yonezawa T, Kohno N, Hasegawa M (2009). The monophyletic origin of sea lions and fur seals (Carnivora; Otariidae) in the Southern Hemisphere. Gene 441(1): 89-99

Tables

Island	Longitude	Latitude	Sea Surface Temp (C)	Proportion of year with sea ice	Depth (m)	Land Elevation (m)	Chl*a* mg m^{-3}	Krill m^{-2} (N)
South Shetland Islands	60.82 W	62.61 S	1.42	0.05	1310	283	0.48	31.3 (1128)
South Georgia	38.21 W	54.01 S	3.40	0.00	1594	422	1.05	87.0 (258)
Bouvetøya	3.41 E	54.42 S	1.26	0.03	2680	184	0.30	4.2 (3)
Marion Island	37.74 E	46.9 S	6.82	0.00	3248	325	0.25	0.0 (12)
Iles Crozet	51.76 E	46.4 S	6.52	0.00	2246	215	0.33	0.0 (0)
Iles Kerguelen	69.39 E	49.36 S	4.87	0.00	285	204	0.61	0.0 (4)
Heard Island	73.58 E	53.09 S	3.24	0.00	965	512	0.34	0.0 (2)
Macquarie Island	158.87 E	54.64 S	6.92	0.00	3889	115	0.19	0.0 (1)

Table 1: *A. gazella* colony locations and relevant metadata. All environmental data are mean values for all area within 160 km of the centre of the island. Sea surface temperature and chlorophyll *a* are austral summer values only. Krill net catch data (N = number of net samples) is year round, mainly austral summer.

Gene region	Contig	Start position	End position	# of SNPs	Fst deviation (mean ±SD)	Peak intensity	Clustering pattern	Putative function
2	9	4515734	4578699	5	0.02±0.12	3.95	A	ambiguous
3	15	2628551	2628789	11	0.08±0.13	3.99	F	ambiguous
4	17	10728754	10798855	7	0.08±0.14	5.38	E	ambiguous
6	25	12203131	12203305	4	0.01±0.11	4.72	A	ambiguous
8	36	8361196	8361203	2	0.14±0.24	5.05	C	ambiguous
9	54	6344925	6501282	2	0.06±0.05	3.64	D	ambiguous
11	63	2823203	2972391	16	0.06±0.10	4.42	B	ambiguous
12	72	3641161	3641214	4	0.08±0.18	8.74	C	mitochondrial ribosome
13	101	2571278	2653080	17	-0.02±0.07	3.39	A	ambiguous
15	152	2169948	2170102	5	0.01±0.10	3.65	A	metabolism (nicotinamide N-methyltransferase)
16	153	1368298	1549212	10	0.03±0.10	10.19	B	mRNA regulation (cytoplasmic polyadenylation element)
18	155	401365	671711	14	0.08±0.10	3.86	A	ambiguous
20	162	2473460	2473493	2	0.04±0.15	3.99	C	ambiguous
21	208	1521256	1521261	2	0.17±0.24	2.93	B	ambiguous
22	230	908557	929347	9	-0.02±0.07	5.10	G	ambiguous
24	312	251312	567172	8	-0.02±0.09	2.55	A	ambiguous
25	321	3240827	3355099	7	0.08±0.16	5.09	B	ambiguous
26	328	1254157	1254159	3	ID	2.51	B	ambiguous
27	339	382874	397210	8	0.00±0.10	2.89	C	ambiguous
28	346	548277	572323	6	0.09±0.16	3.44	A	protein degradation (uniquitin-conjugating enzyme)
29	360	114221	114387	21	0.04±0.11	4.17	A	ambiguous
30	365	1269651	1343708	6	-0.02±0.08	2.51	A	ambiguous
33	458	1170419	1318061	12	0.02±0.07	4.12	C	extracellular signal-regulated kinase
34	482	1870405	2372108	13	-0.02±0.07	3.03	A	cellular differentiation (Ras-associating and dilute domains)
35	544	81603	81825	3	-0.03±0.11	1.99	A	ambiguous
36	544	1054342	1157011	7	0.01±0.11	3.36	D	ambiguous
37	601	249857	249859	3	ID	3.39	C	skeletal development (homeobox NK3-2)

Table 2: Gene regions under selection, position (relative to SRP148937), clustering pattern and putative function. Peak Intensity indicates the strength of selection, as determined by the local scores analysis, where higher values indicate stronger selection. Fst deviations indicate the extent to which the region under selection differs from the overall structure of the genome (ID=insufficient data for this calculation). Putative functions are noted as "ambiguous" if different functions were identified in comparisons with GenBank and KEGG. Clustering patterns are as shown in figure 4, specifically, A=South Shetland Islands, South Georgia,

Bouvetøya, Marion Island and Iles Crozet vs Iles Kerguelen, Heard Island, and Macquarie Island, B= South Shetland Islands, South Georgia, Bouvetøya, Marion Island, Iles Crozet and Macquarie Island vs Iles Kerguelen and Heard Island, C= South Shetland Islands, South Georgia, Bouvetøya, and Marion Island vs Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island, D= South Shetland Islands, South Georgia, and Bouvetøya vs Marion Island Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island, E=, South Georgia and Bouvetøya vs South Shetland Islands, Marion Island, Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island, F= South Shetland Islands, South Georgia, Bouvetøya and Iles Crozet vs Marion Island, Iles Kerguelen, Heard Island, and Macquarie Island, G= South Shetland Islands, South Georgia, and Marion Island vs Bouvetøya, Iles Crozet, Iles Kerguelen, Heard Island, and Macquarie Island

Figures

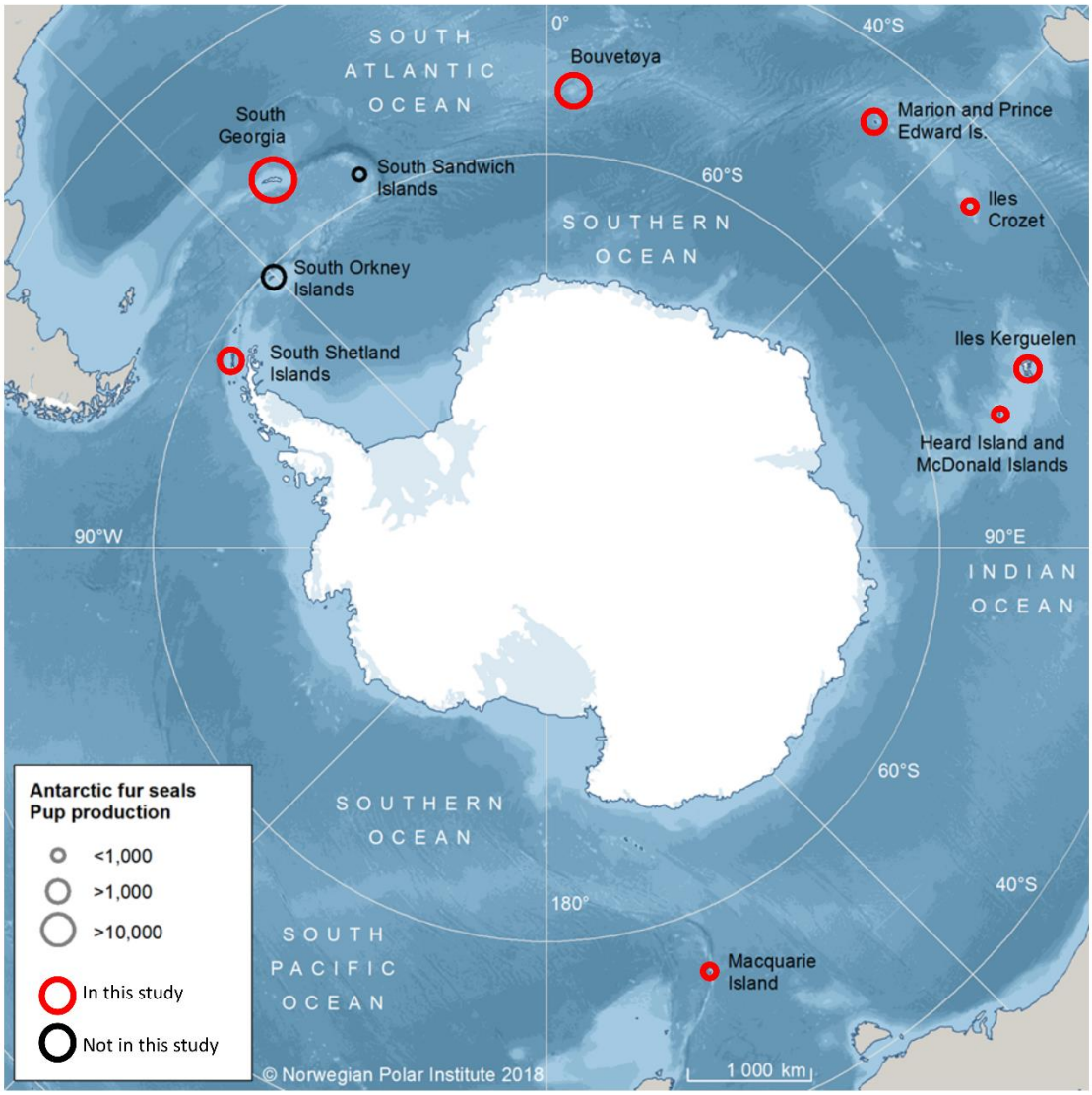

Figure 1: All extant colonies of *A. gazella*, study locations are indicated in red, circle size shows the magnitude of annual pup production (a proxy for population size).

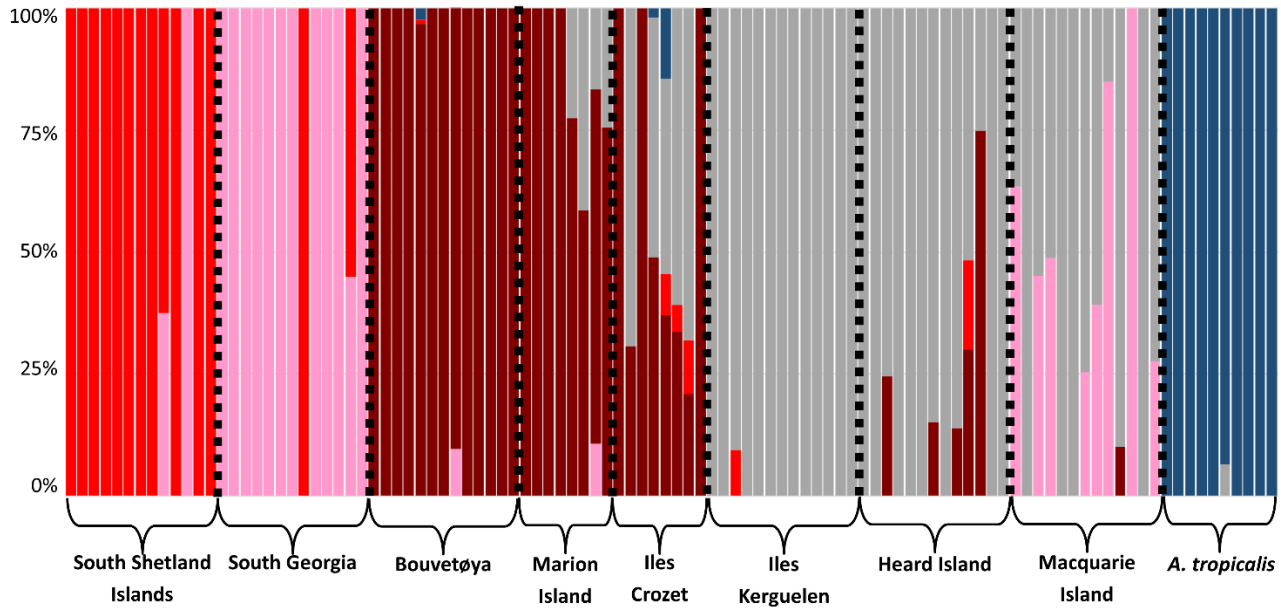

Figure 2: Overall genetic structuring of *A. gazella* populations. Groups and admixture proportions determined in NGSadmix. Each vertical bar indicates one individual seal, with color indicating the genetic population to which that individual belongs. Seals composed of two colors are indicative of mixed ancestry between the two different genetic populations. The islands where each seal was sampled is indicated along the x-axis, with *A. tropicalis* individuals at the far right.

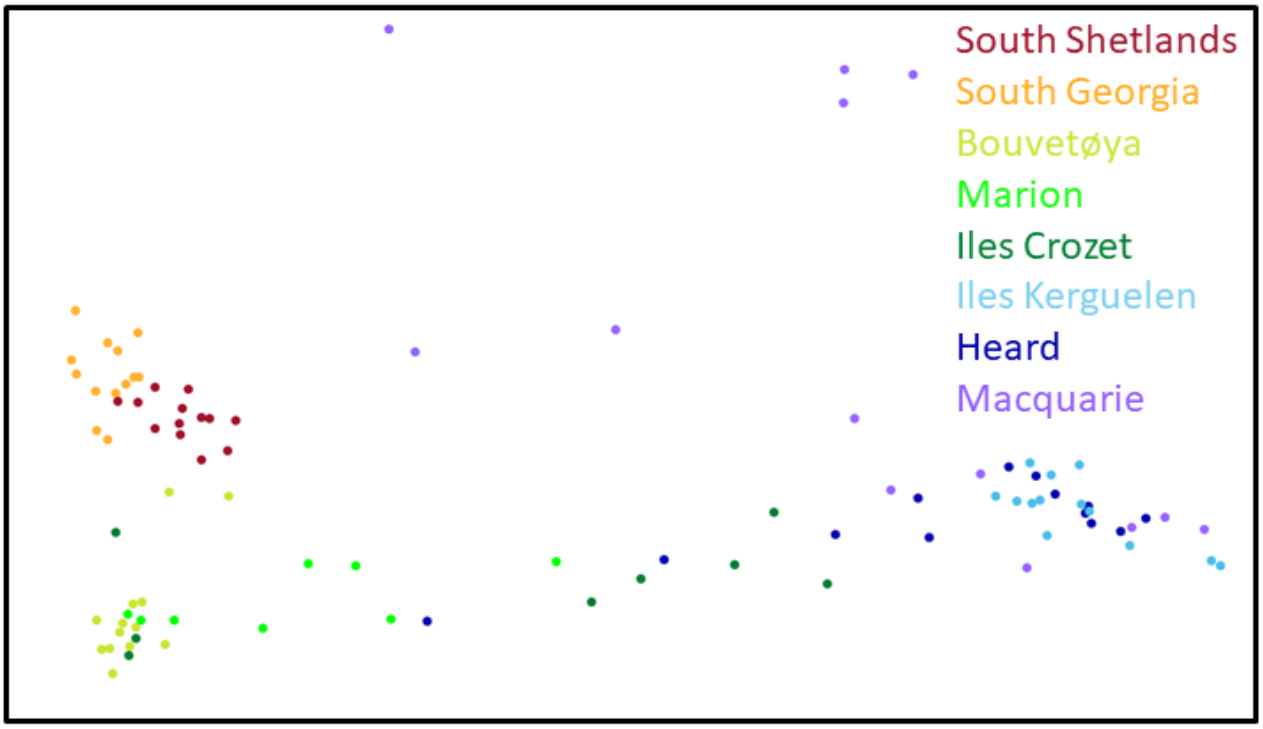

Figure 3: Principle Components Analysis of all genotype likelihoods for all analysed *A. gazella* individuals, coloured by sampling location, with PC1 indicated along the x-axis, and PC2 indicated along the y-axis.

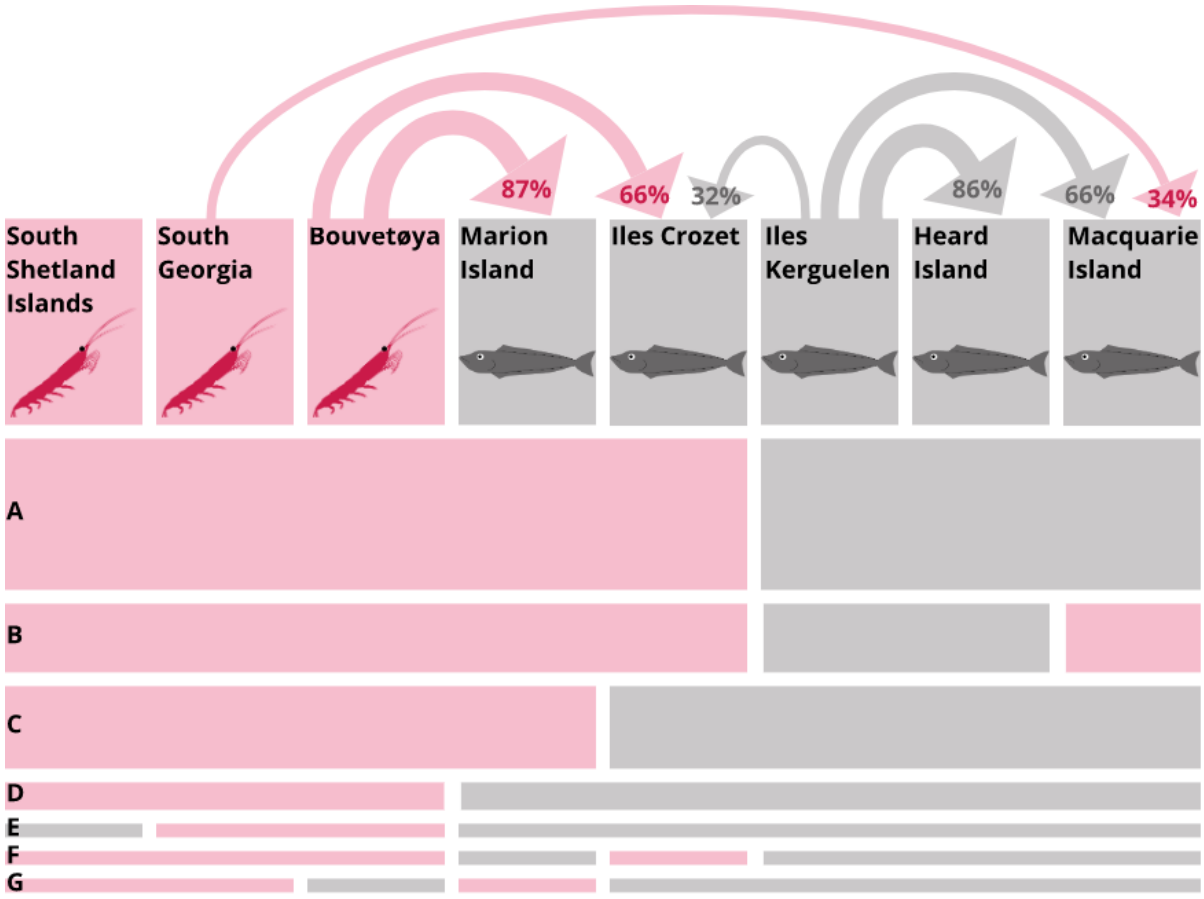

Figure 4: Clustering patterns of regions of the *A. gazella* genome identified as under selection. Islands are indicated as the top row of rectangles, with pictograms indicating the dominant prey item (krill or fish) at each island. Clustering patterns are labelled with letters as in table 2, with line width indicating the number of genetic regions displaying each clustering pattern. Arrows above the rectangles indicate major (>15% of total population) post-harvest migration and founder events, as inferred from overall population structure, with colour indicating the prey regime of the source population, with the percent admixture reflected in the arrow width and noted in the arrowhead.

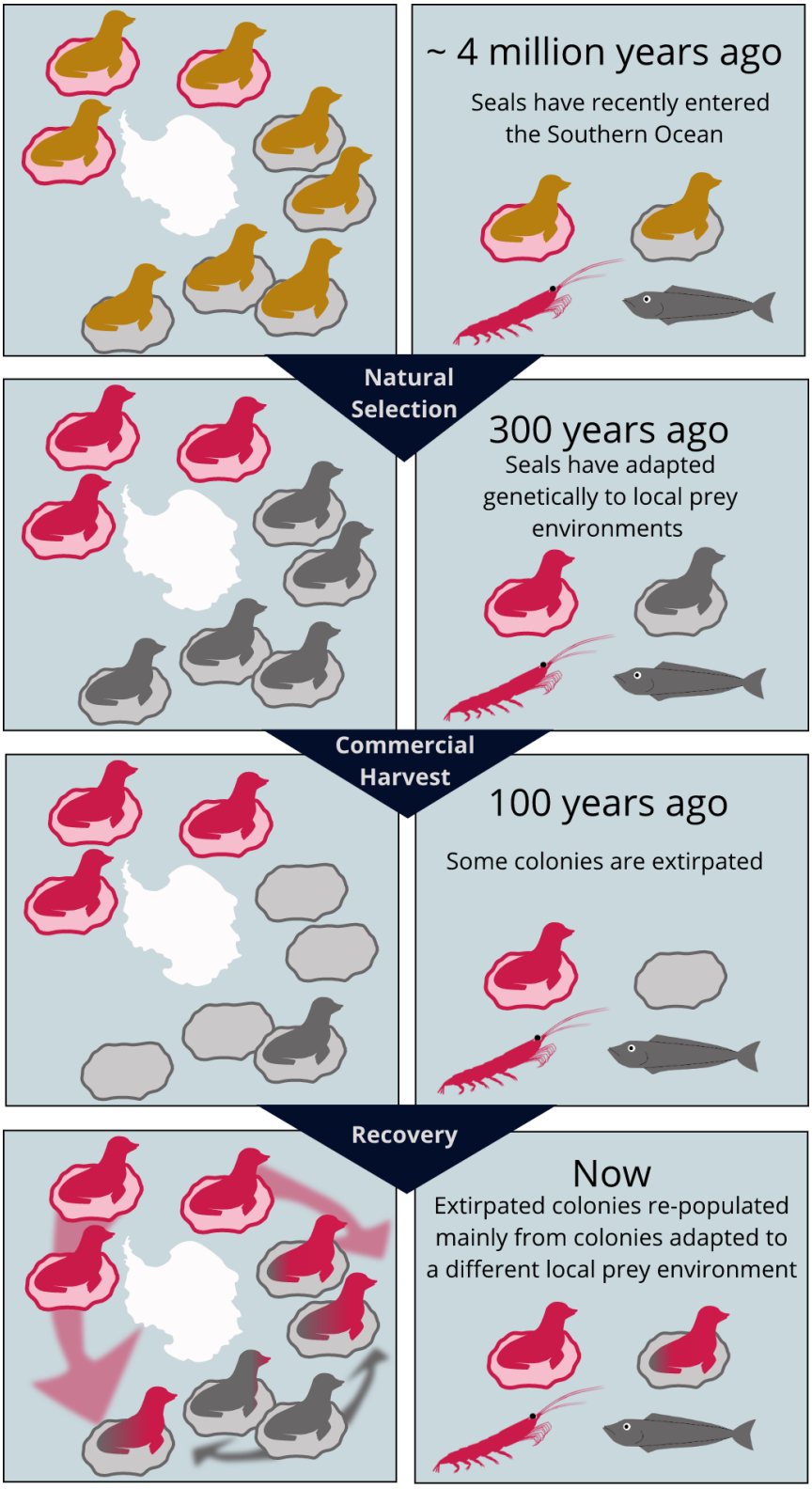

Figure 5: Interaction between adaptation and migration. Boxes illustrate specific points of time, while the triangles indicate processes which occurred between these illustrated time points. Pink coloration indicates krill-dominated habitat and krill-adapted seals, while grey coloration indicates fish-dominated habitat and fish-adapted seals.