Molecular analysis of the diet of Gentoo penguins: developing methods for monitoring programs

Introduction

Investigating the diet of an organism is of primary importance for understanding its ecological requirements and its functional role in the ecosystems that it inhabits. However, despite its central role in ecology, the study of the diet of top predators, such as seabirds, relies mostly in traditional invasive methods developed over 40 years ago (e.g. stomach flushing). It was only relatively recently that alternative biochemical methods, such as fatty acids signatures or stable isotopes analyses, became widespread. These have improved the knowledge of the foraging ecology of many species of animals (e.g. Hobson et al. 2002, Caut et al. 2008, Mancina and Herrera 2010) and have the great advantage of being relatively non-invasive and providing data about diet composition over long time scales. However, these techniques provide information only on the overall trophic level or broad geographical regions in which birds have been foraging, from which can be inferred only broad dietary shifts or large-scale changes in foraging location. The visual identification of prey remains from faeces or pellets provides information on actual prey consumed and is even less invasive but a serious limitation arises from biased recovery of the remains due to differential digestion and the difficulties of identifying well-digested prey (e.g. Seefelt and Gillingham 2006, Tollit et al. 2007).

There is therefore a major requirement for developing a widely applicable, non-invasive and objective method for the study of animal diet in the wild (Barrett et al. 2007). Molecular analysis of prey DNA in the guts, regurgitations or faeces of foragers potentially fulfils this requirement (reviewed by Symondson 2002, King et al. 2008). Prey DNA can be identified from even well-digested, amorphous remains in these samples. These techniques have not yet been extensively explored, particularly in vertebrates, but are a highly promising tool to improve the study of multiple trophic links, including in marine ecosystems (e.g. Jarman et al. 2002, Blankenship and Yayanos 2005, Deagle et al. 2007, Deagle et al. 2010).

Penguins (Spheniscidae) are a key group of the Antarctic marine ecosystem and are important consumers of marine resources in the Southern Ocean (Croxall and Prince

1987). Information on the diet and foraging ecology of penguins is vital to parameterise consumption models that lead to ecologically sensitive fisheries management via international agreements, such as with the Commission for the Conservation of Antarctic Marine Life Resources in the Antarctic (CCAMLR) and with the Scientific Committee for Antarctic Research (SCAR). Almost all studies since the mid 1980s have used stomach flushing as the main source of dietary information for most seabirds, including penguins (Wilson 1984), but this method causes distress and sometimes even deaths of the study animals. Thus, the assessment of non-invasive yet effective methods to study the diet of penguins is now one of the priorities for penguin ecologists (Ratcliffe and Trathan, pers. comm.).

Gentoo penguins *Pygoscelis papua* have broad geographical ranges across the Southern Ocean, are considered Near Threatened according to criteria set by the International Union for the Conservation of Nature (IUCN), are dispersed in a large number of small colonies and have very short foraging ranges from land throughout the year (Tanton et al. 2004; IUCN 2010; Wilson 2010). Therefore, Gentoo penguins are a convenient model to calibrate diet assessment methods for other penguin species that might not be available for sampling outside breeding, during the Austral winter. The aim of this study was to evaluate the potential of molecular methods a suitable alternative method for investigating penguin diet.

Methods

Sample collection

Faecal samples were collected in 2009 during the month of October from a Gentoo Penguin colony in South Georgia. The samples were collected randomly from the floor either immediately after defecation (fresh scats) or after one to two days from being produced (old scats). A total of 30 samples were collected (15 fresh and 15 old) and preserved in 70% ethanol. A further 48 random samples (both old and fresh) were collected from the same colony during the month of December. Instead of being preserved in ethanol, these samples were frozen immediately after collection.

DNA extraction and amplification

DNA from Gentoo penguin faecal samples was extracted using the QIAamp DNA Stool Mini Kit (Qiagen), following the manufacturer's standard protocol. Prior to extraction the samples were homogenised by mixing for at least 30 min and a sub-

sample (approximately 0.5ml) was used for extraction immediately after. Fresh and old samples were extracted separately and two blank extractions, using only water, were included for each batch of extractions to test for any cross-over contamination.

In order to test for the success of the extraction process, DNA extracts were screened using general primers for Bilateria species (Table I). Any samples giving a negative result were tested three times, to confirm that they were indeed negative. Successful extracts were then screened with Osteichthyes (boney fish), Cephalopoda (cephalopods) and Euphausiidae (Antarctic krill) specific primers (Table I) in order to investigate the proportion of birds testing positive for these two prey types.

Amplifications were performed separately for each primer pair, using the Multiplex PCR Kit (Qiagen) in 20 µl reactions containing 1x Multiplex PCR Master Mix, 0.2 µM of each primer and 0.1 mg/ml of BSA (New England Biolabs). The template was 2 µl of the DNA extract. Thermal cycling conditions were as follows: 95ºC for 15 min, 35 cycles (94ºC for 30 s followed by the primer specific annealing temperature for 90 s followed by 72ºC for 90 s), concluding with 72ºC for 10 min. A minimum of three negative controls (the extraction control, plus at least two distilled water blanks) were included in each set of PCR amplifications. PCR products were separated by electrophoresis in 1.5% agarose gels and visualised by staining with ethidium bromide, visualised by transillumination with UV light.

Preparation of DNA libraries for pyrosequencing

Three different pyrosequencing libraries were made for each set of faecal samples (old and fresh): 1) A general prey library using Bilateria primers for all the successful DNA extractions; 2) A fish library using Osteichthyes primers for the subsets of the samples which tested positive for Osteichthyes DNA; 3) A krill library using Euphausiidae primers also on the subsets of sample which tested positive for Euphausiidae DNA. Cephalopoda libraries were not made due to very small sample sizes.

The DNA concentration of the individual PCR products was measured using Qubit and the samples from each sample type (old and fresh) were pooled according to their concentration so that each pool contained an equal contribution from each individual bird. The six sample pools were sent to Eurofins MWG Operon for amplicon sequencing with the Roche GS-FLX Titanium series chemistry (454). Each sample pool was labelled with a unique 3 base pair long tag on the forward and reverse primers so that they could run together in the same platform.

Data analysis

The proportion of faecal samples from which DNA was successful extracted and the proportion of those containing each of the prey types tested (Osteichthyes, Euphausiidae and Cephalopoda) were compared between old and fresh samples using a Chi-square contingency table test or a Fisher's exact test.

For the pyrosequencing data, sequences were analysed using the Galaxy platform (https://main.g2.bx.psu.edu/root, Goecks et al. 2010, Blankenberg et al. 2010, Giardine et al. 2005). Sequences were separated by MID codes. Primers and MID codes were removed and all sequencing reads were collapsed to unique haplotypes. Sequences much longer or shorter than expected length (260-380bp amplicon for 12s, 110-180bp amplicon for 16s, 180-250bp amplicon for 18s) were removed.

The sequences were clustered into molecular operational taxonomic units in the program jMOTU (Jones et al. 2011) and tested thresholds from 1-10bp. A Xbp threshold for 12s, an Xbp threshold for 16s and an Xbp threshold for 18s were used. We determined the thresholds by analysing known data extracted from genbank (Supplemental files X, Y, Z) for the same analysis and selecting thresholds which minimize over-splitting of MOTUs (molecular operational taxonomic units) without losing significant taxonomic diversity. In all cases we erred on the side of conservatisms preferring to lump taxa in the same genus rather than over-splitting which would artificially inflate diversity. All sequences were compared to the NCBI database using a basic local alignment search (BLAST) and saved the resulting score files. These scores were visualized in MEGAN (Huson et al. 2011) using default settings.

Table I. Primers used in the present study.

Target (taxon, gene)	**Primer name**	**Sequence 5'-3'**	**Band size**	**Annealing Temp.**	**Reference**
Bilateria, nuclear 18S rDNA	BilSSU1100_f BilSSU1300_r	AGAGGTGAAATTSTTGGAYCG CCTTTAAGTTTCAGCTTTGCA	~245	62º	Jarman et al. 2004
Cephalopoda, nuclear 28S rDNA	Squid28SF Squid28SR	CGCCGAATCCCGTCGCMAGTAA AMGGCTTC CCAAGCAACCCGACTCTCGGAT CGAA	~180	60º	Deagle et al. 2005
Osteichthyes, mtDNA 12S	FishF1 FishR1	CGGTAAAACTCGTGCC CCGCCAAGTCCTTTGGG	~300	56º	Jarman unpubl.
Euphausiidae, mtDNA 16S	EuphMLSUF EuphMLSUR	TTTATTGGGGCGATAAAAAT TCGAGGTCGYAATCTTTCTTGT	~169	54º	Deagle et al. 2007

Results

DNA was successfully amplified from all the 30 faecal samples (15 old and 15 fresh) using general Bilateria primers. The success of DNA amplication using fish and krill primers was, respectively, 53.3% and 73.3% for fresh samples and 73.3% and 86.7% for old samples. Squid DNA was found only in fresh samples and only in two out of the 15 samples, therefore these primers were not used for the pyrosequencing analysis. These differences in prey detection between fresh and old samples are not statistically significant (all $p > 0.4828$) suggesting that one to two days old scats are as good as fresh samples for dietary analysis of penguins. DNA amplification from frozen samples was less successful with only 50% of the samples amplifying with either primer pair, indicating a successful DNA extraction. Amongst the 24 samples for which DNA was successful extracted, 91.7% were positive for Bilateria (surprisingly two samples amplified with either the fish or krill primers but not with Bilateria general primers), 75% were positive for fish, 79.2% were positive for krill and 4.2% were positive for squid. Squid primers were also not used for pyrosequencing analysis of frozen samples due to the very low occurrence of this prey group. Fresh and old samples from October were pooled to compare prey detection rates between months. No significant differences were found in the proportion of samples testing positive for fish, krill or squid between October and December (all $p > 0.5337$).

Results from the pyrosequencing analysis are shown in Table II. All the fish and krill taxa identified were present in both old and fresh scats. The fish primers enabled the identification of *Krefftichthys anderssoni*, *Champsocephalus gunnari* and other non-identified Nototheniidae. *Euphausia superba* and *Thysanoessa* sp. were the only crustacean taxa that could be identified. Prey identification from general Bilateria primers was harder to achieve at low taxonomic levels. The groups of taxa identified with these primers were: flatworms (Eucestoda), teleostei fish (Elopocephala) and crustaceans (Eumalacostraca). The later were only identified from fresh samples though other non-identified Panarthropoda (an unranked taxon that includes the Arthropods) were present in old samples. *Krefftichthys anderssoni* and an unidentified Euphausiidae MOTU were the only taxa found in Fresh and Old samples from October but not in Frozen samples from December. DNA from Nematodes and Tardigrades were found only in Frozen samples. These are likely to represent contamination from the soil rather

than consumed prey (unlike the Fresh and Old samples, these were collected during the summer, on soil instead of ice).

Table II. Taxa identified from pyrosequencing analysis of Gentoo penguin (*Pygoscelis papua*) faecal samples using DNA fragments from three different genes. The "✓" indicates the presence of each taxa in samples collected in October 2009 (Fresh and Old, preserved in ethanol) and December 2009 (Frozen, a mixture of fresh and old samples immediately frozen after collection). Sequences not identified to the species level could include more than one species though, supposedly, from the same genus.

Target Gene	Filo	Class	Order	Family	Genus/Species	Fresh	Old	Frozen
18S	Platyhelminthes	Cestoda	Unknown	Unknown	Unknown	✓	✓	✓
	Nematoda	Chromadorea	Unknown	Unknown	Unknown			✓
	Tardigrada	Unknown	Unknown	Unknown	Unknown			✓
	Arthropoda	Malacostraca (Eumalacostraca)	Unknown	Unknown	Unknown	✓		✓
		Unknown (Panathropoda)	Unknown	Unknown	Unknown		✓	✓
	Chordata (Euteleostomi)	Actinopterygii (Elopocephala)	Unknown	Unknown	Unknown	✓	✓	✓
		Unknown	Unknown	Unknown	Unknown	✓	✓	✓
	Unknown	Unknown	Unknown	Unknown	Unknown	✓	✓	✓
12S	Chordata	Actinopterygii	Myctophiformes	Myctophidae	*Krefftichthys anderssoni*	✓	✓	
			Perciformes (Notothenioidei)	Nototheniidae	*Champsocephalus gunnari*	✓	✓	✓
					Unknown	✓	✓	✓
				Unknown	Unknown	✓	✓	✓
16S	Arthropoda	Malacostraca	Euphausiacea	Euphausiidae	Euphausia superba	✓	✓	✓
					Thysanoessa sp.	✓	✓	✓
					Unkown	✓	✓	

Discussion

Molecular techniques were successful in obtaining dietary information from fresh and old faecal samples from Gentoo penguins. This is useful since sometimes is it harder to

obtain fresh samples (e.g. no time at a colony to collect fresh samples) or being able to clearly differentiate fresh from old samples. Although freezing is a standard way of preserving DNA, DNA yield can be adversely affected by freezing (Ross, Haites and Kelly 1990). Considering the low DNA yield typically obtained from faecal samples, this could explain why the amplification success from Frozen samples was much lower when compared to samples stored in ethanol.

This study suggests that Gentoo penguins in South Georgia feed primarily on crustaceans and. Also on fish (mainly nototheniids), relying very little on cephalopods. This is in accordance with results from the various studies available for this region using dissections or stomach flushing (Croxall and Prince, 1980; Croxall, Davis and O' Connell, 1988; Croxall, Prince and Reid, 1997; Williams, 1991; Kato et al., 1991; Croxall, Reid and Prince, 1999; Berrow, Taylor and Murray, 1999). Although no direct quantitative conclusions can be drawn from the pyrosequencing analysis, dietary shifts could be identified if comparing data from different years; comparisons on diet composition among colonies or individuals are also valid.

It was possible to identify some prey sequences to the species level, namely the Antarctic krill *Euphasia superba* and the Antarctic icefish *Champsocephalus gunnari*, consistently identified as the most common prey items for this species (Croxall et al., 1980; Croxall, Davis and O' Connell, 1988; Croxall, Prince and Reid, 1997; Williams, 1991; Kato et al., 1991; Croxall, Reid and Prince, 1999; Berrow, Taylor and Murray, 1999). The myctophid *Krefftichthys anderssoni* and species of the genus *Thysanoessa* were also identified in this study. However, the overall resolution of taxa was poor and some expected prey taxa, known from stomach flushing and visually identification from scats to be common prey items, could not be identified: e.g. Fish: *Gymnoscopelus braueri* and *Lepidonotothen larsenii*; Crustaceans: *Themisto gaudichaudii*, *Antarctomysis maxima* and *Gondogeneia georgiana*. Bioinformatics are still being carried out to improve this resolution based on more detailed investigation of the DNA fragments used to amplify the prey sequences. An extensive reference collection of DNA sequences of identified potential prey would greatly aid these analyses in the future. Indeed, data from this study can be revisited as more reference sequences become available for comparison.

Despite the many advantages of molecular methods to study animals' diet, these techniques (pyrosequencing in particular) are usually costly when compared to more

conventional methods. However, such costs are steadily decreasing as the new-generation sequencing technology evolves and its use becomes more widespread.

A less costly alternative would be to design specific primers to target key prey species to routinely investigate the proportion of Gentoo penguin's faecal samples testing positive for these prey types. This is a simple way of investigating yearly dietary shifts potentially caused by changes in prey abundance without the need for a detailed dietary study (using pyrosequecing) which could be done less frequently.

Data from this study will be combined with information based on stable isotopes, fatty acids analysis and visual identification of prey from scats in order to validate a non-invasive method to characterize the diet of penguins, to be implemented into Antarctic penguin monitoring programs.

References

Barrett RT, Camphuysen K, Anker-Nilssen T, Chardine JW, Furness RW, Garthe S, Huppop O, Leopold MF, Montevecchi WA, Veit RR (2007) Diet studies of seabirds: a review and recommendations. ICES Journal of Marine Science, 64: 1675-1691.

Berrow SD, Taylor RI, Murray AWA (1999) Influence of sampling protocol on diet determination of gentoo penguins Pygoscelis papua and Antarctic fur seals Arctocephalus gazella. Polar Biology, 22: 156-163.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: A web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology, 19:Unit 19.10.1-21.

Blankenship LE, Yayanos AA (2005) Universal primers and PCR of gut contents to study marine invertebrate diets. Molecular Ecology 14: 891-899.

Caut S, Angulo E, Courchamp F (2008) Discrimination factors (Delta N-15 and Delta C-13) in an omnivorous consumer: effect of diet isotopic ratio. Functional Ecology, 22: 255-263.

Croxall JP, Davis RW, O' Connell MJ (1988) Diving patterns in relation to diet of gentoo and macaroni penguins at South Georgia. Condor, 90: 157-167.

Croxall JP, Prince PA (1980) The Food of Gentoo Penguins *Pygoscelis papua* and Macaroni Penguins *Eudyptes chrysolophus* at South Georgia. Ibis, 122: 245-253.

Croxall JP, Prince PA (1987) Seabirds as predators on marine resources, especially krill, at South Georgia. In: Croxall JP (ed) Seabirds: Feeding Ecology and Role in Marine Ecosystems. Cambridge University Press, Cambridge, pp 347-368.

Croxall JP, Prince PA, Reid K (1997) Dietary segregation of krill-eating south Georgia seabirds. Journal of Zoology, 242: 531-556.

Croxall JP, Reid K, Prince PA (1999) Diet, provisioning and productivity responses of marine predators to differences in availability of Antarctic krill. Marine Ecology-Progress Series, 177: 115-131.

Deagle BE, Tollit DJ, Jarman SN, Hindell MA, Trites AW, Gales NJ (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. Molecular Ecology, 14:1831-1842.

Deagle BE, Chiaradia A, McInnes I, Jarman S (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? Conservation Genetics, 11:2039-2048.

Deagle BE, Gales NJ, Evans K, Jarman SN, Robinson S, Trebilco R, Hindell MA (2007) Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). PLoS ONE, 2(9): e831. doi:10.1371/journal.pone.0000831.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. Genome Research, 15: 1451–1455.

Goecks J, Nekrutenko A, Taylor J, The Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology, 11: R86 doi:10.1186/gb-2010-11-8-r86.

Hobson K, Fisk A, Karnovsky N, Holstd M, Gagnone J-M, Fortierf M (2002) A stable isotope (Delta 13C, 15N) model for the North Water food web: implications for evaluating trophodynamics and the flow of energy and contaminants. Deep Sea Research, 49: 5131-5150.

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. Genome Research, 17: 377−386.

IUCN (2010) IUCN Red List of Threatened Species. Version 2010.4 (www.iucnredlist.org).

Jarman SN, Deagle BE, Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. Molecular Ecology, 13: 1313-1322.

Jarman SN, Gales NJ, Tierney M, Gill PC, Elliot NG (2002) A DNA-based method for identification of krill species and its application to analyzing the diet of marine vertebrate predators. Molecular Ecology, 11: 2679-2690.

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and Taxonerator: Turning DNA Barcode Sequences into Annotated Operational Taxonomic Units. PLoS ONE 6(4): e19259. doi:10.1371/journal.pone.0019259.

Kato A, Williams AJ, Barton KJ, Rodwell S (1991) Short-term variation in the winter diet of gentoo penguins Pygoscelis papua at South Georgia during July 1989. Marine Ornithology, 19: 31-38.

King RA, Read DS, Traugott M, Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. Molecular Ecology, 17: 947-63.

Mancina CA, Herrera LG (2010) Disparate feeding strategies used by syntopic Antillean nectarivorous bats to obtain dietary protein. Journal of Mammalogy, 91: 960-966.

Ross KS, Haites NE, Kelly KF (1990) Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. Journal of Medical Genetics, 27: 569-570.

Seefelt NE, Gillingham JC (2006) A comparison of three methods to investigate the diet of breeding double-crested cormorants (*Phalacrocorax auritus*) in the Beaver Archipelago, northern Lake Michigan. Hydrobiologia, 567: 57-67.

Symondson WOC (2002) Molecular identification of prey in predator diets. Molecular Ecology, 11: 627-641.

Tanton JL, Reid K, Croxall JP, Trathan PN (2004) Winter distribution and behaviour of gentoo penguins Pygoscelis papua at South Georgia. Polar Biology, 27: 299-303

Tollit DJ, Heaslip SG, Barrick RL, Trites AW (2007) Impact of diet-index selection and the digestion of prey hard remains on determining the diet of the Steller sea lion (*Eumetopias jubatus*). Canadian Journal of Zoology-Revue Canadienne De Zoologie, 85: 1-15.

Williams TD (1991) Foraging Ecology and Diet of Gentoo Penguins Pygoscelis-Papua at South Georgia During Winter and an Assessment of Their Winter Prey Consumption. Ibis, 133: 3-13.

Wilson RP (1984) An improved pump for penguins and other seabirds. Journal of Field Ornithology, 55: 109-112.

Wilson RP (2010) Resource partitioning and niche hyper-volume overlap in free-living Pygoscelid penguins. Functional Ecology, 2010: 649-657.