

Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out?

Bruce E. Deagle · André Chiaradia ·
Julie McInnes · Simon N. Jarman

Received: 2 December 2009 / Accepted: 25 May 2010 / Published online: 17 June 2010
© Springer Science+Business Media B.V. 2010

Abstract DNA barcoding of faeces or stomach contents is an emerging approach for dietary analysis. We pyrosequenced mtDNA 16S markers amplified from faeces of captive little penguins (*Eudyptula minor*) to examine if recovered sequences reflect the proportions of species consumed. We also analysed wild little penguin faeces collected from 100 nests in southeast Australia. In the captive study, pilchards were the primary fish fed to the penguins and DNA sequences from pilchard were the most common sequences recovered. Sequences of three other fish fed in constant mass proportions (45:35:20) were all detected, but proportions of sequences (60:6:34) were considerably different than mass proportions in the diet. Correction factors based on relative mtDNA density in the fish did not improve diet estimates. Consistency between replicate samples suggests that the observed bias resulted from differences in prey digestibility. Detection of DNA from fish consumed before the penguins were brought into captivity indicates that a DNA signal in faeces can persist for at least 4 days after ingestion. In the wild-collected faeces, 24 distinct fish and 1 squid were identified; anchovy, barracouta and pilchard accounted for over 80% of these sequences. Our results highlight that DNA

sequences recovered in dietary barcoding studies can provide semi-quantitative information on diet composition, but these data should be given wide confidence intervals.

Keywords *Eudyptula minor* · DNA barcoding · Seabird diet · GS-FLX · 454 Sequencing · Non-invasive

Introduction

Knowledge of what animals eat is one of the most basic requirements for understanding their biology and the functioning of the ecosystems in which they exist. Therefore, obtaining accurate dietary information is often an integral, but challenging, part of conservation planning. Over the past few years, there has been considerable interest in using DNA-based methods for studying animal diet through the analysis of food DNA recovered from stomach contents or faeces (reviewed in King et al. 2008). One approach being applied is dietary DNA barcoding. This involves PCR amplifying suitable markers from dietary samples, sequencing the resultant amplicons and identifying the sequences by comparison to a reference database (e.g. Hofreiter et al. 2000; Passmore et al. 2006; Bradley et al. 2007; Deagle et al. 2007; Clare et al. 2009). This field of research is currently taking a large leap forward due to technological innovations such as the availability of high-throughput sequencing platforms (Hudson 2008) and the application of blocking oligos to suppress amplification of non-target DNA (Vestheim and Jarman 2008). In addition, ever-increasing access to DNA sequences in public databases has enabled better primer design and identification of the recovered sequences. These advances have drastically increased the scale of DNA-based dietary studies (e.g. Deagle et al. 2009; Soininen et al. 2009; Valentini et al.

B. E. Deagle (✉)
Department of Biology, University of Victoria,
PO Box 302, Victoria, BC V8W 3N5, Canada
e-mail: bdeagle@uvic.ca

A. Chiaradia · J. McInnes
Research Department, Phillip Island Nature Parks,
PO Box 97, Cowes, VIC 3922, Australia

S. N. Jarman
Australian Marine Mammal Centre, Australian Antarctic
Division, Channel Highway, Kingston, TAS, Australia

2009a) and will reduce the effort required to carry out analyses with coverage reflecting the full diversity of food items eaten by animal populations.

The ultimate goal of many diet studies is to provide a quantitative estimate of food species being consumed. In dietary DNA barcoding, it is hoped that proportions of DNA sequences recovered will provide an indication of the importance of various food items in the diet. However, there are a number of reasons why these proportions may be misleading. Recovering quantitative data from complex pools of DNA is prone to several possible methodological biases (e.g. von Wintzingerode et al. 1997; Forney et al. 2004; Acinas et al. 2005). In addition, food species may differ in amount of DNA present per unit biomass and/or in tissue digestibility. One way to assess the accuracy of the quantitative data obtained using the dietary barcoding approach is through studies of food DNA in faeces from captive animals with known diets (e.g. Deagle and Tollit 2007; Weber and Lundgren 2009).

Seabirds are major consumers in many marine ecosystems (Brooke 2004) and long-term diet data of these top-level predators provide a powerful means of monitoring changes in the availability of marine resources at lower trophic levels (Boyd and Murray 2001; Hedd et al. 2006; Boersma et al. 2009). This is particularly relevant given increasing human impacts and difficulty accessing the biological consequences of environmental changes in marine systems (Ropert-Coudert et al. 2009). Since seabird faeces contain few hard components useful for prey identification, their diet is generally studied through stomach flushing or biochemical means (Barrett et al. 2007). Neither of these approaches is ideal: stomach flushing is invasive, overlooks soft prey and is labour intensive; stable isotope analysis gives poor taxonomic resolution and can be confounded by species with similar isotope signatures (Sydeman et al. 1997). It has been shown that prey DNA is recoverable from penguin faeces (Jarman et al. 2002; Deagle et al. 2007), suggesting dietary DNA barcoding could be applied in large-scale minimally invasive studies of seabirds. Despite this potential, no captive trials have been done with seabirds to validate the method. Little penguins (*Eudyptula minor*) provide a good model system for evaluating faecal DNA barcoding. They are central place foragers with a small foraging range (<20 km, Hoskins et al. 2008) which means they return ashore often, providing frequent faecal material from adults and chicks. They are also generalist predators and their breeding success is strongly related to food availability (Cullen et al. 1992; Chiaradia and Nisbet 2006). This makes ongoing dietary studies invaluable to monitor changes in food web structure and to understand the influence of diet on population size fluctuations (Ropert-Coudert et al. 2009; Chiaradia et al. 2010).

In the current study, we pyrosequence prey DNA recovered from faeces of little penguins. We analysed faeces of experimentally fed captive penguins to determine whether an accurate quantitative dietary signature could be recovered from faecal DNA. We also determined the relative DNA content of the fish being fed to the penguins to see if derived correction factors would be helpful in refining quantitative estimates of diet. Finally, we present some data from penguin faeces samples collected in a preliminary field-based study.

Methods

Captive feeding trials

The penguins in the feeding trials ($n = 30$) were 8–9 week old fledglings that were removed from nesting burrows located in high-human use areas of Phillip Island, Australia ($38^{\circ}15'S$, $145^{\circ}30'E$) as part of a translocation program. Penguins were kept in captivity for about 1 week before being returned to the wild at a protected beach site. During an initial acclimatization period the birds were fed a diet of 100% whole pilchards *Sardinops sagax*. For the final 4 days of captivity the penguins were fed a 'test diet' (a constant mass of blended fish tissue; 100 g twice daily) and then fed until satiated with a variable portion of whole pilchards (during this time pilchards made up $47.8 \pm 8.9\%$ of total daily food). The test diet consisted of 45% tuna (Scombrinae sp.), 35% tommy ruff (*Arripis georgianus*), and 20% whiting (*Sillago flindersi*) by mass. Penguins were housed in pairs and faecal samples were collected from their pens on each of the final 3 days of a trial (i.e. when we expected DNA from the test diet to be present). The test diet tissue mix being fed to the penguins was made in large batches; samples from two tissue mixes were preserved in 95% ethanol for DNA analysis.

Field collection

Collections of faecal samples from wild penguins were made between 22 November 2007 and 22 January 2008 at three breeding colonies in Victoria, Australia: north ($n = 15$) and south ($n = 70$) of Phillip Island; ($38^{\circ}28'S$, $145^{\circ}13'E$); and Rabbit Island ($37^{\circ}22'S$, $149^{\circ}45'E$, $n = 15$). Moist faeces were collected at the entrance to burrows, or in the case of artificial nest boxes with removable lids, from inside the nest (details in Chiaradia and Kerry 1999). Faeces could have been from chicks and/or adults of either sex since all are present at burrows during this time of year. All fresh pellets present in a single nesting site were pooled, and faeces from each of the 100 nests were stored in separate tubes in 95% ethanol.

DNA extraction and PCR amplification

For faecal samples from the feeding trial, DNA was extracted from a blend of the faeces collected each day from each pair of penguins (3 days × 15 pairs = 45 DNA template samples). From each of the tissue mixes, four replicate DNA extractions were carried out (8 DNA template samples). For the 100 wild collected faeces, to increase the likelihood of obtaining prey DNA in each sample we pooled equal volumes of homogenized faecal material from five nesting sites before extraction (20 DNA template samples). Before DNA extraction, faeces (or tissue mix) were thoroughly homogenised in ethanol using a micro-blender. DNA was recovered from 50 to 100 mg of material using the QIAamp DNA Stool Mini Kit (Qiagen), as described previously (Deagle et al. 2005). The DNA was eluted in 100 µl Tris buffer (10 mM) and diluted 1:5 in distilled water. Extraction blanks were included with each batch of extractions to monitor for cross-contamination.

In dietary barcoding of faeces, a suitable barcoding marker must be short due to the degraded nature of the recovered DNA (Deagle et al. 2006) and PCR priming sites must be highly conserved to enable relatively unbiased amplification of a heterogeneous mix of DNA templates (Valentini et al. 2009b). While cytochrome oxidase I has become the standard barcoding marker used to identify metazoans, suitably conserved PCR priming sites do not exist in this protein-coding gene. Therefore, we used part of the mtDNA 16S rRNA gene as a barcoding marker. Primers have been designed to amplify short regions of mtDNA 16S DNA from diverse metazoan taxa (Dunshea 2009), the amplified region is variable enough to allow species or genus level identification (Vences et al. 2005) and mtDNA 16S sequences are available in GenBank for many southern Australian fish species due to submissions from two previous dietary studies (Deagle et al. 2009;

Braley et al. 2010). In the current study, all DNA samples were PCR amplified using a primer set targeting a short mtDNA 16S fragment (~100 bp) from chordates (16S SHORT). In addition, DNA from field collected faecal samples were amplified using an alternative reverse primer producing a slightly longer (~250 bp) mtDNA 16S fragment (16S LONG). To amplify the 16S LONG fragment, degenerate primers were used allowing amplification of DNA from cephalopods as well as chordates. Both primer sets are described in Deagle et al. (2009). A penguin-specific blocking oligo was included in all PCR mixes to suppress amplification of penguin DNA templates (see Vestheim and Jarman 2008). All oligos (PCR and blocking) were purified via polyacrylamide gel electrophoresis to remove truncated products.

PCR amplification profiles were determined for the primer sets in a subset of samples using the QuantiTect SYBR Green PCR Kit (Qiagen) and the Chromo4 detection system (MJ Research). Based on these data, final PCR amplifications were stopped during the exponential stage of amplification (cycle 32: faecal DNA and cycle 25: tissue DNA) to minimise differential amplification bias (but see Acinas et al. 2005). PCR amplifications used to produce template for pyrosequencing were performed using the Multiplex PCR Kit with HotStarTaq DNA Polymerase (Qiagen). Each PCR (10 µl) contained 5 µl Multiplex PCR Master Mix, 0.25 µM of each primer, 2.5 µM blocking oligo and 4 µl template DNA. Thermal cycling conditions were: 95°C for 15 min followed by cycles of: 94°C for 20 s/primer specific annealing temperature (see Table 1) for 90 s/72°C for 45 s. Followed by a final extension (72°C) for 2 min. Aerosol-resistant pipette tips were used with all PCR solutions and negative control reactions (extraction control and water template) were included with each set of PCR amplifications. All products were checked on 1.8% agarose gels and DNA quantified by

Table 1 Sequences of primers used to amplify DNA samples prior to pyrosequencing

PCR primer set	Primer name	Primer sequence (5'–3') ^a	Product size (bp) ^b	Annealing temp.
16S SHORT	Chord_16S_F	Primer A <i>TCA</i> or <i>ATGCGAGAAGACCCTRTGGAGCT</i>	~155	57°C
16S SHORT	Chord_16s_R_Short	Primer B <i>TATCCTNGGTCGCCCAAC</i>	~155	57°C
16S LONG	Chord_16S_F	Primer A <i>GATCGAGAAGACCCTRTGGAGCT</i>	~260–310	52°C
16S LONG	Ceph_16S_F	Primer A <i>GATGACGAGAAGACCCTAWTGAGCT</i>	~260–310	52°C
16S LONG	Chord_16S_R	Primer B <i>GGATTGCGCTGTTATCCCT</i>	~260–310	52°C
16S LONG	Ceph_16S_R	Primer B <i>AAATTACGCTGTTATCCCT</i>	~260–310	52°C
Used with both	Blocking oligo ^c	<i>GTGGAACCTGAAAATCAGCGACCACCA</i> C3		

The *italic* regions represent variable nucleotide tags used to discriminate between pooled PCR products from different templates. Primer A and B represent adapter sequences necessary in the GS-FLX sequencing process (sequences below)

^a Primer A: GCCTCCCTCGCGCCATCAG; primer B: GCCTTGCCAGCCCGCTCAG

^b Includes primer sequence

^c Blocking oligo inhibits the annealing of forward primers to little penguin DNA template

fluorescence of PicoGreen in a PicoFluor fluorometer (Turner Designs).

Pyrosequencing was carried out on six distinguishable pools of PCR products (metasamples):

- (1) Faecal DNA, feeding trial day 1 of collection, amplified with 16S SHORT primers
- (2) Faecal DNA, feeding trial day 2 of collection, amplified with 16S SHORT primers
- (3) Faecal DNA, feeding trial day 3 of collection, amplified with 16S SHORT primers
- (4) Fish tissue mix amplified with 16S SHORT primers
- (5) Faecal DNA, field collected, amplified with 16S SHORT primers
- (6) Faecal DNA, field collected, amplified with 16S LONG primers.

Each metasample was composed of equimolar mixes of constituent PCR products. PCR failed in one faecal DNA template from feeding trial day 1 and two from day 3, these were not included in the metasamples. The six metasamples were sent to Australian Genome Research Facility for amplicon sequencing using the Roche GS-FLX (454) platform. Sequencing data were obtained from four regions of a small PicoTitre Plate. Metasamples 1–3 were run in separate plate regions, metasamples 4–6 were pooled and run together. Amplicons within each metasample were labelled with a unique 3 base pair tag present on the forward and reverse primers (16S SHORT only) to allow post-sequencing bioinformatic sequence sorting.

Processing and analysis of 454 sequence data

The sequence sorting and clustering procedures were executed by purpose-written software available from the authors and described in Deagle et al. (2009). For 16S SHORT amplicons, only sequences containing exact matches to both the forward and reverse PCR primers were used in final analysis. For 16S LONG amplicons, only sequences containing exact matches to the forward PCR primers and with reads > 140 bp were analysed. Once sequences had passed through this preliminary screening they were sorted by similarity to pre-defined reference sequences (e.g. sequences from fish prey fed to penguins), or clustered by similarity to each other (see Deagle et al. 2009 for details). Representative sequences from clusters of the unknown sequences were used to search GenBank with BLASTn (Altschul et al. 1990). Final taxonomic classification of each sequence was based on the closest blast match, but several other factors were also considered. These included: the geographic distribution of species identified by the closest blast hit and the diversity of closely related species in southeastern Australia (including related species with no sequences in GenBank). Species

level classifications were made when reference sequences had >99% identity to the query and local congeneric species were present in the database (if they exist) and had lower sequence similarity. When these conditions were not met, final classifications were made at higher taxonomic levels based on available information.

Results

Overview of pyrosequencing data

The final sequence dataset produced from the six metasamples sent for GS-FLX pyrosequencing contained >9,000 DNA sequences (Tables 2, 3). A total of 4,216 sequences were recovered from faecal DNA samples collected from captive penguins, 1,352 from test tissue mix and 4,265 from faeces collected in the field. Penguin sequences made up between 10 and 28% of amplicons generated from faecal DNA template with the 16S SHORT primers, but 80% of the sequences generated with 16S LONG primers matched the penguin reference sequence. This disparity may be due to: changes in efficiency of the blocking oligo in combination with different reverse PCR primers, differential ability of the reverse primers to amplify penguin mtDNA, or may just reflect the differential degradation of DNA in faeces which results in more predator DNA relative to prey DNA at larger fragment sizes (see Deagle et al. 2006).

Captive feeding trials

Pilchard DNA sequences were the most common sequences recovered from faeces of captive penguins, representing between 61 and 71% of all prey sequences (Table 2; Fig. 1a). This follows expectations as pilchards made up 100% of the penguins' diet when they were initially brought into captivity and roughly half of their diet during the faecal collection period. Given our detection of dietary DNA from meals eaten before the penguins were captured (see below), the recovered pilchard sequences likely originated from pilchards fed throughout captivity (i.e. both from meals where they represented 100% and ~50% of the diet).

A total of 825 sequences were recovered from the fish test species fed in constant proportions in the tissue mix at the end of each trial. All three test species were detected in each daily sample, but the proportions of sequences differed considerably from respective mass proportions in the diet (Fig. 1b). Overall, tuna sequences made up 60% of these sequences (vs. 45% by mass in tissue mix) tommy ruff 6% (vs. 35% by mass in tissue mix) and whiting 34% (vs. 20% by mass in tissue mix). We also recovered several sequences that match additional fish species. These included

Table 2 Composition of the test diet fed in fixed proportions to captive little penguins and summary of DNA sequences recovered from penguin faeces and the fish tissue mix

	Test diet (TD)	Captive faeces 1		Captive faeces 2		Captive faeces 3		TD tissue mix	
	Mass TD (%)	No. seqs	TD (%)	No. seqs	TD (%)	No. seqs	TD (%)	No. seqs	(%)
Penguin	–	378	–	244	–	425	–	–	–
Pilchard	–	816	–	570	–	671	–	–	–
Tuna	45	176	67.7	199	64.4	121	47.3	968	71.6
Tommy ruff	35	31	11.9	5	1.6	15	5.9	300	22.2
Whiting	20	53	20.4	105	34.0	120	46.9	84	6.2
Other fish	–	78	–	51	–	158	–	–	–
Total		1532		1174		1510		1352	

Pilchards were fed in variable proportion (~50–100% of diet—see “Methods” for details). Other fish are remnants from food eaten before penguins were brought into captivity

sequences from small-mouth hardyhead (*Atherinason hepsetoides*; n = 145), jack mackerel (*Trachurus* sp.; n = 66), anchovy (*Engraulis australis*; n = 29), leatherjacket (Monacanthidae; n = 22), barracouta (*Thyrsites atun*; n = 11) and gurnard (*Neosebastes* sp.; n = 10). It is unlikely DNA from these additional fish came from the stomach contents of fish fed during trial (i.e. secondary ingestion) because these sequences were not detected in the tissue mix (see below). All of the additional fish are potential prey of little penguins in the wild (Cullen et al. 1992) and were probably consumed by the penguins before they were brought into captivity. This indicates that a DNA signal can persist in penguin faeces for at least 4 days after ingestion (this was the minimum time penguins were kept in captivity before the relevant faecal collections).

In the analysis of fish tissue fed to the penguins, the proportions of mtDNA recovered from the test species did not match their mass in the prepared mixture, but the relative ranking of fish components was maintained (i.e. tuna > tommy ruff > whiting; Table 2; Fig. 1b). These data indicate that the amount of mtDNA per gram of tissue varies between the fish species. However, correction factors derived from information on mtDNA density of the fish (using formula in Deagle and Tollit 2007) do not improve our estimates of diet composition from the penguin faeces. This is because we measured a relatively low mtDNA density in the tissue from whiting, but mtDNA from this species was recovered in relatively large amounts in the penguin faeces (mean corrected whiting values are inflated to 68%, vs. 20% by mass in tissue mix).

Field collected samples

Two separate amplicons from the field-collected little penguin faeces were sequenced (16S SHORT and 16S LONG). The 16S SHORT prey sequences (n = 2,029) matched 23 distinct fish (Table 3). Fewer 16S LONG prey sequences were recovered (n = 394) and the diversity of fish was

proportionally lower, but one additional species of fish and a single squid sequence was recovered (Table 3). In total, 25 distinct prey sequences were recovered and 18 of these could be identified to genus or species level; lack of reference sequence data was the primary limitation on level of taxonomic classification. Three fish (Anchovy, Barracouta and Pilchard) accounted for more than 80% of the sequences from each amplicon, and these species are generally the most common diet items based on stomach content analysis in the study sites (Cullen et al. 1992). Contaminating human DNA was detected but made up <0.5% of the data. DNA from shearwaters (which also nest in burrows in the region) was also detected, but was a very minor contaminant (1:1,800 relative to little penguin sequences).

Comparison between metasamples

Comparisons between datasets from various metasamples can provide some insight into the repeatability of the technical components of the dietary barcoding procedure. The three metasamples from faeces of captive penguins represent data from separate PCR amplifications using a single primer set and different template DNA from a group of penguins. These three metasample were processed in separate GS-FLX pyrosequencing reactions. The results show a high level of congruence between metasamples (mean pairwise Pearson’s r = 0.97; Fig. 2a). This suggests that the composition of DNA in the faeces of the penguins did not change appreciably over the 3 days. It also indicates that the amplicons from separate PCR amplifications (in this case pooled PCR products) can give consistent measurements of species composition, and likewise GS-FLX amplicon sequencing can provide precise (though not necessarily accurate) quantitative data.

It is also informative to compare the amplicon sequences recovered from the two wild penguin metasamples. In this case the metasamples represent PCR amplifications of common DNA templates with different primer sets. There

Table 3 Taxonomic assignment of mtDNA 16S sequences amplified from little penguin faeces collected at nesting burrows in Victoria, Australia

	Common name (scientific name)	16S SHORT		16S LONG	
		No. seq	% Prey	No. seq	% Prey
Osteichthyes	Anchovy (<i>Engraulis australis</i>)	1045	51.5	202	51.3
	Barracouta (<i>Thyrsites atun</i>)	442	21.8	45	11.4
	Pilchard (<i>Sardinops sagax</i>)	186	9.2	78	19.8
	Jack mackerel (<i>Trachurus</i> sp.)	134	6.6	17	4.3
	Warehou (<i>Serirolella</i> sp.)	53	2.6	7	1.8
	Blenny (Blenniidae)	51	2.5	9	2.3
	Blue-spotted goatfish (<i>Upeneichthys vlamingii</i>)	42	2.1	23	5.8
	Leatherjacket 1 (<i>Meuschenia</i> sp.)	16	0.8	–	–
	Unknown Teleostei 1	13	0.6	–	–
	Southern sea garfish (<i>Hyporhamphus melanochir</i>)	9	0.4	3	0.8
	Small-mouth hardyhead (<i>Atherinason hepsetoides</i>)	8	0.4	7	1.8
	Blue mackerel (<i>Scomber australasicus</i>)	7	0.3	–	–
	Leatherjacket 2 (<i>Acanthaluteres</i> sp.)	6	0.3	–	–
	Unknown Teleostei 2	5	0.2	–	–
	Unknown Teleostei 3	3	0.1	–	–
	Leatherjacket 3 (<i>Eubalichthys</i> sp.1)	2	0.1	–	–
	Leatherjacket 4 (<i>Eubalichthys</i> sp.2)	1	<0.1	–	–
	Damsel fish (Pomacentridae)	1	<0.1	–	–
	Gurnard perch (<i>Neosebastes</i> sp.)	1	<0.1	–	–
	Australian salmon (<i>Arripis</i> sp.)	1	<0.1	–	–
	Big-belly seahorse (<i>Hippocampus abdominalis</i>)	1	<0.1	–	–
	Goatfish (<i>Mullidae</i>)	1	<0.1	–	–
	Flathead (<i>Platycephalidae</i>)	1	<0.1	–	–
Australian Mado (<i>Atypichthys strigatus</i>)	–	–	2	0.5	
Cephalopod	Southern Calamari Squid (<i>Sepioteuthis australis</i>)	–	–	1	0.3
	Total number of prey sequences	2029		394	
Other	Little penguin (<i>Eudyptula minor</i>)	242		1556	
	Human (<i>Homo sapien</i>)	16		2	
	Shearwater (<i>Puffinus</i> sp.)	–		1	
	Unassigned sequences	24		1	
	Total number of sequences	2311		1954	

Results are from two barcoding markers (16S SHORT and 16S LONG). Data includes prey identity, number of corresponding sequences and percentage of total prey sequences obtained

is a reasonable level of congruence between the 16S SHORT and 16S LONG datasets (Pearson's $r = 0.94$; Fig. 2b). These primer sets share a common forward primer and amplify fragments of the same gene.

Discussion

It is becoming common for DNA-based identification methods to be applied in studies of wild animal diet. One of the outstanding issues in recent dietary DNA barcoding studies is the relationship between amounts of various food

items consumed and the quantitative data recovered from corresponding dietary samples (i.e. the relative number of sequences generated by high-throughput sequencing of DNA amplified from faeces or stomach contents). Discussion of the issues surrounding accuracy of biomass quantification has been included in several recent DNA diet papers (Deagle et al. 2009; Soyninen et al. 2009; Valentini et al. 2009a, b). In a paper on vole diet using pyrosequencing of plant DNA from stomach contents (Soyninen et al. 2009) the authors urge caution in the quantitative interpretation of their DNA barcoding results. However, they also conclude that the approach gives a relatively

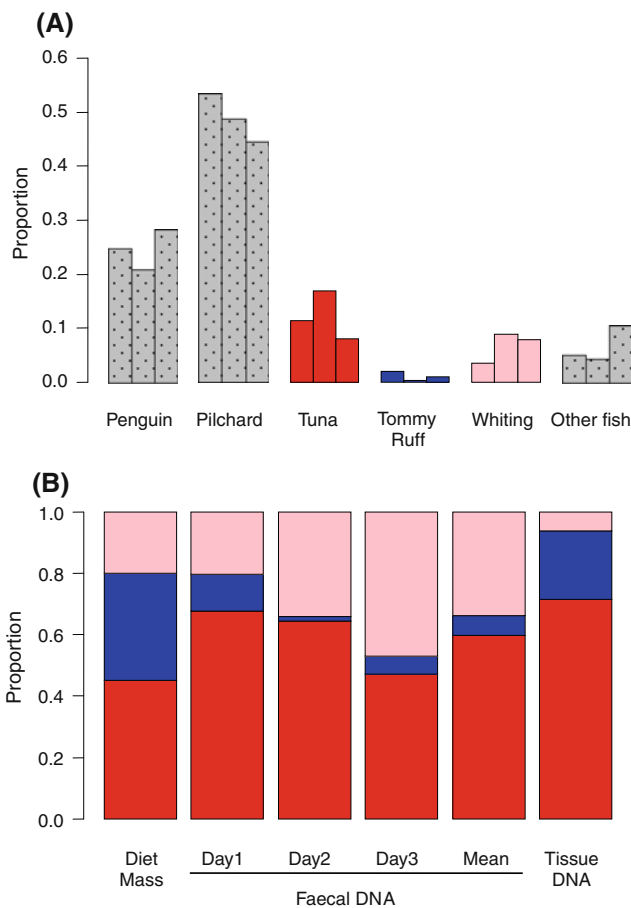


Fig. 1 Proportions of various sequences obtained from the captive feeding trial (16S SHORT mtDNA primer set). **a** Summary of all sequences recovered from three faecal metasamples collected on separate days of the trial. **b** Data from the three test species fed in constant proportions during the trial. Bars represent: mass proportions of the three fish in the diet; proportions of DNA recovered in penguin faeces (day 1–3 and mean); and proportions of DNA recovered from the blended tissue fed to the penguins

unbiased picture of food utilization of herbivores (Soininen et al. 2009). Similarly, in a study of fur seal diet by pyrosequencing prey DNA in faeces (Deagle et al. 2009) the authors outline reasons why a quantitative signature could be inaccurate, but still interpret their tempting data in a semi-quantitative fashion. The level of DNA signal recovered from a consumed food species could be influenced by many factors such as: the amount of DNA per gram of prey tissue, digestive processes, sample pooling, DNA purification, PCR amplification and DNA sequencing. Feeding trials could help reveal biases that may be introduced by some of these factors. If the most important biases can be defined, then methods could be improved, or correction factors might be applied, to recover more accurate estimates of relative biomass consumed.

The results from the current feeding trial produced a somewhat skewed quantitative signature. The dominant

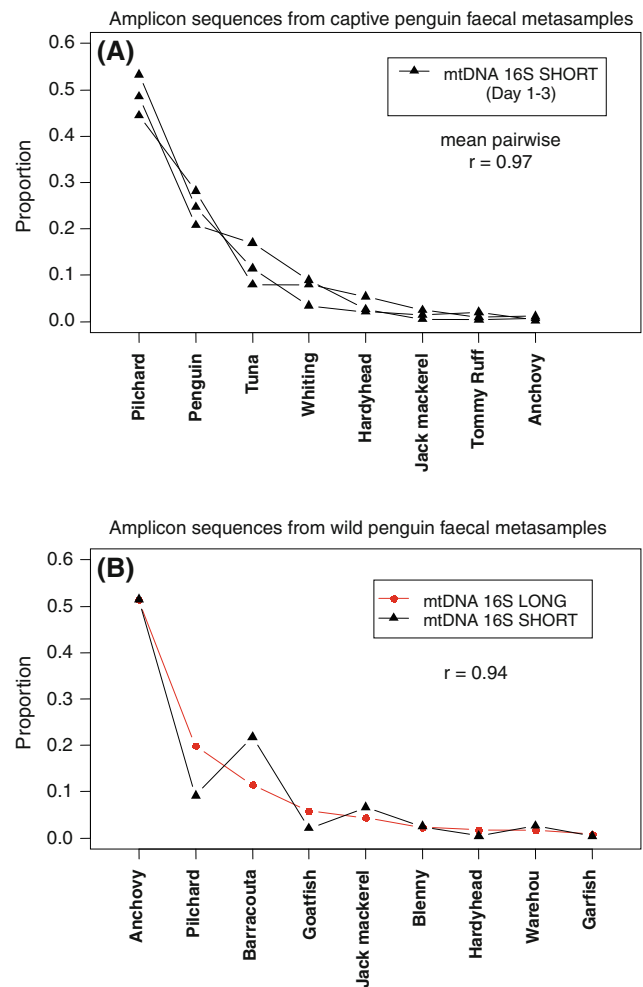


Fig. 2 Comparison of the proportion of sequence reads generated in separate metasamples. **a** Amplicon sequences from captive penguin faecal metasamples. Data are from separate PCR amplifications on different DNA template using a single primer set. **b** Amplicon sequences from wild penguin faecal metasamples. Data are from PCR amplifications of common DNA templates with different primer sets. Plots and correlation coefficients include sequences making up at least 0.5% of data. Penguin sequences were excluded from comparison between primer sets

food item in the penguins’ diet was pilchard (fed in variable quantities) and this was the most common amplicon recovered. However, our ‘test diet’ consisted of three other fish species which were blended and fed in constant proportion. The mtDNA proportions from these fish did not match proportions by mass in the diet. The reasons for this bias can be evaluated to some degree. First, consistency between data from different metasamples indicates that the technical steps of our analysis were reasonably precise. Very similar estimates of sample composition from the faecal samples collected on separate days in the feeding trial provide confidence that both the PCR and GS-FLX pyrosequencing provide repeatable quantitative measurements. This repeatability is likely due in part to the pooling

of several faecal samples before extraction and pooling of several PCR products in each metasample (Acinas et al. 2005). A comparison can also be made between data from the two PCR primer sets used to analyse the same DNA samples from penguin faeces collected in the field. These results are also congruent and suggest the absence of strong primer-specific biases in amplification or sequencing. While exact technical details will be experiment specific, these types of methodological issues are receiving attention in many fields of study (von Wintzingerode et al. 1997; Acinas et al. 2005; Huber et al. 2009; Porazinska et al. 2009) and this should allow diagnosis of problems and development of increasingly robust assays.

In DNA-based dietary studies, the most obvious biological factor to consider when correcting for biases in quantification is prey-specific differences in tissue DNA density. In the current study, the proportions of mtDNA sequences recovered from the fish tissue fed to the captive penguins did not match mass proportions of fish in the mixture, indicating differences in mtDNA density between these fish species. However, the biases in the tissue dataset were not reflected in the faeces of the captive penguins. This incongruence means that correction factors derived from fish tissue mtDNA content did not improve the quantitative estimates of diet derived from faecal mtDNA. Given these results, it appears that the most significant cause of bias in the captive feeding trials is differential survival of tissue during the process of digestion. In particular tommy ruff was significantly under-represented in the amplicons from faeces, even though mtDNA from this fish was recovered at a reasonable level in the tissue mix. The relative digestibility of the different fish may not reflect their true digestibility since our experiment made use of blended fish tissue, rather than the whole fish that would be consumed in the wild. Regardless, our overall findings are consistent with results from a previous study that used quantitative real-time PCR to analyse prey mtDNA in the faeces of captive sea lions (Deagle and Tollit 2007). In the sea lion study, differences in recovery of mtDNA from Pacific herring and sockeye salmon were observed and attributed to differential digestion (Deagle and Tollit 2007). Clearly data from these feeding trials offer only preliminary insights and cannot be extended to other predators and prey species. However, these data do highlight that direct relationships between the number of prey DNA molecules identified in dietary studies and the relative biomass of different prey items consumed are unlikely to exist even if technical biases are minimal.

The measured DNA recovery rates in captive feeding trials could be used to calculate correction factors accounting for differential digestion, but this is not a feasible option for most studies. Even if feeding trials could be carried out on species of interest, prey digestibility and

consumer digestive processes will not only differ between species, but are likely to vary among individuals due to factors such as physiological state and age. Given this, it is doubtful that correction factors derived from feeding trials would be generally useful. An alternative approach to account for differential digestion of food species might be possible if tissue digestibility is reflected in the level of DNA degradation. If this is the case, then the level of DNA degradation could be determined for food species in dietary samples (Deagle et al. 2006) and correction factors applied (i.e. highly digested DNA would be given more weight than less digested DNA). This type of analysis has been used to correct for DNA loss due to digestion in barcoding markers recovered from the gut of copepods (Troedsson et al. 2009).

While our results from captive penguins, and those in the previous study on captive sea lions, show that recovered DNA are not absolutely representative of the true diet, they do broadly reflect the relative contributions of fish prey in the fed diets. Given that all methods of diet analysis are affected by biases, these uncorrected quantitative DNA results may still be as accurate as those obtainable with alternate techniques (Gales 1987; Pierce and Boyle 1991; Barrett et al. 2007).

Dietary barcoding analysis can also clearly provide useful information on identification of prey species and dietary diversity. In many cases, it is the only viable method for generating species-level information on the prey consumed. Measurements of the stable isotope ratios of carbon ($\delta^{13}\text{C}$) and particularly nitrogen ($\delta^{15}\text{N}$) are often used to provide trophic-level information about consumers in marine food webs, but they cannot identify prey at species level especially when potential prey have similar isotope signatures (Inger and Bearhop 2008). For seabirds, stomach flushing and morphological identification of prey remains can provide species-level information, but the invasive and laborious nature of the procedure usually precludes the approach in contemporary diet studies. A previous study of little penguin diet in southeast Australia recovered stomach contents from 1,669 birds over a 3 year period (1985–1988) and identified roughly 34 fish prey (Cullen et al. 1992). Our data from faeces of wild penguins identified 24 fish prey in 100 faecal samples collected over a 2 month period. While it is difficult to compare these studies directly, the relatively high fish diversity in our DNA pilot study indicates that the range of the penguins' diet could be captured through dietary barcoding. The primacy of anchovy in both the stomach content and DNA datasets, and the overlap in the other fish prey identified in the two studies, further indicate that dietary barcoding would be suitable for future less-invasive dietary monitoring. Our results from the analysis of captive penguins' faeces show that DNA was recovered from prey consumed before the birds were brought into captivity. This means prey DNA is detectable for at least 4 days after

ingestion, considerably longer than the 48 h detection period for prey DNA observed in pinnipeds (Deagle et al. 2005; Casper et al. 2007). This extended period of detection means that DNA from prey captured throughout most little penguin foraging trips during breeding will be present in their faeces when they return to land (Chiaradia and Nisbet 2006).

Overall, the current study indicates that DNA barcoding is a promising approach for documenting the diet of seabirds. The challenge now will be moving from the exemplar studies that have been carried out so far to more broad application of the methods. With current high-throughput sequencing technology, it is possible to carry out detailed analysis of the large number of samples typically examined in ecological studies. However, it is important to point out that these PCR-based analyses are limited absolutely by the specificity of PCR primers employed. The continued development of reliable, conserved primer sets amplifying DNA from a range of food species is critical. Ideally multiple primer sets with overlapping specificity should be employed, such as in our analysis of the field collected little penguin faeces. This allows comparisons of results generated by different PCR primer sets, providing a useful control. Of course, the development of complementary DNA barcoding databases to allow identification of recovered sequences is also important. If the limitations of DNA-based analysis are kept in mind, and diet estimates are given realistically wide confidence intervals, then semi-quantitative interpretations of dietary barcoding data can be justified. This is particularly true if studies are comparative (e.g. documenting spatial or temporal variation in diet) and a similar range of dietary species are being consumed within each sample. Further captive feeding trials are likely to be the best approach for refining this potentially powerful method of diet analysis. This will provide further confidence in field-collected data and allow the method to be applied to address significant ecological questions.

Acknowledgments We thank M. Healy and T. Murray for help keeping the temporarily captive penguins happy, and L. Renwick, E. Alamo and P. Wasiaik for collecting samples in the field. We are grateful for funding from Phillip Island Nature Park and support from the Australian Antarctic Division. BED was supported by an NSERC Postdoctoral Fellowship during the later stages of this study. The project was approved by the Phillip Island Nature Park Animal Experimentation Ethics Committee and by the Department of Sustainability and Environment of Victoria, Australia.

References

- Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* 71:8966–8969
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Barrett RT, Camphuysen K, Anker-Nilssen T et al (2007) Diet studies of seabirds: a review and recommendations. *ICES J Mar Sci* 64:1675–1691
- Boersma PD, Rebstock GA, Frere E, Moore SE (2009) Following the fish: penguins and productivity in the South Atlantic. *Ecol Monogr* 79:59–76
- Boyd IL, Murray AWA (2001) Monitoring a marine ecosystem using responses of upper trophic level predators. *J Anim Ecol* 70:747–760
- Bradley BJ, Stiller M, Doran-Sheehy DM et al (2007) Plant DNA sequences from feces: potential means for assessing diets of wild primates. *Am J Primatol* 69:699–705
- Braley M, Goldsworthy SD, Page B, Steer M, Austin JJ (2010) Assessing morphological and DNA-based diet analysis techniques in a generalist predator, the arrow squid *Nototodarus gouldi*. *Mol Ecol Res* 10:466–474
- Brooke MD (2004) The food consumption of the world's seabirds. *Proc R Soc Lond B Biol Sci* 271:S246–S248
- Casper RM, Jarman SN, Deagle BE, Gales NJ, Hindell MA (2007) Detecting prey from DNA in predator scats: a comparison with morphological analysis, using *Arctocephalus* seals fed a known diet. *J Exp Mar Biol Ecol* 347:144–154
- Chiaradia AF, Kerry KR (1999) Daily nest attendance and breeding performance in the little penguin *Eudyptula minor* at Phillip Island, Australia. *Mar Ornithol* 27:13–20
- Chiaradia A, Nisbet ICT (2006) Plasticity in parental provisioning and chick growth in little penguins *Eudyptula minor* in years of high and low breeding success. *Ardea* 94:257–270
- Chiaradia A, Forero MG, Hobson KA, Cullen JM (2010) Changes in diet and trophic position of a top predator ten years after a mass mortality of a key prey. *ICES J Mar Sci*. doi:10.1093/icesjms/fsq067
- Clare EL, Fraser EE, Braid HE, Fenton MB, Hebert PDN (2009) Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Mol Ecol* 18:2532–2542
- Cullen JM, Montague TL, Hull C (1992) Food of little penguins *Eudyptula minor* in Victoria—comparison of 3 localities between 1985 and 1988. *Emu* 91:318–341
- Deagle BE, Tollit DJ (2007) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conserv Genet* 8:743–747
- Deagle BE, Tollit DJ, Jarman SN et al (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Mol Ecol* 14:1831–1842
- Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. *Front Zool* 3:e11
- Deagle BE, Gales NJ, Evans K et al (2007) Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS ONE* 2:e831
- Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol Ecol* 18:2022–2038
- Dunnshea G (2009) DNA-based diet analysis for any predator. *PLoS ONE* 4:e5252
- Forney LJ, Zhou X, Brown CJ (2004) Molecular microbial ecology: land of the one-eyed king. *Curr Opin Microbiol* 7:210–220
- Gales RP (1987) Validation of the stomach-flushing technique for obtaining stomach contents of penguins. *Ibis* 129:335–343
- Hedd A, Bertram DF, Ryder JL, Jones IL (2006) Effects of interdecadal climate variability on marine trophic interactions: rhinoceros auklets and their fish prey. *Mar Ecol Prog Ser* 309:263–278

- Hofreiter M, Poinar HN, Spaulding WG et al (2000) A molecular analysis of ground sloth diet through the last glaciation. *Mol Ecol* 9:1975–1984
- Hoskins AJ, Dann P, Ropert-Coudert Y et al (2008) Foraging behaviour and habitat selection of the little penguin *Eudyptula minor* during early chick rearing in Bass Strait, Australia. *Mar Ecol Prog Ser* 366:293–303
- Huber JA, Morrison HG, Huse SM et al (2009) Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* 11:1292–1302
- Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Resour* 8:3–17
- Inger R, Bearhop S (2008) Applications of stable isotope analyses to avian ecology. *Ibis* 150:447–461
- Jarman SN, Gales NJ, Tierney M, Gill PC, Elliott NG (2002) A DNA-based method for identification of krill species and its application to analysing the diet of marine vertebrate predators. *Mol Ecol* 11:2679–2690
- King RA, Read DS, Traugott M, Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Mol Ecol* 17:947–963
- Passmore AJ, Jarman SN, Swadling KM et al (2006) DNA as a dietary biomarker in Antarctic Krill, *Euphausia superba*. *Mar Biotechnol* 8:686–696
- Pierce GJ, Boyle PR (1991) A review of methods for diet analysis in piscivorous marine mammals. *Oceanogr Mar Biol* 29:409–486
- Porazinska DL, Giblin-Davis RM, Faller L et al (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Mol Ecol Resour* 9:1439–1450
- Ropert-Coudert Y, Kato A, Chiaradia A (2009) Impact of small-scale environmental perturbations on local marine food resources: a case study of a predator, the little penguin. *Proc R Soc B Biol Sci* 276:4105–4109
- Soininen E, Valentini A, Coissac E et al (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front Zool* 6:e16
- Sydeman WJ, Hobson KA, Pyle P, McLaren EB (1997) Trophic relationships among seabirds in central California: combined stable isotope and conventional dietary approach. *Condor* 99:327–336
- Troedsson C, Simonelli P, Nägele V, Nejtgaard J, Frischer M (2009) Quantification of copepod gut content by differential length amplification quantitative PCR (dla-qPCR). *Mar Biol* 156:253–259
- Valentini A, Miquel C, Nawaz MA et al (2009a) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Mol Ecol Resour* 9:51–60
- Valentini A, Pompanon F, Taberlet P (2009b) DNA barcoding for ecologists. *Trends Ecol Evol* 24:110–117
- Vences M, Thomas M, Van Der Meijden A, Chiari Y, Vieites DR (2005) Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front Zool* 2:e5
- Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. *Front Zool* 5:e12
- von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21:213–229
- Weber DC, Lundgren JG (2009) Detection of predation using qPCR: effect of prey quantity, elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA. *J Insect Sci* 9:e41