

Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates

Julie C. McInnes^{1,2*}, Rachael Alderman³, Bruce E. Deagle², Mary-Anne Lea¹, Ben Raymond^{1,2} and Simon N. Jarman^{2,4}

¹Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 129, Hobart, TAS 7001, Australia; ²Australian Antarctic Division, 203 Channel Highway, Kingston, TAS 7050, Australia; ³Department of Primary Industries, Parks, Water and Environment, 134 Macquarie Street, Hobart, TAS 7000, Australia; and ⁴CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão 4485-661, Portugal

Summary

1. DNA metabarcoding of food in animal scats provides a non-invasive dietary analysis method for vertebrates. A variety of molecular approaches can be used to recover dietary DNA from scats; however, many of these also recover non-food DNA. Blocking primers can be used to inhibit amplification of some non-target DNA, but this may not always be feasible, especially when multiple distinct non-target groups are present.

2. We have developed scat collection protocols to optimise the detection of food DNA in vertebrate scat samples. Using shy albatross *Thalassarche cauta* as a case study, we investigated how DNA amplification success and the proportion of food DNA detected are influenced by both environmental and physiological parameters. We show that both the amount and type of non-target DNA vary with sample freshness, the collection substrate, fasting period and developmental stage of the consumer.

3. Fresh scat samples yielded the highest proportion of food sequences. Collecting scats from dirt substrates reduced the proportion of food DNA and increased the proportion of contaminating DNA. Food DNA detection rates changed throughout the albatross breeding season and related to the time since feeding and the developmental stage of the animal. Fasting albatross produced scats dominated by parasite amplicons in universal PCR analysis, with little food DNA recovered. Samples from very young animals also produced reduced food DNA proportions.

4. Based on our observations, we recommend the following procedures for field scat collections to ensure high-quality samples for dietary DNA metabarcoding studies. Ideally, (i) collect fresh scats; (ii) from surfaces with minimal contamination (e.g. rock or ice); (iii) collect scats from animals with minimum time since feeding and avoid fasting animals; (iv) avoid young animals that are not feeding directly (e.g. not weaned or fledged) or target larger/older individuals. The optimised field sampling protocols that we describe will improve the quality of dietary data from vertebrates by focusing on samples most likely to contain food DNA. They will also help minimise contamination issues from non-target DNA and provide standardised field methods in this rapidly expanding area of research.

Key-words: albatross, cestoda, faeces, food, molecular, non-invasive, parasites, prey, procellariiformes, seabird

Introduction

Scat samples provide an important source of DNA that can be utilised in a wide range of molecular ecology studies (e.g. Davison *et al.* 2002; Prugh *et al.* 2005). Food DNA present in scats provides a non-invasive and increasingly popular tool for studying vertebrate diet and can be applied to both predators and herbivores (e.g. Deagle, Kirkwood & Jarman 2009; Zeale *et al.* 2011; Bowser, Diamond & Addison 2013; Kartzinel *et al.* 2015). Dietary DNA metabarcoding uses high-throughput sequencing of small, highly variable DNA regions that survive digestion to identify food species (Pompanon *et al.* 2012).

This may involve identification of a particular food species using species-specific markers (Jarman & Wilson 2004); food within a taxonomic group using group-specific markers (Jarman, Deagle & Gales 2004; Murray *et al.* 2011; Zeale *et al.* 2011); identification of all food taxa using universal metazoan markers (O'Rourke *et al.* 2012; Jarman *et al.* 2013); or a combination of these approaches (Deagle, Kirkwood & Jarman 2009; Bowser, Diamond & Addison 2013). However, characterising the entire diet requires 'universal' markers that are capable of amplifying DNA from any food species (King *et al.* 2008; Jarman *et al.* 2013).

Universal metazoan polymerase chain reaction (PCR) primers amplify from all eukaryotic DNA, but will inevitably also amplify unwanted DNA from non-food items (Deagle,

*Correspondence author. E-mail: julie.mcinnnes@utas.edu.au

Kirkwood & Jarman 2009; O'Rourke *et al.* 2012). Non-target DNA within the scat may originate from the animal being sampled, its parasites, gut flora or contamination from external organisms such as insects and vegetation. These sources of DNA can dominate the sequences amplified from a sample, making detection of DNA from food items less effective. Sample sizes must consequently be increased to address the underlying questions of a study, increasing processing costs. In some cases, non-target DNA amplification can be reduced using a blocking primer to suppress amplification of specific DNA types, such as DNA of the defecating animal (O'Rourke, Lavery & Jeffs 2012). However, development of blocking primers is challenging and food sequences may be inadvertently blocked with this approach. The use of blocking primers becomes more complex when there are multiple non-target DNA groups present. Improved sampling procedures are another approach for increasing the proportion of food DNA identified in a scat.

Selective scat sampling to improve DNA amplification success in genotyping studies has been investigated (Lucchini *et al.* 2002; Piggott 2004; Panasci *et al.* 2011; Vynne *et al.* 2012), but studies to optimise scat collections for DNA dietary analysis are rare (Oehm *et al.* 2011). Genotyping studies have investigated how the age of scats (Farrell, Roman & Sunquist 2000; Lucchini *et al.* 2002; Piggott 2004; Panasci *et al.* 2011; Vynne *et al.* 2012), habitat type (Vynne *et al.* 2012) and season (Lucchini *et al.* 2002; Piggott 2004) affect DNA detection and genotyping accuracy. Fresh scats collected in dry and cool conditions typically provided the highest amplification success and lowest genotyping error rate. However, the time since an animal defecated is seldom known and proxies for scat age are often required. For example, in maned wolf *Chrysocyon brachyurus* scats, higher moisture content and odour were found to be positively correlated with amplification success (Vynne *et al.* 2012). Similarly in brush-tailed rock-wallaby scats *Petrogale penicillata*, colour, consistency and odour correlated well with DNA amplification success (Piggott 2004).

Only one dietary DNA study has examined how field conditions can influence the detection of food DNA. In carrion crow *Corvus corone corone* scats, exposure to sunlight and rain over a 5-day period caused significantly lower amplification success of food DNA (Oehm *et al.* 2011). This was exacerbated by dirt, which may increase the degradation of extracellular DNA (Levy-Booth *et al.* 2007). This study used species-specific markers, which do not amplify non-food DNA. There are currently no studies that investigate whether targeted sample collections improve the detection of food DNA by universal metazoan markers.

We used shy albatross *Thalassarche cauta* as a model to develop optimised field protocols for dietary DNA metabarcoding of scats. Albatross are a good example as they follow predictable behavioural patterns, where they return to the colony after feeding and fast on the nest during incubation. This makes scat samples accessible and tests of fasting effects possible. Albatross are known to eat a diverse range of food items, including jellyfish, cephalopods, fish and carrion (Cherel & Klages 1998). Universal metazoan PCR primer sets which amplify from all potential prey groups are therefore needed to

screen for all food items. Albatross colonies present far from ideal laboratory conditions. Colonies are typically exposed to extremes of weather, with little or no vegetation cover. Sample degradation by UV and rain is likely to reduce PCR amplification success of exposed scats (Oehm *et al.* 2011). Contamination from non-food DNA, such as insects, parasites and fungi, will also reduce the proportion of food DNA detected. Colonies are often remote and expensive to access, on trips that are generally short and/or infrequent, so effective scat collection is imperative.

The optimised field protocols that we developed increase the detection of food DNA by considering the effect of sample freshness; the substrate it was collected from; the bird's breeding and developmental stage; and fasting time. The effects that these factors have on the detection of food DNA are significant enough to be an important consideration when designing dietary DNA studies of vertebrates.

Materials and methods

CASE STUDY SPECIES

Shy albatross lay one egg from early September to early October. The egg is incubated for 10 weeks (incubation stage), and the hatched chicks are brooded for 3–4 weeks (brood stage). During these two breeding stages, parents alternate nest attendance and foraging trips. After brooding, chicks are left unattended while both parents forage independently at sea to complete chick rearing (chick-rearing stage; Hedd & Gales 2005). During incubation, foraging trips may last from 1 to 10 days, with an average of 3 days (Hedd, Gales & Brothers 2001); therefore, an incubating bird could be fasting for this period or longer. Foraging trip durations during the brood stage are short at around 1 day and increase slightly during chick rearing to 2–3 days (Brothers *et al.* 1998; Hedd & Gales 2005).

FIELD METHODOLOGY

Shy albatross scat samples were collected at Albatross Island, Tasmania, Australia (40°23'S, 144°39'E). Scat samples were collected during the breeding period over two seasons: 2013/2014 austral summer, chick rearing (late March) only; and 2014/2015 austral summer: incubation (late September), brood stage (mid December) and chick rearing (late March). Samples were collected during the daytime from albatross observed defecating. A small fragment of the non-uric acid portion of the scat (dark part) was collected using tweezers or a plastic straw. The sample was stored in 80% ethanol and shaken on collection to mix with the ethanol. The only time fresh scats were not collected was when sample freshness was investigated.

Sample freshness

To determine the effect of sample freshness on DNA amplification rates and the proportion of food DNA detected, scats were collected during the chick-rearing period in 2013/2014 and 2014/2015. The amount of time a scat had been present was unknown when a scat was found. Consequently, we wanted to provide a proxy measure for freshness to allow selection of higher quality dietary material. To test this, scat samples were categorised as follows: (i) 'Fresh' when the bird was seen defecating, (ii) 'Recent' when the scat was still wet but the bird was

not seen defecating (there was often a skin forming on these scats) or (iii) 'Dry' when scats were old and had no apparent moisture.

Substrate type

The dominant substrate from which the scat was collected was recorded for all fresh scats collected during chick rearing. Substrate categories included the following: dirt, rock and vegetation.

Breeding stage

To determine whether collecting at different stages of breeding affected the results, we randomly collected from birds in the colony that we saw defecating during incubation, brood guard and chick rearing of the 2014/2015 breeding season.

Developmental stage

When known, the breeding cohort of the bird was recorded as either 'breeder' a bird on an active nest or seen feeding a chick; 'non-breeder' a bird at an empty nest; or 'chick' which could have been a brooded chick <2 weeks old, or a pre-fledged chick c. 3–5 months old.

Fasting

To test the effect that fasting had on dietary results, additional scats were collected during incubation. Two study sites within the colony were set up, each containing c. 100 nests. Each bird was marked on the chest with a small dot of non-toxic stock-marker, with a different colour used to identify their partner to monitor the amount of time a bird had been incubating. Nests were numbered and checked daily at each site and the bird incubating recorded. When birds were observed defecating, the scat sample was collected and the nest number and bird colour was recorded. The incubation time was categorised as <1 day, 1–2 days and >2 days.

MOLECULAR METHODOLOGY

Sample storage and extraction

Samples were stored at 5–10 °C for 1 week while in the field, and then, –20 °C until DNA was extracted. DNA was extracted within 2 weeks of collection using a Promega 'Maxwell 16' instrument and a Maxwell® 16 Tissue DNA Purification Kit (Madison, WI, USA). Samples were vortexed prior to extraction, and c. 30 mg of each sample was used. The quantity was consistent across extractions, which were all performed by the same person. PCR inhibitor concentrations were reduced in the DNA by mixing this subsample in 250 µL of STAR buffer (Roche Diagnostics, Basel, Switzerland) prior to extraction.

Table 1. Oligonucleotides used in this study

PCR round	Primer name	Primer sequence (5'–3')
1	18s_SSU3_F	<u>T</u> CGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTGTGATGCCCTTAGATG
1	18s_SSU3_R	<u>G</u> TCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTGTGTACAAAGGGCAGGG
2	SSU3_Tag_F1	AATGATACGGCGACCAACGAGATCTACACAGTT C GGACTTCGTCGGCAGCGTC
2	SSU3_Tag_R1	CAAGCAGAAGACGGCATACGAGATAGCTTAGGCTGTCTCGTGGGCTCGG

Underlined bases in PCR Round 1 are the Miseq tag primer. Bolded bases in PCR Round 2 are an example of the unique tags attached to each sample. A full list can be found in Appendix S1.

PCR amplification and amplicon sequencing

A PCR primer set for amplifying c. 170 bp of the V7 region of the nuclear small subunit ribosomal DNA gene (18 s; Hadziavdic *et al.* 2014) was designed manually on an alignment of the region that incorporated representatives from all major animal lineages. A two-stage PCR process was used to enable amplification of the DNA region and attachment of unique 'tag' sequences to each sample which allows amplified samples to be pooled (Binladen *et al.* 2007).

Stage one PCRs (10 µL) were performed with 5 µL 2× Phusion HF (NEB), 1 µL 100× bovine serum albumin (NEB, Ipswich, MA, USA), 0.1 µL 5 µM of each 18s_SSU amplification primer (Table 1), 0.5 µL of Evagreen, 2 µL faecal DNA and 1.3 µL of water. Thermal cycling conditions were 98 °C, for 2 min; followed by 35 cycles of 98 °C for 5 s, 67 °C for 20 s, 72 °C for 20 s, with an extension of 72 °C for 1 min. Each sample was run in triplicate on a LightCycler 480 (Roche Diagnostics). A negative control containing no template DNA and positive control containing fish DNA were included in each PCR amplification run. If either the negative amplified or the positive failed to amplify, the PCR was rerun. Samples from each experiment were split among different PCR runs to avoid run-specific biases.

If ≥2 replicates of each sample had a 'crossing threshold' (ct) score <30, they were combined to reduce biases produced by amplification from low template concentration samples (Murray, Coghlan & Bunce 2015). Pooled samples were diluted 1 : 10 for the second stage PCR. A unique tag was attached to each sample (Table 1) in 10 µL PCRs with 5 µL 2× Phusion HF (NEB), 1 µL 100× bovine serum albumin (NEB), 1 µL of 1 µM of each tag primer (Appendix S1, Supporting Information) and 2 µL of diluted PCR product from stage one. Thermal cycling conditions were 98 °C, for 2 min; followed by 10 cycles of 98 °C for 5 s, 55 °C for 20 s, 72 °C for 20 s, with an extension of 72 °C for 1 min. Four microlitres of PCR product from each sample ($n = 511$) and the negative controls were pooled and purified from unincorporated reaction components by washing utilising reversible binding to Agencourt Ampure (Beckman Coulter, Brea, CA, USA) magnetic beads, with 1.8 µL of Ampure per microlitre of DNA product. Sequencing of PCR products was performed with a MiSeq genome sequencer, using the MiSeq reagent kit V2 (Illumina, San Diego, CA, USA) (300 cycles) with paired-end reads.

Bioinformatics

Amplicon pools were demultiplexed based on unique 10-bp Multiplex IDentifiers on the MiSeq and fastq files processed using USEARCH v8.0.1623 (Edgar 2010). Reads R1 and R2 from the paired-end sequencing were merged using the fastq_mergepairs function, retaining only merged reads flanked by exact matches to the 18s_SSU primers and primer sequences were trimmed. Reads from all samples were pooled and dereplicated, then clustered into broad operational taxonomic units (OTUs) using the cluster_otus command (-otu_radius_pct = 10). Potentially chimeric reads are discarded during this

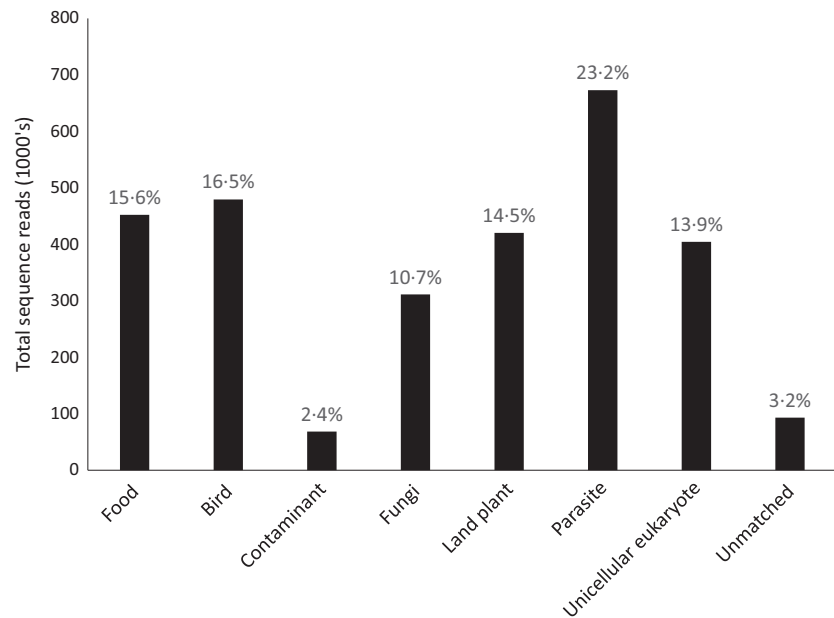


Fig. 1. Total sequence reads obtained and categorised using the SILVA SSU data base. Contaminants included insects (31 628 reads), ectoparasites (31 578 reads) and human DNA from handling (5168 reads).

step. Reads for each sample were assigned to these OTUs (usearch_global -id 0.9) and a summary table generated using a custom R script. Each OTU was assigned to a taxon by BLAST against a local data base derived from the SILVA SSU data base release 118 (Quast *et al.* 2013) with a 0.95 similarity used as a cut-off for identification. OTUs were categorised into seven groups based on their assumed origin: food, bird, parasite, fungi, plant, contaminant and unicellular (Appendix S2; Jarman *et al.* 2013). The contamination category included human, insect and ectoparasite sequences. Any sequences that did not match the Silva data base were excluded from analysis (3.2% of the total). Although some species of albatross are known to eat birds, this has rarely been recorded in shy albatross (Hedd & Gales 2001); therefore, in this study, the bird category represented DNA belonging to the albatross.

STATISTICAL ANALYSIS

We assessed whether DNA amplification success was affected by the specific variables (sample freshness, substrate, breeding stage, cohort and fasting length). Amplification was deemed successful if the total number of DNA sequences was >500 for a sample. We then examined whether there was a significant difference in the proportion of food DNA detected for each of the variables. Generalised Linear Models (GLMs) were used to test the difference in amplification success, and quasibinomial GLMs (to account for overdispersion in the data) were used to test differences in the proportion of food DNA detected (McCullagh & Nelder 1989). Analysis of deviance (with chi-squared test) was used to test for significance of predictor terms, with *post hoc* multiple comparisons by Tukey's method. Analyses were carried out using the R 'stats' package (R Core Team 2013), with multiple comparisons using the package 'MULTCOMP' (Hothorn, Bretz & Westfall 2008) and plots created using the package 'GGPLOT2' (Wickham 2009).

Results

DNA was extracted from 598 scat samples, with 511 of these producing ct values <30, with 458 (89%) producing >500 DNA sequence reads. A total of 2.9 million sequence reads were obtained from the single sequencing run, which included 452 305 (15.6%) food sequences (Fig. 1).

SAMPLE FRESHNESS

The freshness of scat samples significantly affected the DNA amplification success ($\chi^2_{2,254} = 7.61$, $P = 0.02$), with fresh scats amplifying better than recent scats, but not better than dry scats (Table 2). Sample freshness also significantly affected the proportion of food DNA in the samples, ($\chi^2_{2,192} = 31.808$, $P = 0.02$), with fresh scats containing a greater proportion of food DNA than dry scats, but not significantly more than recent scats (Table 2, Fig. 2). Fungi DNA proportions were higher for dry scats than either fresh or recent (Fig. 3).

SUBSTRATE TYPE

Only a small number of scats were collected from vegetation; therefore, substrate comparisons were only analysed using the two most common surfaces: dirt and rock. The substrate did not significantly affect amplification success ($\chi^2_{1,194} = 0.001$, $P = 0.97$), but did significantly affect the proportion of food DNA detected ($\chi^2_{1,148} = 14.805$, $P = 0.04$). Scats obtained from rock contained a higher proportion of food DNA than those obtained from dirt (Table 2, Fig. 2), which contained a higher proportion of unicellular DNA (Fig. 3).

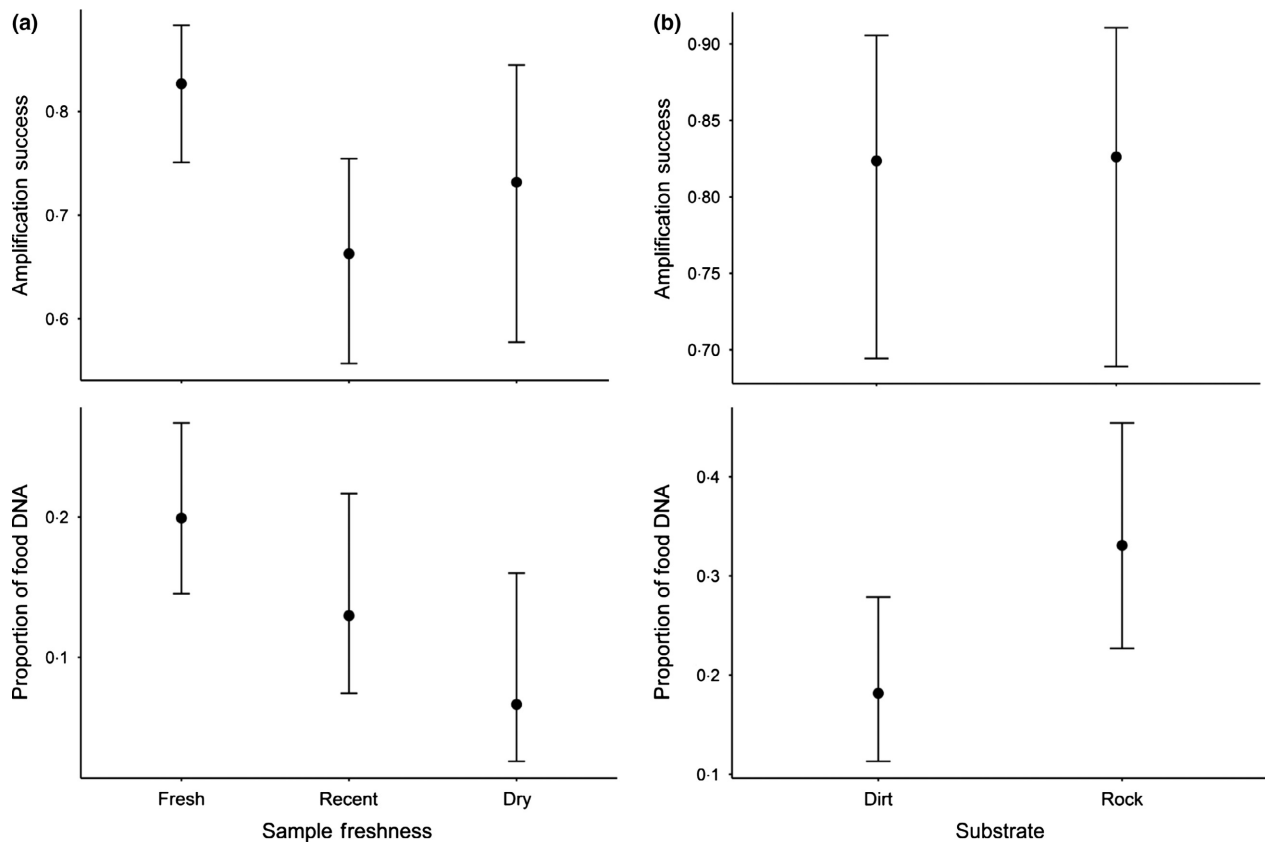
BREEDING STAGE

There was a significant difference observed in the DNA amplification success between breeding stages ($\chi^2_{2,308} = 7.988$, $P = 0.02$), with scats collected during the brood stage having lower amplification success (Table 2, Fig. 4). The breeding stage greatly affected the proportion of food DNA detected ($\chi^2_{2,237} = 115.863$, $P < 0.001$), with scats collected randomly during incubation producing significantly lower proportions of food DNA than scats from brood or chick-rearing stages (Table 2, Fig. 4). Scats collected during incubation were dominated by parasites (98% cestoda; Fig. 5).

Table 2. Generalised Linear Model (GLM) outputs for comparisons of DNA amplification success and the proportion of food DNA detected for sample age, substrate, breeding stage, developmental stage and fasting time

		Scats obtained	Scats with DNA amplified	DNA Amplification			Proportion of food DNA		
				Amplification success	Estimated values	SE	Estimated values	SE	Fitted
Sample freshness	Fresh	127	105	82.7	1.563*	0.234	-1.391*	0.193	0.20
	Recent	86	57	66.2	-0.887*	0.327	-0.513	0.369	0.13
	Dry	41	30	73.2	-0.560	0.423	-1.253*	0.536	0.07
Substrate	Dirt	90	70	77.8	1.540	0.367	-1.505*	0.278	0.18
	Rock	104	78	75.0	0.017	0.535	0.800*	0.382	0.33
Breeding stage	Incubation	79	69	87.3	1.931*	0.338	-2.768 ⁺⁺	0.346	0.59
	Brood	166	119	71.7	-1.002*	0.380	1.893 ⁺	0.388	0.29
	Chick Rearing	63	49	77.8	-0.678	0.454	1.755 [#]	0.423	0.27
Incubation cohort	Breeder	50	44	88.0	1.992	0.435	-3.439	0.649	0.03
	Non-Breeder	29	25	86.2	-0.159	0.692	1.306	0.803	0.11
Brood cohort	Breeder	60	49	81.7	1.494*	0.333	0.180 ⁺⁺	0.230	0.54
	Non-Breeder	40	31	77.5	-0.257	0.505	-1.659 ⁺	0.386	0.19
Incubation time	Chick	66	39	59.1	-1.126*	0.417	-2.117 [#]	0.429	0.13
	Random	79	69	87.3	1.932	0.338	-2.769 ⁺	0.318	0.06
	<1 day	52	41	78.8	-0.616	0.479	1.639 ⁺	0.415	0.24
	1-2 days	18	13	72.2	-0.976	0.626	-1.928	2.108	0.01
	>2 days	29	24	82.7	-0.363	0.597	-1.087	1.174	0.02

DNA amplification was analysed using binomial GLMs, the proportion of food DNA using quasibinomial GLMs. Superscript symbols indicate significantly different values (Tukey's multiple comparison test) * $P < 0.05$, ⁺⁺ $P < 0.001$.

**Fig. 2.** Generalised Linear Model fitted plots for: (a) sample freshness (fresh, recent or dry) and (b) substrate (dirt or rock). All scat samples collected during chick rearing, with only fresh scats included in the analysis of substrate. Points represent means, and bars show 95% confidence intervals.

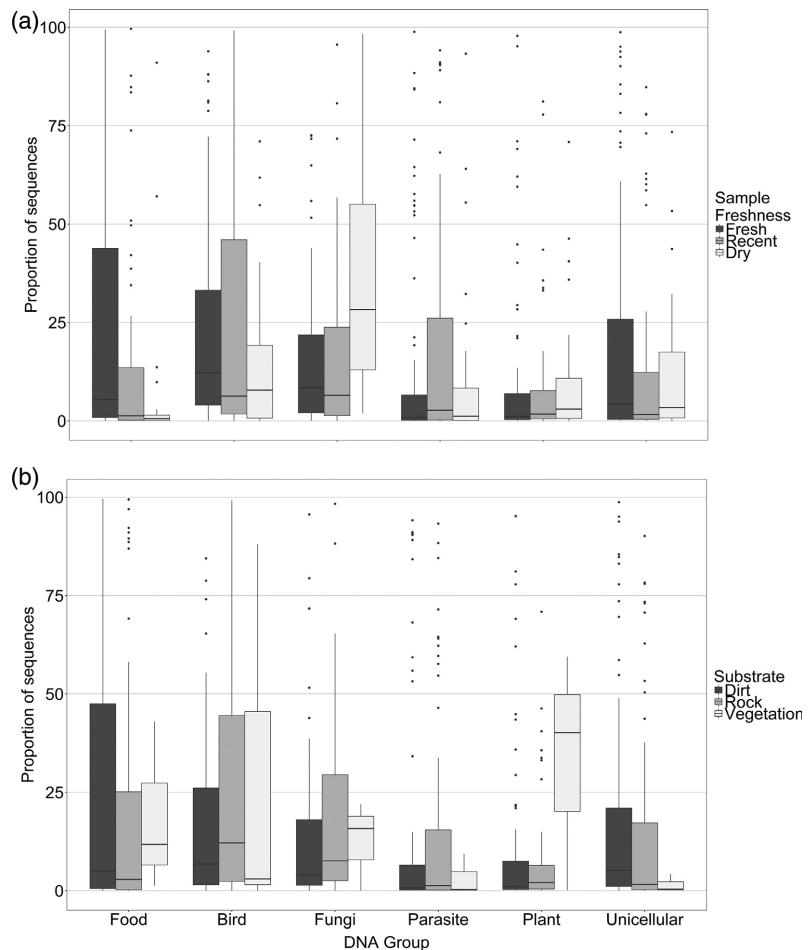


Fig. 3. DNA proportions for each: (a) sample freshness (fresh, recent and dry) and (b) substrate (dirt, rock and vegetation). All samples collected during chick rearing, with only fresh scats included in the analysis of substrate. To improve readability, the category 'contaminant' was excluded from the graph as it contained a very small proportion of DNA sequences.

DEVELOPMENTAL STAGE

During incubation, there was no significant difference between breeders and non-breeders in DNA amplification success ($\chi^2_{1,79} = 0.053$, $P = 0.82$; Table 2), or the proportion of food DNA detected in scats ($\chi^2_{1,69} = 11.502$, $P = 0.09$; Table 2, Fig. 4). However, during brood guard, the developmental stage of birds did significantly affect the DNA amplification success ($\chi^2_{2,166} = 8.711$, $P = 0.01$). Scats from chicks had a lower amplification success than those from breeders (Table 2, Fig. 4). The proportion of food DNA detected was also significantly affected by the developmental stage during brood guard, ($\chi^2_{2,119} = 88.972$, $P < 0.001$), with scats from breeders containing a much higher proportion of food DNA than those from chicks or non-breeders (Table 2, Fig. 4). During the brood stage, chick scats had a higher proportion of bird, fungi and plant DNA than breeders, whereas non-breeder scats were dominated by parasites (Fig. 5).

FASTING

The time a bird spent fasting did not significantly affect the DNA amplification success of the scat ($\chi^2_{3,178} = 3.01$, $P = 0.39$), but did strongly affect the proportion of food DNA detected within the scat ($\chi^2_{3,147} = 70.165$, $P < 0.001$). Scats collected from birds incubating for less than a day had a far

greater proportion of food DNA detected than scats collected randomly; however, this was not the case for any other incubation length category (Table 2, Fig. 5). Scats from birds that had been incubating longer than 1 day contained predominantly parasite DNA (Fig. 5).

Discussion

Our case study clearly indicates that sample freshness, the substrate the scat was collected from, breeding stage, developmental stage and fasting can all impact the amount of food DNA available for dietary DNA metabarcoding. The scat collection protocol presented here contributes to optimising the amount of food DNA that is identified in vertebrate dietary studies.

SAMPLE FRESHNESS

Scat freshness was found to affect both the DNA amplification success and the proportion of food DNA detected. Fresh scats exhibited a higher DNA amplification success than recent scats, but not dry scats. We had expected that both recent and dry scats would have less amplifiable DNA than fresh scats due to degradation during environmental exposure (Oehm *et al.* 2011). However, dry scats have also had more potential exposure to external contamination, particularly from fungi, which was reflected in the non-food DNA sequences

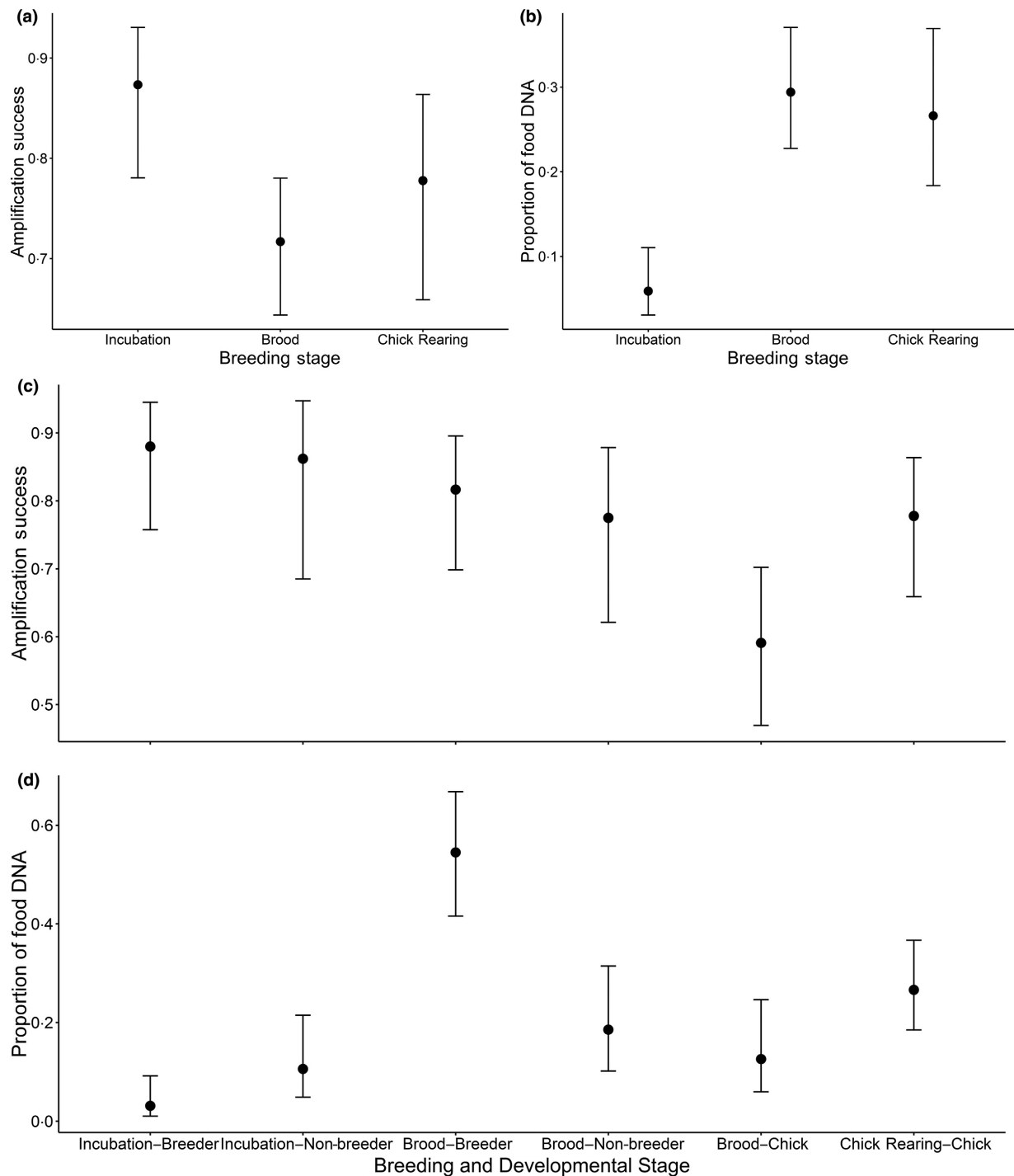


Fig. 4. Generalised Linear Model fitted plots of the (a) amplification success and (b) proportion of food for each breeding stage (incubation, brood and chick rearing); and (c) the amplification success and (d) the proportion of food for each age cohort within each breeding stage. Incubation samples included only scats collected randomly where incubation length was unknown. Points represent means, and bars show 95% confidence intervals.

recovered. When we look specifically at the amount of food DNA amplified from dry scats, this component was significantly less than for fresh scats. Although recent scats had a lower amplification success, the proportion of food DNA detected was not significantly lower than that of fresh scats. This 'recent' category contained a wide range of scats, from those defecated within minutes (but not seen), to those exposed

for many hours. Therefore, using scats that are still wet may produce dietary information, but larger sample sizes would be required and reliance on small amounts of DNA may reduce data quality (Murray, Coghlan & Bunce 2015).

Samples in this study were collected during the day from a species breeding in an exposed habitat with little protection from UV and rain. Scats collected in protected conditions such

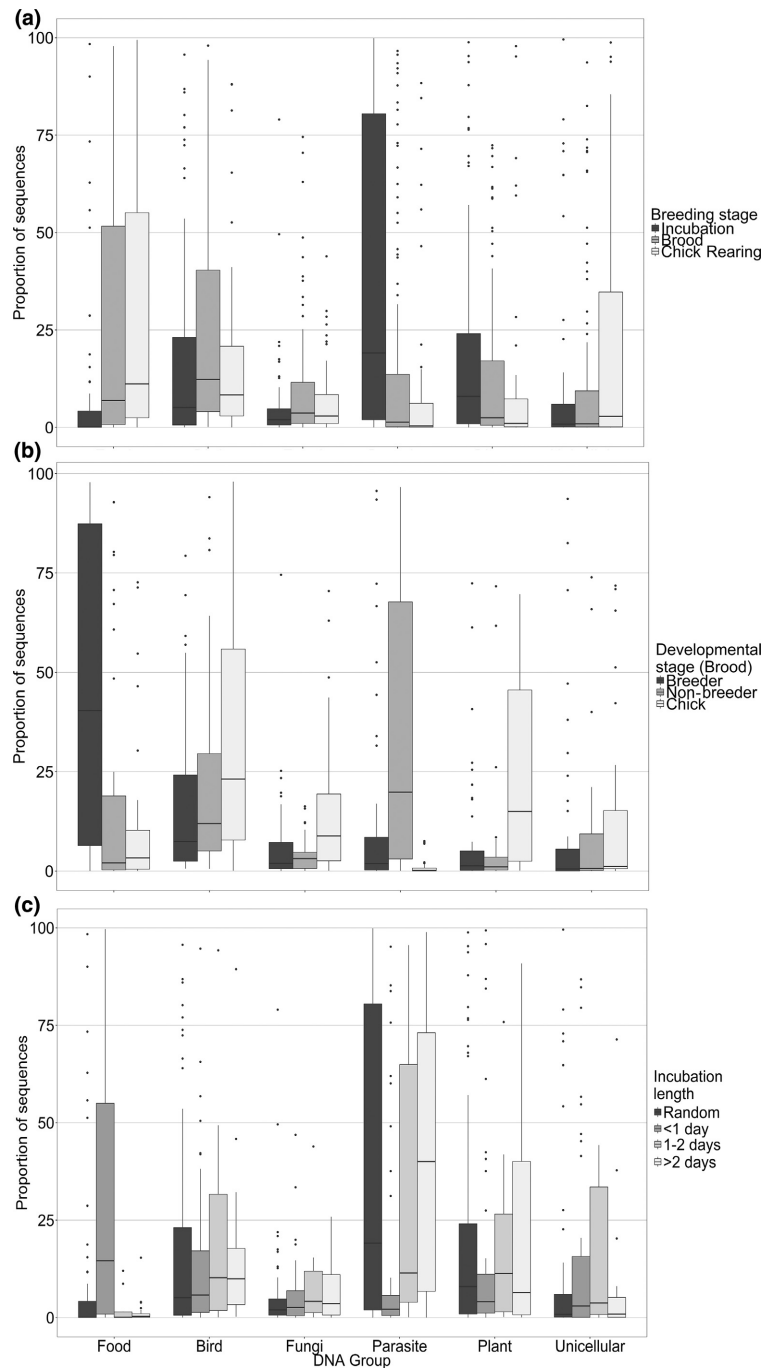


Fig. 5. DNA proportions for each: (a) breeding stage (incubation, brood and chick rearing); (b) developmental stage during brood (breeder, chick and non-breeder); and (c) incubation length (random, <1 day, 1–2 days and >2 days). Only fresh scats were analysed. To improve readability, the category ‘contaminant’ was excluded from the graph as it contained a very small proportion of DNA reads. The incubation category in ‘A’ included scats collected randomly where incubation length was unknown.

as from a shaded area, at night or collected in the early morning may allow amplifiable DNA to persist for longer. For example, in carrion crows, food DNA could be detected for up to 5 days when protected from UV and rain exposure (68% success); however, this was dramatically reduced when scats were left in exposed areas (17.5% success; Oehm *et al.* 2011). Similarly, Steller sea lion *Eumetopias jubatus* scats also produced detectable prey DNA for up to 5 days in some samples (Deagle *et al.* 2005). However, in both of these studies, group-specific markers were used that detected only food items. In our study, some dry scats still contained food DNA, so it is possible that with the use of group-specific markers, dietary information may be detectable for longer.

To ensure fresh scats are collected in the field, some studies have captured or contained animals (Kartzinel & Pringle 2015; Lopes *et al.* 2015), or placed sheets to collect fresh faeces (Deagle *et al.* 2010; Vesterinen *et al.* 2016). When manipulation of the surrounding environment is not physically or ethically possible, alternative sampling strategies are required. In optimising scat collections, we did not seek to determine the amount of time in hours or days that a scat could be collected, as this information is unknown when a scat is found. Instead, we wanted to provide a proxy that allows field biologists to selectively collect scats that will provide high-quality dietary material. Unfortunately, wet recent scats did not provide as much data as fresh scats, which meant that observing defecating

animals will still be best practice in exposed locations. However, this is often not possible and other proxies may be required to determine sample freshness (e.g. odour, colour, consistency), as well as understanding how these may change between species, seasons and environments (Piggott 2004; Vynne *et al.* 2012; Demay *et al.* 2013).

Given the proportions of broad categories of DNA change as the sample ages, it is possible that the measured proportions from various diet species in the samples may change too if DNA from different species degrades at different rates. This should be examined with experimental studies and care taken to ensure consistent collection methods between sites.

SUBSTRATE TYPE

Scats collected from rock and dirt had similar amplification success, but scats from rock had a higher proportion of food DNA detected. This is partially consistent with Oehm *et al.* (2011) who also found that carrion crow scat samples collected from dirt had reduced food DNA detectability, in both protected and exposed samples. However, they found that DNA detection was hampered by an increase in PCR inhibitors. This did not appear to affect the samples in our case study as amplification success was similar between dirt and rock samples. Instead, the presence of non-food DNA was higher in scats obtained from dirt. Our scats were fresh, compared to 5-day-old scats from carrion crows; therefore, the DNA in our samples may not have been as degraded. Shy albatross scat samples from dirt contained a higher proportion of unicellular DNA than from rock. Unicellular eukaryotes are common in soil and these sequences are likely to represent contamination. It is often difficult to separate scat samples from dirt, especially for very liquid samples that have been mixed into the dirt. Seabird colonies can be home to greater densities of microbial communities within the soil (Wright *et al.* 2010), which may exacerbate the presence of non-food DNA.

The three scats collected from plants were dominated by plant DNA. Any surface that contains DNA is likely to decrease the amount of food DNA due to increased contamination. An additional complication occurs when the substrate could be incorrectly assigned as a food item. This is particularly relevant for dietary studies on herbivore species when scats are collected from vegetation (Kartzinel *et al.* 2015), or marine species when scats are collected from the water (Jarman & Wilson 2004). If collecting from vegetation, the substrate species should be recorded to allow appropriate categorisation when interpreting sequencing results.

BREEDING AND DEVELOPMENTAL STAGE

Digestion rates are likely to vary for numerous reasons, for example predator species, metabolic rate, meal size, food type and feeding frequency (Hilton, Houston & Furness 1998). These may all in turn impact the detectability of food DNA in scats. Understanding how feeding behaviour may change throughout the year or breeding season for different

developmental stages will impact how and when samples can be collected.

Collections from young animals are likely to pose problems for DNA dietary analysis depending on the way they obtain food. In this case study, young chicks had a lower proportion of food DNA detected than breeding adults and a higher proportion of bird DNA. In many avian species, juvenile food is delivered by regurgitation; therefore, food items are likely to be partially digested before they are fed to the chick. This was the case in white-chinned petrels *Procellaria aequinoctialis*, where food in chicks' stomachs was more digested than that of adults (Connan *et al.* 2007). Consequently, digestive processes may excessively degrade food DNA in chick scat samples. Additionally, there is presumably crossover of parental DNA to the chick during regurgitation, which may cause the amount of bird DNA to be inflated, thereby reducing the food DNA proportionately. Interestingly, the converse results were seen with Adélie penguins *Pygoscelis adeliae*, with scats collected from chicks more successful than those from breeders, especially during brood guard when chicks were small. Although a similar marker region was used in both studies, a blocking primer was used to suppress bird DNA amplification, which may explain this result (McInnes *et al.* 2016).

Scat samples from young vertebrates should ideally be collected when they are directly feeding on the food themselves, rather than through secondary means. For birds fed by regurgitation, this may not be possible during the nestling period; however, samples from older chicks did contain more food DNA. Older shy albatross chicks had a higher food proportion than small chicks, which may reflect larger meals or a reduction in stomach oil. Procellariiforme (albatross and petrel) stomachs contain oil that is obtained from digested prey (Imber 1976). This oily liquid can contribute up to 80% of the sample mass in some albatross stomachs (Thompson 1992). In shy albatross, there is a greater mass of oily liquid in younger chicks than older chicks (Hedd & Gales 2001), which may dilute the food DNA. Young animals with diet supplemented by suckling milk could also have the same issue.

We also observed differences in food detection between breeding cohorts, with lower proportions of food DNA and higher proportions of parasite DNA detectable from scats of non-breeding animals during brood guard. A non-breeder was identified by its presence at an empty nest and is likely to be either a failed breeder or subadult bird defending a nest. As these individuals do not need to forage to feed a chick, they may have been ashore longer and therefore could fall into a similar category to fasting/incubating birds. This finding highlights the need to understand not only the biology of the study species, but also awareness of which breeding cohorts may be present during scat collections and how this may affect results.

FASTING

The detection of food DNA throughout the season was strongly linked to fasting. Longer periods of fasting during incubation resulted in a low detection of food DNA in scats, whereas food DNA detection was much higher for breeding

birds during brood. This is likely to be linked to more frequent feeding trips during this stage. During periods of fasting, non-target DNA was dominated by endoparasites, rather than external contamination. Cestodes are the main endoparasites in pelagic seabirds, and their presence is largely driven by diet and the availability of intermediate hosts, for example zooplanktonic organisms and fish (Hoberg 1996). Interestingly, there was an apparent increase in parasite DNA during fasting. If the food DNA proportion alone had decreased, then it would be expected that all other DNA groups would increase proportionally. However, there appeared to be a greater increase in the parasite DNA than other groups, suggesting there was an increase in prevalence, not just detection. The exact cause of the increase during fasting is unknown; however, care should be taken when obtaining scats, targeting animals with minimal time since feeding.

Fasting periods occur in many species for many reasons, including territory defence, hibernation, meal availability, migration, incubating or suckling young, moult or limited mobility, for example during pregnancy. Understanding when these fasting periods occur in the study species is important for detection of dietary DNA in scats. Although defecation rate does slow during fasting, it often will not cease. Therefore, the risk of collecting scats that contain no dietary information needs to be taken into consideration when planning a study.

FIELD PROTOCOLS FOR DNA SCAT COLLECTION

We have developed a method to allow high-quality dietary information to be obtained using universal metazoan markers by optimising collection protocols, enabling a reduction in signal from non-target DNA.

Careful planning of DNA dietary metabarcoding studies prior to sample collection is imperative for overall project success. Researchers should consider the dietary question they are targeting and focus on which scat samples will inform this. This includes marker selection, seasonal changes, fasting and the age of animals. These considerations, especially animal behaviour and developmental stage, are likely to be important to a broad array of molecular ecology studies reliant on DNA in scat samples, or those using eDNA. To improve the quality and quantity of dietary information obtained from scat samples, the following collection protocols should be followed when possible.

- 1 Collect fresh scats where the animal is seen defecating. If this is not possible, try to collect only scats that still have moisture or develop species-specific proxies that correlate to sample age.
- 2 Give serious consideration to the scat substrate type, as contamination from substrate can overwhelm the food DNA signal. Ideally, collect scats from surfaces with minimal sources of DNA contamination (e.g. rocks or ice). If collecting from dirt or vegetation, try to minimise the collection of foreign material and record the substrate (and species where applicable) to cross-check and validate results.
- 3 Take into consideration, the seasonal behaviour and feeding ecology of the study animal prior to sample collection.

4 Avoid collections from animals that may not have fed recently, such as periods of fasting.

5 Collect from animals that are directly feeding themselves and avoid secondary feeding where possible (including suckling young). Samples from young animals that are being fed by regurgitation may be problematic due to partially digested food passed on by the parents or large amounts of parental DNA. For such species, collection from older animals may be preferable.

6 If only a single collection is available and the seasonal timing and cohort are not the focus of the study, target the time period with the shortest time since feeding and focus on adult animals.

7 If multiple study sites are used, keep collection protocols and timing consistent between sites

These optimised scat collection protocols provide a basis for future experimental designs and will enable ecologists to collect high-quality diet samples and reduce non-target DNA amplification. They also provide standardised field methods which will be important in this rapidly expanding area of research.

Authors' contributions

Experimental design was carried out by J.M., S.J., R.A. and M.-A.L. Sample collection was made by J.M. and R.A. Molecular marker testing was conducted by J.M. and S.J. Laboratory experiments were conducted by J.M. Bioinformatics was performed by J.M., S.J., B.D. Statistical analysis was performed by J.M. and B.R. All authors wrote the manuscript.

Acknowledgements

This project used University of Tasmania Animal Ethics Permit A13745 and Tasmanian DPIPWE Scientific Permits TFA 14049 and TFA 15081. Funding was provided by Australian Antarctic Science Grant (4014) the Winifred Violet Scott Charitable Trust. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668981. Thanks to James Marthick and Menzies Institute (UTAS) for Miseq use; Kris Carlyon, Sam Thalmann (DPIPWE) and Alistair Hobday (CSIRO) for field assistance; Andrea Polanowski and Cassy Faux (AAD) for laboratory advice; and Simon Wotherspoon (UTAS) for statistical advice.

Data accessibility

Data are accessible from the Australian Antarctic Data Centre (doi:10.4225/15/57EDACF067ADF).

References

- Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*, **2**, e197.
- Bowser, A.K., Diamond, A.W. & Addison, J.A. (2013) From puffins to plankton: a DNA-based analysis of a seabird food chain in the northern Gulf of Maine. *PLoS One*, **8**, e83152.
- Brothers, N., Gales, R., Hedd, A. & Robertson, G. (1998) Foraging movements of the Shy Albatross *Diomedea cauta* breeding in Australia; Implications for interactions with longline fisheries. *Ibis*, **140**, 446–457.
- Cherel, Y. & Klages, N. (1998) A review of the food of albatrosses. *Albatross: Biology and Conservation* (eds G. Robertson, & R. Gales), pp. 113–136. Surrey Beatty & Sons, Chipping Norton, UK.
- Connan, M., Cherel, Y., Mabile, G. & Mayzaud, P. (2007) Trophic relationships of white-chinned petrels from Crozet Islands: combined stomach oil and conventional dietary analyses. *Marine Biology*, **152**, 95–107.

- Davison, A., Birks, J.D.S., Brookes, R.C., Braithwaite, T.C. & Messenger, J.E. (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *Journal of Zoology*, **257**, 141–143.
- Deagle, B.E., Kirkwood, R. & Jarman, S.N. (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.
- Deagle, B.E., Tollitt, D.J., Jarman, S.N., Hindell, M.A., Trites, A.W. & Gales, N.J. (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, B.E., Chiaradia, A., McInnes, J. & Jarman, S.N. (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.
- Demay, S.M., Becker, P.A., Eidson, C.A., Rachlow, J.L., Johnson, T.R. & Waits, L.P. (2013) Evaluating DNA degradation rates in faecal pellets of the endangered pygmy rabbit. *Molecular Ecology Resources*, **13**, 654–662.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
- Farrell, L.E., Roman, J. & Sunquist, M.E. (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology*, **9**, 1583–1590.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M. & Troedsson, C. (2014) Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS One*, **9**, e87624.
- Hedd, A. & Gales, R. (2001) The diet of shy albatrosses (*Thalassarche cauta*) at Albatross Island, Tasmania. *Journal of Zoology London*, **253**, 69–90.
- Hedd, A. & Gales, R. (2005) Breeding and overwintering ecology of shy albatrosses in southern Australia: year-round patterns of colony attendance and foraging-trip durations. *Condor*, **107**, 375–387.
- Hedd, A., Gales, R. & Brothers, N. (2001) Foraging strategies of shy albatross *Thalassarche cauta* breeding at Albatross Island, Tasmania, Australia. *Marine Ecology-Progress Series*, **224**, 267–282.
- Hilton, G.M., Houston, D.C. & Furness, R.W. (1998) Which components of diet quality affect retention time of digesta in seabirds? *Functional Ecology*, **12**, 929–939.
- Hoberg, E.P. (1996) Faunal diversity among avian parasite assemblages: the interaction of history, ecology, and biogeography in marine systems. *Bulletin of the Scandinavian Society for Parasitology*, **6**, 65–89.
- Hothorn, T., Bretz, F. & Westfall, P. (2008) Simultaneous inference in general parametric models. *Biometrical Journal*, **50**, 346–363.
- Imber, M.J. (1976) The origin of petrel stomach oils: a review. *The Condor*, **78**, 366–369.
- Jarman, S.N., Deagle, B.E. & Gales, N.J. (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, **13**, 1313–1322.
- Jarman, S.N. & Wilson, S.G. (2004) DNA-based species identification of krill consumed by whale sharks. *Journal of Fish Biology*, **65**, 586–591.
- Jarman, S.N., McInnes, J.C., Faux, C., Polanowski, A.M., Marthick, J., Deagle, B.E., Southwell, C. & Emmerson, L. (2013) Adélie penguin population diet monitoring by analysis of food DNA in scats. *PLoS One*, **8**, e82227.
- Kartzinel, T.R. & Pringle, R.M. (2015) Molecular detection of invertebrate prey in vertebrate diets: trophic ecology of Caribbean island lizards. *Molecular Ecology Resources*, **15**, 903–914.
- Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W. & Pringle, R.M. (2015) DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences, USA*, **112**, 8019–8024.
- King, R.A., Read, D.S., Traugott, M. & Symondson, W.O.C. (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*, **17**, 947–963.
- Levy-Booth, D.J., Campbell, R.G., Gulden, R.H. *et al.* (2007) Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, **39**, 2977–2991.
- Lopes, C.M., De Barba, M., Boyer, F. *et al.* (2015) DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents. *Heredity*, **114**, 525–536.
- Lucchini, V., Fabbri, E., Marucco, F., Ricci, S., Boitani, L. & Randi, E. (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Molecular Ecology*, **11**, 857–868.
- McCullagh, P. & Nelder, J.A. (1989) *Generalised Linear Models*, 2nd edn. CRC Press, Boca Raton, FL, USA.
- McInnes, J.C., Emmerson, L., Southwell, C., Faux, C. & Jarman, S.N. (2016) Simultaneous DNA-based diet analysis of breeding, non-breeding and chick Adélie penguins. *Royal Society Open Science*, **3**, 150443.
- Murray, D.C., Coghlan, M.L. & Bunce, M. (2015) From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLoS One*, **10**, e25776.
- Murray, D.C., Bunce, M., Cannell, B.L., Oliver, R., Houston, J., White, N.E., Barrero, R.A., Bellgard, M.I. & Haile, J. (2011) DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS One*, **6**, e0124671.
- Oehm, J., Juen, A., Nagiller, K., Neuhauser, S. & Traugott, M. (2011) Molecular scatology: how to improve prey DNA detection success in avian faeces? *Molecular Ecology Resources*, **11**, 620–628.
- O'Rorke, R., Lavery, S. & Jeffs, A. (2012) PCR enrichment techniques to identify the diet of predators. *Molecular Ecology Resources*, **12**, 5–17.
- O'Rorke, R., Lavery, S., Chow, S., Takeyama, H., Tsai, P., Beckley, L.E., Thompson, P.A., Waite, A.M. & Jeffs, A.G. (2012) Determining the diet of larvae of western Rock Lobster (*Panulirus cygnus*) using high-throughput DNA sequencing techniques. *PLoS One*, **7**, e42757.
- Panasci, M., Ballard, W.B., Breck, S., Rodriguez, D., Densmore Iii, L.D., Wester, D.B. & Baker, R.J. (2011) Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. *Journal of Wildlife Management*, **75**, 1616–1624.
- Piggott, M.P. (2004) Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research*, **31**, 485–493.
- Pompanon, F., Deagle, B.E., Symondson, W.O., Brown, D.S., Jarman, S.N. & Taberlet, P. (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.
- Prugh, L.R., Ritland, C.E., Arthur, S.M. & Krebs, C.J. (2005) Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology*, **14**, 1585–1596.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Thompson, K.R. (1992) Quantitative analysis of the use of discards from squid trawlers by black-browed albatrosses *Diomedea melanophris* in the vicinity of the Falkland Islands. *Ibis*, **134**, 11–21.
- Vesterinen, E.J., Ruokolainen, L., Wahlberg, N. *et al.* (2016) What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Molecular Ecology*, **25**, 1581–1594.
- Vynne, C., Baker, M.R., Breuer, Z.K. & Wasser, S.K. (2012) Factors influencing degradation of DNA and hormones in maned wolf scat. *Animal Conservation*, **15**, 184–194.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, NY, USA.
- Wright, D.G., van der Wal, R., Wanless, S. & Bardgett, R.D. (2010) The influence of seabird nutrient enrichment and grazing on the structure and function of island soil food webs. *Soil Biology and Biochemistry*, **42**, 592–600.
- Zeale, M.R., Butlin, R.K., Barker, G.L., Lees, D.C. & Jones, G. (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236–244.

Received 21 June 2016; accepted 22 September 2016

Handling Editor: Michael Bunce

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Second round PCR tag combinations.

Appendix S2. Categories of sequences assigned to each group.

Appendix S3. Field collection protocols for DNA dietary analysis of seabird scats. This video provides an overview of the equipment, methods and guidelines for scat collections based on results presented in this manuscript (https://youtu.be/CQ_6bUX91ls).