Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition?

Bruce E. Deagle · Dominic J. Tollit

Abstract We investigate using relative amounts of prey DNA recovered from pinniped faeces to obtain diet composition data. Faeces were obtained from captive sea lions being fed a diet containing three fish species (50%, 36% and 14% by mass). Real-time PCR was used to quantify mtDNA in undigested tissue and in the faecal samples. The percent composition of fish mtDNA extracted from tissue corresponded reasonably well to the mass of fish in the mixture. In faecal samples the absolute amount of fish mtDNA recovered varied 100-fold, but the percent composition of the samples was relatively consistent ($57.5 \pm 9.3\%$, $19.3 \pm 6.6\%$ and $23.2 \pm 12.2\%$). These results indicate there are prey-specific biases in DNA survival during digestion. However, the biases may be less than those commonly observed in conventional diet studies.

Keywords Non-invasive · Molecular scatology · Steller sea lion · Faecal DNA · Seal diet

B. E. Deagle (✉)
School of Zoology, University of Tasmania,
P.O. Box 252-05, Hobart, Tasmania 7000, Australia
e-mail: bedeagle@utas.edu.au

B. E. Deagle
Australian Antarctic Division, Channel Highway,
Kingston, Tasmania 7050, Australia

D. J. Tollit
Marine Mammal Research Unit, Fisheries Centre,
University of British Columbia, AERL, 2202 Main Mall,
Vancouver, Canada V6T 1Z4

Reliable methods of estimating pinniped diet composition are critical for appropriate management in situations where populations of pinnipeds or their prey are endangered (Sinclair and Zeppelin 2002; Purcell et al. 2004). Diet composition is usually determined by analysis of hard-part remains in faeces, but estimates can be inaccurate due to differential recovery of hard-parts from prey species (Tollit et al. 2003). Recent studies have successfully employed DNA-based methods to identify prey hard remains in pinniped faeces (Purcell et al. 2004; Kvitrud et al. 2005), or to identify prey from DNA present in the soft matrix of faeces (Deagle et al. 2005; Parsons et al. 2005). Quantitative diet estimates could be obtained from DNA-based studies based on prey occurrence. Estimates would be improved through quantification of the DNA present in faeces if the amount of DNA from prey species in the samples is proportional to their mass in the diet. We have previously found that the proportions of fish DNA in faeces from captive sea lions were roughly proportional to the mass of the prey items consumed (Deagle et al. 2005). These results were based on a small number of samples, and showed a bias in the relative amount of DNA recovered from different fish species. This bias could be an artefact of the clone library method we employed—caused by variation in amplification efficiencies between prey species (von Wintzingerode et al. 1997). Alternatively, the bias could be due to prey-specific differences in DNA density, or differential DNA survival during digestion. If variation in prey DNA density can explain the bias, it would be feasible to develop prey-specific correction factors. If there are differences in prey DNA survival during digestion, experimentally derived correction factors would require time-consuming captive feeding studies.

Here, we use quantitative real-time PCR (qPCR) to investigate these issues. Specifically, we quantify the relative amounts of mitochondrial DNA in the tissue of three fish species being fed to captive Steller sea lions (*Eumetopias jubatus*), then determine the amount of DNA recovered from these prey items in the sea lions' faeces.

The faecal samples analysed here ($n = 23$) were collected from two captive sea lions during a previous study (Deagle et al. 2005). The fish component of their diet (fed in two equal daily meals) consisted of 50% Pacific herring (*Clupea pallasii*), 36% surf smelt (*Hypomesus pretiosus*), and 14% sockeye salmon (*Oncorhynchus nerka*) by mass. Faeces were blended, preserved in 95% ethanol and DNA was extracted from the soft matrix as described previously (Deagle et al. 2005). A tissue mix equivalent to the sea lions' diet was produced by mixing 150 g of herring, 108 g of smelt and 42 g of salmon. Fish were taken from the frozen stock fed to the sea lions; the salmon had been cleaned with head/fins removed, the herring and smelt were whole. The fish were blended and further homogenised using a micro blender. DNA was extracted from aliquots ($n = 10$; 72 ± 7 mg) of this finely homogenised tissue following the faecal DNA protocol. Tissue DNA was diluted to ~1 ng/μl before use in qPCR.

The quantity of extracted DNA was estimated using SYBR® Green qPCR assays. For each fish species we designed PCR primers that amplify products (65–69 bp) from the mitochondrial 16S rDNA gene (sequences in caption of Fig. 1). Amplifications were run using the Chromo4™ system (MJ Research). The PCR mix (20 μl) consisted of 10 μl QuantiTect® SYBR® Green mix (Qiagen), 0.5 μM of each primer, 0.1 μg BSA and 4 μl template DNA (diluted 1:5). Thermal cycling conditions were 35 cycles of: 94°C, 30 s /55°C, 30 s /72°C, 45 s; optical data were acquired following each 72°C step. A subset of samples was separated on 1.8% agarose gels to confirm product size and to check for primer dimers.

To ensure accurate relative quantification of DNA from the prey species we used a single recombinant plasmid containing the relevant mtDNA region from each of the target species as a common standard (Miller et al. 2003). This plasmid was constructed using conserved primers (16SfishF and 16S2R; Deagle et al. 2005), modified to include restriction sites on the 5′ ends (Fig. 1). PCR products generated from genomic DNA were digested to produce cohesive ends and ligated into the KpnI/XhoI sites of the pCR®2.1 TOPO® vector (Invitrogen). The resultant plasmid was sequenced to confirm identity, and the concentration of plasmid DNA determined using a PicoFluor fluorometer (Turner Designs). Standard curves were generated using concentrations of the three-fish plasmid encompassing the range of target (2-fold dilution series for tissue and 5-fold dilution series for faecal DNA). For individual DNA extractions, copy numbers for all three target species were quantified in a single run (using a PCR mix differing only in primer composition). Separate standard curves were constructed for each primer set; since there were no consistent differences in these curves, the data were pooled to create one standard curve per run with a minimum of 15 reference points. To assess inter-run variability, two independent runs were carried out. The threshold cycle (C_t) was set at 10 standard deviations above the mean fluorescence over cycle range 1–10. In the standard curves produced there was a linear relationship between the log of the plasmid DNA copy number and the C_t value (R^2 values ≥ 0.989).

Likelihood ratio tests (LRT) (Hogg and Tanis 2005) were used to test whether the proportions of mtDNA were statistically different than a specified set of proportions. It was assumed that the proportions of mtDNA followed a Dirichlet distribution (a multivariate generalization of the beta distribution
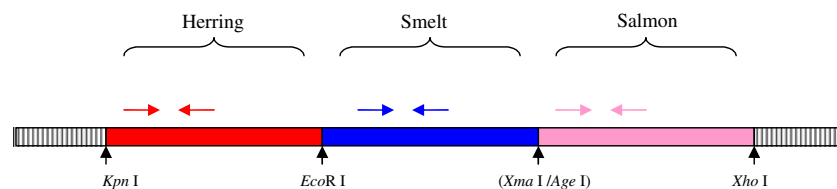

Fig. 1 Schematic of the plasmid used as a standard in qPCR. The plasmid insert contains mtDNA 16S gene fragments from three fish species ligated into the polylinker region of pCR®2.1 TOPO® (Invitrogen). Arrows represent the binding sites of the species-specific primers used during qPCR: Herring-1F (5′-ACCAATCACGAAAAGCAGGT-3′) and Herring-69R (5′-CGAAGACGTTTGTGCCAGTA-3′); Smelt-1F (5′-ACGT-CAAACTTCCCCTTTCA-3′) and Smelt-65 R (5′-CCAACC-GAAGACAGGAGAGA-3′); Salmon-1F (5′-GGCAGAT CACGTCAAAAAC-3′) and Salmon-65R (5′-AGACATATG-GGCTAGGGGTC-3′). These primers were designed with reference to aligned sequences from sea lion and all diet species to ensure they were specific to the intended target

Fig. 2 Box plots showing a summary of the percentage DNA composition data (median, range and upper/lower quartiles) based on quantitative real-time PCR estimates: **(a)** results from tissue mixtures ($n = 10$); **(b)** results from faecal samples ($n = 23$). Filled circles show percentage of each fish prey species (by mass) in the diet. Data for two independent replicate runs (see Table 1 for raw data) and the mean values are shown.
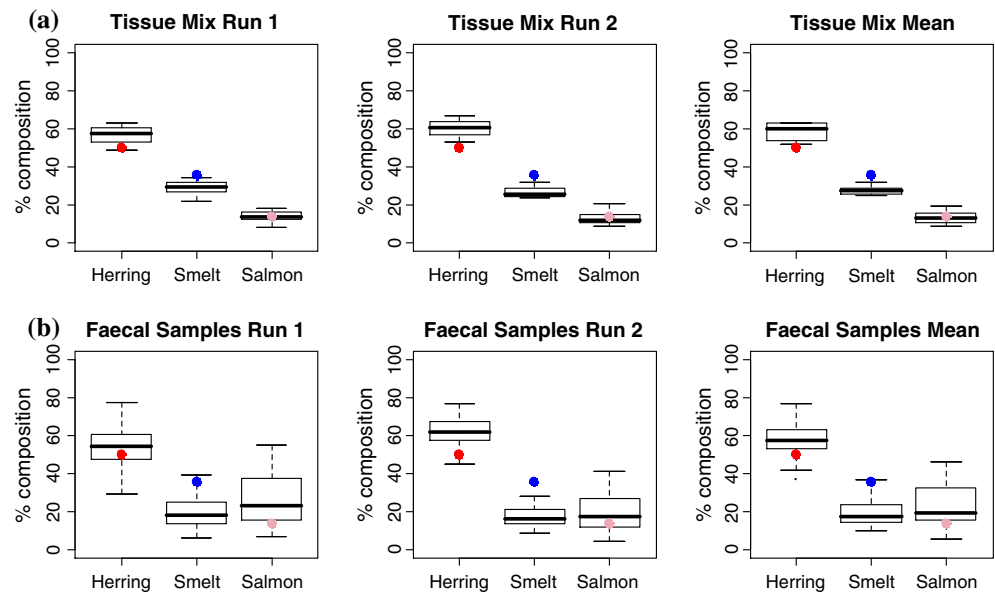
used to describe a set of proportions that sum to one). We parameterized the Dirichlet distribution in terms of expected proportions and an additional variance parameter. In calculating the likelihood under the null hypothesis, we allowed the variance parameter to be estimated freely.

Within tissue samples, the copy number estimates were relatively consistent in the two replicate measurements (data in Table 1a). Overall, the percent composition of mtDNA in the tissue was 58.6 ± 4.6% for herring, 27.9 ± 2.2% for smelt and 13.5 ± 3.1% for salmon (Fig. 2a). The tissue DNA proportions are statistically different from the mass proportions (LRT, P value < 0.001); nevertheless the percent composition of fish mtDNA in undigested tissue is a reasonable proxy for the relative mass of the corresponding fish (Fig. 2a). In the faecal samples there was a large range (100-fold) in the total amount of fish DNA in different samples (data in Table 1b). There were also slightly larger errors in replicate qPCR measurements, possibly due to the wide range covered by the standard curves. Despite this, the percentage composition of fish DNA within the faeces was relatively consistent (57.5 ± 9.3% herring, 19.3 ± 6.6% smelt and 23.2 ± 12.2% salmon; Fig. 2b).

If there is no differential digestion of mtDNA for the different fish species then the mtDNA proportions found in faeces should match the mtDNA proportions in the tissue mix. Our data show this is not the case—the faecal proportions are significantly different from the mean proportions in the tissue (LRT, P value < 0.001). This is due to an overestimation of the proportion of salmon and an underes-

timation in the proportion of smelt mtDNA in the faecal samples. Correction factors based on differences in tissue DNA density improve the estimates (i.e. make them closer to the diet mass proportions), but are not sufficient to account for the observed biases (Table 2). More accurate correction factors accounting for differential digestion could potentially be developed by measuring prey DNA recovery rates in captive feeding trials; however, the recovery of hard-parts has been shown to be affected by numerous variables (Bowen 2000) and developing these correction factors would not be a simple task. Even with the observed bias in DNA recovery, quantification of DNA in faecal samples can clearly provide some informative data. For example, based on our qPCR results, herring was correctly predicted as dominant prey in 21 of 23 samples. Therefore, in situations where large uncertainties surround conventional hard-part faecal analysis (e.g. Laake et al. 2002; Casper et al. 2006), measuring prey DNA amounts in faeces may be a useful approach for determining pinniped diet composition. In the current study we used species-specific primers to compare consumption of individual species, but more general comparisons could be made (e.g. amount of salmonid DNA versus total fish DNA) by using group-specific primers (Jarman et al. 2004). The quantitative DNA-based faecal analysis we have proposed could also be used to examine the diet of species whose faeces do not contain hard-parts (e.g. seabirds or cetaceans) and could be modified to quantify prey DNA in stomachs of invertebrates where few non-molecular approaches to studying diet exist (Symondson 2002).

Table 1 Estimated copy numbers of DNA template in PCR amplifications (two replicate measurements)

Sample Number	Herring		Smelt		Salmon	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
(a) *Tissue mix*						
1	66144	72683	29390	28062	13944	13304
2	23909	24026	11416	12177	7647	6068
3	77211	73359	35145	29778	10221	11837
4	25800	20041	9110	9852	6254	4145
5	68303	69280	37568	28094	15879	12677
6	83087	89079	41434	32177	12964	14991
7	55454	65952	25952	24176	12067	8844
8	18734	18009	11337	10818	5259	5196
9	22654	26951	15380	12279	8380	10293
10	16579	22448	11518	10853	5416	5972
(b) *Faecal samples*						
1	1343	1246	290	401	540	368
2	18787	32492	9424	9225	8430	5543
3	14763	21744	3681	3353	10511	11433
4	84590	77566	38252	41454	30456	28803
5	21347	35755	11850	14802	11337	8748
6	428	658	188	136	116	170
7	4105	4155	1060	1082	817	1056
8	26800	39488	9530	8724	17480	14461
9	5528	8983	2121	2208	7292	4688
10	621	783	441	464	83	59
11	959	1368	370	286	134	126
12	15870	23383	6627	6126	15179	9401
13	6772	6723	1708	1959	2250	2204
14	1681	2069	397	441	1356	1193
15	12531	19324	5076	3785	21657	13606
16	13570	13887	1951	4189	13226	12667
17	3420	4208	1417	1436	908	656
18	2508	2952	823	655	504	569
19	31792	33851	5237	6590	4044	4089
20	27548	58398	29084	15406	16932	9843
21	19421	28447	12077	10526	34086	24150
22	15124	13772	6395	6636	5235	2936
23	557	733	236	293	205	163

It should be noted that the differential survival of DNA from different prey species will not necessarily translate into significant biases in quantitative diet estimates derived from presence/absence genetic data (since prey DNA can still be detected when present at low levels). The potential benefits of using DNA quantification, versus a DNA-based frequency of occurrence approach, will depend on the composition of faecal samples collected in the field. If most samples contain DNA from only one species, or similar amounts of DNA from several species, then further qPCR analysis would be uninformative. However, if most samples contain DNA from several prey species consumed in different amounts, quantification of the relative amount of DNA could be critical for accurate diet composition estimates (see Laake et al. 2002 for discussion)

There are several limitations to the quantitative DNA-based faecal analysis we have proposed. The technical difficulty and cost of quantifying DNA from multiple potential prey species will restrict its application to situations where traditional methods of diet analysis have proven to be inadequate. The approach

Table 2 Summary of percent composition data with faecal composition estimates corrected to account for differences in tissue DNA density

	Diet species (i)		
	Herring	Smelt	Salmon
Diet mass % (m_i)	50	36	14
Tissue DNA % (t_i)	58.6	27.9	13.5
Faecal DNA % (x_i)	57.5	19.3	23.2
Corrected faecal DNA % ($\hat{x}_i$)[a]	50.1	25.4	24.5

[a] Percentage composition corrected for differences in prey tissue DNA density:

$\hat{x}_i = \frac{x_i c_i}{\sum_{i=1}^{n}(x_i c_i)}$ where $c_i = \frac{m_i}{t_i}$ and n = number of prey items

also requires *a priori* knowledge of diet diversity, and a focused question about prey of interest. Finally, to obtain an average view of diet, prey DNA would need to be quantified in a large number of samples; our previous suggestion (Deagle et al. 2005) that DNA from multiple faecal samples could be pooled to reduce the amount of laboratory analysis should be reconsidered due to the large variation in the total amount of prey DNA in different samples.

The use of genetic analysis of faeces for studying predator diet is likely to increase since the approach allows assessment of potential biases from traditional approaches, can provide improved taxonomic identification of prey (e.g., Purcell et al. 2004), and can be used to address questions beyond the scope of traditional methods of diet analysis. This study demonstrates the feasibility of quantifying the amount of DNA from several prey species in predators' faeces and provides an initial assessment of the possibility of using this data to obtain quantitative diet composition data.

References

Bowen WD (2000) Reconstruction of pinniped diets: accounting for complete digestion of otoliths and cephalopod beaks. Can J Fish Aquat Sci 57:898–905

Casper RM, Gales NJ, Hindell MA, Robinson SM (2006) Diet estimation based on an integrated mixed prey feeding experiment using *Arctocephalus* seals. J Exp Mar Biol Ecol 328:228–239

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

Hogg RV, Tanis EA (2005) Probability and statistical inference, 7th edn. Prentice Hall, New Jersey

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

Kvitrud MA, Riemer SD, Brown RF, Bellinger MR, Banks MA (2005) Pacific harbor seals (*Phoca vitulina*) and salmon: genetics presents hard numbers for elucidating predator–prey dynamics. Mar Biol 147:1459–1466

Laake JL, Browne P, DeLong RL, Huber HR (2002) Pinniped diet composition: a comparison of estimation models. Fish Bull 100:434–447

Miller FJ, Rosenfeldt FL, Zhang CF, Linnane AW, Nagley P (2003) Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res 31:e61

Parsons KM, Piertney SB, Middlemas SJ, Hammond PS, Armstrong JD (2005) DNA-based identification of salmonid prey species in seal faeces. J Zool 266:275–281

Purcell M, Mackey G, LaHood E, Park L, Huber H (2004) Molecular methods for the genetic identification of salmonid prey from Pacific harbor seal (*Phoca vitulina richardsi*) scat. Fish Bull 102:213–220

Sinclair EH, Zeppelin TK (2002) Seasonal and spatial differences in diet in the western stock of Steller sea lions (*Eumetopias jubatus*). J Mammal 83:973–990

Symondson WOC (2002) Molecular identification of prey in predator diets. Mol Ecol 11:627–641

Tollit DJ, Wong M, Winship AJ, Rosen DAS, Trites AW (2003) Quantifying errors associated with using prey skeletal structures from fecal samples to determine the diet of Steller's sea lion (*Eumetopias jubatus*). Mar Mamm Sci 19:724–744

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