MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 3706–3718 doi: 10.1111/mec.12523

SPECIAL ISSUE: MOLECULAR DETECTION OF TROPHIC INTERACTIONS

Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias

AUSTEN C. THOMAS,* SIMON N. JARMAN,† KATHERINE H. HAMAN,* ANDREW W. TRITES* and BRUCE E. DEAGLE†

Marine Mammal Research Unit, Fisheries Centre, and Department of Zoology, University of British Columbia, 2202 Main Mall, Vancouver, BC, V6T 1Z4, Canada, †Australian Antarctic Division, Channel Highway, Kingston, Tas. 7050, Australia

Abstract

Ecologists are increasingly interested in quantifying consumer diets based on food DNA in dietary samples and high-throughput sequencing of marker genes. It is tempting to assume that food DNA sequence proportions recovered from diet samples are representative of consumer's diet proportions, despite the fact that captive feeding studies do not support that assumption. Here, we examine the idea of sequencing control materials of known composition along with dietary samples in order to correct for technical biases introduced during amplicon sequencing and biological biases such as variable gene copy number. Using the Ion Torrent PGM©, we sequenced prey DNA amplified from scats of captive harbour seals (*Phoca vitulina*) fed a constant diet including three fish species in known proportions. Alongside, we sequenced a prey tissue mix matching the seals' diet to generate tissue correction factors (TCFs). TCFs improved the diet estimates (based on sequence proportions) for all species and reduced the average estimate error from 28 ± 15% (uncorrected) to 14 ± 9% (TCF-corrected). The experimental design also allowed us to infer the magnitude of prey-specific digestion biases and calculate digestion correction factors (DCFs). The DCFs were compared with possible proxies for differential digestion (e.g. fish protein%, fish lipid%) revealing a strong relationship between the DCFs and percent lipid of the fish prey, suggesting prey-specific corrections based on lipid content would produce accurate diet estimates in this study system. These findings demonstrate the value of parallel sequencing of food tissue mixtures in diet studies and offer new directions for future research in quantitative DNA diet analysis.

Keywords: correction factors, diet analysis, next-generation sequencing, pinniped

Received 18 April 2013; revision received 30 August 2013; accepted 9 September 2013

Introduction

Many ecological studies attempt to identify and accurately quantify trophic interactions between species in food webs to enhance understanding of food web structure (Lindeman 1942; Pomeroy 1974). For decades, the primary tool available to accomplish this task has been the morphological identification of hard food structures that can be identified from the scats and stomach contents of consumers (Scheffer & Sperry 1931; Duffy & Jackson 1986). However, there are major limitations and biases associated with quantifying diets from hard food remains, such as the inability to detect foods without hard structures and the differential survival of diagnostic hard structures during the digestive process (Gales & Cheal 1992; Cottrell *et al.* 1996). As a result, ecologists are turning to molecular-based alternatives to quantify species interactions (Bowen & Iverson 2012). Among those, DNA-based diet analysis is a rapidly evolving tool with quantitative capabilities that are just beginning to be explored (Pompanon *et al.* 2012).

Correspondence: Austen C. Thomas, Fax: (604) 822 8180;
E-mail: a.thomas@fisheries.ubc.ca

An emerging diet quantification technique involves the PCR amplification and sequencing of food DNA using highly diagnostic semi-universal DNA markers such as those used by the Consortium for the Barcode of Life (Hebert *et al.* 2003). Many recent studies take advantage of next-generation sequencing technology to generate thousands of food DNA sequences per dietary sample, which allows for semi-quantitative estimates of diet to be obtained from the sequence proportions (see review by Pompanon *et al.* (2012)).

Despite enthusiasm about the potential for quantitative diet analysis using this approach, the method relies on the substantial assumption that quantities of food DNA detected from dietary samples equate to the biomass proportions of food consumed. However, few studies have attempted to test that assumption. Quantitative analyses of DNA from scats of captive Steller sea lions (*Eumetopias jubatus*) and little penguins (*Eudyptula minor*) found consistent food species DNA proportions in the scats of animals fed the same diet (Deagle & Tollit 2007; Deagle *et al.* 2010; Bowles *et al.* 2011). This implies a numerical relationship does exist between amounts of food consumed and proportions of food DNA detected in scat samples of predators—and indicates that quantitative techniques are reasonably precise. This is further supported by the observation that similar results can be obtained when applying both qPCR and next-generation sequencing to the same set of dietary samples of unknown composition (Murray *et al.* 2011). Other subfields have also reported consistency in sequence read proportions between replicate next-generation sequencing runs (Marioni *et al.* 2008; Kauserud *et al.* 2012).

Unfortunately, the ability to produce consistent diet estimates from sequence counts does not mean estimates are an accurate reflection of diet biomass percentages. In all three captive feeding studies (Deagle & Tollit 2007; Deagle *et al.* 2010; Bowles *et al.* 2011), the mass proportions of food consumed did not match the proportions of species DNA detected in dietary samples. The combination of high precision and low accuracy for these techniques implies that there are systematic biases influencing proportions of food DNA detected in diet samples. However, systematic biases such as these can often be quantified and accounted for with numerical correction factors (Tollit *et al.* 1997; Phillips & Harvey 2009; Cheung *et al.* 2011).

The potential biases likely to influence quantitative DNA diet assessment can be broadly categorized into those that are biological in origin (and therefore inherent to the study system), versus those that are introduced via the methodological protocol.

Documented methodological biases include PCR bias (e.g. differential amplification of food species due to preferential primer binding), primer tag bias (i.e. short identification sequences attached to primers causing preferential species DNA amplification) and sequencing bias (e.g. when sequences from particular species are preferentially sequenced) (Sipos *et al.* 2007; Berry *et al.* 2011; Quail *et al.* 2012). Recently, methodological biases have also been identified as a result of sequence quality filtering, sequencing read direction and interactions between several biasing factors (Deagle *et al.* 2013). When possible, such biases should be minimized with careful study design; however, not all methodological biases are feasible to mitigate for every possible food species.

Biological biases can also be very challenging to mitigate in DNA-based diet quantification. There are likely two primary sources of biological bias in these studies: (i) mass-specific differences in target gene copy number between food species (Deagle & Tollit 2007; Darby *et al.* 2013) and (ii) differential digestion of food species DNA in the alimentary canal of the consumer (Greenstone *et al.* 2010; Leal *et al.* 2014). Although little research has been carried out to look directly at these biological biases, they must be considered if one intends to use food DNA sequence proportions to infer quantitative information about the mass proportions of food ingested by consumers.

The microbial ecology community is beginning to use microbial standards or 'control materials' of known composition to account for similar quantification biases to those encountered in diet studies (Huggett *et al.* 2013; Kembel *et al.* 2012). Control materials can be sequenced along with samples of unknown composition, and the differences between the sequence proportions of the controls and their known compositions can then be used to generate correction factors. The correction factors are applied to the sequence counts of the unknown samples to increase the accuracy of the quantitative estimates. Similar spike-in standards are also applied to account for biases in studies using next-generation sequencing to look at differential gene expression (Jiang *et al.* 2011; Zook *et al.* 2012).

If the control materials and unknown samples are both treated in an identical fashion during the methodological protocol, this approach should account for many of the species-specific methodological biases in a single correction (e.g. DNA extraction bias, PCR bias, sequencing bias, quality-filtering bias, etc.). In addition to accounting for methodological biases, the use of controls can also account for species differences in target gene copy number for all species represented in the controls (Darby *et al.* 2013). As such, the application of food species control materials in DNA diet studies has the potential to vastly improve the accuracy of diet estimates based on food species sequence proportions.

The purpose of our study was to determine whether the accuracy of next-generation sequencing diet analysis can be increased by sequencing DNA of control materials (a food tissue mix of known proportions) along with diet samples taken from animals fed a known diet. We therefore performed a feeding trial using captive harbour seals (*Phoca vitulina*) fed known quantities of prey, and sequenced prey DNA amplified from seal scats and a prey tissue mix.

The study design also allowed for quantification of prey-specific digestion biases, because any remaining bias not accounted for by the prey tissue mix should be attributable to differential prey digestion (i.e. if we know the sequence proportions the methodology produces from a tissue mix that goes into the seal, and the sequence proportions that come out in the scats, the difference between the two represents prey-specific differences in recovery due to digestion). As a secondary component of the study, we compared the prey-specific biases to the proximate compositions of the seal prey (e.g. %protein, %Lipid, %moisture). In particular, we wanted to determine whether these prey characteristics were correlated with the observed digestion bias, in the hopes of identifying potential proxies for digestion bias that can be used when feeding trial data are not available.

Materials and methods

An overview of the study design and laboratory workflow is available in Fig. 1.

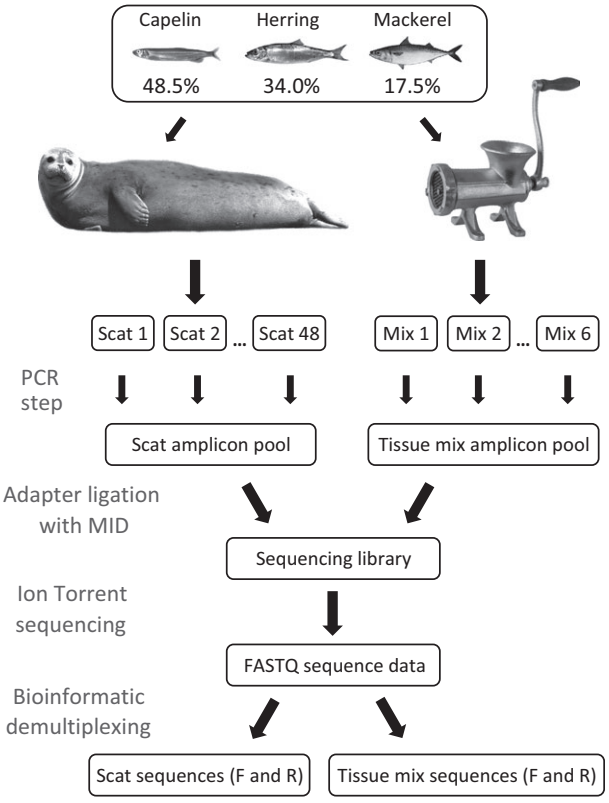

Fig. 1 Overview of the study design and laboratory workflow. Captive harbour seals were fed fixed mass proportions of three fish species (capelin, herring and mackerel), and a fish tissue mix was prepared from whole fish that matched the diet mass proportions. DNA was extracted and amplified from 48 seal scats and 6 fish tissue mix subsamples to form two separate amplicon pools. The amplicon pools received unique Ion Torrent adapter sequences with MIDs, and then sequenced on the Ion Torrent PGM©. Sequence data were demultiplexed by MID and forward/reverse primer sequences, and then assigned to a prey fish species or harbour seal using strict sequence matching criteria. See text for details.

Feeding trial, scat sampling and preservation

The scat samples we analysed were from a feeding trial previously described by Deagle *et al.* (2013). Briefly, the trial involved five adult female harbour seals fed a constant diet of four species in fixed proportions: capelin (*Mallotus villosus*) (40%), Pacific herring (*Clupea pallasii*) (30%), chub mackerel (*Scomber japonicus*) (15%) and market squid (*Loligo opalescens*) (15%). The total daily food intake varied based on seal body mass and their interest in food, but the diet proportions were maintained at the target proportions within the range of measurement precision (2.0% SD per species). During the feeding trial, harbour seal scat samples were collected from both the pool and haul-out areas as a prior study found no significant differences in genetic composition between pool or haul-out collected scats (Bowles *et al.* 2011). Scat samples were generally collected within 2–4 h of deposition, and put into Ziploc bags and immediately frozen at −20 °C. DNA extraction was performed on approximately 20 mg of scat sediment (i.e. hard parts were removed) material using QIAamp DNA Stool kit (Qiagen) according to the protocol described in Deagle *et al.* (2005) with elution in 100 μL elution buffer (10 mM Tris–Cl, 0.5 mM EDTA; pH 9.0).

Preparation of food tissue mixture

A fish tissue mix was prepared based on the mean proportions of fish consumed by the captive seals. Four whole individual fish of each species from the same lot fed to the seals were homogenized using an electric blender, and homogenates were combined by species. Whole fishes were used to ensure that mtDNA variability between prey fish species would be represented in the tissue mix. A 100 g fish tissue mix was created by combining the four species homogenates by wet mass (41.0 g capelin, 29.0 g herring, 15.0 g mackerel and

15.0 g squid). Six approximately 10 g subsamples of the tissue mix were further blended using a tissue homogenizer and DNA was extracted from approximately 20 mg of each of the six tissue mixes, using the DNeasy Blood & Tissue kit (Qiagen) as per the manufacturer's instructions for animal tissues. Subsamples of the 100 g tissue mix were used to diminish the potential influence of laboratory error (e.g. in homogenization, extraction or PCR amplification) on the final tissue mix sequence percentages.

Amplicon library preparation

The barcoding marker we used was a mitochondrial 16S fragment that is approximately 155 bp in length and has been used previously for differentiating fish species (see Deagle *et al.* (2009)). We amplified this marker with primers Chord_16S_F (CGAGAAGACCCTRTGGAGCT) and Chord16S_R (CCTNGGTCGCCCCAAC) which bind to sites that are almost completely conserved in chordates (see Deagle *et al.* (2013) for primer alignments against feeding trial fish species). This primer set does not amplify DNA from squid – therefore diet proportions were recalculated for the three fish species and applied in later calculations.

A blocking oligonucleotide was included in the PCR of all reactions to limit amplification of seal DNA (Vestheim & Jarman 2008). The oligonucleotide (32 bp: ATGGAGCT TTAATTAACTAACTCAACAGAGCA-C3) matches harbour seal sequence (GenBank Accession no. AM181032) and was modified with a C3 spacer, so it is nonextendable during PCR (Vestheim & Jarman 2008). This oligo selectively blocks amplification of seal DNA because it overlaps with the 3′-end of the Chord_16S_F primer and adjoining seal sequence, but has little homology to fish species.

All PCR amplifications were performed in 20 μL volumes using the Multiplex PCR kit (QIAGEN). Reactions contained 10 μL (0.5 X) master mix, 0.25 μm of each primer, 2.5 μm blocking oligonucleotide and 2 μL template DNA. Thermal cycling conditions were as follows: 95 °C for 15 min followed by 34 cycles of: 94 °C for 30 s, 57 °C for 90 s and 72 °C for 60 s. All PCR products were checked on 1.8% agarose gels.

We prepared two separate amplicon pools for sequencing on the Ion Torrent platform. The first contained amplicons from 48 scat samples that were each individually amplified prior to pooling. The pool was created by combining 2 μL of each resultant PCR product to form a single scat metasample for sequencing (scat amplicon pool). The second pool contained amplicons from the six individually amplified tissue mix subsamples that were designed to match the seal diet proportions (tissue mix amplicon pool). The concentration of a subset of samples was quantified using fluorometry (Qubit system; Life Technologies) to ensure approximately equal concentration of the PCR products prior to pooling. To differentiate the two pooled samples, we used the Ion Barcoding kit (Life Technologies; part no. 4468654 Rev. B, 08/2011) that ligates unique multiplex identifier sequences (MIDs) onto amplicons post-PCR along with the necessary Ion Torrent capture sequences. The full amplicon library also contained four other amplicon pools from an unrelated study that each received a unique MID sequence by post-PCR ligation.

Sequencing

We used the Ion OneTouch™ System (Life Technologies) to prepare the amplicon library for sequencing following the user's guide protocol (part no. 4468660 Rev. C, 10/2011). The resultant enriched Ion Sphere™ particles were loaded onto a 314 Ion semiconductor sequencing chip and sequencing (65 cycles) carried out on the Ion PGM sequencer. Bidirectional sequencing was performed (i.e. sequence reads started from either forward or reverse PCR primers), but reads were not paired. Each sequencing run was expected to produce about 10 Mb of sequence data, or 100 000 sequence reads with typical read length of 100 bp (approximately 75 bp being target-specific sequence).

Bioinformatics

The Ion Torrent platform automatically sorted sequences based on the MIDs, removed the MID sequence and output a single FASTQ file for each MID and thus each amplicon pool. We performed the sequence preparation steps using a local installation of the open source Galaxy bioinformatics tools (Giardine *et al.* 2005; Blankenberg *et al.* 2010; Goecks *et al.* 2010). Sequences with less than 100 bp were removed from the data set, and all sequences were trimmed to 100 bp in length to avoid comparability issues with variable length sequences. No quality filtering was applied to the data set to avoid any additional bias that may result from preferential species sequence removal during filtering (Deagle *et al.* 2013).

Sequence assignment to read direction (forward or reverse) and species was performed using the Linux-based open source software package QIIME with sequences from both amplicon pools (Caporaso *et al.* 2010). For a sequence to be assigned to a read direction, it had to match the first 15 bases of the primer (forward or reverse), allowing for up to 2 mismatches in the primer sequence. After assignment to read direction, a local nucleotide BLAST search was performed for each sequence against a reference database containing 16S

sequences for the three fish species and harbour seal (Altschul *et al.* 1990). The accession numbers of the reference sequences are available in the supporting material (Table S2) of the companion study (Deagle *et al.* 2013). The match of each Ion Torrent sequence to reference sequences was assessed based on having a BLASTN e-value less than a relatively strict threshold value of E < 1e-20 and a minimum identity of 0.9. It is worth noting here that the mtDNA marker differs by more than 20% sequence divergence between the three prey fish species. The minimum identity score and our predefined reference sequences prevented assignment of chimeric sequences. To ensure that the species assignment was accurate, a BLAST search was performed in GenBank using a subset of the assigned sequences and the results were 100% congruent with the local database assignment.

Proximate composition analysis of prey species

To help determine whether there are suitable proxies for the calculated biases, we analysed the proximate composition of the prey species and compared the results to the respective correction factors (see correction factors section below). Five individual fish of each prey species from the same lot fed the seals were submitted for full proximate analysis (% moisture, % ash, % protein, % lipid, % carbohydrates). In brief, the % moisture was measured by desiccation of prey tissue, the % ash was measured by combustion of known prey mass, % protein was measured by nitrogen analysis and the % lipid was measured by petroleum ether extraction. Percent carbohydrate was not reported because only negligible levels of carbohydrate were detected in the prey fish.

Tissue mix correction factors

The tissue mix was sequenced along with the diet samples to account for potential differential amplification or sequencing between species and species differences in mtDNA template copy number. Thus, based on sequence proportions from the tissue mix amplicons, we calculated a tissue correction factor (TCF) for each fish species in the diet using

$$\mathrm{TCF}_i = \frac{D_i}{T_i}$$

where i is the prey fish species (capelin, herring or mackerel), D_i is the proportion of species i in the tissue mix and T_i is the proportion of species i detected in the tissue mix amplicon pool. TCFs were then applied to the species sequence counts generated from the scat amplicon pool, and corrected scat proportions were calculated from the corrected sequence counts (later referred to as TCF-corrected scat sequences %).

Digestion correction factors

Our working hypothesis was that any bias that remained after accounting for methodological biases (involved in amplicon sequencing) and biological biases (differential mass-specific target gene copy number between prey species) was attributable to differential digestion of the prey species by the predator. Therefore, the difference between the tissue mix sequence proportions (which account for the aforementioned biases) and the scat sequence proportions should reflect any differential prey digestion—we thus calculated the inferred digestion correction factor (DCF) for each prey species using

$$\mathrm{DCF}_i = \frac{T_i}{S_i}$$

where S_i is the proportion of species i detected in the scat amplicon pool. DCF can only be calculated when both diet and TCFs are known for the particular consumer (which is not possible for field studies). We therefore compared the DCFs to the proximate composition of the prey fish to determine whether a composition component could be used as a proxy for the digestion bias (see Statistical analyses for details).

Statistical analyses

The correction factors were log-transformed to a linear scale prior to comparing them to the results of the proximate composition analysis. Thus, a fourfold correction factor in the positive direction would be 4.00 (or 0.60 when $\log_{10}$ transformed), and a fourfold correction factor in the negative direction equals 0.25 (or −0.60 when $\log_{10}$ transformed). We used coefficients of determination and p-values from general linear models to determine whether there were strong relationships between the $\log_{10}$ transformed correction factors and each component of the proximate composition analysis (i.e. % moisture, % ash, % protein, % lipid). The best-fitting models for the DCFs therefore indicated which properties of prey composition could potentially be used to independently calculate digestion correction factors.

We also compared the TCFs to the proximate composition data and published values of red/white muscle ratios in fishes. Our thought was that if there is a strong relationship between indicators of mitochondrial DNA density (e.g. red muscle ratio) and the TCFs, it would indicate that the methodological biases of the protocol are less influential than are differences in target gene copy number between fish species.

Table 1 Accounting of all sequences produced by Ion Torrent sequencing of the harbour seal scat amplicon pool and the tissue mix amplicon pool for three prey species (capelin, herring and mackerel)

	Scat pool		Tissue mix	
	Sequence count	Percent of total	Sequence count	Percent of total
Total sequences	64 831	100.0	36 393	100.0
Less than 100 bp	21 248	32.8	10 752	29.5
Homopolymer filtered	264	0.4	165	0.5
No primer match	9790	15.1	7219	19.8
No BLAST assignment	7290	11.2	3647	10.0
Forward Capelin	390	0.6	390	1.1
Forward Herring	10 464	16.1	4224	11.6
Forward Mackerel	3514	5.4	2778	7.6
Forward Harbour seal	142	0.2	0	0.0
Reverse Capelin	1237	1.9	1300	3.6
Reverse Herring	5757	8.9	2511	6.9
Reverse Mackerel	3987	6.1	3407	9.4
Reverse Harbour seal	748	1.2	0	0.0

Results

Sequencing and bioinformatics

The Ion Torrent sequencing run that included the scat amplicon pool, tissue mix amplicon pool and four unrelated amplicon pools produced a total of 311 635 amplicon reads or 33.6 Mbp of data. The quality of base calls was 13.7 Mbp Q20 bases, 17.6 Mbp Q17 bases and an average read length of 108 bp. Of the total reads, 64 831 were assigned to the MID for the scat amplicon pool and 36 393 were assigned to the MID for the tissue mix amplicon pool. A complete accounting of all sequences and species assignment for both amplicon pools is contained in Table 1, and all sequences have been deposited in Dryad in FASTQ format. For a discussion of the disparity between forward and reverse read counts, see Deagle *et al.* (2013).

After recalculating the diet proportions excluding the squid component, the expected proportions of sequences from the scat pool and the tissue mix pool were 48.5% capelin, 34.0% herring and 17.5% mackerel. However, after summing the assigned sequence counts for forward and reverse reads and converting these to proportions, neither amplicon pool matched the diet proportions (Table 2). In the tissue mix amplicon pool, capelin was highly under-represented (11.6%), while herring was moderately over-represented (46.1%), and mackerel was highly over-represented (42.3%) (Fig. 2). In the scat amplicon pool, capelin was even more under-represented (6.4%) than it was in the tissue mix amplicons, herring was more over-represented (64.0%), and mackerel was somewhat less over-represented (29.6%).

Table 2 Data used in the calculation of tissue correction factors (TCFs) and digestion correction factors (DCFs)

	Capelin	Herring	Mackerel
Diet biomass %	48.5	34.0	17.5
Tissue mix sequence count	1690	6735	6185
Tissue mix sequence %	11.6	46.1	42.3
Scat sequence count	1627	16221	7501
Scat sequence %	6.4	64.0	29.6
TCF	4.19	0.74	0.41
DCF	1.80	0.72	1.43
TCF corrected scat sequence %	31.2	54.7	14.2

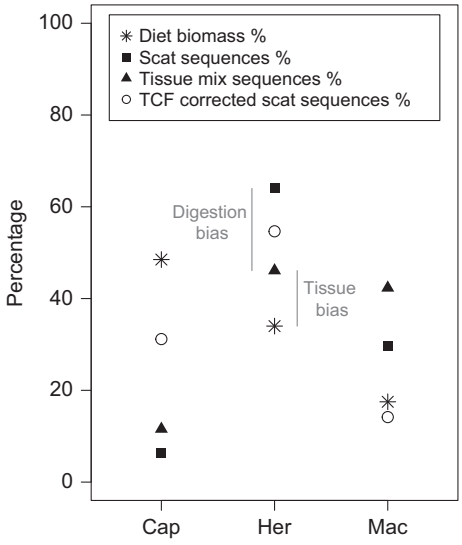

Fig. 2 Comparison between mass percentages of three fish species fed to seals (*) and sequence percentages obtained from scats (■) and the tissue mix (▲). The scat sequence percentage diet estimates adjusted with tissue correction factors (TCF) are also shown (○). For explanatory purposes, the magnitudes of the tissue bias and digestion bias are shown for herring (Cap = capelin, Her = herring, Mac = mackerel).

Tissue correction factors

Using the data from the tissue pool we calculated species-specific correction factors (TCFs) to adjust the sequence counts of the scat amplicon pool to take into account technical biases and differences in DNA density between fish species. The correction was largest for capelin (4.19, $\log_{10}$ transformed = 0.62), followed by mackerel (0.41, $\log_{10}$ transformed = −0.38), and then herring (0.74, $\log_{10}$ transformed = −0.13). Based on these correction factors, capelin is expected to be under-represented

Table 3 Proximate composition analysis results for the three prey fish in the feeding trial, displaying mean percentages and standard errors

	Capelin	Herring	Mackerel
Lipid	2.4 ± 0.9	9.8 ± 1.0	4.3 ± 0.4
Protein	13.4 ± 0.2	16.5 ± 0.3	18.0 ± 0.1
Ash	2.3 ± 0.1	2.5 ± 0.2	3.1 ± 0.2
Moisture	81.3 ± 0.8	71.8 ± 0.8	74.6 ± 0.4

in the scats and the other two species over-represented. This is in fact what we observed in the amplicons recovered from the scats before corrections. After applying the TCFs to the scat DNA sequence counts, the average difference between the percentages of prey DNA contained in the scats and the diet biomass percentages was substantially reduced from 28 ± 15% (uncorrected) to 14 ± 9% (TCF-corrected). The TCF-corrected scat percentages were as follows: capelin = 31.2%, herring = 54.7%, mackerel = 14.2% (Fig. 2). It is noteworthy that even after tissue correction, the scat sequence proportions did not correctly rank the importance of the different prey species in the diet.

Linear models showed a relatively strong negative relationship between the log-transformed TCFs and the percentage of protein (slope = −0.22, intercept = 3.56, $R^2 = 0.99$, $P = 0.05$; see Fig. 3). This indicates that higher protein fishes were over-represented in the tissue mix amplicon pool (Table 3). Weak relationships were observed between TCFs and the percent ash in prey (slope = −1.17, intercept = 3.12, $R^2 = 0.74$, $P = 0.34$), and the percent moisture in prey (slope = 0.09, intercept = −6.91, $R^2 = 0.74$, $P = 0.34$). No relationship was observed between TCFs and the percent lipid (slope = −0.07, intercept = 0.41, $R^2 = 0.25$, $P = 0.67$).

Digestion correction factors

The DCFs were generally smaller in magnitude than the TCFs, indicating that prey-specific digestion was the lesser source of bias in this study. Herring was again over-represented as a product of digestion bias (DCF = 0.72, $\log_{10}$ transformed = −0.14), and capelin was again highly under-represented (DCF = 1.80, $\log_{10}$ transformed = 0.26). Mackerel, however, which was strongly over-represented based on the tissue mix, produced a positive digestion correction (DCF = 1.43, $\log_{10}$ transformed = 0.16), indicating that it was under-represented as a result of digestion bias (Table 2). This result implies that in the case of mackerel, the two sources of bias identified (tissue bias and digestion bias) have opposite biasing effects.

A very strong relationship was detected between the log-transformed DCFs and the percent lipid content of the prey fishes when linear models were fit between the DCFs and the prey proximate composition components (slope = −0.05, intercept = 0.39, $R^2 = 1.00$, $P = 0.001$; Fig. 4). This indicates a negative relationship between prey fish lipid content and the $\log_{10}$ DCF (i.e. higher lipid prey fish require negative correction as a result of digestion bias and lower lipid prey fish require positive correction). A weaker relationship was also observed between the log-transformed DCFs and the percent moisture in prey (slope = 0.04, intercept = −2.71, $R^2 = 0.76$, $P = 0.32$). No relationship was observed between the transformed DCFs and the percent protein in prey fish (slope = −0.04, intercept = 0.70, $R^2 = 0.18$, $P = 0.72$) or the percent ash in prey fish (slope = 0.01, intercept = 0.05, $R^2 = 0.00$, $P = 0.99$).

As previously stated, the digestion correction factors were only calculated to evaluate whether there are suitable proxies for digestion bias in this study system. In this case, we chose not to apply the DCFs to scat sequence counts because they are only calculable when the diet of the consumer is known, and therefore not useful in the typical applications of the technique. However, the strong correlation between the DCFs and the lipid content of the prey fish indicates that a correction simply based on prey lipid percentage would exactly match the DCFs, and therefore would produce scat sequence percentages that perfectly estimate the diet when combined with TCFs.

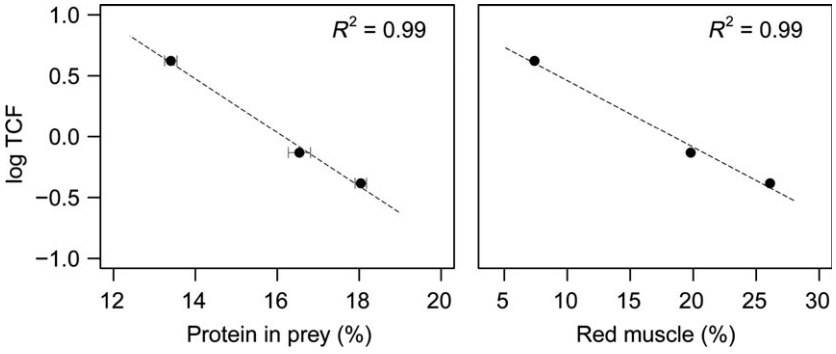

Fig. 3 The relationships between the log-transformed tissue correction factors (log TCF) and the percent whole body protein of the prey fish (Left), and between logTCF and the family-specific percentage of red muscle fibres documented in Greek-Walker & Pull (1974) (Right). Error bars represent standard error.

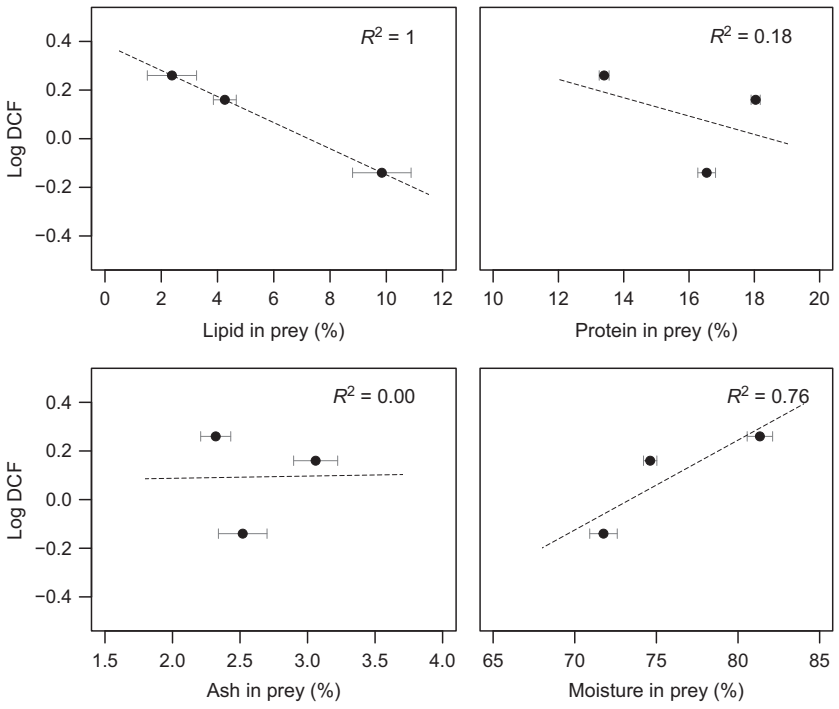

Fig. 4 The relationships between the log-transformed digestion correction factors (log DCF) and the proximate composition analysis components of the three prey species (top left = % lipid, top right = % protein, bottom left = % ash, bottom right = % moisture). Digestion correction factors calculated based on the inferred digestion bias (i.e. the difference between the scat sequence proportions and the tissue mix sequence proportions). Error bars represent standard error.

Discussion

In an ideal situation, dietary studies using next-generation sequencing to characterize diagnostic DNA markers from stomach contents or scats of consumers could assume a direct relationship between the sequence proportions of food items recovered and the proportions of food eaten. If this was the case, the relative importance of species in a consumer's diet could be determined with some certainty – the ultimate goal of most diet studies. Unfortunately, while past captive feeding studies have demonstrated there is a relationship between consumer diet and prey DNA quantity (i.e. scats of animals fed the same diet yield consistent prey sequence proportions), the sequence proportions do not accurately reflected the diet (Deagle *et al.* 2010, 2013). Thus, DNA diet techniques making use of sequence proportions can presently produce consistent but incorrect diet estimates.

Other DNA-based diet studies have taken a variety of approaches to avoid the problems involved in direct DNA quantification. Some researchers have chosen to focus on the overall diet breadth of consumers and identification of the prey field (e.g. Valdez-Moreno *et al.* 2012). This type of approach is robust when contaminants are minimal, and useful in situations where consumer's diet is poorly characterized; however, the level of information produced is not sufficient for many ecological investigations. An alternative approach is to calculate the percent frequency of occurrence of prey items (i.e. summarizing the proportion of samples containing a particular diet item). Frequency of occurrence summaries have been used to make comparisons between sampling sites (e.g. Kowalczyk *et al.* 2011; Shehzad *et al.* 2012) and between the diets of different species (e.g. Razgour *et al.* 2011). While occurrence summaries may be useful as a relative measure of the importance of food species for a consumer population, they have limited utility for the quantification of prey biomass. Furthermore, the importance of minor diet items is often exaggerated using occurrence indices, and small numbers of contaminating sequences or secondary predation DNA can have major impacts on diet estimates. Finally, some researchers have suggested that rather than dismissing the quantitative information contained in food DNA sequence counts, the proportions of those sequences can be useful for comparative studies or ranking of food species importance – even if sequence proportions do not accurately reflect diet biomass (see Pompanon *et al.* (2012)).

Our goal in the current work was to investigate the factors causing the mismatch between scat sequence proportions and diet biomass proportions, and to evaluate the feasibility of correcting for these biases using an approach that has been tested in other subfields. The biases herein likely result from multiple factors, including differential PCR amplification or sequencing of food species DNA, differences in template DNA density between food species and differences in survival of DNA during digestion. We isolated and examined

sources of bias by sequencing scat DNA from captive harbour seals fed known quantities of prey and a tissue mix of the same prey species. Proximate composition analysis of the prey allowed us to explore potential proxies for the isolated biases that could be used when the biases cannot be measured. Due to the limited scope of the feeding trial, and taxa represented, our study can be viewed as a hypothesis generating experiment designed to guide future research efforts in quantitative DNA diet analysis.

Food tissue control materials

The tissue mix we sequenced in parallel with the scat DNA should account for several sources of bias. First, the tissue mix should account for technical biases introduced during the methodological protocol such as the possibility of preferential primer binding or DNA synthesis in PCR and the possibility of selective amplicon sequencing (see Pompanon *et al.* (2012) for discussion). It should also correct for potential bias that would occur if different prey fish contain different densities of mitochondrial DNA in their tissue. In this case, fish that contain higher mtDNA density would yield more PCR amplicons due to increased template availability and would be over-represented by the sequence proportions relative to biomass proportions.

Based on sequences from the tissue mix amplicon pool, capelin DNA was highly under-represented, herring DNA was slightly over-represented and mackerel DNA was highly over-represented. This may indicate that there is a strong methodological bias against recovery of capelin sequences relative to mackerel sequences, or that capelin mtDNA density (i.e. amount of mtDNA per gram of tissue) is substantially lower than for mackerel. One piece of evidence suggesting mtDNA density is more important than methodological biases in the current study is the negative relationship between the tissue mix correction factors and the amount of whole body protein in fish tissue. This indicates that the over-represented fish (mackerel) is higher in protein content than the under-represented fish (capelin). The intuitive explanation is that increased levels of whole body protein are associated with high muscle density and therefore increased levels of mitochondrial DNA (Weatherley *et al.* 1998; López-Albors *et al.* 2008; Fernández-Vizarra *et al.* 2011).

However, the relationship may be more direct if we examined the ratio of red to white muscle fibres in the fish tissue, because red muscle has particularly high mitochondria density (Battersby & Moyes 1998). Chub mackerel belongs to the tuna family *Scombridae*, which is known for having a very high proportion of red muscle fibres and may explain why mackerel are over-represented in this data set. In a survey of red muscle content in marine fishes, the average percentage of red muscle from the fish families included in our study was 7.4% for *Osmeridae*, 19.8% for *Clupeidae* and 26.1% for *Scombridae* (Greek-Walker & Pull 1974). Plotting these red muscle percentages against the tissue correction factors shows virtually the same relationship we observed between the tissue correction factors and protein percentage (slope = −0.05, intercept = 1.01, R^2 = 0.99, P = 0.06; Fig. 3).

In this specific study system, it may be possible to correct for the mtDNA tissue biases simply by taking advantage of the linear relationship between red muscle percentages, or protein percentages and the TCFs. However, correction based solely on a proxy for mtDNA density would not account for the methodological biases also captured by the TCFs and therefore only useful in situations such as this where methodological biases appear to be minimal.

It is often inconvenient in DNA-based diet studies to account for variable gene copy number between prey species or tissues, and it could be possible to mitigate the problem by targeting a single copy genomic DNA marker instead of a mitochondrial gene. However, this approach would only be effective if cell density (and therefore the marker density) is more consistent between food species than mtDNA density. Furthermore, a single copy genomic marker is much less likely to amplify from a scat sample due to the degradation of prey DNA during the digestive process. Therefore, it appears worthwhile to continue pursuing creative methods for dealing with variability in gene copy number between food species, despite the challenges that it poses.

Due to the inherent variability involved in amplicon sequencing diet analysis, a logical next step is to begin sequencing food tissue mixes in other study systems to better understand the magnitudes of system-specific biases. This approach will make it clear whether a quantitative interpretation of amplicon sequence proportions is justified and/or accurate for each study system. In our study, we knew the consumers' diet and could therefore create a tissue mix that corresponded directly to the expected sequence proportions of the scat samples. Studies of wild animals will require an alternative approach. One possibility could be to create a set of tissue mix standards for the consumer, in which 50% of each tissue mix is made up of a variable food species that occurs in the diet, and 50% is made up of a control species that is common to all of the standards. For example, using pollock as a control species, we could create three tissue mix standards for this study system: (50% capelin, 50% pollock), (50% herring, 50% pollock) and (50% mackerel, 50% pollock). In this case, any deviance in the variable fish sequence proportions from 50%

would be indicative of a species-specific bias, and the difference could be used to create a species correction factor. In cases when there are many different food species, a representative of each food family could potentially be used for the tissue mix standards to create family-specific corrections. The use of two species in equal proportions should increase the accuracy of correction factors as deviations can be measured more accurately when a food item is not a minor component of the mix. However, this design would not account for any potential interactive effects between food species DNA.

The effectiveness of a food tissue mixture for bias correction is reliant on the tissue mix and scats being treated identically during the methodological protocol. While we maintained consistency in most aspects of our protocol, it is important to note that the two amplicon pools (scats and tissue mix) received different MID sequences during sequencing adapter ligation, which we used to bioinformatically differentiate between amplicon pool sequences. The MIDs may have biased the sequence proportions between the amplicon pools, although preliminary work suggested that MID bias is not highly influential in this study system. Future investigations will determine the preferred approach to differentiate between sequences of different amplicon pools, while minimizing potential biases.

Proxies for digestion bias

The digestion correction factors we derived in this study were based on the bias introduced by differential prey species digestion, which we defined as the difference between the tissue mix proportions (that account for methodological biases and template DNA density) and the scat DNA sequence proportions. Using this approach, it is only possible to calculate digestion bias when consumer diet is known and a tissue mix has also been sequenced with scat samples. Compared with the TCFs, we found the DCFs were relatively small in magnitude, indicating that the digestion bias was the lesser of the two sources of bias and had a smaller impact on proportional diet estimates. This is counter to a previous captive feeding study which determined that digestion bias is likely the largest source of bias in the DNA-based quantification of little penguin diet (Deagle *et al.* 2010). These conflicting results suggest there may be large variation in the impacts of biasing factors between study systems.

In the current study, we detected a strong negative relationship between the digestion bias correction factor and the percentage of lipid in the prey fish tissues. This implies that high lipid content in the fish consumed is associated with reduced breakdown of fish tissue during the digestion process, thereby preventing mtDNA degradation. Two independent harbour seal digestion studies lend support for this idea (Stanberry 2003; Trumble *et al.* 2003). In these studies, captive harbour seals were fed fish species of differing lipid content, and proximate composition analysis was performed on both the prey and the resultant scats to calculate component digestibility. Both studies found a reduction in protein digestibility with increased lipid content of the prey fish, which likely results in diminished tissue DNA degradation (Fig. 5).

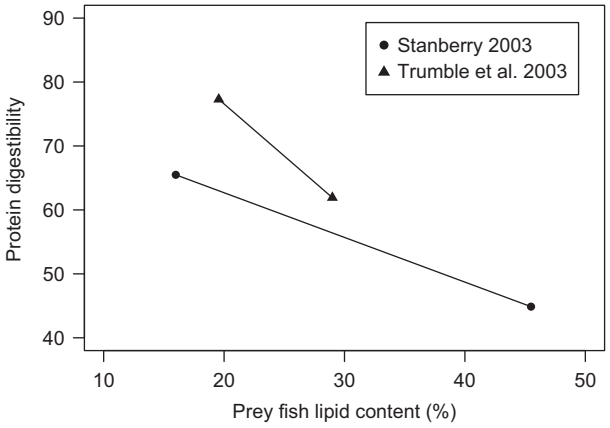

Fig. 5 The relationship between prey fish lipid content and protein digestibility in harbour seals. Data are from two separate digestive efficiency studies in which captive seals were fed fishes of varying lipid content (Stanberry 2003; Trumble *et al.* 2003).

In our experiment, a correction factor derived from the relationship between prey percent lipid and DCFs would make it possible to generate a perfect average diet estimate from the scat sequences. If additional work validates this hypothesis for harbour seals, it will be necessary to evaluate the natural variability in prey fish lipid content, which can fluctuate both seasonally and geographically. Despite this variability, it may be possible in the future to create a categorical correction factor for lipid that improves diet estimate accuracy (e.g. for high, medium and low lipid prey) if the order of lipid percentages is relatively consistent for prey species (e.g. herring > mackerel > capelin, etc.). A similar approach to this has been used to correct for the effects of digestion on the sizes of fish otoliths recovered from pinniped scats (Tollit *et al.* 2004).

Applicability to other study systems

Although the observed relationships between biases and their potential proxies are likely to be specific to this study system, the overall study design and research approach are certainly generalizable to other systems.

The sequencing of food tissue control materials alone can indicate the degree to which quantitative diet estimates based on DNA sequence counts may be biased by factors such as PCR bias and variable template DNA density. In cases where PCR primer binding sites vary considerably between target species (e.g. Razgour *et al.* 2011), or when blocking probes may impact amplification of some prey (Piñol *et al.* 2013), food tissue experiments are particularly relevant in order to assess these potentially strong technical biases. Similarly, this type of analysis seems important when gene copy number varies considerably between target species (Darby *et al.* 2013). If the use of control materials is combined with a captive feeding study, food-specific digestion biases can be deduced in other model systems, and food properties that may influence digestion can be assessed. Clearly, substantial additional work must be conducted before we can confidently use DNA sequence count data to infer food biomass proportions from diet samples. However, this study presents a rational framework to begin identifying the most important sources of bias in each study system and testing creative ways to correct for those biases.

Conclusions

DNA-based diet analysis is a rapidly evolving methodology that offers substantial advantages over existing diet techniques and is being used to address heretofore unanswerable questions in trophic ecology. While the speed and taxonomic accuracy of the methods are clear, the limitations of available tools and potential to collect accurate quantitative data have not been thoroughly examined. Using prey tissue mixes and captive harbour seals fed a known diet, we were able to quantify substantial biases introduced by differences in template DNA copy number between prey species and biases attributable to differential prey digestion. The correction factors we used to account for those sources of bias considerably improved the diet estimates, suggesting that accurate diet estimates can be obtained using this approach. Tissue corrections could feasibly be developed in almost any dietary study using a set of standards derived from food tissue mixes that are sequenced in parallel with diet samples. We have also shown the possibility that proxies based on prey attributes might account for species-specific differences in survival of DNA during digestion. The extent to which differential food digestion affects quantitative diet estimates from amplicon sequences will need to be further evaluated using captive feeding trials in multiple study systems. Given the wide adoption of next-generation sequencing as an approach to study the diets of various taxa, the potential to obtain accurate quantitative data-based on sequence counts deserves further investigation.

Acknowledgements

We thank the Point Defiance Zoo and Aquarium and their staff (Amanda K. Shaffer, Lisa Triggs, Chris Harris, Cindy Roberts and Amanda Chomos) and volunteers for conducting the feeding trials and collecting the seal scats. We also thank Andrea Polanowski for helpful guidance in the laboratory and Stephen Trumble for providing data from previous seal feeding trials. David Rosen, Jerome Spitz, Monique Lance, Carl Walters and Anthony Sinclair provided guidance on the study design and manuscript edits. Two anonymous reviewers and Dr. James Harwood provided constructive feedback on an early version of this manuscript. Funding was provided by the Pacific Salmon Foundation and the Australian Antarctic Division. All animal-related activities were undertaken in accordance with UBC Animal Care Committee guidelines.

References

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

Battersby BJ, Moyes CD (1998) Influence of acclimation temperature on mitochondrial DNA, RNA, and enzymes in skeletal muscle. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **275**, R905–R912.

Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, **77**, 7846–7849.

Blankenberg D, Kuster GV, Coraor N *et al.* (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, Chapter 19, Unit 19.10.1–19.10.21.

Bowen WD, Iverson SJ (2012) Methods of estimating marine mammal diets: a review of validation experiments and sources of bias and uncertainty. *Marine Mammal Science*, doi: 10.1111/j.1748-7692.2012.00604.x.

Bowles E, Schulte PM, Tollit DJ, Deagle BE, Trites AW (2011) Proportion of prey consumed can be determined from faecal DNA using real-time PCR. *Molecular Ecology Resources*, **11**, 530–540.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Cheung M-S, Down TA, Latorre I, Ahringer J (2011) Systematic bias in high-throughput sequencing data and its correction by BEADS. *Nucleic Acids Research*, **39**, e103–e103.

Cottrell PE, Trites AW, Miller EH (1996) Assessing the use of hard parts in faeces to identify harbour seal prey: results of captive-feeding trials. *Canadian Journal of Zoology*, **74**, 875–880.

Darby BJ, Todd TC, Herman MA (2013) High-throughput amplicon sequencing of rRNA genes requires a copy number correction to accurately reflect the effects of management practices on soil nematode community structure. *Molecular Ecology*, **22**, 5456–5471.

Deagle BE, Tollit DJ (2007) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conservation Genetics*, **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. *Molecular Ecology* **14**, 1831–1842.

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.

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

Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN (2013) Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources*, **13**, 620–633.

Duffy DC, Jackson S (1986) Diet studies of seabirds: a review of methods. *Colonial Waterbirds*, **9**, 1–17.

Fernández-Vizarra E, Enríquez JA, Pérez-Martos A, Montoya J, Fernández-Silva P (2011) Tissue-specific differences in mitochondrial activity and biogenesis. *Mitochondrion*, **11**, 207–213.

Gales N, Cheal A (1992) Estimating diet composition of the Australian sea lion (*Neophoa cinerea*) from scat analysis: an unrliable technique. *Wildlife Research*, **19**, 447–455.

Giardine B, Riemer C, Hardison RC *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, **15**, 1451–1455.

Goecks J, Nekrutenko A, Taylor J, Team TG (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, **11**, R86.

Greek-Walker M, Pull G (1974) A survey of red and white muscle in marine fish. *Journal of Fish Biology*, **7**, 295–300.

Greenstone MH, Szendrei Z, Payton ME *et al.* (2010) Choosing natural enemies for conservation biological control: use of the prey detectability half-life to rank key predators of Colorado potato beetle. *Entomologia Experimentalis et Applicata*, **136**, 97–107.

Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.

Huggett JF, Laver T, Tamisak S *et al.* (2013) Considerations for the development and application of control materials to improve metagenomic microbial community profiling. *Accreditation and Quality Assurance*, **18**, 77–83.

Jiang L, Schlesinger F, Davis CA *et al.* (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, **21**, 1543–1551.

Kauserud H, Kumar S, Brysting AK, Nordén J, Carlsen T (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, **22**, 309–315.

Kembel SW, Wu M, Eisen JA, Green JL (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, **8**, e1002743.

Kowalczyk R, Taberlet P, Coissac E *et al.* (2011) Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *Forest Ecology and Management*, **261**, 821–828.

Leal MC, Nejstgaard JC, Calado R, Thompson ME, Frischer ME (2014) Molecular assessment of heterotrophy and prey digestion in zooxanthellate cnidarians. *Molecular Ecology*, **23**, 3838–3848.

Lindeman RL (1942) The trophic-dynamic aspect of ecology. *Ecology*, **23**, 399–417.

López-Albors O, Abdel I, Periago MJ *et al.* (2008) Temperature influence on the white muscle growth dynamics of the sea bass *Dicentrarchus labrax*, L. Flesh quality implications at commercial size. *Aquaculture*, **277**, 39–51.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.

Murray DC, Bunce M, Cannell BL *et al.* (2011) DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS ONE*, **6**, e25776.

Phillips EM, Harvey JT (2009) A captive feeding study with the Pacific harbor seal (*Phoca vitulina richardii*): implications for scat analysis. *Marine Mammal Science*, **25**, 373–391.

Piñol J, San Andrés V, Clare EL, Mir G, Symondson WOC (2013) A pragmatic approach to the analysis of diets of generalist predators: the use of next-generation sequencing with no blocking probes. *Molecular Ecology Resources*, DOI: 10.1111/1755-0998.12156.

Pomeroy LR (1974) The ocean's food web, a changing paradigm. *BioScience*, **24**, 499–504.

Pompanon F, Deagle BE, Symondson WO *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.

Quail MA, Smith M, Coupland P *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

Razgour O, Clare EL, Zeale MRK *et al.* (2011) High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and Evolution*, **1**, 556–570.

Scheffer TH, Sperry CC (1931) Food habits of the Pacific harbor seals, *Phoca vitulina richardsi*. *Journal of Mammalogy*, **12**, 214–226.

Shehzad W, Riaz T, Nawaz MA *et al.* (2012) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, 1951–1965.

Sipos R, Székely AJ, Palatinszky M *et al.* (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, **60**, 341–350.

Stanberry K (2003) *The Effect of Changes in Dietary Fat Level on Body Composition, Blood Metabolites and Hormones, Rate of Passage, and Nutrient Assimilation Efficiency in Harbor Seals*, University of Hawaii, Manoa.

Tollit DJ, Steward MJ, Thompson PM *et al.* (1997) Species and size differences in the digestion of otoliths and beaks: implications for estimates of pinniped diet composition. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 105–119.

Tollit D, Heaslip S, Zeppelin T *et al.* (2004) A method to improve size estimates of walleye pollock (*Theragra chalcogramma*) and Atka mackerel (*Pleurogrammus monopterygius*) consumed by pinnipeds: digestion correction factors applied to bones and otoliths recovered in scats. *Fishery Bulletin*, **102**, 498–508.

Trumble SJ, Barboza PS, Castellini MA (2003) Digestive constraints on an aquatic carnivore: effects of feeding frequency and prey composition on harbor seals. *Journal of Comparative Physiology B, Biochemical, Systemic, and Environmental Physiology*, **173**, 501–509.

Valdez-Moreno M, Quintal-Lizama C, Gómez-Lozano R, del Carmen García-Rivas M (2012) Monitoring an alien invasion: DNA barcoding and the identification of lionfish and their prey on coral reefs of the Mexican Caribbean. *PLoS ONE*, **7**, e36636.

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. *Frontiers in Zoology*, **5**, 12.

Weatherley A, Gill H, Lobo A (1998) Recruitment and maximal diameter of axial muscle fibres in teleosts and their relationship to somatic growth and ultimate size. *Journal of Fish Biology*, **33**, 851–859.

Zook JM, Samarov D, McDaniel J, Sen SK, Salit M (2012) Synthetic spike-in standards improve run-specific systematic error analysis for DNA and RNA sequencing. *PLoS ONE*, **7**, e41356.

A.C.T. and B.E.D. designed research, performed study and wrote the paper; A.C.T. data analysis; S.N.J. conceptual feedback, provided analytical tools, contributed reagents; K.H.H. conceptual feedback, data analysis and interpretation; A.W.T. provided funding, facilitated collaborations, manuscript edits.

Data accessibility

Sequence data (FASTQ) for both amplicon pools and raw data used to create Figs 2–5 have been archived in Dryad repository (doi:10.5061/dryad.h3k4n).