



OPEN

DATA DESCRIPTOR

Data on the diets of Salish Sea harbour seals from DNA metabarcoding

Austen C. Thomas^{1,2}✉, Bruce Deagle^{3,4}, Chad Nordstrom⁵, Sheena Majewski⁵, Benjamin W. Nelson^{2,10}, Alejandro Acevedo-Gutiérrez⁶, Steven Jeffries⁷, Jed Moore⁸, Amelia Loudon⁷, Hassen Allegue^{2,9}, Scott Pearson⁷, Michael Schmidt¹⁰ & Andrew W. Trites²

Marine trophic ecology data are in high demand as natural resource agencies increasingly adopt ecosystem-based management strategies that account for complex species interactions. Harbour seal (*Phoca vitulina*) diet data are of particular interest because the species is an abundant predator in the northeast Pacific Ocean and Salish Sea ecosystem that consumes Pacific salmon (*Oncorhynchus* spp.). A multi-agency effort was therefore undertaken to produce harbour seal diet data on an ecosystem scale using, 1) a standardized set of scat collection and analysis methods, and 2) a newly developed DNA metabarcoding diet analysis technique designed to identify prey species and quantify their relative proportions in seal diets. The DNA-based dataset described herein contains records from 4,625 harbour seal scats representing 52 haulout sites, 7 years, 12 calendar months, and a total of 11,641 prey identifications. Prey morphological hard parts analyses were conducted alongside, resulting in corresponding hard parts data for 92% of the scat DNA samples. A custom-built prey DNA sequence database containing 201 species (192 fishes, 9 cephalopods) is also provided.

Background & Summary

In recent years the fisheries and ecosystem modelling communities have expressed increased interest in harbour seal (*Phoca vitulina*) diet information to help inform prey consumption estimates and ecosystem models^{1–3}. Harbour seals are a common pinniped found in the northeast Pacific Ocean and are particularly abundant in the inland marine waters of southern British Columbia (Canada) and Washington state (USA) – an area defined as the Salish Sea^{4,5}. Growth in the population of seals since their protection in 1972 has led to concerns about their potential predatory impacts on fishes of conservation concern such as Pacific salmon (*Oncorhynchus* spp.)⁶. In addition to being known salmon predators, the inverse population trend between seals and salmon species has led to speculation that seals may be a causal factor influencing salmon populations^{7,8}. More specifically, the marine survival of Chinook (*O. tshawytscha*), coho (*O. kisutch*), and steelhead (*O. mykiss*) salmon in the region decreased dramatically during the same time period when harbour seals grew exponentially^{7,9,10}. However, such correlative evidence is generally not sufficient to mandate management actions. Additional data, such as detailed predator diet information, is needed to establish a potential ecological link between seals and fish populations^{2,11}.

Although harbour seal diet in the region has been studied for nearly a century¹², methodological limitations have prevented accurate quantification of the salmon proportion of seal diet. Previous studies relied primarily on morphological identification of hard remains (e.g., bones) in stomachs or scat samples, and diet was summarized based on the proportion of samples containing bones of certain prey species^{13,14}. Unfortunately, the majority of salmon bones in scats cannot easily be identified to the species level, and simple prey occurrence data are

¹Molecular Division, Smith-Root Inc., 16603 NE 50th Avenue, Vancouver, WA, 98686, USA. ²University of British Columbia, Institute for the Oceans and Fisheries, 247-2202 Main Mall, Vancouver, BC V6T 1Z4, British Columbia, Canada. ³Australian National Fish Collection, CSIRO National Research Collections Australia, Hobart, TAS, Australia. ⁴Australian Antarctic Division, Channel Highway, Kingston, TAS, Australia. ⁵Pacific Biological Station, 3190 Hammond Bay Road, Fisheries and Oceans Canada, Nanaimo, BC, V9T 6N7, Canada. ⁶Department of Biology, Western Washington University, Bellingham, WA, 98225, USA. ⁷Washington Department of Fish and Wildlife, Olympia, WA, 98501, USA. ⁸Nisqually Indian Tribe, Olympia, WA, 98513, USA. ⁹Université du Québec à Montréal, Département des Sciences Biologiques, 141 Avenue du Président-Kennedy, Montréal, H2X 1Y4, QC, Canada. ¹⁰Long Live the Kings, 1326 5th Avenue, 450, Seattle, Washington, 98101, USA. ✉e-mail: athomas@smith-root.com

of limited use for quantitative analyses of predation. Such estimates generally require (as an input) the predator population diet fraction comprised of a particular prey species^{1,2}. A new harbour seal diet method was therefore needed to meet the data requirements of regional modelling efforts.

In 2011, collaborating researchers at the University of British Columbia (UBC) and the Australian Antarctic Division undertook an effort to create a new diet analysis method capable of providing both the salmon species and the proportional biomass of prey contained in harbour seal scat samples. At the time, analysis of predator diet using DNA metabarcoding methods was a new approach that offered the potential to provide both the seal prey species and the relative proportions of prey in scat samples¹⁵. High-resolution taxonomic data can be achieved with DNA metabarcoding using massively parallel sequencing of short (e.g., 200 bp), diagnostic genetic markers amplified from scat samples^{15,16}. Sequences are compared to a database containing known DNA barcodes of potential prey species, and the proportions of DNA sequences assigned to different prey are used as an index of proportional biomass composition for each scat sample¹⁶. The product of the described R&D effort was a DNA metabarcoding diet analysis method for harbour seals that has since been applied to thousands of wild harbour seal scat samples collected throughout the Salish Sea^{8,17}.

Between the years of 2011 and 2019, over four-thousand regional harbour seal scat samples were collected and processed using our standardized DNA metabarcoding diet analysis method. This is the product of a collaborative transboundary research effort including two universities (University of British Columbia, Western Washington University), three government agencies (Washington Department of Fish and Wildlife, Fisheries and Oceans Canada, Australian Antarctic Division), a native American tribe (Nisqually Indian Tribe), a non-profit organization (Long Live the Kings) and private corporation (Smith-Root Inc.). Subsets of the data have been used to address specific questions with respect to pinniped predation on particular prey species or taxa^{1,2,8}. Further, the diet analysis method produces proportional data for all prey species amplified by the semi-universal PCR primers; therefore, these data products are also useful for a broad range of ecological questions and modelling exercises that extend beyond the previous taxon-specific inquiries.

Our objective in publishing this dataset is to make these valuable trophic ecology data broadly available to the fisheries and ecosystem modelling communities. The core of the dataset is a single spreadsheet containing data from 4,625 harbour seal scats representing 52 haulout sites, 7 years, 12 calendar months, and totaling 11,641 prey identifications. We hope that by making these data publicly available they will help facilitate an open and free exchange of ideas regarding harbour seal trophic ecology in the Salish Sea.

Methods

Scat sample collection and preparation. At known harbour seal haulout sites individual scat samples were collected using a standardized protocol (Fig. 1). Disposable wooden tongue depressors were used to transfer deposited scats into 500 ml single-use jars or zip-style bags lined with 126 µm nylon mesh paint strainers¹⁸. Samples were either preserved immediately in the field by adding 300 ml 95% ethanol to the collection jar, or were taken to the lab and frozen at -20°C within 6 hours of collection¹⁹. Later, samples were thawed and filled with ethanol before being manually homogenized with a disposable wooden depressor inside the paint strainer to separate the scat matrix material from hard prey remains (e.g. bones, cephalopod beaks). The paint strainer containing prey hard parts was then removed from the jar leaving behind the ethanol preserved scat matrix for genetic analysis²⁰. The paint strainer containing prey hard parts was refrozen for subsequent parallel morphological prey ID.

Molecular laboratory processing. Scat matrix samples were subsampled (approximately 20 mg), centrifuged and dried to remove ethanol prior to DNA extraction. DNA was extracted from scat with the QIAGEN QIAamp DNA Stool Mini Kit according to the manufacturer's protocols. For additional details on the extraction process see Deagle *et al.*²¹ and Thomas *et al.*²⁰.

The metabarcoding marker we used to quantify fish and cephalopod proportions was a 16S rDNA fragment (~260 bp) previously described in Deagle *et al.*¹⁵ for pinniped scat analysis. We used the combined Chord/Ceph primer sets: Chord_16S_F (GATCGAGAAGACCCTRTGGAGCT), Chord_16S_R (GGATTGCGCTGTTATCCCT), and Ceph_16S_F (GACGAGAAGACCCTAWTGAGCT), Ceph_16S_R (AAATTACGCTGTTATCCCT). This multiplex PCR reaction is designed to amplify both chordate and cephalopod prey species DNA. A blocking oligonucleotide was included in the all 16S PCRs to limit amplification of seal DNA²². The oligonucleotide (32 bp: ATGGAGCTTTAATACTAACAACAGAGCA-C3) matches harbour seal sequence (GenBank Accession AM181032) and was modified with a C3 spacer so it is non-extendable during PCR²².

A secondary metabarcoding marker was used in a separate PCR reaction to quantify the salmon portion of seal diet, because the primary 16S marker was unable to reliably differentiate between coho and steelhead DNA sequences. This marker was a COI "minibarcode" specifically for salmonids within the standard COI barcoding region: Sal_COI_F (CTCTATTTAGTATTTGGTGCCTGAG), Sal_COI_R (GAGTCAGAAGCTTATGTTTRTTTATTCG). The COI amplicons were sequenced alongside 16S such that the overall salmonid fraction of the diet was quantified by 16S, and the salmon species proportions within that fraction were quantified by COI.

To take full advantage of sequencing throughput, we used a two-stage labeling scheme to identify individual samples that involved both PCR primer tags and labeled MiSeq adapter sequences. The open source software package EDITTAG was used to create 96 primer sets each with a unique 10 bp primer tag and an edit distance of 5; meaning that to mistake one sample's sequences for another, 5 insertions, substitutions or deletions would have to occur²³.

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: 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.

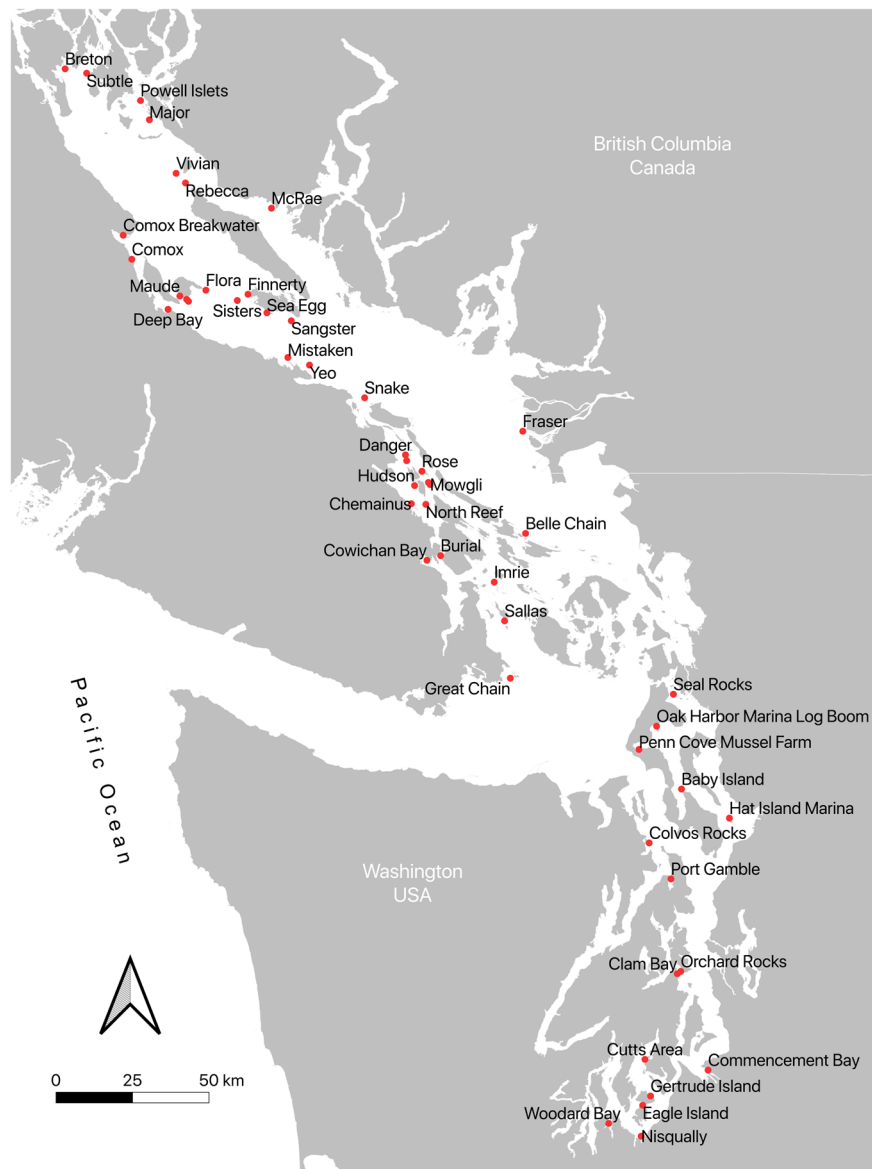


Fig. 1 The 52 harbour seal scat collection sites in the Salish Sea represented in this dataset.

Amplicons from 96 individually labeled samples were pooled by running all samples on 1.5% agarose gels, and the luminosity of each sample's PCR product was quantified using Image Studio Lite (Version 3.1). To combine all samples in roughly equal proportion (normalization), we calculated the fraction of each sample's PCR product added to the pool based on the luminosity value relative to the brightest band. After 2013, amplicon normalization was performed using SequalPrep™ Normalization Plate Kits, 96-well.

Sequencing libraries were prepared from pools of 96 samples using an Illumina TruSeq DNA sample prep kit which ligated uniquely labeled adapter sequences to each pool. Libraries were then pooled and DNA sequencing was performed on Illumina MiSeq using the MiSeq Reagent Kit v2 (300 cycle) for SE 300bp reads. Samples were sequenced on multiple different runs as part of the larger study; however, typically between 4 and 6 libraries (each a pool of 96 individually identifiable samples) were sequenced on a single MiSeq run.

Bioinformatics. To assign DNA sequences to a fish or cephalopod species, we created a custom BLAST reference database of 16S sequences by an iterative process. First, using a list of the fish species of Puget Sound, we searched Genbank for the 16S sequence fragment of all fishes known to occur in the region (71 fish families 230 species)^{24,25}. Reference sequences for each prey species were included in the database if the entire fragment was available, and preference was given to sequences of voucher specimens. When the database was first generated (November, 2012) Genbank contained 16S sequences for 192 of the 230 fish species in the region, and the remaining 38 species were mostly uncommon species unlikely to occur in seal diets. Following a similar procedure, we added to this database sequences for all of the regional cephalopods for which 16S data were available (7 squid species, 2 octopus species). A separate reference database was generated for the COI salmon marker

containing Genbank sequences for the nine salmonid species known to occur regionally: *Oncorhynchus gorbuscha* (Pink Salmon), *Oncorhynchus keta* (Chum Salmon), *Oncorhynchus kisutch* (Coho Salmon), *Oncorhynchus mykiss* (Steelhead), *Oncorhynchus nerka* (Sockeye Salmon), *Oncorhynchus tshawytscha* (Chinook Salmon), *Oncorhynchus clarkii* (Cutthroat Trout), *Salmo salar* (Atlantic Salmon), *Salvelinus malma* (Dolly Varden)²⁴.

To determine if some species in the database cannot be distinguished from each other at 16S (i.e. have identical sequences in the reference database) a distance matrix was performed on the complete database using the DistanceMatrix function in the R package DECIPHER²⁶. Species with identical sequences were identified as having a distance of “0.00”. In some cases, one haplotype for a species was identical to another species but other haplotypes were not. When two species’ sequences were identical, we ultimately reported both species in the prey_ID field.

Sequences were automatically sorted (MiSeq post processing) by amplicon pool using the indexed TruSeq™ adapter sequences. FASTQ sequence files for each library were imported into MacQIIME (version 1.9.1-20150604) for demultiplexing and sequence assignment to species²⁷. For a sequence to be assigned to a sample, it had to match the full forward and reverse primer sequences and match the 10 bp primer tag for that sample (allowing for up to 2 mismatches in either primers or tag sequence).

Next, we clustered the DNA sequences that were assigned to scat or tissue samples with USEARCH (similarity threshold = 0.99; minimum cluster size = 3; de novo chimera detection), and entered a representative sequence from each cluster into a GenBank nucleotide BLAST search^{28,29}. If the top matching species for any cluster was not included in the existing database (or the sequence differed indicating haplotype variation), we put the top matching entry in the reference database. We repeated this procedure with every new batch of sequence data to minimize the potential for incorrect species assignment or prey species exclusion. This process was conducted for both the 16S and COI reference databases with each new batch of samples.

For all DNA sequences successfully assigned to a sample, a BLAST search was performed against our custom 16S or COI reference databases. A sequence was assigned to a species based on the best match in the database (threshold BLASTN e-value < 1e-20 and a minimum identity of 0.9), and the proportions of each species’ sequences were quantified by individual sample after excluding harbour seal sequences or any identified contaminants²⁷. Samples were excluded from subsequent analysis if they contained <10 identified prey DNA sequences (given the current costs of DNA sequencing, a higher threshold is now advisable). Harbour seal DNA diet percentages for individual scats were then calculated using the Relative Read Abundance (RRA) calculation commonly used in metabarcoding studies (Box 1)¹⁶. The RRA formula was used to calculate the “DNA_diet_percent” field in data record “Harbour_Seal_DNA_Diet_Data.csv”.

Box 1 (From Deagle *et al.*¹⁶). Metrics used to summarise sequence data in dietary studies.

Occurrence Data

Frequency of occurrence (FOO) is the number of samples that contain a given food item, most often expressed as a percent (%FOO). Percent of occurrence (POO) is simply %FOO rescaled so that the sum across all food items is 100%. Weighted percent of occurrence (wPOO) is similar to POO, but rather than giving equal weight to all occurrences, this metric weights each occurrence according to the number of food items in the sample (e.g., if a sample contains 5 food items, each will be given weight 1/5). Mathematical expressions are as follows:

$$\%FOO_i = \frac{1}{S} \sum_{k=1}^S I_{i,k} \times 100\%$$

$$POO_i = \frac{\sum_{k=1}^S I_{i,k}}{\sum_{i=1}^T \sum_{k=1}^S I_{i,k}}$$

$$wPOO_i = \frac{1}{S} \sum_{k=1}^S \frac{I_{i,k}}{\sum_{i=1}^T I_{i,k}}$$

where T is the number of food items (taxa), S is the number of samples, and I is an indicator function such that $I_{i,k} = 1$ if food item i is present in sample k , and 0 if not.

Many metabarcoding diet studies make use of both %FOO and POO e.g.⁴². POO provides a convenient view since each food taxon contributes a percentage of total diet (unlike %FOO which does not sum to 100%). In POO summaries samples with a high number of food taxa have a stronger influence, whereas in wPOO each sample is weighted equally (i.e. lower weighting to food taxa in a mixed meal) and this may be more biologically realistic. wPOO is the same as split-sample frequency of occurrence; see⁴³ and references within.

Read Abundance Data

Using the sequence counts, relative read abundance (RRA_i) for food item i is calculated as:

$$RRA_i = \frac{1}{S} \sum_{k=1}^S \frac{n_{i,k}}{\sum_{i=1}^T n_{i,k}} \times 100\%$$

where $n_{i,k}$ is the number of sequences of food item i in sample k .

Prey hard parts analysis. Extraction and identification of hard structures from harbour seal scats was conducted by three different analysts. We used the “all structures” approach to identify harbor seal prey contained in individual scat samples, which make our results comparable to similar studies previously conducted in the region^{8,13,14}. Prey “hard parts” retained in paint strainers were cleaned of debris using either a conventional washing machine or nested sieves. All diagnostic prey hard parts were identified to the lowest possible taxon using a dissecting microscope and reference fish bones from Washington and British Columbia, in addition to published keys for fish bones and cephalopod beaks^{30–33}. Samples containing prey hard parts identifiable only to the family level (e.g., Clupeidae), and bones identifiable to the species level of the same family (e.g., Pacific herring, *Clupea pallasii*) were both tallied.

In previous studies of harbour seal impacts to juvenile salmonids in the Salish Sea^{1,2,8,34}, these diagnostic hard structures (e.g., otoliths, bones) were combined with DNA extracted from each scat sample to estimate the proportion of juvenile and adult salmon (by species) in the seal diet. This approach (see: Thomas *et al.*⁸) integrates separate analyses of hard parts and DNA through an algorithm that apportions the salmonid DNA component in each sample to a “juvenile” or “adult” classification. The decision algorithm is based on the co-occurrence of age-classified salmon bones and salmon DNA in samples, and (when bones were not present but salmon DNA was detected) on known seasonal life-history information. For example, an individual scat sample found to contain 5% Chinook salmon, and a 1:1 ratio of juvenile to adult salmon bones, would be disaggregated into a final classification of 2.5% juvenile Chinook salmon and 2.5% adult Chinook salmon. For individual scat samples that do not contain diagnostic hard structures, the ratio of juveniles to adults in that sample would rely on the ratio of hard parts pooled for the collection month. If no hard structures are available for the collection month—which only occurred for 7% of samples in Thomas *et al.*⁸ a seasonal classification would then be applied to the sample (spring = juvenile, fall = adult). The classification of hard parts as “juvenile” or “adult” was performed by taxonomic experts who differentiated samples visually, and/or according to otolith or vertebral measurements (e.g., Nelson *et al.*³⁴).

Data Records

A summary figure depicting the average diet of the complete Salish Sea harbour seal dataset comprised of 4,625 scat samples is available in Fig. 2.

The following files are available at figshare³⁵.

Harbour_seal_dna_diet_data.csv. This is the primary dataset that contains scat sample composition data for 4,625 harbor seal scats. Field names and descriptions are outlined in Table 1. Multiple entries exist for each sample ID, with each entry representing an identified prey species and its proportional composition within the sample.

Notes for Harbour_Seal_DNA_Diet_Data.csv:

The sample ID nomenclature varied between projects and we present here the IDs used by the respective projects and institutions. This presentation allows for links to draw to other parallel projects or datasets using the common sample IDs.

Similarly, the collection location names were assigned by the separate projects and do not necessarily represent specific cartographic locations. Therefore, the Latitude and Longitude are provided in decimal degree format for each sampling location.

Database duplicates: When the prey ID sequence in the database was a perfect match for another prey species sequence, both prey IDs are provided separated by “OR” in the entry. For example, “Copper_OR_Quillback_OR_China_Rockfish”. The same approach was used for the taxonomic names of the prey species.

DNA_diet_percent is the product of the RRA calculation for each prey species and can be used as an index of proportional prey consumption (see methods).

Harbour_seal_hardparts_diet_data.csv. These are the results of the morphological prey hard parts ID of collected scat samples, with corresponding sample IDs to those in Harbour_Seal_DNA_Diet_Data.csv (Fig. 3). These data can be used to make direct comparisons between the traditional seal diet analysis technique (prey hard parts analysis) and the newly developed diet characterization method (DNA metabarcoding diet analysis).

Notes for Harbour_Seal_Hardparts_Diet_Data.csv:

The.csv file contains data in five fields:

1. Sample ID – the scat sample ID that corresponds to “sample ID” in the DNA dataset.
2. Prey Species – the prey species scientific name.
3. Salmon classification – Binary age classification for salmon structures as noted by the analyst.
4. Sample comments – Additional comments about the specific prey hard structure.
5. Analyst – The name of the prey hard parts identification analyst.

Because the prey hard parts analysis was conducted by three different analysts without an existing standardization protocol, the reported results varied slightly throughout the dataset. Here we attempted to preserve as much of information provided by the analyst as possible in the “Sample comments” field.

Further, given the intense interest in salmon predation by harbour seals and the large magnitude differences in consumption estimates depending on prey life stage, we have included the binary salmon age classification (Juvenile, Adult). We should acknowledge, however, that seals consume salmon prey in a wide range of age/size classes, and this binary scheme is an oversimplification.

All sample data averages (n = 4,625)

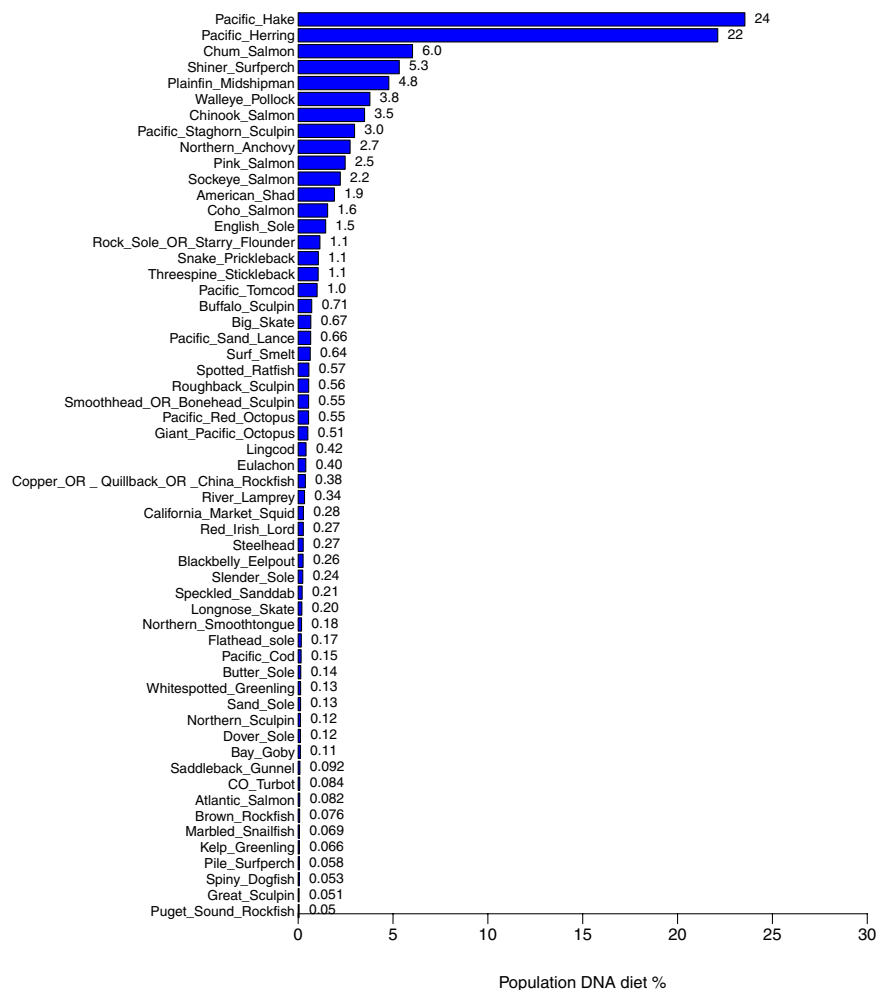


Fig. 2 Harbour seal DNA diet percent averages (RRA) for all of the Salish Sea samples combined. Please note that considerable diet variability exists between haulout sites and regions in the Salish Sea.

Prey_database.txt. The prey reference sequence database (Fig. 4) is a product of the iterative process of adding prey sequences based on the representative sequences of clustered MiSeq data from each successive scat sequencing run. The database started with a Genbank search for 16S sequences using a list of local fish species, and then the clustering processes for each sequencing run added additional species or haplotypes to the database over time.

Notes for prey_database.txt. This version of the database does not contain identified contaminants or predators other than harbour seal that were inadvertently collected.

Prey_species_distance_matrix.xlsx. This distance matrix was used to determine when the reference sequences of two prey species in the database were identical. Both prey species names are given in the Prey_ID field of the diet dataset in those occurrences.

Notes for Prey_Species_Distance_Matrix.xlsx:

Multiple sequence entries exist for some prey species as a byproduct of the iterative sequence entry process. In some cases, one haplotype for a species was identical to another species but other haplotypes were not.

QIIME_mapping_files.tgz. A tar archive of the QIIME mapping files (.txt) needed for demultiplexing the .fastq files in the NCBI SRA database (Online-only Table 1)³⁶. The mapping file names in the tar archive match those of the associated.fastq files. Each mapping file contains the F and R primer sequences along with the unique primer tag for each of the 96 scat samples listed in the sequencing plate.

Field	Example	Description
sample_ID	Pv16.E090	Project assigned sample ID
site	Belle Chain	Collection site name
Lat	48.8323	Site latitude
Lon	-123.1984	Site longitude
day	19	Scat collection day
month	9	Scat collection month
year	2016	Scat collection year
predator	Harbour seal	Predator species
prey_code	ONCTSH	Six-letter prey species code
prey_ID	Oncorhynchus_tshawytscha	Prey species taxonomic ID
common_name	Chinook_Salmon	Prey species common name
count.16S	184	16S DNA sequence count assigned to that prey species
count.COI	4963	COI DNA sequence count assigned to that prey species (only salmon)
DNA_diet_fraction	0.201164835	Fractional component of scat DNA sequences assigned to prey species
DNA_diet_percent	20.11648352	Percentage of scat DNA sequences assigned to prey species
Lab	DFO	The molecular laboratory where work was conducted
Bioinformatician	Nordstrom	The analyst responsible for generating diet data summaries
File	dna_new_BC_2015-16-17.csv	The output project file used

Table 1. Field names and descriptions of the data contained in “Harbour_Seal_DNA_Diet_Data.csv”.

1	Sample ID	Prey species	Salmon classification	Sample comments	Analyst
10787	Pv18.209	<i>Clupea pallasii</i>		, vertebrae, prootic, atlas	W. Walker
10788	Pv18.209	<i>Ammodytes personatus</i>		, vertebrae	W. Walker
10789	Pv18.209	<i>Oncorhynchus</i> sp.	Adult	adult?, large, partial 9.5e mm vertebra	W. Walker
10790	Pv18.210	<i>Merluccius productus</i>		, vertebrae	W. Walker
10791	Pv18.210	<i>Gadus chalcogrammus</i>		,	W. Walker
10792	Pv18.210	<i>Leptacottus armatus</i>		, vertebrae, preopercular spines	W. Walker
10793	Pv18.210	<i>Porichthys notatus</i>		, vertebrae	W. Walker
10794	Pv18.211	<i>Merluccius productus</i>		, vertebrae	W. Walker
10795	Pv18.211	<i>Cymatogaster aggregata</i>		, vertebrae	W. Walker
10796	Pv18.211	<i>Entosphenus tridentatus</i>		, teeth	W. Walker
10797	Pv18.211	<i>Oncorhynchus tshawytscha</i>	Juvenile	juvenile, chinook vertebrae (age classifi	W. Walker
10798	Pv18.213	<i>Clupea pallasii</i>		, vertebrae, prootic, atlas, ceratohyal	W. Walker
10799	Pv18.213	<i>Merluccius productus</i>		, vertebrae	W. Walker
10800	Pv18.214	<i>Clupea pallasii</i>		,	W. Walker
10801	Pv18.214	<i>Alosa sapidissima</i>		, vertebrae	W. Walker
10802	Pv18.214	<i>Gadus chalcogrammus</i>		, vomer, supracleithrum	W. Walker
10803	Pv18.214	<i>Merluccius productus</i>		, vertebrae	W. Walker
10804	Pv18.214	<i>Porichthys notatus</i>		, vertebrae	W. Walker
10805	Pv18.214	<i>Entosphenus tridentatus</i>		, teeth	W. Walker
10806	Pv18.215	<i>Clupea pallasii</i>		, vertebrae, prootic, atlas	W. Walker
10807	Pv18.215	<i>Merluccius productus</i>		, vertebrae	W. Walker
10808	Pv18.215	<i>Oncorhynchus</i> sp.	Juvenile	juvenile, juvenile size pectoral fin ray #	W. Walker

Fig. 3 A screen capture of the harbour seal hard parts diet data file, illustrating example data.

Technical Validation

In parallel to the field sample collection effort, we conducted a series of studies to evaluate the quantitative capabilities of DNA metabarcoding diet analysis for seal diet estimation. Specifically, we wanted to determine if prey DNA sequence percentages recovered from scats (RRA) accurately reflect the proportional biomass of the prey consumed. This work involved a series of experiments including: a captive seal feeding study with known diets, the development of food tissue control materials (homogenized fish tissue) to produce correction factors, and computer simulations to compare RRA to alternative diet indices such as weighted percent of occurrence. Here we will briefly describe those studies and the principal conclusions drawn from each. Lastly, we will outline why we have chosen to report RRA as the principal diet index for this dataset.

Ideally, RRA would be a direct reflection of the proportional biomass of the prey species consumed – i.e. a 1:1 relationship would exist between metabarcoding DNA sequence proportion and diet biomass proportion. However, a large number of potential factors could skew this relationship, including variability in mDNA density (number of mitochondria per gram of tissue) between different prey species, technical biases such as preferential primer binding during PCR caused by primer mismatches, and the many bioinformatic processes that may

```

>Yellowtail_Rockfish gi|144953752|gb|EF446525.1| Sebastes flavidus voucher UW:110468
16S ribosomal RNA gene, partial sequence; mitochondrial
CGAGAAGACCCTATGGAGCTTTAGACACCAAGAAGATCCTGTCAAGTAACCCCAACAA
GGGCTTGAACCTAATGGAATCCTTCCCTAATGTCTTTGGTTGGGGCGACCGCGGGGAAACA
AAAAACCCACAGTGGAAAGGGAGCACCCCTCTACAATAAGAGCCGAGCTCTAATT
AACAGAATATCTGACCAATAAGATCCGGCAATGCCGATCAACGGACCGAGTTACCTAGG
GATAACAGCGCAATCC
>Rosethorn_Rockfish gi|144953798|gb|EF446571.1| Sebastes helvomaculatus voucher UW:
114035 16S ribosomal RNA gene, partial sequence; mitochondrial
CGAGAAGACCCTATGGAGCTTTAGACACCAAGAAGATCCTGTCAAGTAACCCCAACAA
GGGCTTGAACCTAATGGAATCCTTCCCTAATGTCTTTGGTTGGGGCGACCGCGGGGAAACA
AAAAACCCACAGTGGAAAGGGAGCACCCCTCTACAATAAGAGCCGAGCTCTAATT
AACAGAATATCTGACCAATAAGATCCGGCAATGCCGATCAACGGACCGAGTTACCTAGG
GATAACAGCGCAATCC
>Quillback_Rockfish gi|144953826|gb|EF446599.1| Sebastes maliger voucher UW:114064 16S
ribosomal RNA gene, partial sequence; mitochondrial
CGAGAAGACCCTATGGAGCTTTAGACACCAAGAAGATCCTGTCAAATAACCCCTCACAA
GGGCTTGAACCTAATGGAATCCTTCCCTAATGTCTTTGGTTGGGGCGACCGCGGGGAAACA
AAAAACCCACAGTGGAAAGGGAGCACCCCTCTACAATAAGAGCCGAGCTCTAATT
AACAGAATATCTGACCAATAAGATCCGGCAATGCCGATCAACGGACCGAGTTACCTAGG
GATAACAGCGCAATCC
>Quillback_Rockfish gi|661786464|gb|KJ879123.1| Sebastes maliger voucher UW151128 16S
ribosomal RNA gene, partial sequence; mitochondrial
CGAGAAGACCCTATGGAGCTTTAGACACCAAGAAGATCCTGTCAAATAACCCCTCACAA
GGGCTTGAACCTAATGGAATCCTTCCCTAATGTCTTTGGTTGGGGCGACCGCGGGGAAACA
AAAAACCCACAGTGGAAAGGGAGCACCCCTCTACAATAAGAGCCGAGCTCTAATT
AACAGAATATCTGACCAATAAGATCCGGCAATGCCGATCAACGGACCGAGTTACCTAGG
GATAACAGCGCAATCC

```

Fig. 4 A screen capture example of the prey database FASTA file.

select for one prey species' sequences over another. For this reason, our first validation study was an evaluation of the factors that influence the relationship between diet biomass proportion and RRA in a captive seal feeding study wherein aquarium animals were fed a known diet of fixed biomass proportions of three prey species³⁷.

In that study using the Ion Torrent PGM sequencer, we found that DNA sequence quality filtering, direction of sequencing (F vs R), and our chosen minimum read length, all largely impacted the proportions of prey sequences ultimately resulting from RRA analysis³⁷. Furthermore, we detected interactive effects between biasing factors such as variable effects of quality filtering depending on the primer tags used to identify sequences from unique samples. However, despite these effects we found that replicate samples of a common diet produced largely consistent DNA sequence percentages when technical factors were held constant. This consistency implied that biasing factors could potentially be corrected using a set of standards sequenced alongside scat samples, such as homogenized mock communities of known composition (i.e. prey fish tissue mixes).

The next study therefore attempted to correct for biasing factors by sequencing a prey fish tissue mixture that matched the diet biomass proportions of captive seals, allowing us to calculate and apply Tissue Correction Factors (TCFs) to the scat DNA sequence counts²⁰. When applied to seal scat DNA, TCFs substantially improved the relationship between RRA and prey biomass proportion for all prey species. We also surmised that while the TCFs account for technical biases (e.g. quality filtering, differential primer binding) and mDNA density variability, they did not account for any potential effects of differential prey species digestion (e.g., one species being more fully digested than another in the seal's gastrointestinal tract). The study design of the experiment allowed us to quantify the bias introduced by differential prey digestion and determine in that case the magnitude of the required Digestion Correction Factor (DCF) for each species. It is unrealistic to experimentally determine the DCF for all prey when hundreds of prey species are possible, so we explored prey characteristics (proximate composition analysis) that could be used as a proxy for prey digestion. We found that the percent lipid of prey species closely predicted the DCF, implying that highly accurate diet biomass estimates could be obtained by applying both TCFs and lipid-based DCF correction factors. The correction factor design of this experiment however relied on *a priori* knowledge of the seal diets; therefore, a more generalizable approach was needed that could be applied to samples of unknown composition.

Using the lessons learned from the captive feeding studies, we thus created a novel correction factor approach for samples of unknown composition by generating 50/50 biomass percentage mixtures of variable potential prey fish combined with a "control" fish that was held constant in all mixtures³⁸. By holding one of the two prey constant in all 50/50 mixtures, we were able to detect biases in the "test" prey relative to the control species and to calculate Relative Correction Factors (RCFs) that could be applied to DNA sequence counts from samples of unknown biomass composition (i.e. seal scat samples). We then built a prey library of the 50/50 mixtures and sequenced them alongside seal scat samples, calculating RCFs for prey species in the library and applying them to scat samples to determine the magnitude of corrective effect. Similar to previous results, RCFs were highly consistent between sequencing replicates, and RCF values indicated that there is some phylogenetic structure to the magnitude of bias (active swimming fishes were overestimated relative to more sedentary species, likely reflecting mDNA density). The effects of RCF correction were most pronounced on individual samples, although when samples were averaged together (as is common practice when calculating population level diet summaries) the effect of RCF correction were less impactful. We ultimately concluded that RCF correction is

Comox

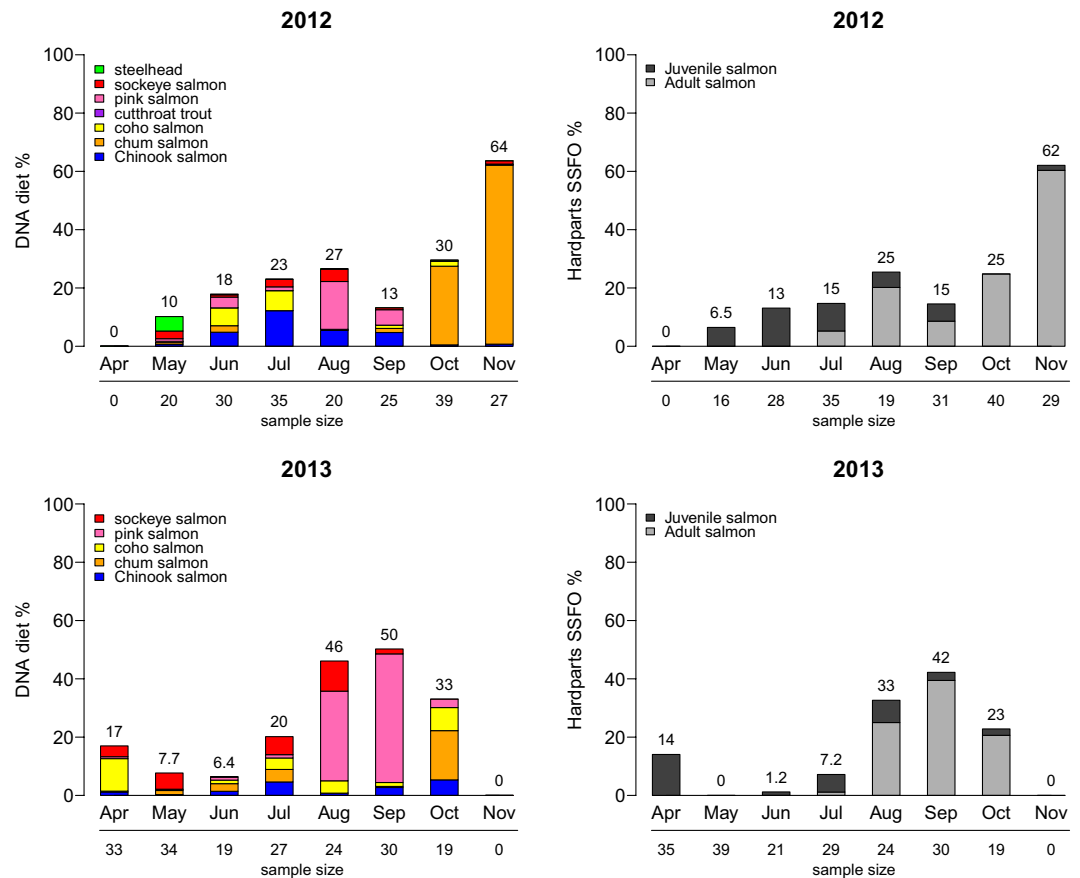


Fig. 5 Figure B-2 from Thomas¹⁷, illustrating the general agreement between scat DNA RRA (Left) and prey hard parts analysis using wPOO (Hard parts SSFO%) (Right) from the same set of scat samples. Data shown are the salmon component of harbour seal diet from the Comox collection site over two years: 2012 (Top), 2013 (Bottom).

only a worthwhile endeavor when a high degree of diet biomass accuracy is needed for a single sample, or when generating population diet summaries from a small number of diet samples.

Despite the known biases, researchers have consistently concluded that RRA is a semi-quantitative tool even without the application of correction factors. When the diet proportion of a prey species increases, there is a corresponding increase in the relative number of DNA sequences for that prey species in diet samples. So that begs the question, are uncorrected “semi-quantitative” RRA estimates good enough? One way to answer this question is to compare the accuracy of the RRA method to other competing methodological alternatives. DNA metabarcoding RRA may be subject to known biasing factors, but if it outperforms alternative diet metrics in terms of taxonomic resolution and biomass estimate accuracy, then it may be the current best option for pinned diet analysis.

The most common practice for pinned diet studies is to treat detections of prey species in diet samples as presence/absence data (i.e., occurrences). Percent of occurrence (POO) tables can be generated for groups of samples using such data, or when proportions must sum to 1 (as is needed for bioenergetics modelling) a weighted percent of occurrence (wPOO) can be calculated (Box 1). The latter has also been called Split Sample Frequency of Occurrence (SSFO). Occurrence-based indices however are prone to overestimating prey eaten in low proportion and underestimating prey eaten in high proportions. In a meta-analysis of DNA metabarcoding diet studies, Deagle *et al.*¹⁶ compared multiple diet indices using an *in silico* simulation experiment. Diet datasets were simulated at multiple sample sizes and the proportional composition was calculated using RRA and occurrence indices, then compared to the theoretical true diet using Bray–Curtis dissimilarity. They found that occurrence indices produce consistent diet estimates (relatively precise) but result in a less accurate reflection of the true diet by comparison to RRA. RRA produced a more accurate estimate than POO even when the simulated level of bias was extreme (20X), suggesting that the quantitative signature of RRA is important for generating accurate diet estimates despite the presence of biases.

Comparing DNA metabarcoding RRA to non-DNA based methods with a proven history of use (e.g., scat morphological hard parts analysis) is another means of assessment. In an extensive methods comparison

between scat DNA and hard parts analysis, Thomas (2015)¹⁷ found a high degree of agreement between the two methods in the proportions of salmon estimated in seal diets (Fig. 5), in addition to other diet species (Online-only Table 2). Furthermore, in a recent report³⁹, pinniped diet experts Dominic Tollit and Ruth Joy compared DNA metabarcoding RRA (Termed “DNA-Fixed” in report) to the current best practices method (hard parts biomass reconstruction – “BR-Fixed”) and found “near identical” diet contributions between methods (see Supplementary Information for full report, posted with permission). This implies that DNA metabarcoding RRA provides not only better taxonomic resolution than hard parts, but also produces comparable diet proportions to the current best practices method while requiring substantially less labor. A copy of the report is available in Supplementary Information.

Given the strengths of RRA for proportional biomass estimation in the absence of correction factors, the cost/labor involved in generating a complete prey library for RCF correction, and fact that harbour seal prey are minimally biased with our 16S marker, we have chosen to present uncorrected RRA as the proportional diet index in the current dataset. Caution should therefore be exercised when generating diet summaries using small numbers of samples (e.g. < 70 per stratum), as is considered best practice with other diet indices⁴⁰. Future efforts may be made to build a 50/50 RCF prey library for harbour seals in the Salish Sea region, in which case the diet database presented here could be RCF corrected for improved sample proportional biomass accuracy. It should also be noted that our methods evaluations did not account for other potential sources of bias (e.g., sampling bias introduced during scat collection, selective influence of blocking oligos, secondary prey consumption, reliance on Genbank for reference database sequences, etc.). Those issues are worth investigating further in future methods refinement.

Usage Notes

As stated previously, we encourage users of these data to follow statistical best-practices when creating diet summaries from this dataset. Ample literature is available on the topic of appropriate size for predator diet studies^{40,41}.

Code availability

The bioinformatic code used to process scat DNA sequence data are available in the “Example Bioinformatic Files. zip” folder on figshare along with example data³⁵. Sample demultiplexing and sequence taxonomy assignment steps were performed using MacQIIME (version: 1.9.1-20150604) and the code used for these steps is in folder “1_Qiime_processing”. Example output files from MacQIIME processing steps containing 16S and COI sequence taxonomy assignments are in the folder “2_Qiime_output_R_input”. Those files were then further processed in R (multiple versions) to generate DNA sequence percentages using the code contained in folder “3_R_processing”. Note: Demultiplexing requires QIIME mapping files for each sample plate which are stored in the figshare folder³⁵ “QIIME_Mapping_Files” for sequence data stored the NCBI SRA database³⁶. A complete list of the bioinformatic functions in MacQIIME is available online <http://qiime.org/scripts/>.

Received: 15 April 2020; Accepted: 13 January 2022;

Published online: 02 March 2022

References

- Chasco, B. E. *et al.* Competing tradeoffs between increasing marine mammal predation and fisheries harvest of Chinook salmon. *Scientific Reports* **7**, 15439, <https://doi.org/10.1038/s41598-017-14984-8> (2017).
- Chasco, B. *et al.* Estimates of Chinook salmon consumption in Washington State inland waters by four marine mammal predators from 1970 to 2015. *Canadian Journal of Fisheries and Aquatic Sciences* **74**, 1173–1194, <https://doi.org/10.1139/cjfas-2016-0203> (2017).
- Li, L., Ainsworth, C. & Pitcher, T. Presence of harbour seals (*Phoca vitulina*) may increase exploitable fish biomass in the Strait of Georgia. *Progress in Oceanography* **87**, 235–241 (2010).
- Jeffries, S., Huber, H., Calambokidis, J. & Laake, J. Trends and status of harbor seals in Washington State: 1978–1999. *The Journal of Wildlife Management*, 207–218 (2003).
- Olesiuk, P. An assessment of population trends and abundance of harbour seals (*Phoca vitulina*) in British Columbia. *DFO Can. Sci. Advis. Sec. Res. Doc* **105** (2009).
- Scordino, J. *West coast pinniped program investigations on California sea lion and Pacific Harbor seal impacts on salmonids and other fishery resources.* (Pacific States Marine Fisheries Commission Portland, 2010).
- Nelson, B. W., Walters, C. J., Trites, A. W. & McAllister, M. K. Wild Chinook salmon productivity is negatively related to seal density and not related to hatchery releases in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences* **76**, 447–462, <https://doi.org/10.1139/cjfas-2017-0481> (2018).
- Thomas, A. C., Nelson, B. W., Lance, M. M., Deagle, B. E. & Trites, A. W. Harbour seals target juvenile salmon of conservation concern. *Canadian Journal of Fisheries and Aquatic Sciences* **74**, 907–921, <https://doi.org/10.1139/cjfas-2015-0558> (2017).
- Zimmerman, M. S. *et al.* Spatial and temporal patterns in smolt survival of wild and hatchery coho salmon in the Salish sea. *Marine and Coastal Fisheries* **7**, 116–134 (2015).
- Kendall, N. W., Marston, G. W. & Klungle, M. M. Declining patterns of Pacific Northwest steelhead trout (*Oncorhynchus mykiss*) adult abundance and smolt survival in the ocean. *Canadian journal of fisheries and aquatic sciences* **74**, 1275–1290 (2017).
- Ward, E. J., Levin, P. S., Lance, M. M., Jeffries, S. J. & Acevedo-Gutiérrez, A. Integrating diet and movement data to identify hot spots of predation risk and areas of conservation concern for endangered species. *Conservation Letters* **5**, 37–47 (2012).
- Sperry, C. C. Food habits of the Pacific harbor seal, *Phoca richardii*. *Journal of Mammalogy* **12**, 214–226 (1931).
- Lance, M. M., Chang, W.-Y., Jeffries, S. J., Pearson, S. F. & Acevedo-Gutiérrez, A. Harbor seal diet in northern Puget Sound: implications for the recovery of depressed fish stocks. *Marine Ecology Progress Series* **464**, 257–271 (2012).
- Olesiuk, P. F. *An Assessment of the Feeding Habits of Harbour Seals Phocia Vitulina in the Strait of Georgia, British Columbia, Based on Scat Analysis.* (Department of Fisheries and Oceans, Biological Sciences Branch, Pacific ..., 1990).
- Deagle, B. E., Kirkwood, R. & Jarman, S. N. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology* **18**, 2022–2038, <https://doi.org/10.1111/j.1365-294X.2009.04158.x> (2009).
- Deagle, B. E. *et al.* Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology* **28**, 391–406, <https://doi.org/10.1111/mec.14734> (2019).

17. Thomas, A. C. *Diet analysis of Pacific harbour seals (Phoca vitulina richardsi) using high-throughput DNA sequencing*, University of British Columbia (2015).
18. Orr, A. J. *et al.* Comparison of processing pinniped scat samples using a washing machine and nested sieves. *Wildlife Society Bulletin* **31**, 253–257 (2003).
19. King, R. A., Read, D. S., Traugott, M. & Symondson, W. O. C. INVITED REVIEW: Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology* **17**, 947–963, <https://doi.org/10.1111/j.1365-294X.2007.03613.x> (2008).
20. Thomas, A. C., Jarman, S. N., Haman, K. H., Trites, A. W. & Deagle, B. E. Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias. *Molecular Ecology* **23**, 3706–3718, <https://doi.org/10.1111/mec.12523> (2014).
21. Deagle, B. E. *et al.* Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology* **14**, 1831–1842, <https://doi.org/10.1111/j.1365-294X.2005.02531.x> (2005).
22. Vestheim, H. & Jarman, S. N. Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs. *Front Zool* **5**, 12, <https://doi.org/10.1186/1742-9994-5-12> (2008).
23. Faircloth, B. C. & Glenn, T. C. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One* **7**, e42543 (2012).
24. DeVaney, S. & Pietsch, T. W. *Key to the Fishes of Puget Sound*, <http://www.burkemuseum.org/static/FishKey/> (2006).
25. Benson, D. A. *et al.* GenBank. *Nucleic Acids Research* **28**, 15–18 (2012).
26. R_Core_Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. **2016** (2019).
27. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
29. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
30. Harvey, J. T. Relationship between fish size and otolith length for 63 species of fishes from the eastern North Pacific Ocean. (2000).
31. Kashiwada, J., Recksiek, C. W. & Karpov, K. Beaks of the market squid, *Loligo opalescens*, as tools for predator studies. *CalCOFI* **20**, 65–69 (1979).
32. Morrow, J. E. *Preliminary keys to otoliths of some adult fishes of the Gulf of Alaska, Bering Sea and Beaufort Sea*. **420** (Department of Commerce, National Oceanic and Atmospheric Administration ..., 1979).
33. Wolff, G. A. A beak key for eight eastern tropical Pacific cephalopod species with relationships between their beak dimensions and size. *Fishery Bulletin* **80**, 357–370 (1982).
34. Nelson, B. W. *et al.* Variation in predator diet and prey size affects perceived impacts to salmon species of high conservation concern. *Canadian Journal of Fisheries and Aquatic Sciences* (2021).
35. Thomas, A. C. *et al.* Harbour seal DNA metabarcoding diet data of the Salish Sea. *figshare* <https://doi.org/10.6084/m9.figshare.c.4910811> (2020).
36. NCB Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP256355> (2020).
37. Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W. & Jarman, S. N. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources* **13**, 620–633, <https://doi.org/10.1111/1755-0998.12103> (2013).
38. Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H. & Trites, A. W. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular ecology resources* **16**, 714–726 (2016).
39. Tollit, D. J. & Joy, R. Validation of diet reconstruction (Predation by Harbour Seals on Salmon Smolts Project). 1–19 (SMRU Consulting North America, 2018).
40. Trites, A. W. & Joy, R. Dietary analysis from fecal samples: how many scats are enough? *Journal of Mammalogy* **86**, 704–712 (2005).
41. Boyd, I. L., Bowen, W. D. & Iverson, S. J. *Marine mammal ecology and conservation: a handbook of techniques*. (Oxford University Press, 2010).
42. Xiong, M. *et al.* Molecular dietary analysis of two sympatric felids in the Mountains of Southwest China biodiversity hotspot and conservation implications. *Scientific Reports* **7** (2017).
43. Tollit, D. *et al.* Diet of endangered Steller sea lions in the Aleutian Islands: New insights from DNA detections and bio-energetic reconstructions. *Canadian Journal of Zoology* (2017).

Acknowledgements

Many people contributed to the production of this dataset beyond the list of named coauthors. We would specifically like to thank the volunteers who helped to collect harbour seal scat samples throughout the Salish Sea, and the laboratory technicians who spent countless hours extracting DNA and performing PCRs for this project. This is Publication Number 62 from the Salish Sea Marine Survival Project (marinesurvivalproject.com). Funding was provided by the Pacific Salmon Foundation, the Pacific Salmon Commission's Southern Endowment Fund via Long Live the Kings, by Washington State with equal in-kind contributions by those participating in the research.

Author contributions

A.C.T. helped develop the DNA diet analysis method, collected samples, performed data analyses, and wrote the manuscript; B.D. was principally involved in the DNA diet analysis method development; C.N. collected samples and performed data analyses; S.M. collected samples and led project management for D.F.O.; B.N. collected samples and contributed to method development; A.A.-G. collected samples and led project management for WWU; S.J. collected samples and led project management for WDFW; J.M. collected samples and led project management for NIT; A.L. contributed to laboratory protocol refinement and sample processing; H.A. collected samples and led project management for U.B.C.; S.P. secured funding and contributed to project management for WDFW; M.S. secured funding and help coordinate the joint USA/Canada research effort; A.T. secured funding, supervised graduate students conducting related work, and facilitated collaborations. All authors contributed to the review and editing of this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01152-5>.

Correspondence and requests for materials should be addressed to A.C.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2022