A Non-Invasive Technique for Tracking a Marine Predator (*Phoca vitulina*) Through Molecular Scat Analysis

By

Nathan Guilford

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Dr. Alejandro Acevedo-Gutiérrez, Chair

Dr. Dietmar Schwarz, Co-chair

Dr. Austen Thomas

Dr. Dan Pollard

GRADUATE SCHOOL

David L. Patrick, Interim Dean

Master's Thesis

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and it does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Nathan Guilford

29 November 2020

A Non-Invasive Technique for Tracking a Marine Predator (*Phoca vitulina*) Through Molecular Scat Analysis

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Nathan Guilford
November 2020

Abstract

As ecosystems are subjected to increased urbanization, habitat loss, and resource depletion, management practices will benefit from higher resolution models of local trophic dynamics. Harbor seals (*Phoca vitulina*), the most abundant marine mammal in the Salish Sea of British Columbia and Washington State, are of great regional interest due to their consumption of species of conservation concern such as Chinook salmon (*Oncorhynchus tshawytscha*) and Pacific herring (*Clupea pallasii*). This ecologically influential diet can vary with season, region, and local sex ratios, creating localized pressures on prey species. Variation in diet has been observed at the individual level, an important consideration for examining the total influence of predators that have previously been treated as species-wide averages. This project aimed to develop a method that allows researchers to track individual specialization rates in a protected marine predator by testing if 1) harbor seal scat represents a suitable source of DNA for individual identification through single nucleotide polymorphism (SNP) genotypes produced by direct sequencing, and 2) prey reads could be detected within the sequence data for simultaneous diet analysis without the need for PCR-based methods. SNP loci identified in this study successfully distinguished individual seals with confidence, however read alignments to prey references indicated potentially erroneous classifications. This indicates prey analyses through direct read counts will benefit from more research such as direct feeding trials and digestion correction factors, or from employing more robust techniques (or a combination of methods). Nonetheless, this direct sequencing pipeline of scat DNA for marker identification, individual identification, and simultaneous prey analysis from one sample type provides important considerations for highly scalable/cost-effective non-invasive investigations of regional trophic dynamics in complex and/or understudied systems.

Acknowledgements

I would first like to thank my advisors, Dr. Alejandro Acevedo-Gutiérrez and Dr. Dietmar Schwarz for allowing me to pursue this project with them. Throughout the entirety of this project, they have provided excellent support, guidance, and advice that made it all possible. They welcomed me into WWU with open arms and I have sincerely enjoyed our time together. Thank you to Dr. Austen Thomas, not only for the guidance from someone who deeply understands this study system, but also for providing a substantial collection of field-collected scat samples, which were truly appreciated. I would also like to thank Dr. Dan Pollard and Dr. Matthew Zinkgraf, for providing additional support that was crucial for the design and completion of this research. Thank you to my friends, classmates, and cohort, for being my partners throughout the highs and lows that come with graduate school.

Thank you to The Seattle Aquarium, Julie Carpenter, and Hogan the Harbor seal, for so willingly participating in this project and providing the much-needed captive samples. Thank you to Evan Forsberg and the University of Minnesota Genomics Center for providing assistance in the design of custom sequencing protocols.

Thank you to the WWU Biology Department, Office of Research and Sponsored Programs, College of Science and Engineering, Dr. Arlan Norman, and the Fraser, Hodgson, Ross, and Flora families for the generous support throughout this process. This project truly would not be possible without these sources of support, and I would like to thank you all for seeing the value in my research and providing me the opportunity to advance the scientific knowledge of our Salish Sea community and beyond.

Finally, I would like to thank my family; I wouldn't be where I am today without your continuous support in all my efforts.

Table of Contents

List of Tables

List of Figures

Introduction

As ecosystems are subjected to increased urbanization, habitat loss, and resource depletion, management practices will benefit from higher resolution models of local trophic dynamics. Lack of data reflecting accurate trophic interactions of influential species such as high-level predators can prevent successful recovery of both predator and/or prey populations (Marshall et al. 2016). Furthermore, unclear depictions of local trophic dynamics can inhibit our ability to anticipate the consequences of future population declines (Matich et al., 2011). In the search for more accurate ecological models, we know that many species exhibit within-population niche variation, challenging the idea of population-average-based management practices (Balme et al., 2020; Beaudoin et al., 1999; Navarro et al., 2017). This variation has been observed all the way down to the individual level, with a broad range of taxa exhibiting distinctions in resource use between individuals of the same population, referred to as individual specialization (Araújo et al., 2011; Bolnick et al., 2003). Individual specialization occurs when an individual's niche width comprises only a portion of the populations total niche width, creating a spectrum of partially overlapping specialists, rather than strict identical generalists. As such, what used to be considered a single-trophic level group for ecosystem models may in fact be a group of disparate functional sub-groups. Factors and environmental conditions such as geographic location, levels of urbanization, and population densities appear to be drivers of dynamic individual specialization rates within populations of even the same species (Bolnick et al., 2003; Jacquier et al., 2020; Svanbäck & Bolnick, 2007; Svanbäck & Persson, 2009; Zango et al., 2019). Not surprisingly, the degree to which individual specialization occurs within populations may have profound impacts on a population's stability and adaptability, in addition to its true impacts on vulnerable prey populations along spatiotemporal timescales (Bolnick et

al., 2003). However, obtaining individual-based data sets can be time-consuming, expensive, and highly invasive, all of which pose logistical and analytical difficulties for researchers. Here, I developed a method that allows researchers to track individual specialization in a marine predator by genotyping individuals and detecting prey species from non-invasive scat samples. This process minimizes the need for costly sample processing by directly sequencing scat DNA isolates, followed by bioinformatic individual identification and prey detection, providing a high-throughput framework for future researchers.

Molecular markers and fecal analysis

Molecular markers (fragments of DNA at known genomic locations) have become incredibly useful for investigating a wide range of phenomena in a wide range of wild populations, and as next generation sequencing (NGS) technologies have lowered in cost, the barrier to perform marker analyses has also dropped (Ackerman et al., 2017; Cammen et al., 2018; Coltman, 2007; Gärke et al., 2012; Morin et al., 2004; Seeb et al., 2011). However, to identify molecular markers, high-quality DNA from the target species is typically employed in order to construct confident genotypes. Previous marker analyses of wild populations have routinely employed invasive blood or tissue samples, as these are sources of pure target DNA (Baird et al., 2008; Cammen et al., 2018; Heaton et al., 2014; Huber et al., 2010, 2012; le Roex et al., 2012; Senn et al., 2013; Valencia et al., 2018). Sourcing tissues can present legal, temporal, logistical, financial, and ethical restraints for some researchers, especially for species under legal protection, like marine mammals. Additionally, the identification of markers from tissue samples may necessitate the collection of respective samples for diet analysis, such as feces or stomach contents, increasing labor cost and effort. Non-invasive sampling techniques can

provide a solution to the limits introduced through invasive study methods, while also increasing sample sizes for inaccessible or protected species (Ferter et al., 2014; Johnson, 2002). With increased interest in reducing environmental impacts of research and advancements in NGS technologies, non-invasive resources such as hair, whiskers, and feces have become practical tools for molecular analyses (Natesh et al., 2019; Reddy et al., 2012; Sigsgaard et al., 2019; Thaden et al., 2018). For typically inaccessible species like marine mammals, scat samples of pinnipeds (seals, sea lions, fur seals, and the walrus) represent an opportune resource, as they can be collected on terrestrial haul-out sites. Molecular markers have proved to be discoverable from field-collected scat samples, available for use in sample genotyping and individual identification (Guertin et al., 2009; Gulsby et al., 2016; Lathuillière et al., 2001; Rothstein et al., 2017). Furthermore, as prey specific sequences can also be detected in scat, these resources may represent promising tools that will allow for simultaneous diet analyses, therefore reducing the number of samples needed per individual at one point in time for individual specialization investigations (Ford et al., 2016; Reed et al., 2003; Schwarz et al., 2018; Thomas et al., 2017; Voelker et al., 2020).

DNA present in scat presents some obstacles for molecular analyses, as endogenous depositor DNA often comprises <5% of the total DNA content, with exogenous bacterial DNA dominating the sample (Perry et al., 2010; Qin et al., 2010; Snyder-Mackler et al., 2016). This small amount can present problems when attempting to repeatedly uncover fine-scale patterns, as the target genetic information in the samples may be obscured or inaccessible. These challenges may be further amplified in methods where specific amplicons are not targeted through PCR prior to sequencing and characterization (as in direct sequencing). FecalSeq, a recently developed method for increasing the proportion of endogenous DNA in fecal samples, presents a

solution for investigating taxa like pinnipeds, for which tissue samples are not easily accessed (Chiou & Bergey, 2018). This technique proved the efficacy of employing microbiome enrichment kits to remove, rather than isolate, microbial DNA from fecal samples. Eukaryotic DNA, with significantly more methylation of CpG sites than prokaryotic DNA, selectively binds to methyl-CpG-binding proteins on immunoprecipitation beads. Washing away the unbound prokaryotic DNA produces samples with much higher eukaryotic DNA concentrations. These enriched samples are compatible with downstream processes such as NGS to identify novel molecular markers, and as prey DNA will remain post-enrichment, individual identification and diet analysis from the same sample in a non-invasive manner is possible.

Another obstacle in employing DNA from non-invasive sources such as scat in marker analyses is the high rate of DNA degradation from both digestion and environmental conditions. Proper choice in marker type may help overcome the challenges of working with low-quality DNA. Microsatellites (MS), regions of DNA motifs repeated 5-50 times, are a common marker type used in identification studies, due to the high informativeness of these multiallelic loci (up to 20 alleles per locus). However, a previous MS analysis of individual identification in harbor seal (*Phoca vitulina*) scat suggests that MS may be an unreliable tool for non-invasive scat samples (Rothstein et al., 2017). This finding is most likely due to the short fragments of degraded DNA found in scat preventing successful amplification of long MS products (Morin & McCarthy, 2007; Sanchez & Endicott, 2006; Seddon et al., 2005). Single nucleotide polymorphisms (SNPs), or variations in a single nucleotide, are much shorter genomic loci than MS, allowing for high functionality with degraded, non-invasive samples (Morin & McCarthy, 2007). SNP genotyping error rates as low as 0.1% have been reported from non-invasive samples, which are much lower than the 0.8%-2.0% often reported for MS (Bonin et al., 2004;

Morin & McCarthy, 2007). SNPs have also increased in popularity due to their even spread throughout genomes, low mutation rates, and standardization possibilities between laboratories (Morin et al., 2004; Kraus et al., 2011). While SNPs are largely bi-allelic, providing much less genotypic information per locus than MS, employing a larger number of SNPs still allows for easier multiplexing of assays while producing higher quality data than MSs (Hauser et al., 2011; Kalinowski, 2002; Liu et al., 2005; Morin et al., 2004; Seddon et al., 2005; Viengkone et al., 2016).

Markers such as SNPs can be scored directly through PCR, sequencing of amplicons targeting known loci, or through the direct sequencing of the genome (often from a reduced representation library). However, the former two methods rely on previous knowledge of markers in the target species, and/or require error-prone/costly lab work. At the same time, direct sequencing of the genome typically employs tissue samples for high-quality DNA (ensuring whole-genome sampling). A crucial point for this study is that this direct sequencing (of a reduced representation library) may be appropriate for non-invasive samples such as scat, despite their metagenomic nature. NGS advancements now allow researchers to identify new marker panels in a timely and affordable manner from metagenomic DNA sources such as scat (Perry et al., 2010; Seeb et al., 2011). Specifically, restriction-site associated DNA sequencing (RADseq) eliminates the need for high sequencing effort of high-quality tissues through genome sub-sampling (Cammen et al., 2018). In RADseq, a restriction enzyme shears a collection of DNA samples at restriction sites, creating fragments that represent a subset of the entire genome. Sequencing this reduced representation library minimizes the sequencing effort needed to compare nucleotides at the same loci across samples, efficiently identifying polymorphic markers like SNPs while reducing costs (Baird et al., 2008; Peterson, B., et al., 2012). Following

RADseq and SNP identification, genotypes are constructed, and samples can be easily distinguished (Davey et al., 2011; Nielsen et al., 2011).

Studies of diet specialization through fecal analysis typically rely on PCR amplification of species-specific target sequences, followed by the sequencing of resulting products to confirm species presence (Deagle et al., 2005; Deagle et al., 2007; Schwarz et al., 2018; Valentini et al., 2008; Voelker et al., 2020). Other techniques for non-invasive prey identification include hard-part, fatty-acid, or isotopic analyses (Bromaghin et al., 2013; Jacquier et al., 2020; Lance et al., 2012; Luxa & Acevedo-Gutierrez, 2013; Voelker et al., 2020). While these methods can be robust, they can require laborious lab work, costly reagents, and/or additional sequencing runs after predator marker identification. However, due to the non-specific enrichment of eukaryotic DNA from methyl-CpG-binding proteins as seen in FecalSeq, examination of sample reads post-sequencing should allow for the identification of prey species present in the original scat in the same workflow for predator identification, requiring only reference genomes for selected prey taxa. Despite these advantages, the direct sequencing approach requires some important considerations. As scat is a metagenomic collection of DNA with varied levels of degradation; alignment to a reference genome for marker identification may cause conserved regions from non-target sources to map, misrepresenting allele frequencies. This is also a caveat for assigning exogenous reads to prey references. The binning of raw reads into various taxa before marker identification will both strengthen the final SNP panel, as well as allow for the identification of prey species without the need for any PCR-based methods or a separate sequencing run.

Case study: Salish Sea harbor seals

The harbor seal is an excellent case study to develop a method to track individual specialization using DNA from scat in a marine predator. Marine mammals play a significant role in aquatic ecosystems due to their large size (and therefore energetic demand), mobility, and top-down regulation of lower trophic levels. Consequently, their distribution and behavior have significant effects on community dynamics and trophic linkages, but investigating fine scale patterns such as individual diets can prove difficult due to their wide distributions and aquatic nature (Bowen, 1997; Williams et al., 2004). The Salish Sea of Washington State, USA, and British Columbia, Canada, represents a 16,925 km^2 inland sea that provides a diverse wealth of habitats and resources for notable marine mammals such as orcas (*Orcinus orca*), harbor porpoises (*Phocoena phocoena*), Steller sea lions (*Eumetopias jubatus*), California sea lions (*Zalophus californianus*), and harbor seals. Recently, the interaction between pinnipeds and Salish Sea fish stocks has increasingly sparked discussions of the intervening role that management agencies should play in predator/prey population dynamics. Harbor seals, the most abundant marine mammal in the Salish Sea, prey heavily on the depressed stocks of commercially and ecologically valuable species such as Chinook salmon (*Oncorhynchus tshawytscha)* and Pacific herring (*Clupea pallasii*) (Lance et al., 2012; Olesiuk, 1993; Thomas et al., 2017). Researchers are therefore increasingly interested in quantifying this predation and determining the significance of harbor seal predation on the recovery of depressed fish populations. Investigating the magnitude of fine-scale trophic pressures such as harbor seal individual specialization will not only provide management agencies with clearer depictions of local dynamics, but also provide a framework for similar analyses within other species and ecosystems of interest.

Washington State-financed seal bounty hunting from 1943-1960, driven by concerns for competition between seals and commercial fishing, contributed to drastic historical harbor seal declines. This hunting reduced the state's seal population to as few as 2,000-3,000 individuals by 1970 (Newby, 1973). Following the Marine Mammal Protection Act (MMPA) of 1972 and the Endangered Species Act (ESA) of 1973, Washington seals rapidly rebounded, reaching a predicted carrying capacity of almost 30,000 individuals by 2002 (Jeffries et al., 2003). Similarly, the Strait of Georgia saw an estimated increase in harbor seal population from 3,570 to 37,300 between 1966 and 1998 (Olesuik, 2009). In 2015, it was estimated that pinnipeds in Washington consumed over six times the combined commercial and recreational catch of Chinook (Chasco et al., 2017). Further, the Washington Department of Fish and Wildlife estimates that inland Washington seals consume about 1,425,000 juvenile Chinook every month (Pamplin et al., 2018). Herring stock assessments in Washington State also indicate a decrease in the total number of stocks classified as 'healthy' or 'moderately healthy,' placing stocks at increasingly higher risks of severe predation effects (Stick et al., 2014). While mammal protections have clearly benefited harbor seals, conservation efforts such as the ESA and MMPA may overlook key ecological connections through single-species protection. When multiple depressed populations co-occur (e.g., Pacific salmon, Pacific herring, and orcas), successful recovery of one top predator may impede further ecosystem recovery, as the top-down forces of predators are complex (Marshall et al., 2016).

High-level predators are disproportionately individual specialists relative to other trophic niches; therefore, it is critical that management agencies examine this phenomenon when working with interactions between species of conservation or ecological concern, such as harbor seals and their prey (Araújo et al., 2011). Recently, it has been found that Salish Sea harbor seal

diets vary on several scales, rejecting the idea of a generalist trophic group. Variation in dive behavior and spatial use (and therefore prey utilization) exists between seals at both the male and female level and individual level across the species' global range (Dietz et al., 2013; Peterson, S., et al., 2012; Wilson et al., 2014). Diet analyses in the Salish Sea have indicated that harbor seal diet composition can vary based on the season and geographic location (Lance et al., 2012; Olesuik, 1993; Schwarz et al., 2018). Additionally, male harbor seal diets consist of more salmonids, while female harbor seals diets consist of more salmonid predators, highlighting the potential for contradictory top-down pressures on vulnerable prey species (Schwarz et al., 2018). This variation creates a confounding yet crucial consideration when evaluating seal diet breadth and total trophic influence in the development of management plans regarding a historically protected species (Jiang & Morin, 2005). While the examination of sex ratios in regional seal communities should help management plans, agencies would further benefit from the evaluation of individual specialization rates in harbor seals, which remain unknown. The existence of individual specialization in harbor seals has not been conclusively demonstrated, but indirect evidence suggests that it is prevalent in the Salish Sea region (Bjorkland et al., 2015; Bromaghin et al., 2013; Matthews et al., 2010; Schwarz et al., 2018; Voelker et al., 2020), posing challenges to the creation of practical management plans (Marshall et al., 2016). If specific seal populations exhibit drastically different diets than expected by species-level averages (even when taking into account the aforementioned sources of variation), broad-scale management may miss another source of localized selection pressure on prey species. Due to the likelihood that this specialization occurs in harbor seals, the development of methods for individual-based studies using this system should provide a helpful framework for high-resolution analysis of predator diets.

This project aimed to develop a novel application of the FecalSeq enrichment protocol for tracking individual predators of management concern, such as pinnipeds. Using harbor seals in the Salish Sea as a study system, a framework for employing non-invasive/low-quality samples in predator trophic dynamics was tested. Specifically, a method to track individual harbor seals in the Salish Sea and their diet through the direct sequencing of scat DNA and analysis of all resulting reads. This approach relied on testing if harbor seal scat represents an adequate resource for identifying robust SNPs for individual identification, and if there is enough intact prey DNA in scat for the simultaneous individual diet analyses, avoiding the need for any additional sequencing and/or PCR-based methods. This study hopes to provide future investigations with information on the performance of samples such as scat in an affordable, scalable framework, increasing the investigative resolution for inaccessible and/or protected species.

Methods

DNA extraction and enrichment

Ten scat samples previously collected at three haul-out sites from 2013-2016 (wild samples) in the Salish Sea and two scat samples collected for this study from a male harbor seal at the Seattle Aquarium (captive samples) were selected for the DNA library (Table 1). Wild samples (PV.04, PV.05, PV.06, PV.07, PV.09, PV.14, PV.15, PV.16, PV.17, PV.18) were selected to assess this genotyping method on a range of collected samples, while captive samples (PV.02, PV.19) were selected to serve as genotype controls. Of the wild samples, PV.04, PV.05, PV.06, and PV.14 were collected in 2013 from a single haul-out site in Comox, BC, and individually homogenized, with the resulting slurries stored in ethanol at -20C as described in Thomas et al. (2017). PV.07 and PV.09 were collected in 2016 from Whidbey Island, WA, and processed in the same manner (Voelker et al., 2020). The remaining 4 wild samples (PV.15, PV.16, PV.17, PV.18) were collected in 2014 from Cowichan Bay, BC (Rothstein et al., 2017). These last four samples were mucous swabs of scat exteriors that were subsequently stored in 95% EtOH at -20C. These swabs were taken in an attempt to increase the proportion of seal epithelial cells in the samples, and therefore the amount of seal DNA for subsequent analyses. PV.15, PV.16, PV.17, and PV.18 were included to analyze any potential performance disparities between the swab and slurry methodology, in addition to the fact that previous MS analyses indicated that these four swabs may be a resampling event of two distinct individuals (PV.15/PV.16 being one individual and PV.17/PV.18 being the other) (Rothstein et al., 2017). The two captive samples (PV.02, PV.19) were collected in May 2019 from Hogan the harbor seal at the Seattle Aquarium in Seattle, WA, and stored in 95% EtOH at -80C, until homogenization and storage at -20C following the protocol described in Thomas et al. (2017).

Collection at Comox was conducted by a team from University of British Columbia under Fisheries and Oceans Canada Marine Mammal Research License (MML 2011-10) and a University of British Columbia Animal Care Permit (A11-0072) awarded to University of British Columbia Marine Mammal Research Unit. Collection of Cowichan Bay samples was conducted by Sheena Majewski, Research Biologist at the Department of Fisheries and Oceans Canada, under "License to Study Marine Mammals for Research Purposes MML-003". Collections at Whidbey Island were conducted by a team from Western Washington University under Federal Permit 18002 from the United States Office of Protected Resources, National Marine Fisheries Service, awarded to Alejandro Acevedo-Gutiérrez.

The swab samples were previously extracted in 2014 with a Qiagen DNeasy® Blood and Tissue kit as it was not thought that there would be a large number of inhibitors present in the mucous matrix; due to the sample type, there was no remaining material to re-extract for this study (Rothstein et al., 2017). For DNA extraction of the other samples, 1500 µL of settled slurry were aliquoted into 2mL microcentrifuge tubes, briefly centrifuged, and drained of excess EtOH. Slurries were dried in a SpeedVac until all EtOH was evaporated. 200 mg of dried wild slurries were extracted with the NucleoSpin® DNA Stool kit, while 200 mg of dried captive slurries were extracted using the MP Biomedical FastDNA™ Spin Kit for Feces (with an increased final elution volume of 60 µL loaded in two consecutive 30 µL aliquots to the filter column). DNA quantification was performed on a Qubit™ 4 Fluorometer using the 1X dsDNA Assay to ensure adequate yield. DNA samples were sent to the University of Minnesota Genomics Center for enrichment using the NEBNext® Microbiome DNA Enrichment Kit, following the protocol described in Chiou & Bergey (2018). Samples were then standardized to 10 ng/µL prior to sequencing.

Sequencing and variant identification

Samples were sequenced following the ddRADseq protocol described in Peterson, B., et al. (2012) using the restriction enzymes SbfI and TaqI. Sequencing was performed on an Illumina NextSeq 550, with 150bp single end reads and 10 μL sample input. Sequencing of all samples excluding PV.19 was initially performed on a 0.25 lane, and these preliminary results required a new aquarium replicate to be sent for sequencing (PV.19). All samples were then sequenced on a 0.75 lane and reads from the first run and second run were included in analysis. Sample reads were demultiplexed, and the first 12 bases were removed from each read to eliminate adaptor sequences, along with adapter trimming and quality trimming (sliding window of 4 bases with a minimum Q=16 and removal of bases with Q<3) from the 3' end with Trimmomatic (Bolger et al., 2014). Reads shorter than 50bp were also removed.

The following methods were applied to four levels of read depth: 100% of all reads per sample, 75% of randomly selected reads, 50% of randomly selected reads, and 25% of randomly selected reads. Subsampling was performed in order to examine the effect sequencing coverage may have on the ability to distinguish between individual samples. Subsample levels were selected based on percentage rather than absolute number of reads due to the wide range of reads returned per sample, which would eliminate samples from analyses when using absolute cutoffs. Prior to SNP identification, all clean reads from each sample were aligned to the harbor seal reference genome (GenBank accession GCA_004348235.1) using BBSplit (BBMap – Bushnell B.). BBSplit was used prior to alignment in an attempt to separate out any potentially erroneous mappings to the seal reference of exogenous sources in highly conserved regions. This step included the simultaneous alignment to potential prey/contaminating taxa references (described

below), reducing the potential for multiple mappings of the same read. Ambiguous reads (reads that partially aligned to multiple references) were assigned to the reference with the first best site. Reads that aligned to the harbor seal genome were then aligned to the harbor seal reference using BWA-MEM V0.7.17 (Li & Durbin, 2009). High-coverage areas were down sampled to <500x using SEQTK (https://github.com/lh3/seqtk). The resulting aligned files were converted to sorted and indexed BAM files with SAMtools V1.9 (Li et al. 2009). Freebayes V0.9.21 (Garrison & Marth, 2012) was used to jointly call variants across all samples simultaneously. A final, filtered set of variants was created for each subsampling level with VCFtools V0.1.16 (Danacek et al. 2011). Variants were filtered to select loci with genotyping rates of 100% across all samples and quality scores >20. Filtering based on minor allele frequency (MAF) was not performed due to the low sample size of the library (with 12 samples the smallest possible MAF is 0.042). These four sets of variants (100%, 75%, 50%, and 25% subsample levels) were used for subsequent genotyping analyses.

Individual identification

Individual genotypes extracted for each locus in the VCF file were compared using NgsRelate2 (Hanghøj et al., 2019). This program allows for the estimation of relatedness, inbreeding coefficients, and other summary statistics between pairs of individuals in NGS data sets where low coverage may produce unreliable genotype calls. Specifically, comparisons between the estimates of pairwise relatedness (RAB, Hedrick & Lacy, 2014), coefficient of kinship (θ, Jacquard, 1974), and zygosity (Ackerman et al., 2017) between samples were used for this analysis. These relatedness measures rely on the proportion of homologous alleles shared by individuals due to identity-by-descent (IBD) and are based on Jacquard's nine

condensed coefficients of identity (Δ_1 - Δ_9) (Jacquard, 1974). These coefficients each describe the probability of individuals sharing a randomly sampled homologous allele through a different relationship (e.g., sibling/sibling, grandparent/grandchild, cousin/cousin, etc.).

RAB is the pairwise relatedness between individuals based on the modes of IBD, with an upper bound of 1, permitting use in small and/or inbred populations, an important consideration for populations of conservation concern with potential historical bottlenecks. With this measure, two samples from the same individual should always return a value of 1, regardless of inbreeding levels, while sibling pairs and parent/offspring pairs should return values around 0.5. This estimate of pairwise relatedness between individuals x and y is represented by

$$\text{RAB}_{xy} = \Delta_1 + \Delta_7 + {}^3/_4(\Delta_3 + \Delta_5) + {}^1/_2(\Delta_8)$$

θ is the probability that 2 homologous alleles randomly sampled between individuals are IBD. Therefore, an identical individual sampled twice should return a value of θ of 0.5, while parent/offspring pairs and sibling pairs should return values around 0.25 (in outbred populations). This estimate of kinship between individuals x and y is represented by

$$\theta_{xy} = \Delta_1 + {}^1/_2(\Delta_3 + \Delta_5 + \Delta_7) + {}^1/_4(\Delta_8)$$

Zygosity (how one individual's status of heterozygosity or homozygosity is correlated to another individual's) is defined as the probability that two individuals through descent, share identical genotypes at a locus. An identical individual sampled twice should exhibit a zygosity correlation of 1. Siblings, able to receive the same alleles from *both* parents, should share a zygosity correlation of 0.25 on average. The zygosity between parent/offspring pairs should be zero, due to the fact that a parent only passes on one haploid genome to each offspring.

Therefore, this coefficient should help distinguish relationships, as the previous two coefficients have the same expected values for sibling pairs and parent/offspring pairs. The estimate of correlation of zygosity between individuals x and y is represented by

$$\Delta_{xy} = \Delta_1 + \Delta_2 + \Delta_7$$

To confirm the ability to distinguish pairs of relatives from pairs of identical samples, four *in silico* crosses between samples were performed using the genotypes at each variant, creating simulated offspring. Loci were assumed to be independent in these simulated crosses and linkage analysis was not performed due to the small size of the library. These offspring were crosses of PV.06 x PV.14, PV.07 x PV.14, PV.09 x PV.16, and an additional PV.06 x PV.14. In addition to parent/offspring and full sibling pairs, half sibling pairs were simulated to examine the effect relationship orders have on these relatedness measures.

Prey identification

In addition to the assignment of reads by BBSplit to the harbor seal genome, reads were simultaneously mapped to representative references for a variety of taxa meant to encompass potential prey (and other exogenous DNA sources) that may occur in seal scat. Reference genomes from 17 potential prey species were selected from 11 orders: Gadiformes, Perciformes, Scorpaeniformes, Scombriformes, Clupeiformes, Salmoniformes, Pleuronectiformes, Gasterosteiformes, Osmeriformes, Cephalopods, and Batrichoidiformes. Five bacterial references were also selected in order to examine the amount of bacterial DNA remaining in the samples (see Supplemental Information for assembly information and accession numbers). Output from BBSplit provided the number of reads per sample that mapped to each reference,

allowing for the estimation of which taxa (other than harbor seals) were present in each sample, and their relative abundances. Prey analysis was performed only for the 100% data set.

Results

DNA extraction and enrichment

All samples that were extracted for this study yielded an adequate amount of DNA (>1000 ng per microbiome enrichment kit manufacturers recommendation) for enrichment (Table 2, Figure 1). Although three of the four previously extracted swab samples fell below this threshold, they were included in this study to assess the previous hypothesis that these swabs were resampled individuals as indicated by microsatellite analyses. There was wide variation in the total amount of DNA recovered from each sample, regardless of extraction method (mean 1979.98 ng, range 573.05-4,671.76ng). The average concentration of DNA across samples was 60.99 ng/μL, with wide range as well (range: 4.86-122.94 ng/μL). All samples retained DNA after enrichment (mean 0.58 ng/μL, range 0.065-2.22 ng/μL) and human DNA controls exhibited >90% recovery efficiency when compared to enrichment kit manufacturers specifications.

Sequencing and variant identification

Samples returned a wide range of total reads, averaging 7,973,811 raw reads per sample (range: 1,261,032-15,672,930) (Table 3, Figure 2). Following adapter trimming and read cleaning, samples averaged 7,659,992 reads (range: 1,166,680-15,265,477), and FastQC (Andrews, 2010) confirmed quality scores >30 across samples. On average, 56.86% of sample reads mapped to a reference genome (range: 9.63-93.81%). Of these reads, on average, 35.24% aligned to the harbor seal genome (range: 8.81-88.28%), and 24.56% aligned to an exogenous reference genome (range: 6.10-62.81%) (Table 3, Figure 3). Four out of the five samples with the highest amount of exogenous DNA were scat surface swabs (PV.15, PV.16, PV.17, and PV.18), as opposed to slurries.

The 100%, 75%, 50%, and 25% subsample levels returned a total number of SNPs of 45,801, 38,274, 29,779, and 19,962, respectively. After filtering, these numbers dropped to 4,709, 4,080, 3,220, and 1,723 (Table 4).

Individual identification

Pairs of all wild samples other than PV.17 and PV.18 (a proposed resampling event from a previous study) displayed extremely low values of both pairwise relatedness and coefficients of kinship (Figure 4, grey). The pair of samples from the individual at the Seattle Aquarium (PV.02 and PV.19, blue), and the pair of proposed resamples (PV.17 and PV.18, orange) conversely displayed elevated measures, approaching the theoretical maxima for identical individuals. These patterns were present in all read subsample levels. PV.02 and PV.19 displayed an average RAB=0.71, average θ=0.36, and average Δ=0.61 across subsample levels. PV.17 and PV.18 displayed an average RAB=0.94, average θ=0.47, and average Δ=0.91. The proposed pair of samples representing a wild resampling event exhibited relatedness measures expected between identical individuals stronger than those between the known pair of samples from the resampled individual at the Seattle Aquarium. Simulated parent/offspring pairs exhibited average measures congruent with the expected parental values (RAB_{obs}=0.50, θ_{obs}=0.26, Δ_{obs}=0.0010). Simulated sibling pairs also exhibited average values close to those of the expected familial relationship (RAB_{obs}=0.436, θ_{obs}=0.22, Δ_{obs}=0.33). One pair of proposed resamples from a previous study (PV.15 and PV.16), displayed relatedness measures indicative of non-related individuals (RAB_{obs}=0.031, θ_{obs}=0.016, Δ_{obs}=0.099).

Prey identification

In the examination of exogenous DNA found across samples, the total exogenous content varied widely between samples (range: 6.10-62.81%). Within samples, all taxa selected exhibited positive mapping, however taxa-specific mapping rates also varied widely within samples (Table 5). As previously mentioned, an average of 24.56% of reads mapped to an exogenous reference across samples. The most prevalent source of exogenous DNA found across samples was Clupeiformes, which comprised on average 51.11% of all exogenous reads (Table 5, range: 0.204-98.99%). The second most prevalent source of exogenous DNA was Salmoniformes, which comprised an average of 13.11% of all exogenous reads (range: 1.12-42.60%). The third most prevalent source of exogenous DNA was Gadiformes, which 11.48% of all exogenous reads on average (range: 2.42-44.28%). The least prevalent source of exogenous DNA within the taxa chosen was Osmeriformes, which comprised an average of 1.35% of all exogenous reads (range: 0.22-2.79%).

When compared to a known captive diet, exogenous reads from aquarium samples (PV.02 and PV.19) followed the same general pattern, with Clupeiformes again dominating (Figure 6). Clupeiformes (*C. pallasii*) comprised 45% by mass of the aquarium diet, with 40.62% and 41.66% of the exogenous reads in PV.02 and PV.19 respectively showing Clupeid origin. Although Osmeriformes and Cephalopods comprised almost half of the aquarium diet (45% by mass), the combined portion of total prey reads (excluding bacterial counts) corresponding to these orders in PV.02 and PV.19 were only 7.00% and 7.61% respectively. Scombriformes comprised the minority of the captive diet (10% by mass) and returned the least amount of reads in PV.02 and PV.19 (2.62% and 2.08% respectively). Reads from PV.02 and

PV.19 aligned to all exogenous references employed, despite a known diet that did not contain all taxa exhibiting alignments.

Discussion

SNPs identified from the enriched seal scat samples successfully distinguished between replicate samples from the same individuals and wild samples from distinct individuals, indicating promise for employing non-invasive samples such as scat for the tracking of predators of conservation and/or management concern. While prey sequences were detected in each sample and prey assemblages could be constructed, further research should investigate the true relationship between relative read abundances and actual diet composition when employing a direct sequencing workflow.

The known pair of samples from the captive seal at the Seattle Aquarium, as well as one previously collected pair of proposed resamples exhibited relatedness measures expected for identical individuals. One pair of previously collected samples thought to be a proposed resample event exhibited relatedness measures expected of non-related individuals, challenging previous MS analysis (Rothstein et al., 2017). Simulated parents, offspring, and siblings exhibited relatedness measures expected of their respective relationships. All these finding support the premise that scat samples represent an adequate source of genetic material for non-invasive harbor seal tracking techniques. However, the highly variable content of endogenous DNA across samples may prevent some future field-collected samples from performing as expected, potentially requiring large sample sizes or pre-screening criteria. Prey sequences were identifiable in scat samples from a variety of taxa; however, the content of exogenous DNA was highly variable between samples and sample types (slurries vs. swabs), and samples for which diet composition was known did not have congruent sequence counts. Scat surface swabs exhibited higher exogenous DNA content than slurries. However, slurries contained higher endogenous DNA content than swabs. Therefore, future studies may benefit from increased

comparisons of sample type on sample performance, as well as combinatorial collection techniques for individual samples.

DNA extraction and enrichment/sequencing and variant identification

While qPCR screening for initial seal DNA concentrations in each sample was not performed, an average mapping rate of 35.93% (range: 8.08-86.39%) to the harbor seal reference genome is commensurate with previous applications of the enrichment method employed in this study. Bergey et al. (2018), who developed the methylation-based capture method for non-invasive fecal samples used in this study, reported an average mapping rate of 28.8% (range: 0.7-81.2%) to the target genome, and a mean 195-fold increase in host DNA concentration. Interestingly, samples comprised of extracts from scat slurries contained an average of 41.59% of reads that mapped to the seal genome, while samples comprised of extracts from scat surface swabs averaged 24.60% reads mapped to the seal genome. Challenging the notion of swabs maximizing the proportion of seal epithelial cells, there appeared to be no advantage over homogenizing whole scats when employing a direct sequencing method. A 16S ribosomal RNA assay was performed on a subset of samples prior to enrichment, however, zero samples received bacterial DNA quantification post enrichment, preventing this study from evaluating the exact performance of the enrichment method. All four swab samples received an initial quantification of bacterial DNA and averaged 135,274 16S molecules per 1 ng of total DNA. Six of the homogenized samples received this quantification and averaged 409,854 16S molecules per 1 ng of total DNA. This result indicates that swabbing scats may provide an initial minimization of unwanted bacterial signals. Should time and funding allow, researchers should include this quantification in all steps of analyses in order to demonstrate the efficacy in different

applications. While prescreening samples for those with higher initial proportions of host DNA (and/or lower proportions of bacterial DNA) would certainly benefit the downstream performance of samples, this study shows that ensuring an initial DNA content of >1 µg (per enrichment kit manufacturer's recommendation) permits the genetic analysis of the host species. Furthermore, the reduction in financial costs and labor associated with prescreening for host concentrations (substituted with screening for total gDNA concentration) may permit for increased sample sizes for sample enrichment, increasing final library sizes. As is the case in this study, initial total DNA concentration was not an accurate predictor of post-enrichment concentrations nor final mapping rates. This may indicate that due to the highly variable nature of fecal samples, fecal analyses may benefit from increased sample sizes and a decreased initial screening.

Individual identification

The amount and quality of loci discovered in this study was sufficient for individual identification, even at the lowest read subsampling level. With the 25% subsampling level identifying a final panel of 1,723 filtered SNPs, the ability to distinguish between individuals and even close relatives should be robust, as identity and close-relative analyses can rely on as few as 32-80 SNPs (Hauser et al., 2011; Heaton et al., 2002; Werner et al., 2003). When employing larger numbers of scat samples than this study (e.g., dozens to hundreds), the number of shared loci will decrease. However, using a highly informative set of SNPs (unlike a general panel identified in this study) and subsampling a panel of 960 loci down to just 15 loci can maintain the integrity of unique genotypes, indicating that even in small panels, probabilities of identities can remain within the range allowing for confident identity analyses (Tokarska et al., 2009). It

should be noted that due to the nature of the data generated in this study, results may be noisier due to false alignments and the inability to select informative loci. However, due to the high number of markers identified, stringent and informative panel creation should pose few obstacles. While relative pairs were created with simulated genotypes, relatedness measures commensurate with the respective relationships indicate that this method may allow for fine-scale individual tracking, preventing the collection of samples from parents and/or siblings from appearing as potential matches. Simulated relative pairs also appear to exhibit measures that would permit the analysis of diet similarities between close relatives, without confounding those comparisons with more distant relative pairs (e.g., sibling pairs with RAB ~0.5 and half siblings with RAB ~0).

While the pair of samples from the captive seal at the Seattle Aquarium (PV.02/PV.19) displayed somewhat dampened relatedness metrics (RAB_{obs}=0.71, θ_{obs}=0.36, and Δ_{obs}=0.61 rather than RAB_{exp}=1, θ_{exp}=0.5, and Δ_{exp}=0.1), all other samples sequenced (excluding PV.17/PV.18 and simulated relatives) exhibited relatedness measures approaching zero. There was clear separation between the sample pairs of non-identical individuals and aquarium samples, and as the only known true resampled individual, the sample pair of PV.02/PV.19 represents the threshold for confidently calling identical samples in this study. It should be noted that PV.02 contained the lowest proportion of harbor seal reads than all other samples (8.08%) and the highest proportion of unmapped reads (90.37%), despite having one of the highest total read counts (11,833,891). The aquarium samples were collected from the bottom of the enclosure pool, after spending 1-2 days submerged, essentially washing the surface of the scats of epithelial cells and degrading any DNA present (Brinkman et al., 2009; Murphy et al., 2007; Piggot 2004). The dampened relatedness measures seen between PV.02 and PV.19 may be due

to the low proportion of harbor seal DNA in PV.02 combined with the large seal genome size (~2.4GB), contributing to significant allelic dropout and inconsistent allele ratios between duplicate samples. Therefore, prescreening samples based on initial host DNA proportion through qPCR may prevent low quality samples from skewing future results. This clustering of non-related individuals near zero with known and proposed matches approaching theoretical maxima is commensurate with a previous study investigating these relatedness coefficients with SNPs in wild water fleas (*Daphnia pulex)* populations, which contained both sexually and asexually (cloned) reproducing individuals (Ackerman et al., 2017).

Two pairs of scat samples previously collected and screened using 9 microsatellite loci were thought to represent two individual seals from a haul out site that were each sampled twice (PV.15/PV.16 representing one unique individual, and PV.17/PV.18 representing another unique individual) (Rothstein et al. 2017). According to the analyses in this study, PV.15 and PV.16 appear not to be a resampled individual, but rather two scats from unrelated individuals (a conclusion further strengthened by the 'lower quality' aquarium samples still exhibiting elevated relatedness measures). Conversely, PV.17 and PV.18 returned relatedness metrics indicative of identical individuals, the values of which were the highest out of all other pairwise comparisons (even between the known matching pair of aquarium samples). As this pattern persisted at all subsampling levels, SNPs may outperform microsatellites for identity and familial analyses when using low-quality noninvasive samples, and sequencing coverage may have been able to be lowered even further to reveal the same patterns. This increased performance is supported by multiple studies indicating that as reference genomes for various species continue to be created, SNP analyses may outperform microsatellite analyses for genome-wide investigations (Gärke et al., 2012; Hauser et al., 2011; Kaiser et al., 2016; Liu et al., 2005; Morin et al., 2004). As fecal

samples contain highly degraded DNA from random genomic locations, techniques optimal for genome-wide studies are likely to promise higher chances of success, in addition to the benefits of a much more direct, cost effective sequencing option.

Simulated parent/offspring pairs (RAB_{avg}=0.50, θ_{avg}=0.26, Δ_{avg}=0.0010) and simulated sibling pairs (RAB_{avg}=0.436, θ_{avg}=0.22, Δ_{avg}=0.33) indicate that this method may accurately separate close relatives from resampled individuals, a crucial component of a technique for tracking individual specialization (particularly in populations with bottlenecks and/or high rates of inbreeding). Replicate simulated crosses of PV.06 x PV.14 failed to cluster with resampled individuals, and this simulation of relative genotypes at every single locus genotyped in the scat samples may not accurately reflect the genotyping rates and downstream relatedness measures achieved if biological samples from parents and/or siblings were included (as no linkage analyses were performed due to small library size). Therefore, future work should include samples from known relationship pairs to evaluate the effect using real, potentially low-quality samples may have on distinguishing between relatives.

Prey identification

The highly variable concentrations of exogenous DNA appeared to correlate with the sample types used in this study. Samples, which upon collection were stored in ethanol followed by homogenization and then extraction averaged an exogenous content of 4.91%. Samples which were comprised of swabs taken from the surface of scats, followed by storage in ethanol and extraction averaged an exogenous content of 49.03% (49.37% if including bacterial reads). In bighorn sheep (*Ovis canadensis*) fecal analysis, PCR results from fecal material on the outermost pellet surface were almost equivalent to blood-derived results (Wehausen et al., 2004).

As the surface of scats will dry first, microbial degradation of DNA in this region may be limited, while DNA within the sample continues to degrade. Therefore, while homogenized scats effectively illuminate harbor seal diets through DNA metabarcoding (Schwarz et al. 2018), swabs may outperform slurries when investigating metagenomic samples where target-sequences are not pre-amplified such as in direct sequencing.

The most prevalent prey taxa encountered was Clupeiformes, with swab sample exogenous reads consisting of up to 93.79% clupeid reads. Swab samples were collected in spring 2014, corresponding to when harbor seals primarily consumed Pacific herring (*C. pallasii*) (Lance et al., 2012). The two other most prevalent taxa were Salmoniformes and Gadiformes, which in addition to Clupeiformes, creates a realistic depiction of the majority of harbor seal diet (Lance et al., 2007; Lance et al., 2012; Schwarz et al., 2018). Interestingly, the exogenous DNA within swab samples was dominated by one taxon, where the exogenous DNA within slurry samples contained a wider range of taxa reads. It is possible that the DNA from slurries, having potentially degraded more than surface DNA due to microbial activity, mapped to more taxa due to small fragments with non-specific mapping affinities. Conversely, fecal DNA degrades with increased exposure to sunlight and elements such as heat and rain (Agetsuma-Yanagihara et al., 2017; Nsubuga et al., 2004; Oehm et al., 2011), allowing for the possibility that swab samples are more degraded than slurries, thus only the most common prey taxa is identified through any resulting reads. Future analyses should take care to collect more samples of both types to investigate this finding, as fecal sampling, storage, and extraction methods vary widely between study systems and workflows (Waits & Paetkau, 2005). Regardless of sample type, scat samples were able to construct exogenous profiles that may be able to illuminate changes in consumption patterns if examined throughout time. For example, PV.15, PV.16, PV.17, and PV.18 returned

most exogenous reads from Clupeiformes, however repeated sampling at this haul out site, followed by relatedness analyses and exogenous mapping could show whether or not the unique genotype that is associated with multiple scat samples consistently returns mostly clupeid exogenous reads or not.

While the diet of the captive seal at the Seattle Aquarium was available, exact consumption patterns of the seal immediately prior to collection and sequencing were not known. However, there was a general trend seen of the most abundant prey species (Clupeiformes) comprising the majority of exogenous reads in PV.02 and PV.19. Previous studies have used read counts of prey taxa to draw conclusions about diet composition (Brown et al., 2012; Deagle et al., 2009; Murray et al., 2011); however, some studies have indicated there may be biases in using read counts that are heavily dependent on experimental design and data analysis (Berry et al., 2012; Deagle et al., 2013). Reads in the aquarium samples were also assigned to species that were known to not be present in the captive diet (e.g., Salmoniformes and Gadiformes), indicating not just erroneous read abundances, but species assignments. Harbor seal diet analyses have employed correction factors that factor in differential digestion, DNA extraction, and PCR amplification, calculated by sequencing in parallel a mixture of prey and scat from an individual fed the same mixture of prey (Thomas et al., 2014). Although this correction allows for a better correlation between actual diet and observed read counts, it relies on knowing the exact diet of the individual immediately before sequencing of a scat sample, which conflicts with the scope of this non-invasive technique. An attempt to create standard correction factors for four fish species in field-collected harbor seal scat samples did find that increased read counts did result from increased diet input for a given species (Thomas et al., 2016). Interestingly, when a species made up >50% of the diet, the resulting read counts underestimated the abundance, while when a

species made up <50% of the diet, read counts overestimated the abundance. However, these biases were predictable and consistent in magnitude, and using these correction factors did improve the relationship between read counts and diet composition.

Thomas et al. (2016) employed targeted amplicon sequencing of mitochondrial DNA (mtDNA) for prey identification, and mtDNA is typically employed for species identification studies due to large number of mtDNA copies per cell, while nuclear DNA is used for individual identification (Frantzen et al., 1998; Kohn et al., 1999; Lucchini et al., 2002; Poole et al., 2001; Schwarz et al., 2018; Waits 2004). This potentially unreliable nature of nuclear DNA read counts from direct sequencing may help explain the discrepancies seen between the Seattle Aquarium's diet composition and the species assignments in PV.02/PV.19. Thomas et al. (2017) also used samples PV.05, PV.06, PV.07, PV.09, and PV.14 for prey analysis through targeted mtDNA amplicons, and discrepancies between the results here and Thomas et al. (2017) also appear (Table 7). For example, Thomas et al. (2017) found PV.07 contained only cephalopod mtDNA amplicons, whereas this study had an extremely low proportion of cephalopod sequences (2.16% of exogenous reads). High water content organisms such as cephalopods have been shown to degrade more completely during digestion than other species, which may account for the lack of detection (Thomas et al., 2014). Additionally, the binning program used in this study may have wrongly sorted some sequences if they lacked diagnostic characteristics or contained fast evolving regions that have lost phylogenetic signals and simply matched a reference assembly by chance (similar to long branch attraction in phylogenetic analyses). Heterogeneity in the quality/completeness of the different reference genomes may have also caused erroneous alignments.

Hard parts analyses of scat samples suggest samples on average contain two to three species, and metabarcoding techniques also support this (Orr et al., 2003; Schwarz et al., 2018; Thomas et al., 2016; Tollit et al., 2009). Therefore, the direct sequencing of prey DNA in scat likely erroneously classifies prey sequences, as it does not pick-up accurate prey levels due to varied digestion rates, low levels of indirect prey, assembly gaps, or a combination of these factors. While more work is warranted in the employment of direct sequencing for accurate prey analysis, increased sequencing efforts, the creation of more complete prey reference genomes, and feeding trials for digestion correction factors over various timescales/conditions may help overcome challenges.

Conclusions

In this study, non-invasively collected scat samples were employed for the successful identification of individual harbor seals through the direct sequencing of scat DNA followed by SNP genotyping. A pair of scat samples from a known individual at the Seattle Aquarium, as well as one pair of field-collected scat samples previously thought to represent a resampling event (through microsatellite analysis) exhibited genotypes consistent with identical individuals, while unrelated samples from separate populations exhibited no risk of misidentifications. A pair of field-collected samples also thought to be a potential resampling event (through microsatellite analysis) appears through SNP analysis to be a pair of samples from two unrelated individuals. Through the simulation of relative pairs from the final, filtered SNP panel, this study shows that with highly informative SNPs (a large number of which are identifiable through the endogenous DNA present in scat samples), samples from close relatives should not erroneously classify as a resample for an individual seal. In fact, this method would allow for

the additional fine-scale investigation of social structure in inaccessible species like harbor seals (e.g., lifetime haul-out site fidelity). Due to a lack of knowledge on the complete *P. vitulina* genome, combined with a small sample size, a linkage analysis on SNP loci was not performed or factored into simulated relatives. Therefore, future work should examine these conclusions, as no biological samples from relatives were included in this study.

These results highlight the ability for these sample types to serve as tools when combined with NGS techniques, for tracking seals throughout time by repeatedly sampling scat at haul-out sites. Additionally, the sequencing of the metagenomic DNA population in these scat samples returned exogenous reads that could be mapped to various prey taxa, indicating that in conjunction with tracking seals, this technique may assist in studies of individual specialization in seal communities. Should correction factors for the differential detectability of prey species be investigated in directly sequenced scat samples, the ability to quantify diet beyond presence/absence will improve this potential. Additional comparisons between direct sequencing, metabarcoding, and hard part analysis within the same workflow including feeding trials would also highlight potential discrepancies between these methods for examining diets. While DNA samples extracted from scat slurries contained the highest amounts of endogenous seal DNA, samples extracted from scat surface swabs contained the highest amounts of exogenous prey DNA (granted swab samples simply are not more degraded due to UV radiation and misassignments of small fragments). Therefore, in order to investigate the reliability of this pattern, combinatorial sampling of both swabs and slurries per sample (or direct sequencing of nuclear DNA for identification and mtDNA amplicon sequencing for diet) may also benefit future non-invasive techniques of tracking individual predators and their diets.

This study attempted to increase the practical knowledge of employing non-invasive and/or low-quality samples such as scat in ecological investigations on fine spatiotemporal scales, using an ecologically influential marine mammal as a case study. The successful identification of harbor seals in combination with the ability to detect the presence of prey species simultaneously from the direct sequencing of scat represents a completely non-invasive pipeline for tracking an inaccessible, protected species. Here, non-invasive techniques and NGS technologies have been shown to increase the accessibility of fine-scale ecological studies, in an affordable and scalable manner. The sufficient performance of easily collected swab samples, along with the relative ease of acquiring non-invasive field permits and the ability to multiplex samples pose a promising future for management agencies. Resulting SNP panels can also be employed on other platforms (e.g., microarrays and PCR), increasing our understanding of these predator populations. The increased emphasis on non-invasive samples such as scat may also benefit future attempts at non-invasive, high resolution multiomics (e.g., genome, epigenome, metabolome, and microbiome investigations), allowing for insights previously unexpected from a single sample-type. This study provides a framework and considerations for analyses of target and/or non-model species as management agencies seek more accurate models of trophic dynamics on increasingly defined local scales, while also highlighting the power of NGS methods in a variety of potential non-invasive investigations.

Works Cited

Ackerman, M.S., Johri, P., Spitze, K., Xu, S., Doak, T.G., Young, K., and Lynch, M. (2017). Estimating seven coefficients of pairwise relatedness using population-genomic data. *Genetics* 206, 105–118.

Agetsuma-Yanagihara, Y., Inoue, E. and Agetsuma, N. (2017). Effects of time and environmental conditions on the quality of DNA extracted from fecal samples for genotyping of wild deer in a warm temperate broad-leaved forest. Mammal Research 62, 201–207.

Araújo, M.S., Bolnick, D.I., and Layman, C.A. (2011). The ecological causes of individual specialisation: The causes of individual specialisation. *Ecology Letters* 14, 948–958.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* 3, e3376.

Balme, G.A., Roex, N. le, Rogan, M.S., and Hunter, L.T.B. (2020). Ecological opportunity drives individual dietary specialization in leopards. *Journal of Animal Ecology* 89, 589–600.

Beaudoin, C.P., Tonn, W.M., Prepas, E.E., and Wassenaar, L.I. (1999). Individual specialization and trophic adaptability of Northern pike (*Esox lucius*): an isotope and dietary analysis. *Oecologia* 120, 386–396.

Berry, D., Ben Mahfoudh, K., Wagner, M., and Loy, A. (2012). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied Environmental Microbiology* 78, 612–612.

Bjorkland, R., Pearson, S., Jeffries, S., Lance, M., Acevedo-Gutiérrez, A., and Ward, E. (2015). Stable isotope mixing models elucidate sex and size effects on the diet of a generalist marine predator. *Marine Ecology Progress Series*, 526, 213–225.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Bolnick, D.I., Svanbäck, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D., and Forister, M.L. (2003). The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist* 161, 1–28.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., And Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13, 3261–3273.

Bowen, W. (1997). Role of marine mammals in aquatic ecosystems. *Marine Ecology Progress Series* 158, 267–274.

Brinkman, T.J., Schwartz, M.K., Person, D.K., Pilgrim, K.L., and Hundertmark, K.J. (2010). Effects of time and rainfall on PCR success using DNA extracted from deer fecal pellets. *Conservation Genetics* 11, 1547–1552.

Bromaghin, J.F., Lance, M.M., Elliott, E.W., Jeffries, S.J., Acevedo-Gutiérrez, A., and Kennish, J.M. (2013). New insights into the diets of Harbor seals (*Phoca vitulina*) in the Salish Sea revealed by analysis of fatty acid signatures. *Fishery Bulletin* 111.

Brown, D.S., Jarman, S.N., and Symondson, W.O.C. (2012). Pyrosequencing of prey DNA in reptile faeces: analysis of earthworm consumption by slow worms. *Molecular Ecology Resources* 12, 259–266.

Cammen, K.M., Schultz, T.F., Bowen, D.W., Hammill, M.O., Puryear, W.B., Runstadler, J., Wenzel, F.W., Wood, S.A., and Kinnison, M. (2018). Genomic signatures of population bottleneck and recovery in Northwest Atlantic pinnipeds. *Ecology and Evolution* 8, 6599–6614.

Chasco, B., Kaplan, I.C., Thomas, A., Acevedo-Gutiérrez, A., Noren, D., Ford, M.J., Hanson, M.B., Scordino, J., Jeffries, S., Pearson, S., Marshall, M.N., and Ward, E.J. (2017). 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.

Chiou, K. L., and Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, *8*(1).

Coltman, D.W. (2008). Molecular ecological approaches to studying the evolutionary impact of selective harvesting in wildlife. *Molecular Ecology* 17, 221–235.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12, 499–510.

Deagle, B.E., Tollit, D.J., Jarman, S.N., Hindell, M.A., Trites, A.W., and Gales, N.J. (2005). Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology* 14, 1831–1842.

Deagle, B.E., Gales, N.J., Evans, K., Jarman, S.N., Robinson, S., Trebilco, R., and Hindell, M.A. (2007). Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLOS ONE* 2, e831.

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

Deagle, B.E., Thomas, A.C., Shaffer, A.K., Trites, A.W., and Jarman, S.N. (2013). Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources* 13, 620–633.

Dietz, R., Teilmann, J., Andersen, S.M., Rigét, F., and Olsen, M.T. (2013). Movements and site fidelity of Harbour seals (*Phoca vitulina*) in Kattegat, Denmark, with implications for the epidemiology of the phocine distemper virus. *ICES Journal of Marine Science* 70, 186–195.

Ferter, K., Hartmann, K., Kleiven, A.R., Moland, E., and Olsen, E.M. (2015). Catch-and-release of Atlantic cod (*Gadus morhua*): post-release behaviour of acoustically pretagged fish in a natural marine environment. *Canadian Journal of Fisheries and Aquatic Sciences* 72, 252–261.

Ford, M.J., Hempelmann, J., Hanson, M.B., Ayres, K.L., Baird, R.W., Emmons, C.K., Lundin, J.I., Schorr, G.S., Wasser, S.K., and Park, L.K. (2016). Estimation of a killer whale (*Orcinus orca*) population's diet using sequencing analysis of DNA from feces. PLOS ONE 11, e0144956.

Frantzen, M.A.J., Silk, J.B., Ferguson, J.W., Wayne, R.K., and Kohn M.H. (1998). Empirical evaluation of preservation methods for fecal DNA. *Molecular Ecology* 7, 1423-1428.

Gärke, C., Ytournel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., and Simianer, H. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations: Population differentiation: SNPs vs SSRs. *Animal Genetics* 43, 419–428.

Garrison E, Marth G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907 [q-bio.GN].

Guertin, D.A., Harestad, A.S., Ben-David, M., Drouillard, K.G., and Elliott, J.E. (2010). Fecal genotyping and contaminant analyses reveal variation in individual River otter exposure to localized persistent contaminants. *Environmental Toxicology and Chemistry* 29, 275–284.

Gulsby, W.D., Killmaster, C.H., Bowers, J.W., Laufenberg, J.S., Sacks, B.N., Statham, M.J., and Miller, K.V. (2016). Efficacy and precision of fecal genotyping to estimate coyote abundance. *Wildlife Society Bulletin* 40, 792–799.

Hanghøj, K., Moltke, I., Andersen, P.A., Manica, A., and Korneliussen, T.S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience* 8, giz034.

Hauser, L., Baird, M., Hilborn, R., Seeb, L.W., and Seeb, J.E. (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild Sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources* 11, 150–161.

Hedrick, P.W., and Lacy, R.C. (2015). Measuring relatedness between inbred individuals. *Journal of Heredity* 106, 20–25.

Huber, H.R., Jeffries, S.J., Lambourn, D.M., and Dickerson, B.R. (2010). Population substructure of harbor seals (*Phoca vitulina richardsi*) in Washington State using mtDNA. *Canadian Journal of Zoology* 88, 280–288.

Huber, H.R., Dickerson, B.R., Jeffries, S.J., and Lambourn, D.M. (2012). Genetic analysis of Washington State harbor seals (*Phoca vitulina richardii*) using microsatellites. *Canadian Journal of Zoology* 90, 1361–1369.

Jacquard, A. (1974). *The Genetic structure of populations*. Berlin: Springer.

Jacquier, M., Simon, L., Ruette, S., Vandel, J.-M., Hemery, A., and Devillard, S. (2020). Isotopic evidence of individual specialization toward free-ranging chickens in a rural population of Red foxes. *European Journal of Wildlife Research* 66, 15.

Jeffries, S., Huber, H., Calambokidis, J., and Laake, J. (2003). Trends and status of Harbor seals in Washington State: 1978-1999. *The Journal of Wildlife Management* 67, 207.

Jiang, L., and Morin, P.J. (2005) Predator diet breadth influences the relative importance of bottom-up and top-down control of prey biomass and diversity. *The American Naturalist* 165, 350-363.

Johnson, D.H. (2002). The importance of replication in wildlife research. *The Journal of Wildlife Management* 66, 919.

Kaiser, S.A., Taylor, S.A., Chen, N., Sillett, T.S., Bondra, E.R., and Webster, M.S. (2017). A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources* 17, 183–193.

Kalinowski, S.T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity* 88, 62–65.

Kraus, R.H., Kerstens, H.H., Van Hooft, P., Crooijmans, R.P., Van Der Poel, J.J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H.H., et al. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics* 12, 150.

Kohn, M., York, E.C., Kamradt, D.A., Haught, G., Sauvajot, R.M., and Wayne, R.K. (1999). *Estimating population size by genotyping feces. Proceedings of the Royal Society of London Series B* 266, 657-663.

Lance, M.M., and S.J. Jeffries. 2006. Estimating importance of rockfish, lingcod and other bottomfish in the diet of Harbor seals in the San Juan Islands. *Contract Report to SeaDoc Society Research Agreement No. K004431-22*. Washington Department of Fish and Wildlife, Olympia WA.

Lance, M., Chang, W., Jeffries, S., Pearson, S., and Acevedo-Gutiérrez, A. (2012). Harbor seal diet in northern Puget Sound: implications for the recovery of depressed fish stocks. *Marine Ecology Progress Series* 464, 257–271.

Lance, M.M., and S.J. Jeffries. (2007) Temporal and spatial variability of Harbor seal diet in the San Juan Island archipelago. *Contract Report to SeaDoc Society Research Agreement No. K004431-25*. Washington Department of Fish and Wildlife, Olympia WA.

Lathuillière, M., Ménard, N., Gautier-Hion, A., and Crouau-Roy, B. (2001). Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in Barbary macaques (*Macaca sylvanus*). *American Journal of Primatology* 55, 151–158.

le Roex, N., Noyes, H., Brass, A., Bradley, D.G., Kemp, S.J., Kay, S., van Helden, P.D., and Hoal, E.G. (2012). Novel SNP discovery in African buffalo, *Syncerus caffer*, using high-throughput sequencing. *PLOS ONE* 7, e48792.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Liu, N., Chen, L., Wang, S., Oh, C., and Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* 6, S26.

Lucchini, V., Fabbri, E., Marucco, F., Ricci, S., Boitani, L., and Randi, E. (2002). *Noninvasive molecular tracking of colonizing wolf (Canis lupus) packs in the western Italian Alps. Molecular Ecology* 11, 857-868.

Luxa, K., and Acevedo-Gutiérrez, A. (2013). Food habits of Harbor seals (*Phoca vitulina*) in two estuaries in the central Salish Sea. *Aquatic Mammals* 39, 10–22.

Marshall, K.N., Stier, A.C., Samhouri, J.F., Kelly, R.P., and Ward, E.J. (2016). Conservation challenges of predator recovery: conservation challenges of predator recovery. *Conservation Letters* 9, 70–78.

Matich, P., Heithaus, M.R., and Layman, C.A. (2011). Contrasting patterns of individual specialization and trophic coupling in two marine apex predators: Specialization in top marine predators. *Journal of Animal Ecology* 80, 294–305.

Matthews B, Marchinko K.B., Bolnick D.I., Mazumder A. (2010) Specialization of trophic position and habitat use by sticklebacks in an adaptive radiation. *Ecology* 91, 1025–1034.

Morin, P.A., Luikart, G., Wayne, R.K., and the SNP workshop group. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* 19, 208–216.

Morin, P.A., and Mccarthy, M. (2007). Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes* 7, 937–946.

Murphy, M.A., Kendall, K.C., Robinson, A., and Waits, L.P. (2007). The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics* 8, 1219–1224.

Murray, D.C., Bunce, M., Cannell, B.L., Oliver, R., Houston, J., White, N.E., Barrero, R.A., Bellgard, M.I., and Haile, J. (2011). DNA-Based Faecal Dietary Analysis: A comparison of qPCR and high throughput sequencing approaches. *PLOS ONE* 6, e25776.

Natesh, M., Taylor, R.W., Truelove, N.K., Hadly, E.A., Palumbi, S.R., Petrov, D.A., and Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution* 10, 853–859.

Navarro, Joan & Grémillet, David & Ramirez, Francisco & Afán, Isabel & Bouten, W. & Forero, Manuela. (2017). Shifting individual habitat specialization of a successful predator living in anthropogenic landscapes. *Marine Ecology Progress Series* 578, 243-251.

Newby, T.C. (1973). Changes in Washington State harbor seal population, 1942-1972. *The Murrelet* 54, 4-6.

Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* 12, 443–451.

Nsubuga, A. M., M. M. Robbins, A. D. Roeder, P. A. Morin, C. Boesch, and L. Vigilant. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* 13, 2089–2094.

Oehm, J., Juen, A., Nagiller, K., Neuhauser, S., and Traugott, M. (2011). Molecular scatology: how to improve prey DNA detection success in avian faeces? *Molecular Ecology Resources*, 11.

Olesiuk, P.F. (1993). Annual prey consumption by harbor seals (*Phoca vitulina*) in the Strait of Georgia, British Columbia. *Fishery Bulletin* 91, 491-525.

Olesiuk, P.F. (2009). An assessment of population trends and abundance of harbour seals (*Phoca vitulina*) in British Columbia. Fisheries and Oceans Canada. *Canadian Stock Assessment Secretariat Research Document* 99/33.

Orr, A.J., Laake, J.L., Dhruv, M.I., Banks, A.S., DeLong, R.L., and Huber, H.R. (2003). *Comparison of Processing Pinniped Scat Samples Using a Washing Machine and Nested Sieves. Wildlife Society Bulletin (1973-2006)* 31, 253–257.

Pamplin, N., Pearson, S., and Anderson, J. (2018). *Impact of pinnipeds on chinook salmon*. Presentation from the 2018 WDFW Commission meeting, Olympia, WA.

Perry, G. H., Marioni, J. C., Melsted, P. & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology* 19, 5332–5344.

Peterson, S.H., Lance, M.M., Jeffries, S.J., and Acevedo-Gutiérrez, A. (2012). Long distance movements and disjunct spatial use of Harbor seals (*Phoca vitulina*) in the inland waters of the Pacific Northwest. *PLOS ONE* 7, e39046.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* 7, e37135.

Poole, K.G., Mowat, G., and Fear, D.A. (2001). *DNA based population estimate for grizzly bears Ursus arctos in northern British Columbia, Canada. Wildlife Biology* 7, 105-115.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sichertiz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., MetaHIT Consortium, Bork, P., Ehlrich, S.D., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.

Reddy, P.A., Bhavanishankar, M., Bhagavatula, J., Harika, K., Mahla, R.S., and Shivaji, S. (2012). improved methods of carnivore faecal sample preservation, DNA extraction and quantification for accurate genotyping of wild tigers. *PLOS ONE* 7, e46732.

Reed, J.Z., Tollit, D.J., Thompson, P.M., and Amos, W. (2003). Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology* 6, 225–234.

Rothstein, A. P., McLaughlin, R., Acevedo-Gutiérrez, A. & Schwarz, D. 2017. WISEPAIR: a computer program for individual matching in genetic tracking studies. *Molecular Ecology Resources* 17, 267-277.

Sanchez, J.J., and Endicott, P. (2006). Developing multiplexed SNP assays with special reference to degraded DNA templates. *Nature Protocols* 1, 1370–1378.

Schwarz, D., Spitzer, S.M., Thomas, A.C., Kohnert, C.M., Keates, T.R., and Acevedo-Gutiérrez, A. (2018). Large-scale molecular diet analysis in a generalist marine mammal reveals male preference for prey of conservation concern. *Ecology and Evolution* 8, 9889–9905.

Seddon, J.M., Parker, H.G., Ostrander, E.A., and Ellegren, H. (2005). SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* 14, 503–511.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources* 11, 1–8.

Senn, H., Ogden, R., Cezard, T., Gharbi, K., Iqbal, Z., Johnson, E., Kamps-Hughes, N., Rosell, F., and McEwing, R. (2013). Reference-free SNP discovery for the Eurasian beaver from restriction site-associated DNA paired-end data. *Molecular Ecology* 22, 3141–3150.

Sigsgaard, E.E., Jensen, M.R., Winkelmann, I.E., Møller, P.R., Hansen, M.M., and Thomsen, P.F. (2020). Population-level inferences from environmental DNA- Current status and future perspectives. *Evolutionary Applications* 13, 245–262.

Snyder-Mackler, N., et al. (2016). Efficient genome-wide sequencing and low coverage pedigree analysis from non-invasively collected samples. *Genetics* 203, 699–714

Stick, K.C., Lindquist, A., and Lowry, D. (2014). *2012 Washington State herring stock status report*. 106.

Svanbäck, R., and Bolnick, D.I. (2007) Intraspecific competition affects the strength of individual specialization: an optimal diet theory method. *Evolutionary Ecology Research* 7, 993-1012.

Svanbäck, R., Persson, L. (2009). Population density fluctuations change the selection gradient in Eurasian perch. *The American Naturalist* 173, 507–516.

Thomas, A.C., Jarman, S.N., Haman, K.H., Trites, A.W., and Deagle, B.E. (2014). Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias. *Molecular Ecology* 23, 3706–3718.

Thomas, A.C., Deagle, B.E., Eveson, J.P., Harsch, C.H., and Trites, A.W. (2016). Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources* 16, 714–726.

Thomas, A.C., Nelson, B.W., Lance, M.M., Deagle, B.E., and Trites, A.W. (2017). Harbour seals target juvenile salmon of conservation concern. *Canadian Journal of Fisheries and Aquatic Sciences* 74, 907–921.

Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J.M., Pertoldi, C., Kristensen, T.N., Loeschcke, V., Gregersen, V.R., and Bendixen, C. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* 103, 326–332.

Tollit, D. J., Schulze, A.D., Trites, A.W., Olesuik, P.F., Crockford, S.J., Gelatt, T.S., Ream, R.R., Miller, K.M. (2009). *Development and application of DNA techniques for validating and improving pinniped diet estimates*. *Ecological Applications* 19, 889–905.

Valencia, L.M., Martins, A., Ortiz, E.M., and Di Fiore, A. (2018). A RAD-sequencing approach to genome-wide marker discovery, genotyping, and phylogenetic inference in a diverse radiation of primates. *PLOS ONE* 13, e0201254.

Valentini, A, Pompanon, F, Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology and Evolution* 24, 110-117.

Viengkone, M., Derocher, A.E., Richardson, E.S., Malenfant, R.M., Miller, J.M., Obbard, M.E., Dyck, M.G., Lunn, N.J., Sahanatien, V., and Davis, C.S. (2016). Assessing polar bear (*Ursus maritimus*) population structure in the Hudson Bay region using SNPs. *Ecology and Evolution* 6, 8474–8484.

Voelker, M. R., Schwarz, D., Thomas, A., Nelson, B. W. & Acevedo-Gutiérrez, A. 2020. Large-scale molecular barcoding of prey DNA reveals predictors of intrapopulation feeding diversity in a marine predator. *Ecology and Evolution* 10, 9867-9885.

von Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T.E., Alves, P.C., Lyons, L.A., Mattucci, F., Randi, E., Cragnolini, M., Galián, J., Hegyeli, Z., Kitchener, A.C., Lambinet, C., Lucas, J.M., Mölich, T., Ramos, L., Schockert, V., Cocchiararo B. (2020). Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources* 20, 662-680.

Wehausen, J.D., Ramey, R.R., and Epps, C.W. (2004). Experiments in DNA extraction and PCR amplification from bighorn sheep feces: The importance of DNA extraction method. *Journal of Heredity* 95, 503–509.

Waits, L.P. (2004). Using noninvasive genetic sampling to detect and estimate abundance of rare wildlife species. Pages 211-228 in W.L. Thomas, editor. *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island Press, Washington D.C., USA.

Waits, L., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, 69(4), 1419-1433.

Williams, T.M., Estes, J.A., Doak, D.F., And, A.M., and Springer, A.M. (2004). Killer appetites: Assessing the role of predators in ecological communities. *Ecology*. 85, 3373-3384.

Wilson, K., Lance, M., Jeffries, S., and Acevedo-Gutiérrez, A. (2014). Fine-scale variability in Harbor seal foraging behavior. *PLOS ONE* 9, e92838.

Zango, L., Reyes-González, J.M., Militão, T., Zajková, Z., Álvarez-Alonso, E., Ramos, R., and González-Solís, J. (2019). Year-round individual specialization in the feeding ecology of a long-lived seabird. *Scientific Reports* 9, 11812.

Supplemental Information

Exogenous references and NCBI GenBank accessions**:** Atlantic herring (*Clupea harengus*, GCA_900700415.1), Chinook salmon (*Oncorhynchus tshawytscha*, GCA_002872995.1), Coho salmon (*Oncorhynchus kisutch*, GCA_002021735.2) Sockeye salmon (*Oncorhynchus nerka*, GCA_006149115.1), European hake (*Merluccius merluccius* GCA_900312545.1), Alaskan pollock (*Gadus chalcogrammus*, GCA_900302575.1), Tiger rockfish (*Sebastes nigrocinctus*, GCA_000475235.3), Atlantic halibut (*Hippoglossus hippoglossus*, GCA_009819705.1), Rhine sculpin (*Cottus rhenanus*, GCA_001455555.1), Lingcod (*Ophiodon elongatus*, GCA_004358465.1), European pilchard (*Sardina pilchardus*, GCA_900499035.1), Pacific bluefin tuna (*Thunnus orientalis*, GCA_009176245.1), Toadfish (*Thalassophryne amazonica*, GCA_902500255.1), Wolf eel (*Anarrhichthys ocellatus*, GCA_004355925.1), Threespine stickleback (*Gasterosteus aculeatus*, GCA_000180675.1), Capelin (*Mallotus villosus*, GCA_903064625.1), California two-spot octopus (*Octopus bimaculoides*, GCA_001194135), and bacteria (*Escherichia coli*, GCA_000005845.2, *Helicobacter pylori*, GCA_900478295.1, *Mycobacterium smegmatis*, GCA_001457595.1, *Bacteriodetes plebeius*, GCA_000187895.1, Firmicutes bacterium, GCA_902375325.1).

Tables

Table 1. Samples included in this study, along with collection locations.

Sample	Sample Type	Collection Location	Notes
PV-02	Slurry	-	Aquarium control (Hogan, Seattle Aquarium)
PV-04	Surry	Comox, British Columbia	-
PV-05	Slurry	Comox, British Columbia	-
PV-06	Slurry	Comox, British Columbia	-
PV-07	Slurry	Whidbey Island, Washington State	-
PV-09	Slurry	Whidbey Island, Washington State	-
PV-14	Slurry	Comox, British Columbia	-
PV-15	Swab	Cowichan Bay, British Columbia	Proposed resampled individual #1
PV-16	Swab	Cowichan Bay, British Columbia	Proposed resampled individual #1
PV-17	Swab	Cowichan Bay, British Columbia	Proposed resampled individual #2
PV-18	Swab	Cowichan Bay, British Columbia	Proposed resampled individual #2
PV-19	Slurry	-	Aquarium control (Hogan, Seattle Aquarium)

Table 2. Initial and final DNA concentrations of all *Phoca vitulina* scat samples after the methylation-based enrichment for eukaryotic DNA.

Sample	Initial DNA Concentration (ng/uL)	Final DNA Concentration (ng/uL)
PV-02	122.94	0.11
PV-04	125.43	0.215
PV-05	97.5	0.146
PV-06	91.7	0.178
PV-07	93.78	0.065
PV-09	72.27	0.178
PV-14	45.47	1.56
PV-15	15.66	2.22
PV-16	7.81	0.07
PV-17	4.86	1.78
PV-18	4.88	0.32
PV-19	49.61	0.12
Avg.	**60.99**	**0.58**

Table 3. Read counts per scat sample, including percentages of reads that mapped to the *P. vitulina* reference, any exogenous reference, or no reference (unmapped).

Sample	Raw reads	Clean reads	% *P. vitulina*	% exogenous	% unmapped
PV-02	12217235	1144674	8.81	8.62	82.56
PV-04	13283560	12367273	81.78	6.51	11.72
PV-05	13143939	12315643	59.79	6.10	34.10
PV-06	1261032	1175904	28.25	19.30	52.45
PV-07	1267335	1163335	62.04	11.99	25.98
PV-09	2343371	2148792	10.32	15.43	74.25
PV-14	14600829	9845392	25.06	22.58	52.36
PV-15	2348172	1323564	23.73	62.81	13.46
PV-16	2308274	2042070	21.45	24.84	53.71
PV-17	9563255	6199648	34.50	52.33	13.17
PV-18	7675806	6519813	25.23	57.51	17.26
PV-19	15672930	14751318	42.04	6.70	51.25
Avg.	**7973811**	**6774957**	**35.25**	**24.56**	**40.19**

Table 4. Mean number of reads in each subsample level, including the number of loci identified by Freebayes, and the final number of loci after filtering for loci present in all samples with quality scores > 20.

Subsample	Mean # reads	# loci	# filtered loci
25%	1693739	19962	1723
50%	3387478	29779	3220
75%	5081218	38274	4080
100%	6774957	45801	4709

Table 5. Percentages of all exogenous reads per *P. vitulina* scat sample (not total reads) that mapped to the specified taxa.

Sample	Clupeiformes	Salmoniformes	Gadiformes	Scorpaeniformes	Pleuronectiformes	Scombriformes	Batrichoidiformes	Perciformes	Cephalopods	Gasterosteiformes	Osmeriformes	Bacteria
PV-02	40.62	17.20	10.21	7.05	2.79	2.51	4.37	2.56	4.49	2.03	2.24	3.94
PV-04	33.74	11.28	44.28	2.52	1.09	0.88	1.87	0.99	1.13	0.75	0.95	0.53
PV-05	37.78	15.97	11.96	7.59	2.62	2.96	3.24	2.52	2.44	1.82	1.64	9.45
PV-06	27.76	42.60	9.03	4.99	2.14	1.88	3.29	1.78	2.55	1.39	1.75	0.84
PV-07	40.93	16.89	13.00	7.92	7.50	1.53	2.96	1.59	2.16	1.37	1.35	2.81
PV-09	28.35	18.06	13.03	13.91	2.57	2.21	4.64	3.24	3.16	2.58	2.79	5.46
PV-14	14.06	6.73	6.65	30.93	8.94	13.47	2.03	9.62	1.11	4.63	1.18	0.63
PV-15	93.79	1.64	2.42	0.52	0.24	0.17	0.30	0.17	0.27	0.15	0.27	0.06
PV-16	71.37	7.40	6.14	3.01	1.18	1.06	1.67	1.21	1.63	0.84	1.11	3.38
PV-17	89.53	1.12	7.13	0.54	0.17	0.11	0.23	0.12	0.22	0.10	0.22	0.52
PV-18	93.75	1.24	2.59	0.46	0.21	0.12	0.25	0.14	0.43	0.12	0.28	0.41
PV-19	41.66	17.14	11.37	5.89	2.37	1.98	3.80	2.13	4.78	1.72	2.48	4.68
Avg.	**51.11**	**13.11**	**11.48**	**7.11**	**2.65**	**2.41**	**2.39**	**2.17**	**2.03**	**1.46**	**1.35**	**2.73**

Table 6. Diet composition (in % mass) of the captive *P. vitulina* scat sample, with the percentage of total exogenous reads that mapped to the specified taxa in samples PV.02 and PV.09 (PV.02 and PV.09 sum <100% due to mapping to taxa not present in actual captive diet).

Order	Diet Composition (% mass)	PV.02 (% exogenous reads)	PV.19 (% exogenous reads)
Clupeiformes	45.0	40.62	41.66
Osmeriformes	25.0	2.24	2.48
Cephalopods	20.0	4.49	4.78
Scombriformes	10.0	2.51	1.98

Table 7. Diet composition results of scat samples PV.05, PV.06, PV.07, PV.09, and PV.14 in terms of proportion exogenous reads from this study and proportion of amplicon reads from Thomas et al. (2017). Dashes indicate no detection.

		Gadiformes	Perciformes	Scorpaeniformes	Clupeiformes	Salmoniformes	Pleuronectiformes	Cephalopods	Batrichoidiformes
Direct sequencing	**PV.05**	0.12	0.03	0.08	0.38	0.16	0.03	0.02	0.03
	PV.06	0.09	0.02	0.05	0.28	0.43	0.02	0.03	0.03
	PV.07	0.13	0.02	0.08	0.41	0.17	0.08	0.02	0.03
	PV.09	0.13	0.03	0.14	0.28	0.18	0.03	0.03	0.05
	PV.14	0.07	0.10	0.31	0.14	0.07	0.09	0.01	0.02
mtDNA meta-barcoding	**PV.05**	-	0.69	-	0.30	0.02	-	-	-
	PV.06	-	-	-	0.01	0.93	0.06	-	-
	PV.07	-	-	-	-	-	-	1.00	-
	PV.09	0.01	-	0.46	-	-	-	-	0.52
	PV.14	-	0.98	-	0.02	-	-	-	-

Figures

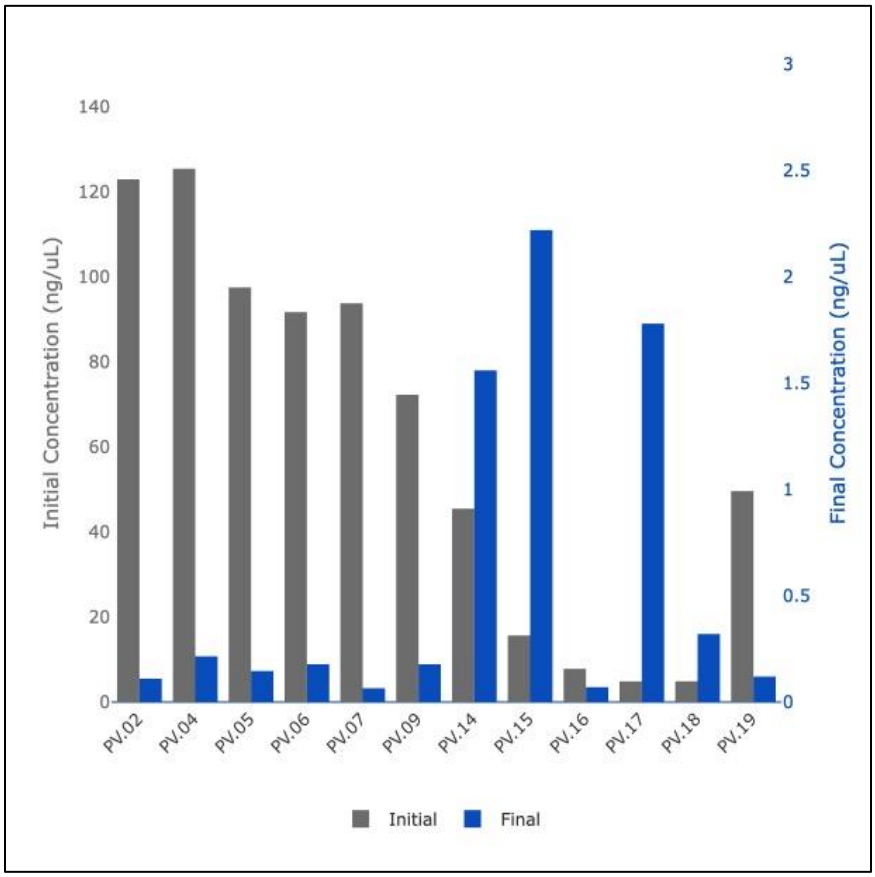

Figure 1. Initial DNA concentrations of all *Phoca vitulina* scat samples (grey) versus corresponding final DNA concentrations (blue), after the methylation-based enrichment for eukaryotic DNA.

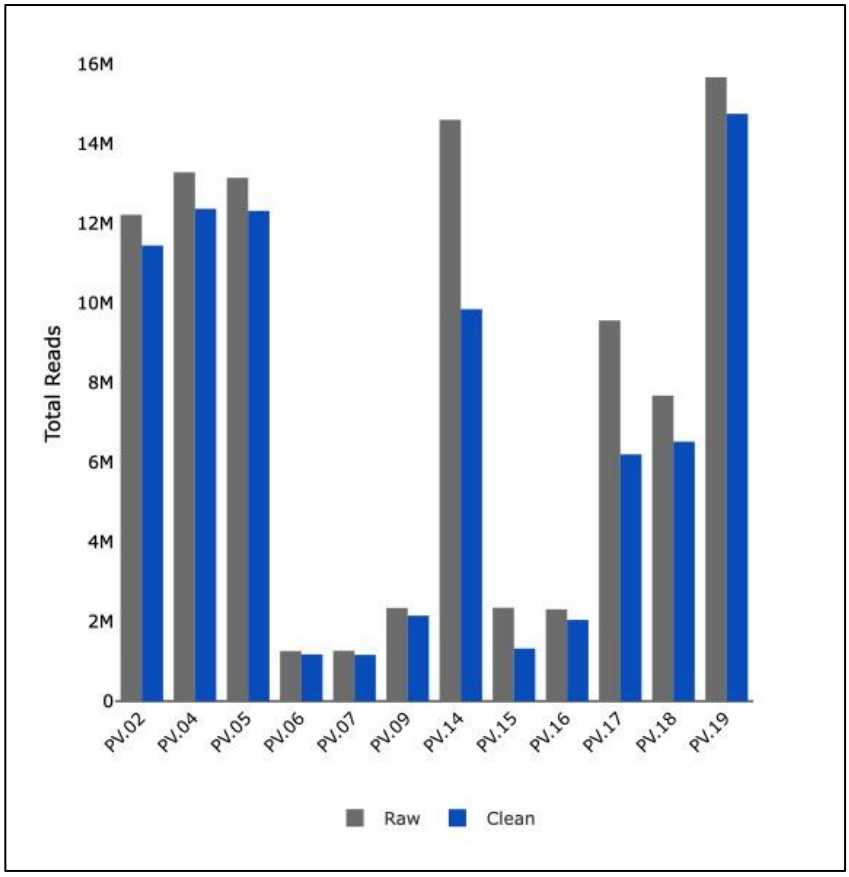

Figure 2. Total number of raw reads obtained from each *P. vitulina* scat sample (grey), versus the final number of reads per sample (blue) following adapter trimming, quality filtering, and size selection for reads >50 bp.

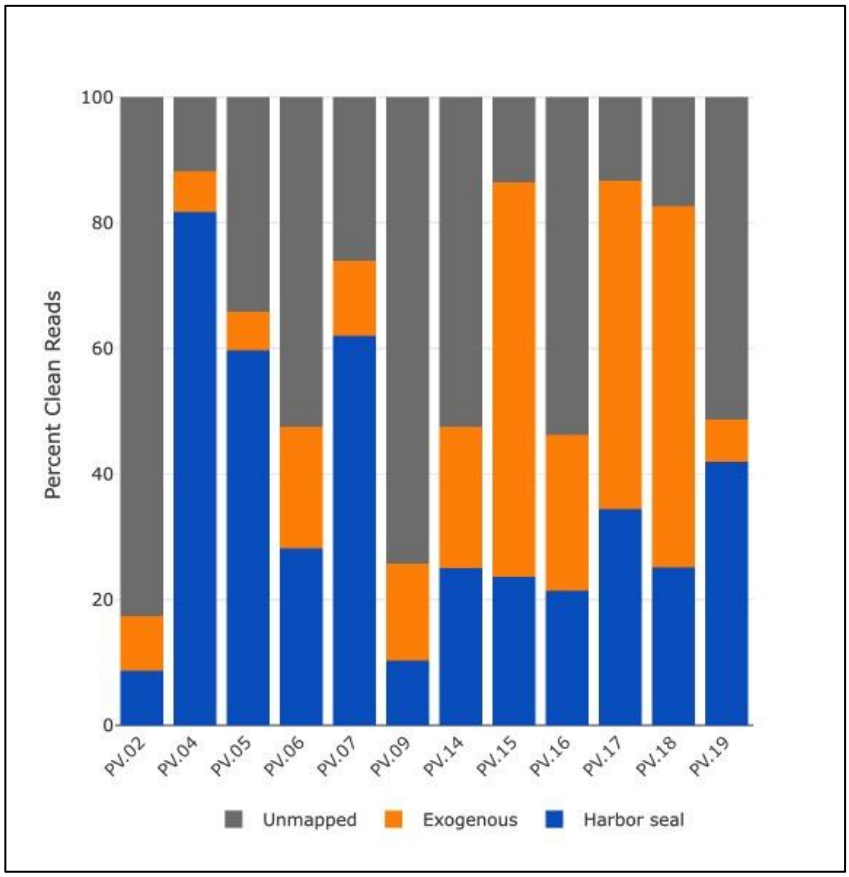

Figure 3. Percentage of clean reads per scat sample that aligned to either no reference (grey), an exogenous reference (orange), or the *P. vitulina* reference (blue).

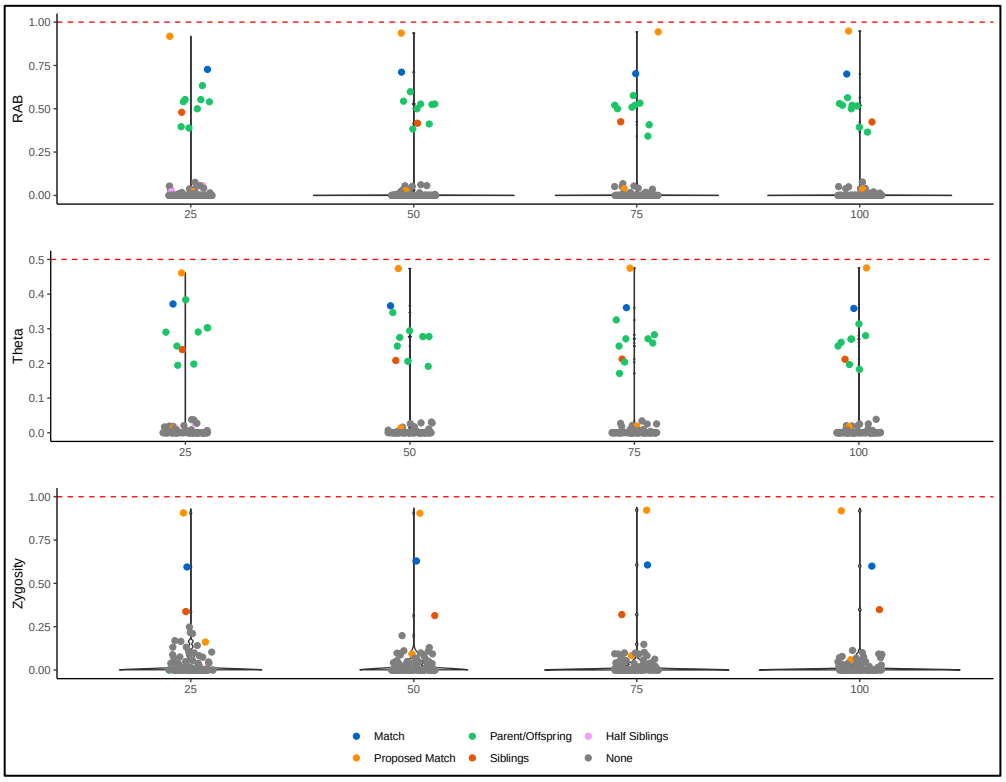

Figure 4. Relatedness values for all pairwise comparisons of all *P. vitulina* scat samples at each subsample level. Color indicates known, proposed, or simulated relationship between sample pairs representing each data point. All comparisons including offspring, siblings, and half siblings are performed between simulated individuals (using all loci identified at each respective subsample level). Dashed red lines indicated thresholds for respective values expected between identical individuals.

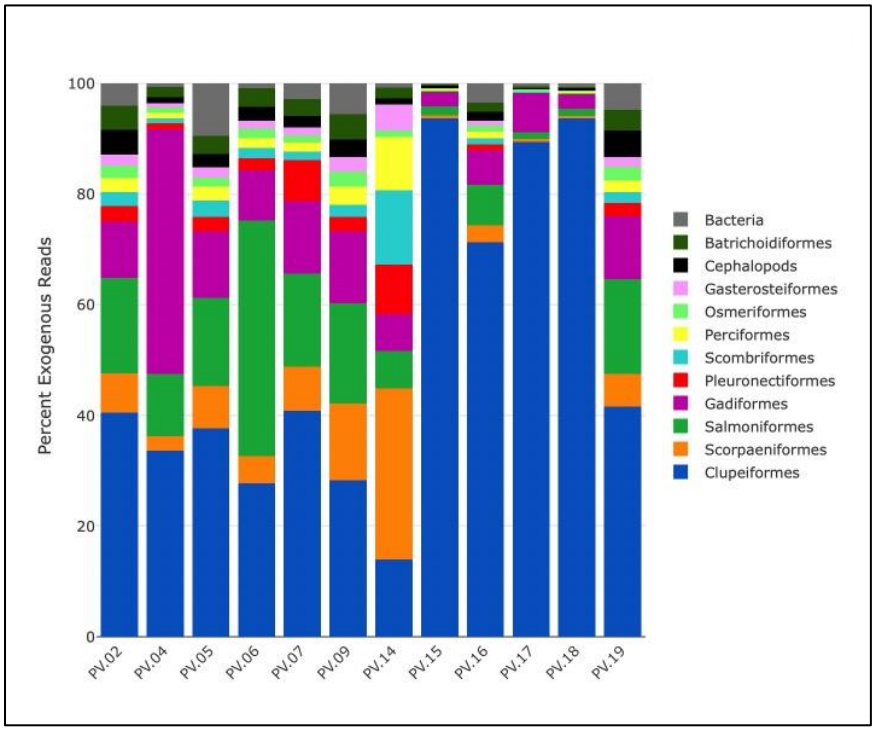

Figure 5. Percentages of all exogenous reads per *P. vitulina* scat sample that mapped to the selected exogenous taxa.

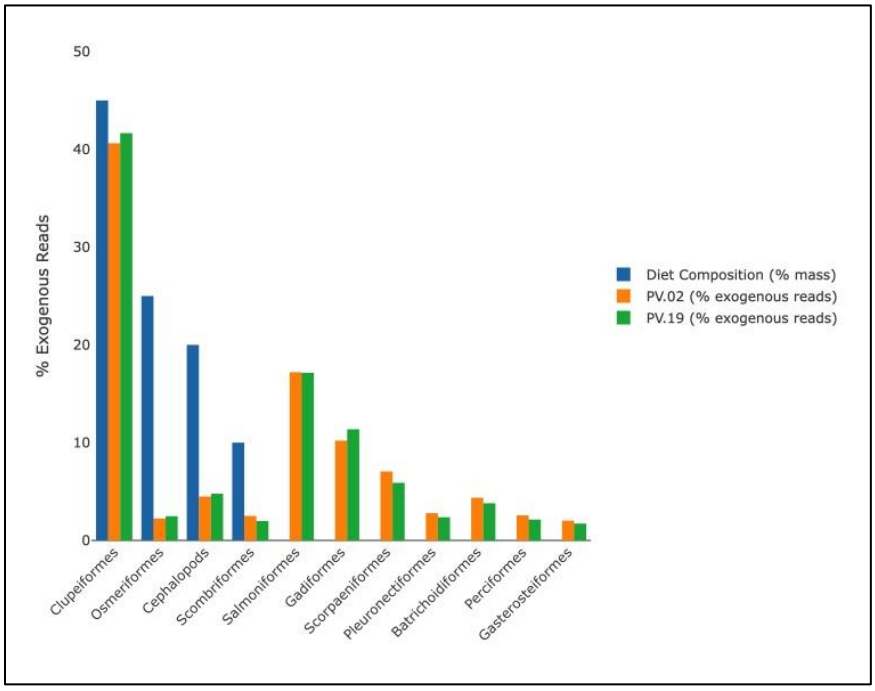

Figure 6. Diet composition of the captive seal (blue) compared to the percentage of exogenous reads per sample that mapped to the selected prey references (orange and green). Note the alignment to prey taxa not present in the actual diet. PV.02 and PV.19 read abundances do not include bacterial counts.