What's in the diet? DNA-based analysis for qualitative and quantitative assessment of animal diet

Groen, K.

6

General Discussion

Interpreting diet data obtained by genetic approaches might seem straightforward as genetic codes are assigned to species which subsequently results in a taxonomic list reflecting the consumed prey/food items. However, inferring a consumer's true diet does not only rely on assigning the right taxa (or the right presence/absence measurements) to the (present) genetic sequences in your samples, unfortunately. A consumer's true diet represents the actual consumption of all potential diet items available in the ecosystem, while the estimated diet represents the diet identified by DNA-based diet analysis (Figure 6.1a). A mismatch between true and estimated diet can't be prevented (as you cannot sample indefinitely), and thus the aim of diet estimation is to maximize the overlap of the true and estimated diet. However, errors arise from, either, true diet not detected by the selected method (methodological bias) or potential diet items in the environment not ingested by the consumer but detected by the method (biological and environmental biases) (Figure 6.1b).

In this thesis, I dove into the realm of DNA-based diet analyses by investigating these biasing factors (in Chapters 2 to 5), that can influence the fate of DNA in dietary samples quantitatively and qualitatively. Through the use of droplet digital PCR, we optimized DNA extraction, mitigated PCR inhibition, and examined the influence of DNA markers on detectability, despite the inherent challenge of DNA degradation due to digestive processes (Chapter 2). Our newly developed approach enabled robust and minimally invasive quantification of small diet constituents in feces. In a realistic field scenario, we tested this method by analyzing the effect of biological factors, meal size, and digestibility on DNA quantification in wood mice feces, establishing calibration curves that allowed for accurate dietary intake estimates (Chapter 3). Additionally, we utilized DNA metabarcoding to investigate the contribution of small prey species to the diet of lions, revealing biases in non-genetic methods and highlighting the prevalence of field contamination (Chapter 4). Finally, applying these insights to an ecological context, we compared lion diet composition and prey preferences across different National Parks, finding distinct patterns related to the presence of fencing (Chapter 5). Our study underscores the potential of DNA-based analysis for ecological research, while also identifying its limitations and the need for complementary methods. Investigating and elucidating these biases further and the implications thereof hopefully aids the research field in moving forward toward estimating a customer's true diet as closely as possible which I discuss in this final chapter. I will start by discussing the wealth of opportunities, but also the limitations,

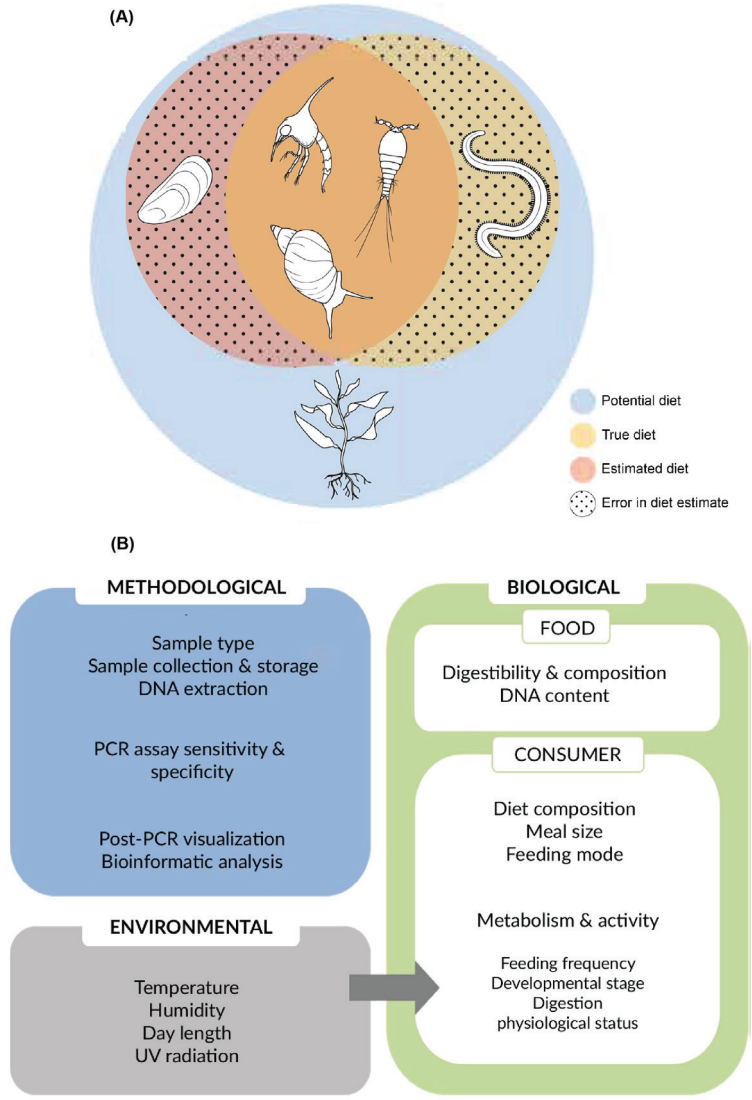

Figure 6.1 | (A) A consumer's true diet (yellow shaded area) represents some fraction of all potential diet items (blue shaded area) available in the ecosystem, while the estimated diet (red shaded area) represents diet identified by DNA-based diet analysis. Mismatches between yellow and red shaded areas represents error in the diet estimation (black dotted area), thus the aim of diet estimation is to maximize the overlap of the true and estimated diet (orange shaded color) qualitatively as well as quantitatively. Errors can arise from the true diet not being detected by the selected method (right black dotted area) and potential diet items in the environment not ingested by the consumer but detected by the method (left black dotted area). Figure adapted from Nielsen et al., 2017. **(B)** A range of factors affect the fate of food DNA in dietary samples and influence the interpretation of molecularly derived trophic data. These factors can be grouped into methodological, biological and environmental aspects. Figure adapted from Traugott et al., 2020.

presented by this innovative approach. Next, I will continue by linking qualitative and quantitative diet assessments, and discussing where the research field might be heading in the near future. Next, I will debate the complementarity of genetic methods and non-genetic approaches, as current literature suggests that the animal diet can be best inferred by combining approaches. Lastly, the implications of novel approaches in diet analyses for ecological research are explored.

6.1 | Combining qualitative and quantitative DNA-based analysis

Combining high throughput sequencing (HTS) techniques, such as DNA metabarcoding and diagnostic PCR, such as quantitative PCR (qPCR) and droplet digital PCR (ddPCR), have already been implemented in studies to combine information on both prey and consumer DNA For example, it has been utilized to assess sex-specific dietary choices; Schwarz et al. (2018) examined feces of male and female harbor seals in Canada, identifying prey DNA via metabarcoding and using qPCR to determine the sex of the seals. The novel combination of molecular techniques demonstrated that diet differences between males and females were consistent across sites and years suggesting that seals differ in their foraging strategies depending on their sex. Furthermore, individual identification of carnivores from fecal DNA (genotyping) enables individuals to be linked to their diets, and provides the means to assess intra-population and temporal variation in foraging behavior and prey consumption patterns (Fedriani & Kohn 2001).

Although the above combination of techniques opens up new ecological study possibilities (see section 6.3 below), it still remains largely qualitative. Currently, research combining qualitative research with quantitative techniques linking species identification to species abundance remains challenging and scant mostly due to DNA concentrations not always reflecting abundances as we also demonstrated in Chapter 3 (Alberdi et al., 2018). Murray et al. (2011) did compare HTS approaches with qPCR approaches and found that qPCR shows higher sensitivity in terms of amplifying low quantity DNA compared to HTS (although dependent on marker length). This is an important advantage of species-specific qPCR over HTS, in that it can detect species at very low DNA abundances, whereas the nature of universal primers, such as those used in HTS, renders them less specific and less likely to efficiently amplify low copy number targets in the presence of abundant targets (Thomas et al., 2015). However, the higher qPCR detection success did not drastically affect the overall diet estimates of both methods, due to the low abundance of prey species in the study of Murray et al. (2011). Although even more sensitive PCR techniques, with a higher detection rate, such as ddPCR (compared to qPCR) might have enhanced these results (Verhaegen et al., 2016).

In theory, HTS could be first used to study diet breadth and the importance of species of interest in a diet could be then identified and subsequently quantified using diagnostic PCR (as knowledge ***a priori*** of diet composition is needed for diagnostic PCR) (as did Wang et al., 2023). However, quantified DNA copy numbers still need to be coupled to actual biomass estimates by feeding studies for each prey item as we did in Chapter 3. More importantly, these feeding studies need to be performed separately for each diet item considered. Moreover, these feeding studies get more complicated when incomplete diet items are consumed such as a pride of lions feeding together on one large herbivore (e.g. zebra) carcass. In these cases DNA copy numbers cannot be easily translated into biomass consumed, as quantification is subject to tissue-specific differences in DNA content (e.g. spleen or liver do not contain the same number of DNA copies) which might result in highly varying DNA copy numbers and faulty biomass estimates if not properly calibrated (discussed in Chapter 2).

The utility of using only DNA metabarcoding for quantifying relative species abundances is currently limited by both biological and technical biases which influence sequence read counts as discussed in section 1.5 of the General Introduction. However, Thomas et al. (2015) found promising results by correcting for multiple sources of potential biases by simultaneously sequencing DNA extracts of field-collected diet samples as well as 50/50 mixtures of target species (i.e. mock samples) and a control species in order to generate relative correction factors (RCFs) which greatly improved relative abundance estimates. While 50/50 RCFs may provide a solution to multiple sources of bias in a single correction, there are other sources of bias that are not accounted for using this approach that require consideration. Most notably are biases introduced by differential degradation of species DNA due to either digestion (in the case of diet studies), or other degenerative (environmentally related) processes responsible for degrading environmental DNA (Alberdi et al., 2018). A metabarcoding diet study

with penguins suggested that differential DNA degradation due to digestion was the most significant cause of bias in the study system (Deagle et al., 2010). In those cases, additional bias correction efforts (e.g. lipid correction; Thomas et al., 2014) may be needed in order to achieve a highly accurate representation of mass proportions from DNA sequence counts of environmental samples (Thomas et al., 2015). Furthermore, it should be emphasized that the species-specific correction factors calculated using this approach are also specific to the experimental conditions of the methodological protocol. Therefore, the control materials used to produce RCFs should be re-sequenced each time a different experimental setup is used or an alteration is made to the methodological protocol, such as a change in PCR conditions or the transfer of protocols between laboratories (Thomas et al., 2015).

In summary, it remains challenging to enable accurate reconstructions of the physical diet as estimates are currently confounded by such a range of factors including differential digestion rates of prey between species, DNA per unit biomass variability between tissues and the developmental stage of the prey species to name but a few issues (Thomas et al., 2015). Furthermore, combining HTS and diagnostic PCR approaches and conducting feeding studies or calculating RCFs for each diet item is very time-consuming and therefore hardly feasible (although not impossible). Taking this into consideration together with the many caveats listed above suggests that other methods and techniques need to be considered to obtain the accurate quantification of prey DNA actually contained in fecal matter (but see Shelton et al., 2022 for a statistic modelling approach on quantification of metabarcoding data). Luckily, genetic techniques advance rapidly, and new approaches to estimate species abundance from diet samples are already emerging, which are primarily targeting and analyzing intraspecific genetic variation to infer the abundance of a species (see Wagemaker et al., 2020 and Andres et al., 2023).

Inferring species abundance from intraspecific genetic variation also relies on the use of HTS techniques. However, the difference to conventional metabarcoding is the focus on using different genetic markers (mostly targeting nuclear genomes or (full) organelle genomes) that contain enough intraspecific genetic variation within individuals to decipher genotypes (instead of using DNA concentrations to estimate abundances) from which true species numbers and biomass can potentially be inferred (Andres et al., 2021).

For example, Sethi et al. (2018) applied a likelihood-based framework to data of 14 microsatellite loci (nuclear markers) from stomach content samples of a predatory fish (largemouth bass) to estimate the number of contributors to these DNA mixtures. This analysis indicated a 5-fold increase in estimated predation rates of yellow perch by largemouth bass relative to conventional visual assessment of diet contents which can miss partially digested prey items (Sethi et al., 2018).

With the current knowledge, quantification approaches in Chapters 2 and 3 could instead have been better focused on detecting and analyzing genetic variation within seeds to quantify their abundance in wood mouse fecal samples. Of course, this is dependent on sufficient DNA quality (which is generally poor in fecal samples due to digestion but needs further study) and sufficient genetic diversity within individual seeds to discriminate between haplotypes of individual seeds. Furthermore, optimizing extraction protocols would not have been as easy to conduct as compared to diagnostic PCRs as HTS approaches generally take more time, resources and bioinformatics before results can be inferred. Nonetheless, estimating species abundance from genetic diversity sounds promising and future research will likely expand on this field of research.

While this field is now limited to obtaining single species counts from environmental samples, it would benefit to move towards multispecies abundance estimates, only then metabarcoding of fecal samples can truly quantify full dietary breadth. Hence, see the method of Wagemaker et al. (2020), genotyping by sequencing, which shows promising results for quantifying relative species abundance of plant roots in multispecies plant communities. However, it is not clear whether this method would work in highly degraded DNA mixtures, like in feces or soil or water eDNA (Piñol, 2021). Identifying haplotypes for entire communities from environmental samples may present additional challenges over single-species approaches. Deiner et al., 2017 show that it is possible to amplify whole mitochondrial genomes (> 16 kb) from environmental samples using HTS sequencing. This opens up the possibility of linking species assignments to genetic diversity indices. However, to make use of this method, there is a need for continued research focused on parsing out sequencing noise from real variation to determine intra-species haplotype diversity collected from environmental samples and high-throughput sequencing (Deiner et al., 2017; Andres et al., 2023). Furthermore, mitochondrial DNA (mtDNA) shows different mutation rates

than nuclear DNA (nuDNA) among phyla of animals and is exclusively inherited through either the paternal or maternal line (Allio et al., 2017). Consequently, it remains challenging to genotype individuals only based on mtDNA. Despite these difficulties, organelle metagenomics (e.g., mitochondria for animals or chloroplasts for plants) is gaining popularity, as it is easier to assemble organelle genomes then whole genomes (Piñol, 2021).

The continued advancement of single molecule and long-read technologies, such as the Oxford Nanopore MinION (Laszlo et al., 2014), might improve this approach and it is expected that long-read sequencing technologies, once they are cost effective and error rates are reduced, will become the method of choice for sequencing long-range PCR products, such as mitogenomes, and will allow population genetic analysis of eDNA samples (Denier et al., 2017). Moreover, Nanopore opens up the possibility of directly sequencing in the field, which prevents further degradation of field-collected samples (Wang et al., 2021). In the coming decade it may be possible to monitor genetic diversity for hundreds of species across the tree of life using eDNA sequence data from diet samples and estimate a consumer's true diet qualitatively and quantitatively (directly in the field) (Deiner et al., 2017; Shum & Palumbi, 2021; Weitemier et al., 2021).

6.2 | Complementarity with non-genetic methods

In recent decades, estimates of dietary habits have predominantly relied on visually inspecting undigested remnants present in feces or stomach contents (Steenweg et al., 2015). This method's popularity stems from its rapid application and the ability to gather both (semi)quantitative and qualitative data about the diet, including insights into food source characteristics such as prey age, size, or developmental stage (Klare et al., 2011). Nevertheless, the current array of available dietary tracing techniques, encompassing morphological examination of remains in feces and stomach contents, DNA barcoding, and the use of biomarkers like stable isotopes and fatty acid ratios (see section 1.2.1 of the General Introduction), complicates the selection of the most suitable method for specific research or management inquiries. The concept of employing integrated approaches has recently gained traction, particularly to address the inherent limitations of these dietary tracing methods when used in isolation (Horswill et al., 2018; Matley et al., 2018; Nielsen et al., 2018; Bonin et al., 2020).

Our study of Chapter 4 would likely have benefited from adding an extra dietary method, as extra information on prey age, size, or developmental stage obtained by morphological approaches would increase qualitative dietary information substantially especially elevating issues with prey size related detection rates between molecular and morphological techniques (Hayward & Kerley, 2005; Pereira et al., 2019). While molecular barcoding facilitates time efficient and mostly unbiased species identification across various life stages, a task often challenging based on morphological traits (Packer et al., 2009; Pereira et al., 2019), it may limit insights into the foraging strategies of predators. Take insectivores, for instance, where a predator's preference for eggs, larvae, or adult stages can offer valuable information about foraging strategies as early life stages of insects tend to be less mobile than adults. Therefore, a combination of molecular analysis and a morphological examination of (the hard part of) prey remains is particularly recommended in such cases (see for examples studies like Thalinger et al., 2018 and Pereira et al., 2019).

Moreover, old fecal samples are generally known to contain lower quality DNA and hamper accurate species identification using genetic approaches (Tollit et al., 2009). However, the age of a sample affects morphological identification to a much lower extent (Bonin et al., 2020). For instance, Bonin et al. (2020) demonstrated a consistent and unexpected finding that some food sources were detected in a lower proportion of feces by molecular analyses compared to a morphological approach in the diet of bears. They suggested that even though DNA within samples might be poorly distributed (see Alberdi et al., 2018 and Mumma et al., 2016), DNA quality could also have been an issue here and could have contributed to failures of DNA amplification or the increased occurrence of false negatives. Tollit et al. (2009) observed a reduced detection rate in molecular analyses compared to visual identification of pinniped remains in feces, particularly in cases involving aged feces. They attributed these false negatives to DNA degradation over time and/or exposure to abiotic conditions. Similarly, McInnes et al. (2016) encountered a comparable limitation in the use of molecular tools for studying seabird food habits in free-ranging conditions, highlighting a logistical challenge. Given the relatively high percentage of feces samples in which molecular analyses could not be completed, and considering that some feces were collected up to 14 days after excretion, it underscores the importance of cautiously considering the impact of environmental factors on DNA degradation rates and sampling time for molecular dietary analyses

(Alberdi et al., 2019). Moreover, this highlights the necessity of obtaining feces close to excretion (as fresh as possible) and applying rapid DNA extraction after collection (to stop degradation as soon as possible) for molecular analysis, and implementing correction metrics and models to address potential false negatives or imperfect detection (Alberdi et al., 2019; Monterroso et al., 2019; Morin et al., 2019). Recognizing that time and access constraints are common challenges, particularly in remote areas with free-ranging populations, the combined use of morphological and molecular tools during specific sampling periods or seasons can help overcome logistical limitations.

Furthermore, false positives are prone to occur in molecular analysis of fecal samples collected in the field due to contamination (Traugott et al., 2020). These false positives might be recognized when morphological analysis would have been used to check true consumption events. For instance, DNA of mesopredators was frequently obtained in fecal samples of lions (Chapter 4). However, if no such hairs or bones (of mesopredators) would have been found in morphological analysis, these occurrences could have assuredly been attributed to overmarking of lion scats by mesopredators to mark their own territories. Likewise, cannibalistic interactions remain challenging to detect, as existing molecular diet methods enable the differentiation of prey DNA among species but not yet within the same species (but see section 6.1). In most cases, dietary studies continue to depend on the morphological identification and counting of prey remnants alongside molecular trophic analyses to evaluate intraspecific feeding interactions (Eigaard et al., 2014).

In a study by Bonin et al. (2020), stable isotopes, morphological, and molecular analyses were compared to investigate the diet of free-ranging consumers with distinct diet types, namely carnivores and omnivores. Their results showed that the different methods led to a consistent diet description for carnivores, which have a relatively simple diet mixture, but their outcomes somewhat differed for omnivores, which have a more complex diet. For instance, fish were missed in the diet of bears using the morphological approach but did get picked up by the molecular analysis. Additionally, Bonin et al. (2020) and Tollit et al. (2009) both reported a lower detection rate for molecular analyses compared to visual identification of diet remains in feces in cases involving old feces. Furthermore, they demonstrated that the taxonomic discriminative power of stable isotope analysis could not go as far as for molecular tools, and for morphological analysis to a lower extent. However, Bonin et al. (2020) still suggested that stable isotope analysis is a powerful tool for dietary studies, especially those addressing the contribution of distinct food sources or food categories over long time scales and for longitudinal monitoring but they also caution because reliable estimates with stable isotope analysis require a minimum of a priori information about the consumer's foraging patterns. In general, the combined use of morphological and molecular analyses enhanced the diversity of food sources detected compared to the use of a single method independent of diet (Bonin et al., 2020). Bonin et al. (2020) emphasized that the choice of an appropriate approach to diet description should be a key step when planning dietary studies of free-ranging populations and recommend using more than one dietary determination method, especially for species with complex diet mixtures.

Another possible complementarity between genetic diet analysis and morphological approaches might lie within the field of calibration studies, like our study in Chapter 3. These kind of studies have also been conducted for several predator species using morphological methods (Klare et al., 2011). And it might be possible to use these morphological-based calibration curves, where they relate kilograms fed to a predator versus the biomass collected in fecal samples and use these curves to be applied to genetic diet studies. It would be interesting to see if these calibration curves which have already been acquired for multiple predator species (e.g. wolf, coyote, side-striped jackal, red fox, badger, American mink, European polecat, grizzly bears, black bears, cougars, cheetahs, bobcats and lynxes to only name a few, see Klare et al., 2011 for references) compare to obtained calibration curves with genetic methods. If these relate, diet quantification (as a new feeding study based on genetic methods might not be needed for each prey item) might benefit from the usage of genetic approaches as these have been demonstrated to be more time and cost-efficient than traditional methods (Ando et al., 2020).

6.3 | Implications for ecological research

Ecologists are increasingly interested in performing community-wide assessments in space and time (Bartley et al., 2019; Daam et al., 2019; Eisenhauer et al., 2019; Meyer et al., 2020). However, many contemporary ecological studies face challenges in considering entire communities or trophic networks due to practical

constraints (Creer et al., 2016). Immediate obstacles include the difficulty in identifying taxonomically complex communities, the extensive diversity that needs sampling, and the lack of resources, such as funding for taxonomists, required for species identifications (Creer et al., 2010). In essence, the task of conducting comprehensive community-wide assessments is substantial and challenging, with a shortage of skilled personnel and time to carry out these ongoing tasks (Creer et al., 2016). By focusing on a range of genetic source material (e.g. environmental DNA), habitats, time and spatial scales, we can now comprehensively characterize entire communities and begin to unpick their biotic/abiotic interactions from dietary samples. Moreover, genetic samples can be analyzed in bulk and eliminate the need for taxonomic specialists, which saves time and allows this (spared) time to be reinvested in collecting more samples generally potentially enhancing knowledge of the system and interactions (Baird & Hajibabaei, 2012).

Referred to as a transformative technology, harnessing the information held in DNA, from dietary samples, potentially has the power to overcome many limitations of classical biodiversity assessment and thereby answer unresolved ecological questions (Baird & Hajibabaei, 2012). Chapter 5 is a great example of such an ecological study which would have been difficult to conduct using only traditional ways of biodiversity assessment (as this would cost at least a year to gather the same amount of data). The power of DNA made it feasible to study the diet of lions in different National Parks in the same period and also made the comparison between parks possible (e.g. no presence of observer bias as genetic samples are analyzed together).

Currently, individual identification from fecal DNA allows researchers to link individuals to their diets, enabling the assessment of intra-population and temporal variations in foraging behavior and prey consumption patterns (Monterosso et al., 2019). Advancements in library preparation and sequencing technologies offer the potential, in theory, for even extended genomic coverage (Chiou & Bergey, 2018). This opens the exciting new possibility of non-invasively studying the genetic basis of specific traits, such as factors that influence dietary choices by individuals (e.g. physical condition, aggressive behavior, boldness) (Carroll et al., 2018). Additionally, metagenomic approaches applied to dietary samples theoretically allow for the identification of infectious agents providing insights into health in response to environmental and management changes (such as fencing) (Bahrndorff et al., 2016). These approaches hold even more promise for addressing questions in evolutionary biology and contributing to the conservation management of natural populations. It isn't hard to imagine that in the near future when metagenomic approaches have been well-developed, DNA extracts from the study in Chapter 5 could be reanalyzed to study not only a lions' diet and the identification of lion (including sex and age) individuals but also a lions' individual health and link that to the effect of fencing (and all that information just from scat samples).

Another compelling aspect that can be explored through DNA-based analysis of dietary samples is the microbiome of animals. Presently, studies on the microbiome of wild terrestrial carnivores are limited, but recent research in both model and non-model systems has revealed the profound impact of the gut microbiome on various aspects, including metabolism, nutrition, immune response, adaptation, tolerance to environmental changes, behavior, anxiety levels, and overall fitness (see Monterosso et al., 2019). Furthermore, examinations of wild populations have highlighted that factors such as phylogeny, physiology, diet, habitat quality, spatial location, sex, social system and status, reproductive status, and age influence the microbiome composition of both species and individuals (Bahrndorff et al., 2016). These studies, along with a growing body of literature emphasize the role of microbiomes in shaping individual fitness, underscore the significance and utility of fecal samples for characterizing and comprehending the gut microbiome in populations of wild animals (see for example Gomes et al., 2020). It is anticipated that fecal samples will play an increasingly prominent role in future quantitative and qualitative DNA-based diet studies (Bahrndorff et al., 2016; Monterosso et al., 2019).

As mentioned previously feces can be used to investigate infectious agents/pathogens, this feces serve as carriers for various pathogens posing potential threats to other wildlife, domestic animals, and ultimately, humans (Monterosso et al., 2019). High-throughput sequencing has marked a significant advancement in epidemiologic surveillance. While a limited number of studies have employed HTS technologies to screen fecal samples of mammalian carnivores for viruses (e.g. Li et al., 2010; Bodewes et al., 2014; Conceição-Neto et al., 2017), integrating these tools into surveillance programs is crucial, given the profound impacts infectious diseases can have on both animal and human populations (Scott, 1988; Pedersen et al., 2007).

Inevitably, future ecological studies will steer towards understanding complete ecosystems (instead of species or single interactions) and Meyer et al., 2020 have already demonstrated that using environmental DNA from feces to identify the plant and mammal diet of species on three trophic levels (two carnivores, two omnivores and two herbivores) they could build a fully functioning ecological network. These authors show that genetic approaches form a solution to transform ecological research and thereby opens up possibilities for humans to study complete ecosystems and thereby manage these systems in a way that serves biodiversity conservation and ultimately enhances human society (Meyer et al., 2020).

6.4 | Final thoughts

While discussing the prospects of DNA-based analysis for qualitative and quantitative assessment of animal diet, it occurred to me what amazing things genetic approaches are likely capable of in the (near) future. Combined, with the recent advent of research into real-time monitoring of ecosystem biodiversity with CRISPR technology, which allows for complex community identification from eDNA samples in less than 1 hour (Phelps, 2019; Williams et al., 2019; Baerwald et al., 2020; Shashank et al., 2023) and research such as genotyping individuals of three large carnivore species from just snow tracks (De Barba et al., 2023) I am in awe at what DNA techniques can already achieve. I have no doubt that in the future quantitative and qualitative DNA-based assessment of diet will become (or already is) the new traditional/conventional way of studying an animal's diet.