Integrative and Comparative Biology

Integrative and Comparative Biology, volume 00, number 0, pp. 1–12
https://doi.org/10.1093/icb/icac086

Society for Integrative and Comparative Biology

SYMPOSIUM

Integrating DNA-Based Prey Occurrence Probability into Stable Isotope Mixing Models

Brandon D. Hoenig [*,1], Brian K. Trevelline [†‡], Steven C. Latta [§] and Brady A. Porter [¶]

[*]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA; [†]The Cornell Laboratory of Ornithology, Cornell University, Ithaca, NY 14850, USA; [‡]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14850, USA; [§]The National Aviary, Pittsburgh, PA 15282, USA; [¶]Department of Biological Sciences, Duquesne University, Pittsburgh, PA 15282, USA

From the symposium "DNA metabarcoding across disciplines: sequencing our way to greater understanding across scales of biological organization" presented at the annual meeting of the Society for Integrative and Comparative Biology virtual annual meeting, January 3–February 28, 2022.

[1]E-mail: brandonhoenig@gmail.com

Synopsis The introduction of laboratory methods to animal dietary studies has allowed researchers to obtain results with accuracy and precision, not possible with observational techniques. For example, DNA barcoding, or the identification of prey with taxon-specific DNA sequences, allows researchers to classify digested prey tissues to the species-level, while stable isotope analysis paired with Bayesian mixing models can quantify dietary contributions by comparing a consumer's isotopic values to those derived from their prey. However, DNA-based methods are currently only able to classify, but not quantify, the taxa present in a diet sample, while stable isotope analysis can only quantify dietary taxa that are identified *a priori* as prey isotopic values are a result of life history traits, not phylogenetic relatedness. Recently, researchers have begun to couple these techniques in dietary studies to capitalize on the reciprocal benefits and drawbacks offered by each approach, with some even integrating DNA-based results directly into Bayesian mixing models as informative priors. As the informative priors used in these models must represent known dietary compositions (e.g., percentages of prey biomasses), researchers have scaled the DNA-based frequency of occurrence of major prey groups so that their normalized frequency of occurrence sums to 100%. Unfortunately, such an approach is problematic as priors stemming from binomial, DNA-based data do not truly reflect quantitative information about the consumer's diet and may skew the posterior distribution of prey quantities as a result. Therefore, we present a novel approach to incorporate DNA-based dietary information into Bayesian stable isotope mixing models that preserves the binomial nature of DNA-based results. This approach uses community-wide frequency of occurrence or logistic regression-based estimates of prey occurrence to dictate the probability that each prey group is included in each mixing model iteration, and, in turn, the probability that each iteration's results are included in the posterior distribution of prey composition possibilities. Here, we demonstrate the utility of this method by using it to quantify the prey composition of nestling Louisiana waterthrush (*Parkesia motacilla*).

Introduction

The bulk of animal dietary characterizations to date, have been performed by detecting and identifying prey based on the presence of characteristic prey tissues found within diet samples, such as fecal (Ralph et al. 1985) or stomach samples (Sherry 1984), or from observations of foraging (Croxall 1976; Collis et al. 2002), feeding (Fleischer et al. 2003), or provisioning (Snyder and Wiley 1976). However, in recent years, researchers have begun to rely more heavily on laboratory-based techniques, which can indirectly describe an animal's diet by tracing biomolecules or characterizing the chemical compositions that consumers derive from their prey (Hoenig et al. 2021). One such laboratory technique is dietary DNA metabarcoding, which often uses high-throughput sequencing of prey

Advance Access publication June 9, 2022

DNA, followed by the pairing of these prey DNA sequences to known, reference sequences for the taxonomic classification of dietary taxa (Pompanon et al. 2012). While researchers have applied this method to understand many facets of an animal's trophic ecology, including interspecific competition (Trevelline et al. 2018) and dietary response to pollutants (Trevelline et al. 2018), DNA metabarcoding is hindered by its inability to provide accurate estimates of prey count or biomass (Piñol et al. 2015; Jusino et al. 2019) due to the biases inherent to DNA amplification and sequencing (Hoenig et al. 2021). Therefore, when researchers require quantitative information about an animal's trophic ecology, they frequently turn to stable isotope analysis, a laboratory method which quantifies the ratio of elemental isotopes found in consumer tissues to determine many facets of an animal's trophic niche, such as their basal nutrient source (DeNiro and Epstein 1978), their trophic level (Wassenaar 2019), and even the areas where these animals forage (Bradshaw et al. 2017). Following the principle, "you are what you eat, plus a few per mille" (DeNiro 1976), researchers can also use stable isotope mixing models (Phillips 2012), which compare predator and prey isotopic values, to determine the relative contribution of prey-derived elements, and as a result, the relative contribution of prey to the diets of their consumers. Furthermore, recent iterations of these mixing model approaches have begun adopting a Bayesian framework (Stock et al. 2018), which has since allowed the inclusion of prior information to increase the precision of these prey contribution estimates (Franco-Trecu et al. 2013).

Typically, informative priors in Bayesian stable isotope mixing models are derived from dietary characterizations from stomach samples (O'Donovan et al. 2018) or feeding observations (Robinson et al. 2018), which can provide estimated dietary contributions in the form of relative biomass or count of each prey group. These estimated dietary proportions, as well as variation associated with these estimates, are then used as informative priors which limit the range of possible draws in Markov chain Monte Carlo (MCMC) mixing model iterations, and thus improve the precision of prey contribution estimates. With recent advances in DNA-based dietary techniques, which offer improved taxon detection and identification (Braley et al. 2010), researchers have begun incorporating their DNA-based results directly into stable isotope mixing models as informative priors by summing the frequency of occurrence of potential prey groups, and dividing each prey group by this sum to create a pseudo-compositional understanding of diet from presence-absence, DNA-based data (referred to here as the "Scaled Priors" strategy; Bonin et al. 2020). Unfortunately, such an approach is theoretically flawed as this pseudo-composition does not truly reflect prey composition for each group (e.g., two prey groups, each detected in 90% of diets but are consumed in drastically different biomasses, would have the same informative priors), and instead may skew higher the prey contribution of frequently detected, but less abundantly consumed prey, while diluting prey groups that may be truly consumed in higher abundances. Because an accurate understanding of a species' dietary niche is vital for characterizing its ecological role (Eaton 1958), it is imperative that fundamental nature of the data from each diet study technique is preserved, as the resulting dietary characterizations are only as good as the data they are built on (Phillips et al. 2014).

Therefore, in this study we present a novel approach, termed the "Occurrence Probability" strategy, which integrates DNA metabarcoding data into stable isotope mixing models, while still preserving the probabilistic nature of the DNA-based data as well as the compositional nature of stable isotope mixing models. Instead of formally incorporating scaled, DNA-based estimates as informative priors, the "Occurrence Probability" strategy, instead, performs mixing model analyses on all possible prey group iterations (e.g., all four prey groups, each permutation involving three prey groups, and each permutation involving two prey groups), and weights the results from each models' iterations, so that they are proportional to their occurrence probability. Put simply, the probability that a prey group is detected in a diet is the probability that the iterations involving that prey group will be included in the aggregated posterior distribution of prey composition possibilities.

Here, we apply our novel, "Occurrence Probability" approach to characterize the diet of nestling Louisiana waterthrush, a species of Neotropical migratory wood-warbler (Family: Parulidae), and compare this strategies' prey contribution estimates to those from two other common methods of parameterizing Bayesian mixing models: the "Scaled Priors" approach discussed above and the "Diffuse Priors" approach, which sets prior distributions based solely on the number of prey groups included in the model (e.g., each group in a five-group model is centered around 20%, four-group around 25%, and so on). We hypothesize that the "Occurrence Probability" strategy will provide prey contribution estimates, which are most consistent with the Louisiana waterthrush's known breeding ecology than either of the other two methods. However, as the "Occurrence Probability" strategy does not formally include informative priors, we also hypothesize that prey contribution estimates returned by this method will be less precise, as the formal inclusion of priors has been shown to lower the variation of isotope-based

prey contribution estimates (Franco-Trecu et al. 2013). To our knowledge, this is the first study to incorporate DNA-based prey occurrence probabilities into Bayesian stable isotope analysis, and presents a novel approach to incorporate additional variables, either dietary or non-dietary, within a unified methodological framework.

Materials and methods

Sample collection

Fecal and blood samples were collected from 7-day-old nestling Louisiana waterthrush on four headwater streams in the Laurel Highlands of western Pennsylvania (Camp Run, $n = 17$ across four nests; Linn Run, $n = 27$ across six nests; Loyalhanna Creek, $n = 10$ across two nests; Powdermill Run, $n = 18$ across three nests) during the 2017 breeding season (May 14–June 15). Commonly detected aquatic-stage arthropod prey taxa were collected from Powdermill Run using D-net sampling in December 2019, while terrestrial arthropods were collected opportunistically from the surrounding riparian areas with light trapping from 2017 to 2019. Fecal samples, red blood cells, and aquatic arthropods were stored at −20°C, while terrestrial arthropods were pinned and stored at room temperature until downstream DNA-based or isotopic analyses.

DNA metabarcoding and stable isotope analysis

Following Hoenig (2022), total DNA was extracted from nestling fecal samples using a Qiagen Mini Stool Kit and the resulting DNA extracts were subjected to triplicate PCR using primers targeting a hyper-variable region of the arthropod cytochrome c oxidase 1 gene (Trevelline et al. 2016). These amplicons, along with positive (*Acroneuria carolinensis*; Order: Plecoptera) and negative (DNA-free water), were then indexed following Illumina guidelines and pooled to an equimolar concentration and sequenced on an Illumina MiSeq platform with a PhiX spike-in of 20%. Resulting, DNA sequencing reads were subjected to a bioinformatic pipeline described by (Hoenig et al. 2021), which removed low-quality sequencing reads and assigned a taxonomic classification to each exact sequence variant (ESV). All ESVs detected in either the positive or negative control were removed from downstream analyses.

Prior to stable isotope analysis, nestling red blood cells (Supplementary Table S1) and arthropod tissue samples (Supplementary Table S2) were dried, ground, and homogenized before ∼1 mg of each sample was loaded into separate tin capsules. Stable isotope analysis of δ^{13}C and δ^{15}N for each sample was performed on a Thermo Scientific Delta V Advantage Isotope Ratio Mass Spectrometer coupled with a PN150 autosampler and a Carlo Erba NC2500 elemental analyzer housed at Cornell University (Ithaca, NY, USA). Isotope values were expressed in parts per million deviations from each standard (‰), which were Vienna PeeDee Belemnite and atmospheric air for δ^{13}C and δ^{15}N, respectively. Tissue from deer was used as an internal standard and indicated that the standard deviation for measurements of δ^{13}C and δ^{15}N were 0.06 and 0.07‰, respectively. Prior to downstream stable isotope mixing model analysis, nestling isotopic values were adjusted to account for trophic discrimination using intercept-only Bayesian linear mixed effect models with nested random effects (Goodrich et al. 2020, ;package—rstanarm, function—stan_glmer). These models included one of the element's isotopes as a response variable (e.g., δ^{13}C) and the interactions between tissue type (e.g., arthropod or nestling) and the other elemental isotope values (e.g., δ^{15}N). The trophic discrimination factor (TDF) for each elements' isotope values is defined as the difference in intercept between each tissue type, which was 2.31‰ for δ^{13}C and 0.376‰ for δ^{15}N. Although, we agree that the use of experimentally-derived TDFs will likely provide the most accurate results (Martínez del Rio et al. 2009), we elected to use the TDFs from our models for five reasons. 1) To our knowledge, no TDFs exist for nestling songbirds. 2) TDFs for δ^{13}C were similar with expected values of insectivorous songbirds. (3) The use of typical TDFs for δ^{15}N (3.4‰) implied that the nestlings fell below the trophic level of their known prey. 4) In over 1800 provisioning observations, nestlings were observed consuming non-arthropods fewer than 10 times, with each of these being salamanders. 5) The slopes of the nestling-derived and arthropod-derived isotopic values did not differ significantly; all suggesting that similarities between nestling and common arthropod prey isotopic values were a result of consumption and not erroneous TDFs. Additionally, research has found that younger, rapidly developing birds exhibit lower δ^{15}N TDFs (Sears et al. 2009; Micklem et al. 2021), further supporting the likelihood that the nestlings in our study would exhibit similar effects.

Stable isotope mixing models

Bayesian stable isotope mixing models were designed using the nestling isotopic values as mixtures, the mean and standard deviation of arthropod isotopic values as sources, the *simmr_mcmc* function within the *SIMMR* R package (Parnell et al. 2010) and one of three strategies, termed "Diffuse Priors," "Scaled Priors," and "Occurrence Probability" (Fig. 1). The "Diffuse Priors"

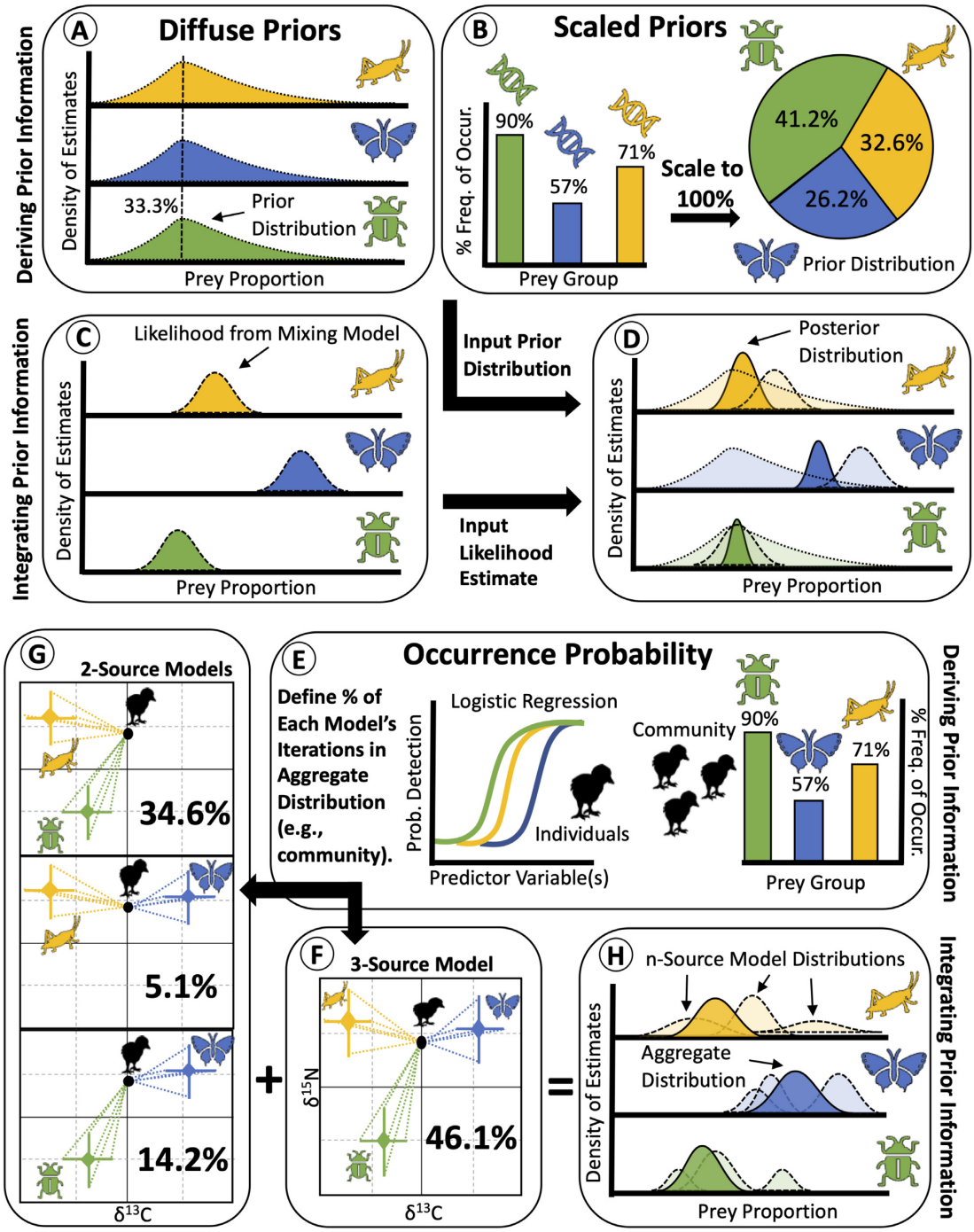

Fig. 1 Flow chart of methods to incorporate prior information into stable isotope mixing models (SIMMs) in the absence of *a priori* information of quantitative prey composition. *Formal Incorporation as Priors*: Diffuse Priors (**A**) assume that all prey included in the mixing model are consumed in equal proportions (1/*n* sources, e.g., 33% for a three-source model). Scaled Priors (**B**) scale DNA-based frequency of occurrence (FOO) of major prey groups by dividing FOO of each prey group by the sum of their collective FOOs so that they sum to 100%. These prior distributions are then coupled with likelihood estimates from SIMMs (**C**) to define the posterior distribution of possible prey contribution estimates (**D**). *Informal incorporation using Occurrence Probability:* The probability of detecting each prey group for individuals or at the community-level (**E**) is then used to define the percentage of each multi-source SIMM's iterations (F–G) that are included in the final aggregate distribution (H)

strategy, which was used to determine community-wide and individual dietary prey proportions, assigned identical priors for each prey group using a multivariate beta distribution (Fig. 1A) and default arguments for mixing model analysis. The "Scaled Priors" strategy, which was only applicable for the community-wide analysis—as frequency of occurrence data cannot be obtained from nestlings sampled a single time—follows the same

general methodologies as the Diffuse Priors strategy (Fig. 1B and C). However, the "Scaled Priors" strategy derived informative priors from the frequency of occurrence of the four main prey groups in this study: Aquatic Diptera, Terrestrial Diptera, Ephemeroptera and *Plecoptera*, and Lepidoptera. The mean frequency of occurrence and standard deviation for each prey group were divided by the summed frequency of occurrence of the prey groups to get a scaled DNA-based estimate of prey proportion. These estimates were supplied to the *simmr_elicit* function within the *SIMMR* R package, which specifies the prior multivariate beta distribution by using the centralized log ratio to transform the estimated dietary proportions. This prior distribution was then used in the Bayesian mixing model analysis identical to that used for the "Diffuse Priors" strategy. The resulting prey composition estimates from each iteration were rarefied to 1800 iterations and these values were used in downstream analyses.

Unlike the "Scaled Priors" strategy, which attempts to transform binomial occurrence frequency data into pseudo-compositional dietary proportion estimates to specify prior distributions, the "Occurrence Probability" strategy, instead, uses the probability of prey occurrence to determine which prey groups are included in each iteration of the stable isotope mixing model (Fig. 1E–H). For our community-wide mixing models, these probabilities were derived from the DNA-based, dietary frequency of occurrence data for each of the specified prey groups. As each nestling was sampled only a single time, we were unable to determine the dietary frequency of occurrence for each nestling in this study. Therefore, we derived the occurrence probability of each prey group in the diet of each nestling using a logistic regression approach. For each nestling, the *glm* R function used the presence or absence of each major prey group as a response variable, and the predictor variables were the frequency of occurrence of each prey group among other nestlings in the same brood. Stream was also included as a predictor variable as previous studies suggest that nestlings on different streams tend to have different prey compositions (Trevelline et al. 2018). These models were then used in the *predict.glm* function (*type* = "response"), which returned predictions for the probability of occurrence of each prey group in each nestling sample.

Once prey group occurrence probabilities were determined for the community-wide and individual data (Fig. 1E), we were then able to use these data along with the *sample* function (package—base) to create the prey group lists (e.g., prey lists with two, three, or all four prey groups; Fig. 1F and G) for the community-level analysis (1000 prey group permutations across all nestlings) and individual nestling analysis (1000 prey group permutations for each nestling). Using a for loop in R, mixing models were run on each of the prey group lists (community-level: 100,000 iterations per permutation, 1000 burn-in, 100 thinning; individual nestling: 100,000 iterations per permutation, 1000 burn-in, 100 thinning), and the resulting prey composition estimates were aggregated across the entire community or among each nestling (Fig. 1H). The resulting prey contribution estimates for, are presented as median values with 89% credible intervals. The credible intervals were computed using the highest density interval (HDI) as the aggregate posterior distributions were non-symmetrical. The code and files required to repeat mixing model and statistical analyses can be found in the Supplementary Analysis Files.

Results

Determining occurrence probability of major prey groups

According to dietary DNA metabarcoding, the most frequently detected prey groups for nestling Louisiana waterthrush in the 2017 breeding season were: Aquatic Diptera (95.3% of sampled diets), Ephemeroptera (76.6%), *Plecoptera* (71.9%), Lepidoptera (64.1%), and Terrestrial Diptera (56.2%). These frequency of occurrence data were used to represent the probability of occurrence within the "Occurrence Probability" strategy as well as the informative priors in the "Scaled Priors" mixing model strategy for all prey groups except Plecoptera and Ephemeroptera. As *Plecoptera* and Ephemeroptera were isotopically indistinguishable and share many life history traits (e.g., aquatic larval and terrestrial adult stages; Brittain 1990), their pooled frequency of occurrence (96.1%) was used in all community-wide analyses. After normalizing the frequency of occurrence and standard deviation so that the frequencies summed to 1 , the values for each prey group used in the "Scaled Priors" strategy were as follows: Ephemeroptera and Plecoptera (mean: 0.299; standard deviation: 0.088), Aquatic Diptera (0.309 +/- 0.069), Lepidoptera (0.208 +/- 0.157), and Terrestrial Diptera (0.182 +/- 0.162).

Logistic regressions for determining the probability of occurrence for each prey group in each nestling sample, found that the occurrence probabilities differed among the four prey groups with Aquatic Diptera (93.7% +/- 24.3%) exhibiting the highest probability, followed by Ephemeroptera and Plecoptera (93.5% +/- 21.3%), Lepidoptera (74.6% +/- 13.7%), and Terrestrial Diptera (42.4% +/- 28.9%). In addition to the high degree of inter-nestling variation in occurrence probabilities, there was also a high degree of variation

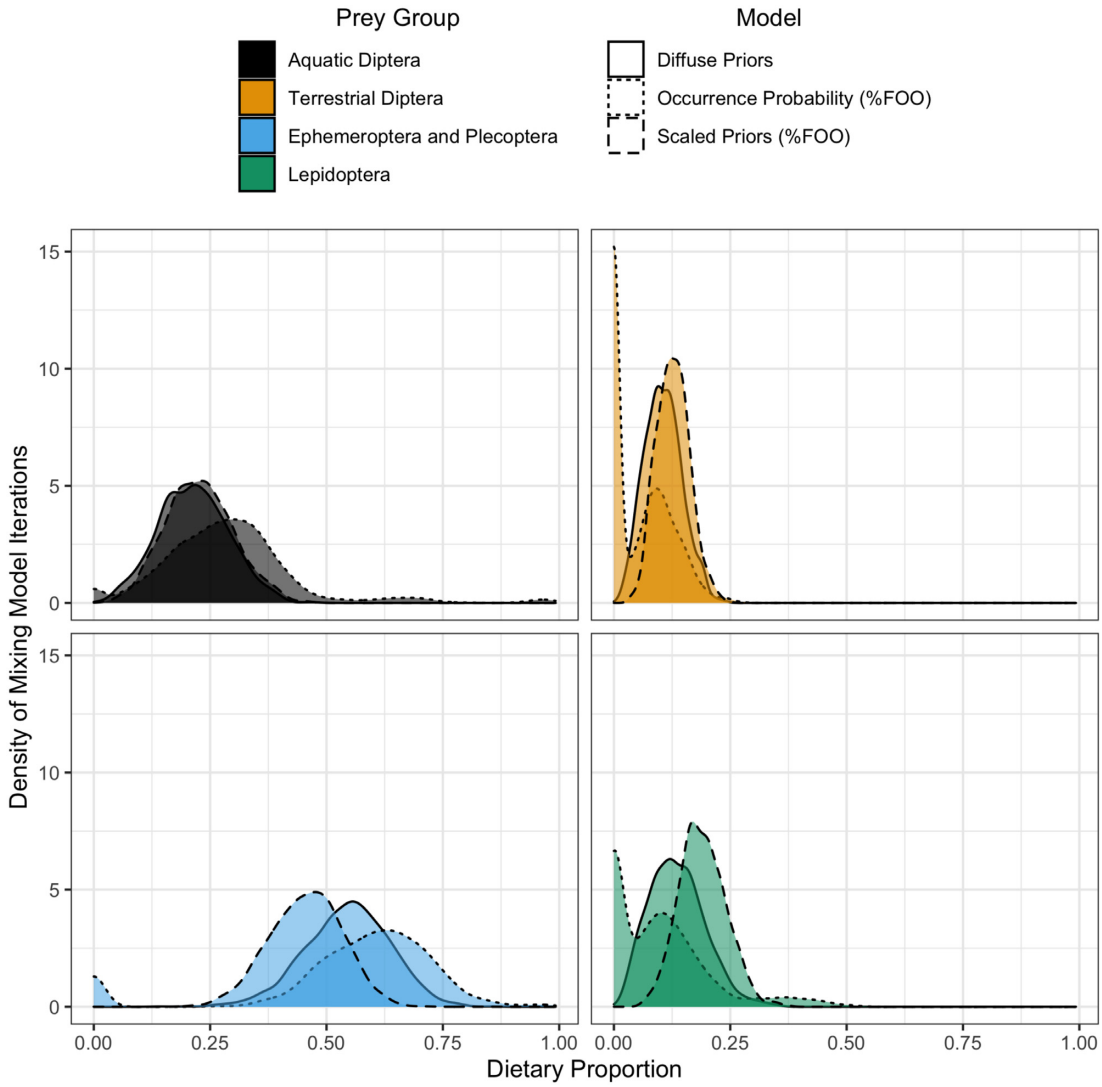

Fig. 2 Comparison of three approaches for estimating community-wide prey composition from stable isotope mixing models. Mixing models were either run with diffuse priors, priors derived from scaled DNA-based dietary frequency of occurrence or with occurrence probability information from DNA-based dietary frequency of occurrence. The results from each model were rarefied to allow for consistent number of iterations across model types

among streams for some prey groups. Though Aquatic Diptera and Ephemeroptera and Plecoptera were detected near-ubiquitously on each of the study streams, Terrestrial Diptera occurrence probability ranged from as low as 5.6% on Camp Run to as high as 91% on Powdermill Run. Similarly, the average Lepidoptera occurrence probability was lowest on Camp Run (68.5% +/- 11.2%), but highest on Loyalhanna Creek (80.7% +/- 3.6%).

Community-wide stable isotope mixing models

Mixing models using each of the three strategies determined that Ephemeroptera and *Plecoptera* exhibited, the highest contribution to nestling Louisiana waterthrush diets (Diffuse Priors: 0.548 [0.405–0.686]; Occurrence Probability: 0.601 [0.373–0.836]; Scaled Priors: 0.459 [0.331–0.575]), followed by Aquatic Diptera (Diffuse Priors: 0.212 [0.099–0.340]; Occurrence Probability: 0.280 [0.063–0.447]; Scaled Priors: 0.225 [0.106–0.340]), Lepidoptera (Diffuse Priors: 0.131 [0.045–0.230]; Occurrence Probability: 0.083 [0–0.214]; Scaled Priors: 0.187 [0.118–0.276]), and Terrestrial Diptera (Diffuse Priors: 0.104 [0.044–0.171]; Occurrence Probability: 0.038 [0–0.140]; Scaled Priors: 0.126 [0.069–0.178]) (Fig. 2). The "Scaled Priors" strategy yielded the most precise prey contribution measurements for each prey group (average prey contribution estimate interquartile range: 0.081; average prey contribution estimate total range: 0.375) followed by the "Diffuse Priors" strategy (average interquartile range: 0.091; average total range: 0.441), and the "Occurrence

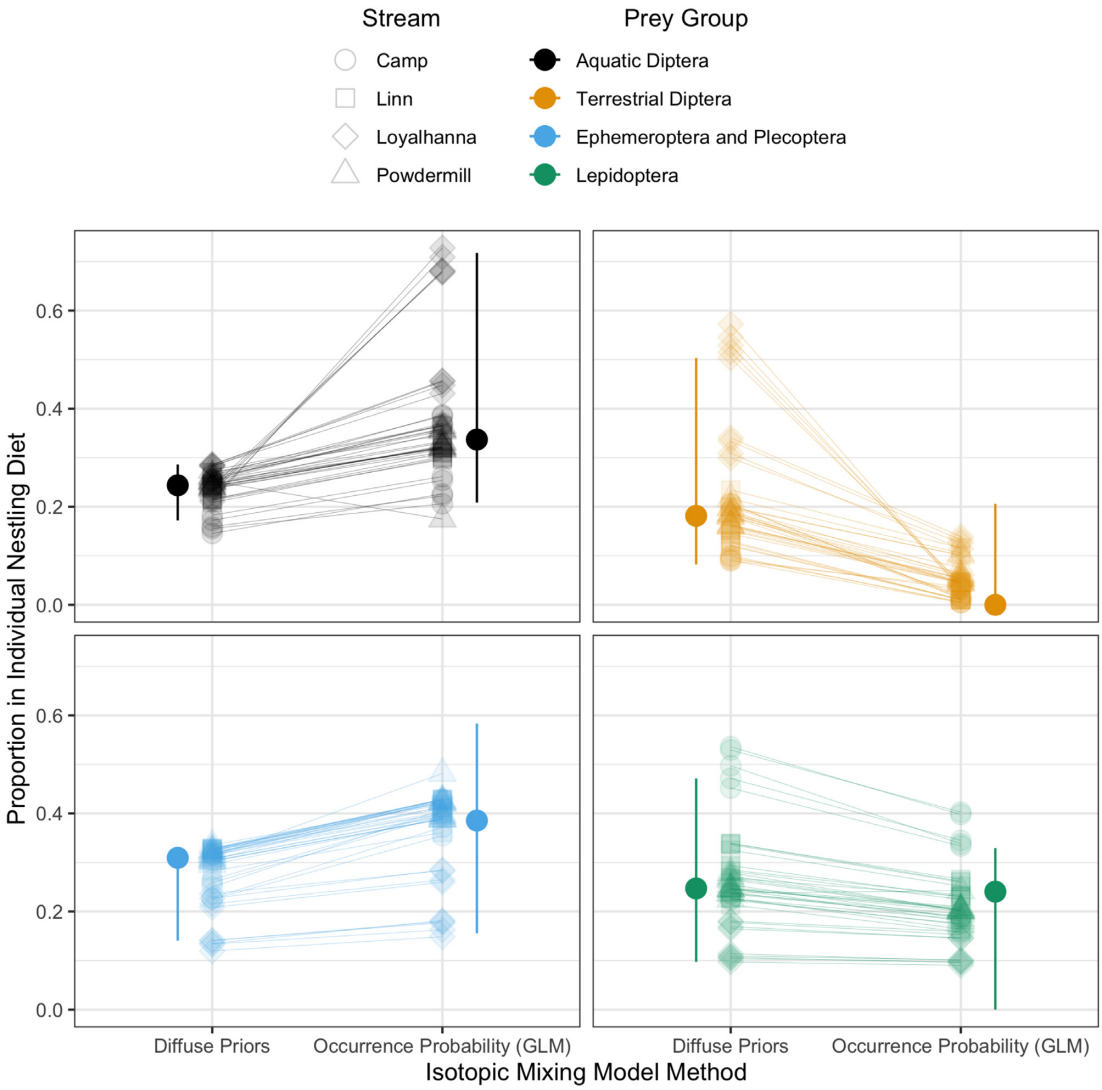

Fig. 3 Individual prey contributions derived from stable isotope mixing models with diffuse priors or DNA-based occurrence probability information. Individual mean dietary contributions for each prey group (transparent shapes) were derived from mixing models using diffuse priors or with DNA-based occurrence probabilities derived from logistic regression models. Median values and 89% credibility intervals across all nestlings are shown on the side of individual mean values for each method

Probability" strategy (average interquartile range: 0.143; average total range: 0.690).

Individual nestling stable isotope mixing models

Mixing models using the "Diffuse Priors" strategy to quantify relative prey contributions for individual nestlings, found that the average prey contribution differed very little between prey groups, though terrestrial prey contribution estimates exhibited a relatively high degree of variation across nestlings (Lepidoptera: 0.247 [0.097–0.471]; Terrestrial Diptera: 0.181 [0.083–0.503]; Fig. 3) compared to aquatic sources (Ephemeroptera and *Plecoptera*: 0.310 [0.141–0.330]; Aquatic Diptera: 0.243 [0.172–0.286]). Using the "Occurrence Probability" strategy, however, indicated a high degree of variation among prey groups with Aquatic Diptera (0.337 [0.208–0.718]; Fig. 3) and Ephemeroptera and *Plecoptera* prey contribution (0.385 [0.156–0.583]) exhibiting higher prey contribution than either Lepidoptera (0.241 [0–0.330]) or Terrestrial Diptera (0 [0–0.206]).

Discussion

Community-wide stable isotope mixing models

When used to quantify the community-wide prey composition for nestling Louisiana waterthrush, each of the three tested stable isotope mixing model strategies—"Diffuse Priors," "Occurrence Probability," and "Scaled Priors"—returned the same rank-order of prey

contribution: (1) Ephemeroptera and *Plecoptera*, (2) Aquatic Diptera, (3) Lepidoptera, and (4) Terrestrial Diptera. However, the estimated prey contribution of each group and the precision of these estimates often differed based on the selected mixing model strategy. The largest of these discrepancies was observed with the prey contribution estimates of aquatic-derived prey, namely Ephemeroptera and *Plecoptera* and Aquatic Diptera. Previous work by Hoenig (2022) has indicated that breeding adult waterthrush preferentially provision aquatic prey to their offspring, and other research suggests that aquatic prey, which contain higher densities of long-chain polyunsaturated fatty acids (LC-PUFA) than terrestrial prey (Hixson et al. 2015), are important for early songbird development (Twining et al. 2016, 2018). Because the consumption of these prey sources likely has an effect on the development and, as a result, long-term survival of songbirds (Naef-Daenzer and Keller 1999), it is important that these dietary estimates are as accurate as possible to ensure that our understanding of pollution-sensitive songbirds, and the conservation actions that stem from this understanding, are well-informed.

In this study, the "Diffuse Priors" and "Occurrence Probability" strategies each suggested that nestling waterthrush derive more than half of their nutrients from Ephemeroptera and Plecoptera, while the inclusion of informative priors derived from scaled, DNA-based frequency of occurrence data (e.g., "Scaled Priors") lowered the estimated mean prey contribution of this prey to less than half of their dietary composition. This discrepancy is particularly surprising as breeding Louisiana waterthrush are known to preferentially occupy and defend territories on unimpacted streams, which support an abundance of Ephemeroptera and Plecoptera taxa, while streams lacking these pollution-intolerant taxa tend to have fewer breeding pairs, fledge fewer young per kilometer, and are associated with stunted nestling development (Mulvihill et al. 2008). In addition, results from Hoenig (2022) found that the isotopic niche area (a proxy for isotopic niche breadth; Newsome et al. 2007) shrunk significantly as adult waterthrush provisioned increasing amounts of Ephemeroptera and Plecoptera, results, which strongly suggest that pollution-intolerant aquatic prey are an important and likely preferred component of nestling waterthrush diet. While it is possible that the "Scaled Priors" strategy may be highlighting a more generalized prey composition for waterthrush, assessing these results alongside our understanding of waterthrush breeding ecology suggests that the relative decrease in prey contribution of Ephemeroptera and Plecoptera returned by the "Scaled Priors" strategy is more likely a result of high dietary occurrence frequency of alternative prey (e.g., Aquatic Diptera) skewing informative priors, and, as a result, prey contribution estimates.

These results not only highlight how influential informative prior selection can be on the posterior distribution for stable isotope mixing models, but also highlight the importance of appropriately deriving and incorporating prior information into these models. As the presence–absence data returned by DNA-based methods do not represent compositional estimates of prey quantity, their direct incorporation into composition-based mixing models as prior information is inappropriate, even after scaling these estimates to sum to 100%. For example, two consumer populations may exhibit identical frequency of occurrence for the same prey groups (e.g., Aquatic Diptera and Ephemeroptera and Plecoptera in this study), but utilize these prey groups at significantly different rates, thus, making these populations' isotopic values entirely distinct. However, if the "Scaled Priors" strategy were employed for these consumer populations, the prey contribution estimates of these two isotopically distinct groups would more closely resemble informative prior values (e.g., Robinson et al. 2018), as opposed to each consumer population's truly distinct dietary composition, due to the influence that informative priors can have on the posterior distribution. Therefore, if researchers wish to integrate multiple methods within a single analysis, we recommend they do so in a manner that preserves the fundamental nature of the data (e.g., compositional vs. presence–absence), as such an approach will better preserve the accuracy of dietary characterizations. Though this approach may have lowered the precision of our prey contribution estimates as compared to those using presence–absence data in compositional mixing models, the latter option does not properly integrate these data, and in turn, has the potential to provide high-precision, but potentially inaccurate, prey contribution estimates, which may not be able to confidently forward our understanding of a species dietary ecology.

Individual nestling stable isotope mixing models

While the "Diffuse Priors" and "Occurrence Probability" strategies differed very little in their prey composition estimates at the community-level, these strategies differed greatly in their estimation of prey composition for individual nestlings. The results returned by the "Diffuse Priors" strategy for each prey group were closely centered around 25%—an expected value for priors in a four-source mixing model—which further highlights how the selection of priors in mixing model analysis can greatly shape the posterior

distribution of prey contribution estimates. However, when performing mixing model analysis using a GLM-based approach to determine each prey group's probability of occurrence, and thus their probability of being included in each iteration of the mixing model, we observed higher aquatic prey contribution estimates and lower terrestrial prey contribution estimates; findings which are to be expected for a stream-dependent songbird that is hypothesized to specialize on aquatic-derived arthropod prey (Mulvihill et al. 2008). Additionally, the "Occurrence Probability" strategy permitted mixing model iterations to be individually parameterized, thus allowing prey occurrence probabilities to vary among nestlings and shape the estimated prey contribution of each nestling in a manner not possible with the current version of SIMMR (though see applications of the MixSIAR R Package; Stock et al. 2018). As these occurrence probabilities were derived from logistic regression models, researchers using this approach can also incorporate relevant environmental variables (e.g., temporal or spatial information), information on prey availability, or even results returned from other dietary study techniques into these regression models to provide an even more precise estimate of prey occurrence probability than what is offered by frequency of occurrence, alone. We are aware of only one other mixing model approach that can easily incorporate additional, and often non-dietary, variables into mixing model analysis (e.g., MixSIAR; Stock et al. 2018; though see derivations for these methods in "Supplementary Information: MixSIAR model description"), and we recommend that future research attempt to incorporate the "Occurrence Probability" strategy alongside the other, numerous parameters allowed by the *MixSIAR* framework (e.g., random or fixed effects and various sources of error).

In addition to the expected changes in prey contribution for these prey groups, the observed changes between mixing model strategies also highlighted other interesting findings, namely for those prey sources which are likely provisioned when preferred prey are less available. One such prey source, Lepidoptera, is often preferred among breeding, migratory songbird species (Holmes 2007); so much so that one genus of wood-warbler, *Setophaga*, derives its name from the Greek *ses,* meaning "moth" and *phagos*, meaning "eating." Although Louisiana waterthrush belong to a different genus within their shared family, Parulidae, recent DNA-based studies (Trevelline et al. 2016, 2018) as well as the isotopic evidence presented in (Hoenig 2022) suggest that Lepidoptera are also an important prey source for nestling Louisiana waterthrush on the breeding grounds, and may provide compensatory nutrients later in the breeding season when Ephemeroptera availability declines after peaking in early spring (Trevelline et al. 2016). Therefore, the minimal differences in Lepidoptera prey composition estimates returned by each method give further credence that the estimated contributions of prey, which are known to be consumed in high abundances (e.g., Lepidoptera), but detected less frequently than other prey groups (e.g., Aquatic Diptera), are not particularly sensitive to the incorporation of occurrence probability into stable isotope mixing models. Furthermore, as the contribution of Terrestrial Diptera, a prey group that appears to be consumed more often in the absence of preferred prey, decreased, it also appears that the "Occurrence Probability" strategy can provide a more accurate prey contribution estimate for prey groups that are frequently detected, but consumed in lower abundances.

Methodological considerations

Though we believe our approach to integrating DNA-based data into isotopic analyses presents an improvement over previous attempts, there are limitations worth considering. For one, stable isotope mixing models typically work under the assumption that prey-derived isotopes are dispersed evenly throughout a consumer's body. However, research has shown that consumers differentially allocate macromolecular pools (e.g., proteins, lipids, and carbohydrates) to the various tissues within their body, suggesting that even the incorporation of DNA-based information cannot overcome the fact that each tissue type can incorporate prey nutrients at variable rates and thus return different information about a consumer's diet (Schwarcz 1991). Although most dietary studies to date have used bulk stable isotope analysis (e.g., analysis of every elemental isotope of interest in a given tissue), compound-specific isotope analysis (CSIA; O'Brien et al. 2005)—such as that of amino acids (CSIA-AA; Pollierer et al. 2019) or fatty acids (CSIA-FA; De Troch et al. 2012)—may help to better elucidate the source of specific macromolecules found within consumer tissues, and thus improve the accuracy of all mixing models, including those presented here. Another limitation exists when considering isotopic turnover, or the time required to replace the isotopes in active tissues, and how that time frame correlates with information from dietary DNA metabarcoding. Fecal samples, which are one of the most used sample types for dietary DNA metabarcoding, likely reflect the prey composition of an individual over a single day or less, while isotopic analyses, which often use red blood cells or organ tissues, reflect the prey composition over multiple days, weeks or even months (Hoenig et al. 2021). This discrepancy may be particularly problematic if an individual's prey composition drastically

changes over a short time frame, as such a dramatic shift may cause DNA-based evidence to suggest a given dietary frequency of one prey type that may have made up an entirely different amount of the isotopic pool in a given tissue. Researchers could overcome this by sampling individuals for prey DNA over a time frame that matches the isotopic turnover of a given tissue or by performing isotopic analysis on tissue types with turnover times that match dietary DNA retention rates, such as plasma (~1 day) or breathe (~4 h) (Podlesak et al. 2005).

Another limitation of this approach lies in its simplicity. Performing every permutation of an *n-source* mixing model presents a straightforward method of incorporating DNA-based frequency of occurrence into isotopic dietary characterizations. However, in limiting the number of prey sources in the model, the mixing area (i.e., the polygon created by connecting all prey sources in the model) may exclude some consumer isotopic values, and thus return a poor model fit and convergence for any one of the mixing models that contribute to the final aggregate distribution. This effect would be particularly pronounced in 2-source models as the mixing polygon consists only of the distance between and associated errors with the two sources. Furthermore, because each of the *n-source* models are run independently, it is also difficult to assess the overall fit and convergence of the final aggregation of models; though the individual models that contribute to the aggregate distribution could be assessed in manners found in more standard applications. For these reasons, we present this study as one that highlights an important future direction in trophic ecology and demonstrates a simplified approach and an important early step to formally integrating DNA-based and isotopic techniques. Another approach that could conceivably integrate DNA-based and isotopic methods is a recently developed hierarchical Bayesian model, *EcoDiet*, which combines knowledge from the literature, traditional dietary analysis, and biotracers (e.g., isotope or fatty acid) data into a mixing model framework (Hernvann et al. 2022). Though Hernvann et al. (2022) contend that the *EcoDiet* model is not yet fit for DNA-based data, it appears that occurrence probability calculations from this study—which stem from community-wide frequency of occurrence or logistic regression-based estimates for individuals—could be integrated into the *EcoDiet* model with only small modifications (e.g., limiting the number of consumer species and simplifying trophic linkages to include only the most frequent). The quick succession at which the present approach and the *EcoDiet* model were developed suggests a pressing need for incorporating dietary frequency of occurrence—be it from traditional or DNA-based methods—into stable isotope mixing models, and we look forward to seeing how researchers use and improve upon these methods in future dietary studies.

Conclusions

A number of recent reviews have highlighted the importance of combining, and even integrating, multiple techniques within a single study to ensure our understanding of animal diets is accurate as possible (Traugott et al. 2013; Nielsen et al. 2018; Hoenig et al. 2021), the most appealing of which being the integration of DNA metabarcoding and stable isotope analysis. Most studies to date incorporating data from these techniques have done so by including DNA-based, presence–absence data as informative priors in compositional stable isotope mixing models (Chiaradia et al. 2014; Bonin et al. 2020). Because these data reflect fundamentally different units of a consumer's diet, such an approach is analytically inappropriate, and while these methods often increased the precision of prey contribution estimates, the ability of informative priors to skew posterior distributions puts the accuracy of these estimates in doubt. However, the "Occurrence Probability" strategy preserves the probabilistic nature of the DNA-based data and, when using the GLM-based approach, can easily integrate additional variables to improve occurrence probability estimates and, as a result, isotope-based prey contribution estimates. While we believe that this application is an improvement on past efforts to incorporate DNA-based data into stable isotope mixing models, we highly encourage future researchers to improve upon current methods to integrate dietary data, or even develop completely novel ones, so that we may continue to forward our understanding of animal diets.

Acknowledgments

We thank the Carnegie Museum of Natural History, Dr. John Wenzel, and Cokie Lindsay for providing accommodations and access to study sites near the Powdermill Nature Reserve. We thank Danilo Mejía, Maria Paulino, Jason Lacson, and Nancy Ransom for providing field assistance. . All protocols conducted in this study were approved by the Institutional Animal Care and Use Committee of the National Aviary/Pittsburgh Zoo and PPG Aquarium (NA16-001).

Funding

This work was supported by Duquesne University's Bayer School of Natural and Environmental Sciences through teaching assistantships and the Bayer Graduate Research Fellowship and by the Avian Conservation Endowment of the National Aviary and the

American Ornithological Society through a Hesse Award to B.D.H.

Supplementary data

Supplementary data available at *ICB* online.

Conflict of interest

The authors declare no conflicts of interest.

Data availability statement

Data and bioinformatic pipelines can be found at https://osf.io/5v76x/.

References

Bonin M, Dussault C, Taillon J, Lecomte N, Côté SD. 2020. Combining stable isotopes, morphological, and molecular analyses to reconstruct the diet of free-ranging consumers. Ecol evol 10:6664–76.

Bradshaw PJ, Broderick AC, Carreras C, Inger R, Fuller W, Snape R, Stokes KL Godley BJ. 2017. Satellite tracking and stable isotope analysis highlight differential recruitment among foraging areas in green turtles. Mar Ecol Prog Ser 582:201–14.

Braley M, Goldsworthy SD, Page B, Steer M, Austin JJ. 2010. Assessing morphological and DNA-based diet analysis techniques in a generalist predator, the arrow squid Nototodarus gouldi. Mol Ecol Resour 10:466–74.

Brittain JE. 1990. Life history strategies in Ephemeroptera and Plecoptera. In: Mayflies and stoneflies: life histories and biology. Dordrecht: Springer. p. 1–12.

Chiaradia A, Forero MG, McInnes JC, Ramírez F. 2014. Searching for the true diet of marine predators: incorporating Bayesian priors into stable isotope mixing models. PLoS One 9: e92665.

Collis K, Roby DD, Craig DP, Adamany S, Adkins JY, Lyons DE 2002. Colony size and diet composition of piscivorous waterbirds on the lower Columbia River: implications for losses of juvenile salmonids to avian predation. Trans Am Fish Soc 131:537–50.

Croxall JP. 1976. The composition and behaviour of some mixed-species bird flocks in Sarawak. Ibis 118:333–46.

De Troch M, Boeckx P, Cnudde C, Van Gansbeke D, Vanreusel A, Vincx M, Caramujo MJ. 2012. Bioconversion of fatty acids at the basis of marine food webs: insights from a compound-specific stable isotope analysis. Mar Ecol Prog Ser 465: 53–67.

Deniro MJ., Epstein S. 1976. You are what you eat (plus a few per mil): the carbon isotope cycle in food chains.Geological Society of America Abstracts with Programs. 8: 834–835.

DeNiro MJ, Epstein S. 1978. Carbon isotopic evidence for different feeding patterns in two hyrax species occupying the same habitat. Science 201:906–8.

Eaton SW. 1958. A life history study of the Louisiana Waterthrush. Wilson Bull 70:211–36.

Fleischer AL, Jr, Bowman R, Woolfenden GE. 2003. Variation in foraging behavior, diet, and time of breeding of Florida scrub-jays in suburban and wildland habitats. The Condor 105: 515–27.

Franco-Trecu V, Drago M, Riet-Sapriza FG, Parnell A, Frau R, Inchausti P. 2013. Bias in diet determination: Incorporating traditional methods in Bayesian mixing models. PLoS One 8:e80019.

Goodrich B, Gabry J, Ali I, Brilleman S. 2020.rstanarm: Bayesian applied regression modeling via Stan. Available at: https://mc-stan.org/rstanarm/. (Last accessed on 6 June 2022).

Hernvann PY, Gascuel D, Kopp D, Robert M, Rivot E. 2022. EcoDiet: a hierarchical Bayesian model to combine stomach, biotracer, and literature data into diet matrix estimation. Ecol Appl 32:e2521.

Hixson SM, Sharma B, Kainz MJ, Wacker A, Arts MT. 2015. Production, distribution, and abundance of long-chain omega-3 polyunsaturated fatty acids: a fundamental dichotomy between freshwater and terrestrial ecosystems. Environ Rev 23:414–24.

Hoenig BD. 2022. A review of current methods in avian dietary analysis and their integrated application to characterize the trophic niche of Louisiana Waterthrush (*Parkesia Motacilla*).Duquesne University, PhD Dissertation,

Hoenig BD, Snider AM, Forsman AM, Hobson KA, Latta SC, Miller ET, Porter BA. 2021. Current methods and future directions in avian diet analysis. Ornithology 139:1–28, https://doi.org/10.1093/ornithology/ukab077.

Hoenig BD, Trevelline BK, Nuttle T, Porter BA. 2021. Dietary DNA metabarcoding reveals seasonal trophic changes among three syntopic freshwater trout species. Freshw Biol 66: 509–23.

Holmes RT. 2007. Understanding population change in migratory songbirds: long-term and experimental studies of Neotropical migrants in breeding and wintering areas. Ibis 149:2–13.

Jusino MA, Banik MT, Palmer JM, Wray AK, Xiao L, Pelton E, Kawahara AY,Barber JR,Gratton C,Lindner DL,Peery MZ. 2019. An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Mol Ecol Resour 19:176–90.

Martínez del Rio C, Wolf N, Carleton SA, Gannes LZ. 2009. Isotopic ecology ten years after a call for more laboratory experiments. Biol Rev 84:91–111.

Micklem I, Connan M, Stander N, McQuaid CD. 2021. Influence of ontogeny on stable isotope ratios and trophic discrimination factors of African penguin (Spheniscus demersus) tissues. Mar Biol 168:1–12.

Mulvihill RS, Newell FL, Latta SC. 2008. Effects of acidification on the breeding ecology of a stream-dependent songbird, the Louisiana waterthrush (*Seiurus motacilla*). Freshw Biol 53:2158–69.

Naef-Daenzer B, Keller LF. 1999. The foraging performance of great and blue tits (Parus major and P. caeruleus) in relation to caterpillar development, and its consequences for nestling growth and fledging weight. J Anim Ecol 68:708–18.

Newsome SD, Martinez del Rio C, Bearhop S, Phillips DL. 2007. A niche for isotopic ecology. Front Ecol Environ 5:429–36.

Nielsen JM, Clare EL, Hayden B, Brett MT, Kratina P. 2018. Diet tracing in ecology: Method comparison and selection. Methods Ecol Evol 9:278–91.

O'Donovan SA, Budge SM, Hobson KA, Kelly AP, Derocher AE. 2018. Intrapopulation variability in wolf diet revealed using a combined stable isotope and fatty acid approach. Ecosphere 9:e02420.

O'Brien DM, Boggs CL, Fogel ML. 2005. The amino acids used in reproduction by butterflies: a comparative study of dietary sources using compound-specific stable isotope analysis. Physiol Biochem Zool 78:819–27.

Parnell AC, Inger R, Bearhop S, Jackson AL. 2010. Source partitioning using stable isotopes: coping with too much variation. PLoS One 5:e9672.

Phillips DL. 2012. Converting isotope values to diet composition: the use of mixing models. J Mammal 93:342–52.

Phillips DL, Inger R, Bearhop S, Jackson AL, Moore JW, Parnell AC, Semmens BX,Ward EJ. 2014. Best practices for use of stable isotope mixing models in food-web studies. Can J Zool 92:823–35.

Piñol J, Mir G, Gomez-Polo P, Agustí N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Mol Ecol Resour 15:819–30.

Podlesak DW, McWilliams SR, Hatch KA. 2005. Stable isotopes in breath, blood, feces and feathers can indicate intra-individual changes in the diet of migratory songbirds. Oecologia 142:501–10.

Pollierer MM, Larsen T, Potapov A, Brückner A, Heethoff M, Dyckmans J, Scheu S. 2019. Compound-specific isotope analysis of amino acids as a new tool to uncover trophic chains in soil food webs. Ecol Monogr 89:e01384.

Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P. 2012. Who is eating what: diet assessment using next generation sequencing. Mol Ecol 21:1931–50.

Ralph CP, Nagata SE, Ralph CJ. 1985. Analysis of droppings to describe diets of small birds. J Field Ornithol 56:165–74.

Robinson BG, Franke A, Derocher AE. 2018. Stable isotope mixing models fail to estimate the diet of an avian predator. The Auk 135:60–70.

Schwarcz HP. 1991. Some theoretical aspects of isotope paleodiet studies. J Archaeolog Sci 18:261–75.

Sears J, Hatch SA, O'Brien DM. 2009. Disentangling effects of growth and nutritional status on seabird stable isotope ratios. Oecologia 159:41–8.

Sherry TW. 1984. Comparative dietary ecology of sympatric, insectivorous Neotropical flycatchers (Tyrannidae). Ecol Monogr 54:313–38.

Snyder NF, Wiley JW. 1976. Sexual size dimorphism in hawks and owls of North America. Ornithological Monograph 30, American Ornithologist's Union.

Stock BC, Jackson AL, Ward EJ, Parnell AC, Phillips DL, Semmens BX. 2018. Analyzing mixing systems using a new generation of Bayesian tracer mixing models. PeerJ 6:e5096.

Traugott M, Kamenova S, Ruess L, Seeber J, Plantegenest M 2013. Empirically characterising trophic networks: what emerging DNA-based methods, stable isotope and fatty acid analyses can offer. Adv ecol res 49:177–224.

Trevelline BK, Latta SC, Marshall LC, Nuttle T, Porter BA. 2016. Molecular analysis of nestling diet in a long-distance Neotropical migrant, the Louisiana Waterthrush (*Parkesia motacilla*). The Auk 133:415–28.

Trevelline BK, Nuttle T, Hoenig BD, Brouwer NL, Porter BA, Latta SC. 2018. DNA metabarcoding of nestling feces reveals provisioning of aquatic prey and resource partitioning among Neotropical migratory songbirds in a riparian habitat. Oecologia 187:85–98.

Trevelline BK, Nuttle T, Porter BA, Brouwer NL, Hoenig BD, Steffensmeier ZD, Latta SC. 2018. Stream acidification and reduced aquatic prey availability are associated with dietary shifts in an obligate riparian Neotropical migratory songbird. PeerJ 6:e5141.

Twining CW, Brenna JT, Lawrence P, Shipley JR, Tollefson TN, Winkler DW. 2016. Omega-3 long-chain polyunsaturated fatty acids support aerial insectivore performance more than food quantity. Proc Natl Acad Sci 113:10920–5.

Twining CW, Shipley JR, Winkler DW. 2018. Aquatic insects rich in omega-3 fatty acids drive breeding success in a widespread bird. Ecol Lett 21:1812–20.

Wassenaar LI. 2019. Introduction to conducting stable isotope measurements for animal migration studies. In: Hobson KA,Wassenaar LI, editors.Tracking animal migration with stable isotopes.Amsterdam, the Netherlands: Elsevier. p. 25–51.