Title: Predator body size and species identity determine predator-prey interactions in diet DNA metabarcoding data

Potential author list: Ana Miller-ter Kuile, Austen Apigo, An Bui, Bart DiFiore, Elizabeth Forbes, Daniel Preston, Taylor Bogar, Jasmine Childress, Rodolfo Dirzo, Magalay Espinoza, Cora Johnston, Maggie Klope, Kevin Lafferty, Michelle Lee, John McLaughlin, Marisa Morse, Carina Motta, Katherine Plummer, Chelsea Steele, Hillary Young

Lab: anyone with 199?

ProcB: 6 MS pages, 850 words = 1 page and 2-4 small figures = 1 page (this is already currently too long…)

**Abstract (200 words)**

**Keywords (3-6 words)**

**Introduction**

Baby: The field of food web ecology has spent nearly a century observing and predicting patterns in complex communities of multiple species and their interactions (reviewed in Layman et al. 2015). Through multiple evolutions in the thinking in the field, especially on the understanding of how ecologically complex ecosystems maintain stability (e.g. May 1971 to Polis 1991), it has been clear that while there are patterns specific to food webs in different contexts (e.g. environmental; Brose et al. 2006), there seem to be general relationships between the role of consumers in a food web and basic species characteristics, in particular, body size (Woodward 2002). Body size is a fundamental trait that predicts feeding behaviors, including how much an individual eats, and for predators, which prey are available to those predator species based on relative body size differences and optimal foraging theory (e.g. Stouffer 2015). In this way, body size has been used to define the niches of consumers, in particular predators, in many of the food web models that best predict observed patterns in nature (e.g. Gravel et al. 2015, Pomeranz et al. 2019, Rudolf and Lafferty 2011).

Baby 2: One of the food web models best able to predict observed patterns of species interactions (measured as a binary presence-absence of interactions) is the niche model (Williams and Martinez 2000), which has been extended to incorporate predator and prey sizes to build species niches (Gravel et al. 2015, Gravel et al. 2013). The niche model with body sizes defines a predator’s diet based on its body size, the allometric relationship between its size and the size of its prey (Cohen et al. 2003; Brose et al, 2006; Riede et al. 2010), and the range of prey items around that central relationship, with larger predators having a wider range of potential prey sizes but with a lower bound based on optimal foraging (e.g. Nakazawa, SI Figure 1). While the body size niche model predicts empirical feeding interactions with high accuracy (Gravel et al. citations), it is improved even more when feeding interactions are constrained by limiting interactions even more to empirically-observed interactions from phylogenetically-similar species of both predators and prey (Grey et al., Pomeranz et al. 2019, check that this is true). Whether it be species, functional grouping (beyond body size, e.g. locomotion, metabolism), or phylogenetic similarity, incorporating aspects of predator characteristics beyond body size is important (Rudolf et al. 2014 OTHERS).

Werewolf: Both a body size and a body size and species identity approach is promising for describing patterns of predator-prey interactions across consumer groups and environmental contexts. However, for some predators, in particular small-bodied invertebrate predators for which empirical diet methods (e.g. gut dissections) are impossible or often unfeasible (e.g. predators smaller than 1.0g, CITE From methods paper set of papers), we have a dearth of observed interaction data with which to validate extrapolated body size or phylogenetic methods of predicting food web structure. For these consumers, species interactions are often inferred from literature reports of observed interactions from phylogenetically-related species (e.g. ) or based on body size feeding constraints (e.g. Hines et al. 2019). Because these are the very two patterns we are aiming to observe in these datasets, then, without empirical data it is unclear whether any patterns in these consumers related to body size or species identity or general patterns or artefacts of the non-empirical diet assignment methods used to compile them. While body body size and species identity have been observed to be empirically important in some invertebrate food webs (e.g. aquatic invertebrate food webs, Woodward and Hildrew 2002, Rudolf et al. 2014), the biology of many terrestrial invertebrate predators suggests that assuming these patterns will hold may be invalid. For example, many terrestrial invertebrate predators use webs or venom to subdue and catch prey (Nakazawa et al.), thus loosening restrictive body-size based interaction assignments. Furthermore, because, unlike other invertebrate systems, in terrestrial food webs, predators are not necessarily bigger than other organisms in the food web (e.g. Sechi et al. 2015), the numerical majority of prey items may be larger than individual predators, thus necessitating variations by some species in predator-prey body size relationships or prey selection criteria. Thus, relatively smaller predator species that are phylogenetically-similar to relatively larger species may be faced with ecosystems where many potential prey are relatively larger than they are. [or something…].

Werewolf 2: Small-bodied terrestrial invertebrates, including insects and spiders, comprise 50% of the earth’s animal biomass (Bar-On et al. 2018) and the majority of global animal species diversity (Mora et al. 2011 Costello et al. 2013), so understanding their trophic interactions is not only important for finding general food web patterns, but could be crucial for understanding the ecology of the world’s most dominant group of consumers shaping biomass and nutrient cycling dynamics (e.g. Bar-On et al. 2018). Specifically, we have a lack of empirical predator-prey interaction data with which to understand how body size and species identity shape food web patterns, and thus, we are limited in our ability to compare empirical data of this group of predators to compare to previous diet-assignment methods to either validate or refine those approaches (e.g. Hines et al., Brose citations).

Silver Bullet: In this study, we employ novel diet DNA metabarcoding data from 182 individuals of nine terrestrial invertebrate predator species to document predator-prey interactions between these predators and their prey in natural field conditions. Combined with an extensive dataset of body sizes for both predator individuals and the prey groups identified in their diets, we examine the predator-prey body size relationships for these predators across and within species. To understand how predator size and species identity drive empirical predator-prey interactions, we asked: 1) do larger individuals eat larger prey and is this mediated by predator species identity? 2) is prey selection dependent on predator size, species, or both? We then compared the predator-prey body size relationship we observed in this dataset to the predator-prey body size scaling reported in a recent synthetic study (Brose et al. 2019), selecting data from the \_\_\_ terrestrial invertebrate food webs in this study to ask: 3) How does the predator-prey body size scaling relationship in empirical DNA-based diet data compare to the predator-prey body size scaling relationship across multiple studies in which a combination of empirical and literature methods were used to assign interactions? Using empirical interaction data to understand how body size and species identity shape terrestrial invertebrate food webs will be key to building predictive food web models built on these parameters (Gravel et al. 2015, Pompanez et al. 2018). In addition, this diet DNA metabarcoding dataset provides an opportunity to strengthen our understanding of predator-prey interactions for small-bodied organisms while providing a unique opportunity to examine these patterns at the level (individuals) at which predator-prey interactions occur, thus bridging the gap between species-level food webs and individual-level interactions (e.g. Nakazawa 2017, Stouffer 2010).

**Methods:**

*Field site and collections*

We conducted this work on Palmyra Atoll National Wildlife Refuge, Northern Line Islands (5º53’ N, 162º05’W). Palmyra Atoll has a well-characterized species list, and like many atolls, is relatively species poor, allowing for characterization of potential diet items (Handler et al., 2007). Predator individuals were collected across habitat types representing likely loosely-connected compartments in the atoll metaweb (CITE), including different forest types and microhabitats, including understory vegetation, canopy vegetation, and soil habitats. For each of these habitats, we used a combination of methods, including individual collection during visual surveys for understory and soil collections and canopy fogging with insecticide (cite) onto collection sheets for canopy individuals. All individuals were collected individually with sterilized implements (ethanol-burned forceps) in sterilized collection containers containing 95% EtOH to avoid contamination (Greenstone et al., 2011). All individuals were stored in 95% EtOH a -20ºC before DNA extraction.

We identified all predators to morphospecies in a laboratory environment and later validated unique species by DNA metabarcoding sequence data. The predator species used in this study represent the most common predator species found in each food web compartment (habitat location) and span a body size range of 0.2 – 928 mg (wet mass, SI Table 2). These predators included five arachnid species (*Oonopidae* sp., *Neoscona theisi*, *Heteropoda venatoria*, *Smeringopus pallidus*, and *Scytodes longipes*), one dragonfly (*Pantala flavescens*), one predatory katydid (*Phisis holdhausi*), one earwig (*Euborellia annulipes*), and one soil-dwelling centipede species (*Geophilomorpha* sp.). These predators employ various hunting tools, including webs, venom, and grasping forearms and employ several different hunting strategies, including web-building, sit-and-wait, and active hunting. (DO I need a table?)

*DNA extraction, PCR amplification, library preparation, sequencing, and denoising*

Our full DNA extraction, PCR amplification, library preparation, sequencing, and denoising methods can be found in the Supplementary Information, though we provide an abridged version here.

We individually measured the length of each predator (mm) and separated the thorax, opisthosoma, or trunk (depending on predator species, Krehenwinkel, Macias Herndandez) for DNA extraction following a modified CTAB extraction protocol (Fulton et al., 1995). We followed methods in Krehenwinkel et al. (2017) for extracting diet DNA from terrestrial invertebrate predators, using PCR primers (mlCOIintf/Fol-degen-rev; Krehenwinkel et al. 2017, Leray et al. 2013, Yu et al. 2012) targeting the CO1 gene, which is well-represented in online databases (Porter and Hajibabaei 2018). Following amplification of the CO1 gene, we attached Illumina index primers (Nextera XT Index Kit v2) following the standard protocol for these primers (Illumina). To achieve ample sequencing depth (SI Figure 2), we ran samples across four separate sequencing runs (SI Table 1). All individuals within a predator species were sequenced on the same run and each run contained one to five predator species. We ran 19 samples of one predator species (*H. venatoria*) across all runs to quantify run-to-run variation in sequencing (SI Figure 3). We submitted multiplexed samples for sequencing at the University of California, Santa Barbara Biological Nanostructures Laboratory Genetics Core. Samples were run on an Illumina MiSeq platform (v2 chemistry, 500 cycles, paired-end reads) with a 15% spike-in of PhiX. Following sequencing, samples were demultiplexed using Illumina’s bcl2fastq conversion software (v2.20) at the Core facility. We merged, filtered, and denoised our sequences around amplicon sequence variants (ASVs) using the DADA2 algorithm in R (dada2 package version 1.1.14.0; Callahan et al., 2016). We removed samples from analysis that had not been sequenced to sufficient depth using iNEXT (CITE) and a lower quantile cutoff (SI Figures 6 & 7). We rarefied remaining samples (McKnight et al., 2019) based on the sample with the lowest sequencing depth which had been sequenced with 95%+ sampling completeness based on iNEXT (version 2.0.20) interpolation and extrapolation methods (Hsieh & Chao, 2017). We rarefied using the rrarefy() function in the vegan (version 2.5.6) package in R to 15,954 reads per sample.

*ASV taxonomic assignment with BLAST and BOLD*

We performed taxonomic assignment using BLAST and the blastn command (version 2.7.1) and using sequences in the GenBank and BOLD sequence databases (GenBank accessed: November 20, 2019 and BOLD accessed: May 21, 2020). BLAST analyses were run using the computing cluster at UC Santa Barbara. We exported taxonomic assignments using MEGAN Community Edition (version 6.18.0, Huson et al., 2016), using default settings and selecting the subtree with all possible diet items for this species (Kingdom: Animalia, Clade: Bilateria). We chose to combine prey taxonomies at the family level, similar to diet resolution in both metabarcoding and histological methods in this field (e.g. Brose et al. 2019, Kartzinel – find other metabarcoding ones) by summing the cumulative read abundances across the ASVs that corresponded to each diet family in each sample. All DNA matching any predator family present on an individual sequencing run was also removed as a conservative method to account for potential sequence jumping within sequencing runs which could alter prey identity or diversity in favor of predator species on a shared run (CITE SEQUENCE JUMPING).

*Predator and prey size determination*

We converted predator lengths to wet mass using mass-length scaling relationships for each predator species using existing datasets from the literature and the field site (Soehlstrom et al, Su et al. Yaninek et al. 1993, Miller-ter Kuile *unpublished data*, McLaughlin et al. *unpublished data*). Prey masses were taken as the average mass for individuals across species within each family (Supplementary Information).

*Is prey size determined by predator size, species, or both?*

To determine whether individual predator size, species, or both predicted prey size, we used model selection based on AICc to compare a set of nested linear models with the response variable of prey size. The full model included the interaction between log10 predator mass (in mg) and species identity, with a random effect of predator individual. We log10 transformed the response variable (log10 prey mass in wet mg). To perform model selection, we used the dredge() function in the MuMIn package in R (Versions, citations) to compare all nested models and chose the model with the lowest AICc value. We verified model assumptions for best-fitting models using the DHARMa package in R (versions).

*Is prey identity determined by predator size, species, or both?*

To determine whether predator individuals partition prey items based on body size, species identity, or both, we ran a Canonical Correspondence Analysis (CCA) of 73 predator individuals of four species (*P. holdhausi, N. theisi, H. venatoria,* and *E. annulipes*) collected in the same compartment of the food web. We selected one compartment (canopy of the tree species *Pisonia grandis*) because we wanted to examine how individuals partition resources across body sizes and species given one shared potential prey pool. The CCA analysis included the matrix of feeding interactions for the food web compartment (canopy of *P. grandis*) and the variables of species identity and body size (log10 predator mass in mg) as predictors. We determined the likelihood of the model based on an ANOVA comparing the full CCA to a null model. We then extracted the marginal effects of predictor variables if the ANOVA showed a significant difference between the full CCA and a null model (p-value < 0.05).

*How does predator-prey body size scaling compare to literature scaling value?*

Finally, in order to compare our results to those from the literature, we extracted data from four invertebrate terrestrial food webs (compiled in Brose et al. 2019) on predator-prey body size relationships (not within species, so we could not examine the same species effects) to determine how the general scaling of predator-prey body size in our dataset compared to the relationship from previous food webs in which predator-prey interactions are built with a combination of literature documentation, phylogenetic relatedness, body size constraints, and some observation (Cattin Blandenier (2004), Digel et al. (2014), McLaughlin et al. (2010), Piechnik et al. (2008), Simberloff and Wilson (1969)). For this model, we created a full model of log10 prey mass predicted by the interaction between log10 predator mass and web type (DNA versus literature) with a random effect of food web identity, since data were drawn from a collection of food webs. Because our question was whether the scaling relationship was the same or different between our web and that from the literature webs (ANCOVA), we compared a model with and without the interaction between fixed effects (log10 predator mass and web type) to one without the interaction using AICc values to determine whether the scaling relationship we observed was similar to or different from that observed in these literature webs. We verified model assumptions for best-fitting models using the DHARMa package in R (versions). All figures were created in ggplot2 with color palettes from the calecopal package (CITE).

**Results**

*DNA extraction, PCR amplification, library preparation, sequencing, denoising, and ASV taxonomy*

Complete results and QC for each step of the DNA sequencing protocol can be found in the Supplementary Information.

Our final analyses were performed on a total of 182 predator individuals of nine species. Each predator had consumed 1 - \_\_ prey families. Thirty percent (n = 524 of 1,738 total ASVs) of the total ASVs found in samples received taxonomic assignments from GenBank and BOLD corresponding to prey items at the family level or lower and so were used in analyses. These corresponded to 57 prey families (SUPP table?).

*Is prey size determined by predator size, species, or both?*

The best model predicting prey size included the terms of predator mass and predator species identity, but not the interaction between these two terms. (log10(prey mass) = 0.41\*log10(predator mass), R2m = 0.30, R2c = 0.35, with significant variation in by-species intercepts, Figure 1, SI Table 3)

*Is prey identity determined by predator size, species, or both?*

The predictors of predator species identity and body size explained 15.1% of the total variance in prey composition for the 73 predators collected in the same food web compartment (*P. grandis* canopy; R2 = 0.15, adjusted R2 = 0.10). The best CCA model included both species identity and body size (p-value < 0.001), with a significant marginal effect of species identity (marginal p-value < 0.001) and a non-significant effect of predator size (marginal p-value = 0.07).

*How does predator-prey body size scaling compare to literature scaling value?*

They look ~the same but I need to run stats.

The relationship we observed follows a similar pattern as one from the literature, with larger individuals/species within this size range of invertebrate predators eating larger prey with an asymptote. However, our data suggest that there may be an overestimation of “larger” invertebrate predators eating prey larger than themselves and an underestimation of “smaller” invertebrate predators eating prey larger than themselves.

**Discussion**

Predator-prey size: both size and species

Prey identity: more predator species but some size

Scaling vs. literature: TBD

We found that individual predator size predicts the size of prey a predator consumes. However, predator species identity influences this relationship. Furthermore, predator species was a stronger predictor of prey identity, with a weaker influence of predator size. Finally, [what we find with literature highlighting the empirical vs. non-empirical differences]. These results highlight that while food web models predicting the structure of biological communities built on body size alone (both within and across species) explain much of the patterns in predator-prey interactions, these models can be refined based on predator species taxonomy (Brose et al. 2019, Laigle et al. 2018).

[predator-prey size relationships]

There is continued recognition in the field that food web patterns are likely the result of a combination of body size and species- or taxon-specific traits (Raffaieli 2007, Rudolf et al. 2014, Rall et al. 2011). Our results strengthen the need for combining body size with species identity for explaining and predicting food web patterns. For example, in our results, predators that may be more gape-limited (e.g. *Pantala flavescens* and *Geophilomorpha* sp.) have smaller prey items on average compared to predators of similar, or even, smaller size of other species that may not be as gape limited due to “tools” such as webs (e.g. Oonopidae sp). Conversely, the use of tools or evidence of scavenging meant that some predators fed on proportionally larger body size prey (*Euborellia annulipes*; CITE for DERMAPTERA; Wilson and Wolkovich 2011, Berg et al. 2012). While these data represent a subset (albeit of the most abundant predator species in this food web) of predator species, they suggest that approaching predator-prey size scaling with species functional traits beyond body size in mind may help to generalize while also describing more realistic patterns of species interactions (cite functional traits lit). Furthermore, these, and other datasets that describe feeding interactions at the individual level (e.g. Woodward and Hildrew 2002) for small-bodied invertebrates will link patterns that emerge at the food web level with the mechanisms that drive predator-prey interactions at the individual level (Stouffer 2005, Ings et al. 2009).

[Predator species niches for prey identity]

[Literature vs. empirical]

[downsides/pluses/next directions]

New benefits: within-species data demonstrates cross-species differences in scaling

Downsides: prey body size data aggregated at same level as in the past.

[some other stuff that could be put into the paragraph above]

[These data also revealed patterns of body size used to describe food webs in other ways, including predator-prey size ratios and interaction frequency, which can help infer interaction strengths (CITE). In particular, roughly 25% of the interactions in this dataset represent prey items that are the same size or larger than the predator (mostly in the spider species), going against common assumptions that most interactions occur between larger predators and smaller prey. These interactions suggest that this group of species (mostly spiders) may be able to target larger prey due to web-building (Laigle et al., Nakazawa et al. 2013) and suggest that thinking about restrictive rules for predators who are, themselves, on the small end of ecosystem size spectrums, may be a poor assumption. Furthermore, while we used these data to explore patterns in the niche model, which is based on binary interaction occurrence (presence-absence), because diet DNA metabarcoding data can be collected at the individual level, they provide an opportunity to quantify interaction strength (based on interaction frequency and the abundance of prey in the environment CITE) or stage structure (e.g. Rudolf and Lafferty 2011). As methods for diet DNA metabarcoding for these small predators becomes more established (e.g. advances in Krehenwinkel et al. 2017, 2019) and technology is becoming more affordable, these datasets may provide an effective way to quantify diet for all predators within an environment, either at the individual, stage, or population level.]

P: Concluding: not body size only, not species only, individuals acting on interactions at the individual level. A promising validation and refinement of what we already know: body size and species identity are important. Maybe worth considering

**Acknowledgements**

This project was funded by the National Science Foundation (DEB #1457371), National Geographic Society, and a Faculty Research Grant from the UC Santa Barbara Academic Senate. We would like to thank field technicians [list non-authors] and laboratory technicians Emily Lutz and Tessa Chou [others!]. We would like to thank the U.S. Fish and Wildlife Service and Palmyra Atoll Research Consortium for supporting field work for this project. We would like to thank Dr. Ryoko Oono for use of her laboratory space and equipment. We acknowledge the use of the Biological Nanostructures Laboratory within the California NanoSystems Institute, supported by the University of California (UC) Santa Barbara and the University of California Office of the President. We especially thank Dr. Jennifer Smith, manager of the Biological Nanostructures Laboratory for her assistance in preparing and troubleshooting our samples. We acknowledge the use of computational facilities at the Center for Scientific Computing (CSC), which was purchased with funds from the National Science Foundation (CNS-1725797) and is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara. We thank [EVERYONE WHO READS BUT NOT AUTHOR] D. Orr, H. Lowman, C. Jerde, M. Lee, and R. Ramiro for help in aspects of this manuscript, including statistics, bioinformatics, framing, and editing. We thank XX anonymous reviewers for help revising this manuscript. This is publication number PARC-XXX from the Palmyra Atoll Research Consortium.

**Literature Cited**

Will add – all that say “Name et al.” are in my Mendeley and easy to find. Anything with CITE I have in my Mendeley but I can’t remember authors off the top of my head.

**Figures**

****

Figure 1: Predator size distributions across the nine predator species. The x-axis scale depicts absolute values but has been log10 transformed. Predator individuals span from 2x10-1 to 9.3x102 mg in wet weight and the facets in this figure have been ordered by increasing predator species mean size.



Figure 2: Prey size distribution. Prey family average size spans between 3.8x10-4 to 3.1x 102.



Figure 3: Larger predators eat larger prey based on a log10-log10 transformed relationship (panel (a), log10 prey size = 0.41\*log10 predator size), though the effect is mediated by species identity (b). Continuous axis labels represent absolute values but the scale between them has been log10 transformed. In panel (b), “+” and “-“ icons indicate species that either have higher (“+”) or lower (“-“) prey sizes relative to predator body size and the general predator-prey body size patterns. The dashed line in panel (a) represents the 1:1 relationship between predator and prey size.



Figure 4: CCA bi-plot depicting the loadings on CCA1 and CCA2 driven by species identity for four predator species collected in the same food web compartment and so which share a common potential prey pool. Predator identity significantly contributes to prey composition (marginal p-value < 0.001) while predator size does not significantly contribute to prey identity/compostion (marginal p-value = 0.07).



Figure 5: comparison of our predator-prey size relationship to that from four food webs compiled from the literature. Our predator prey body size relationship had a significantly shallower slope than that of food webs in the literature (log10(prey mass) = 0.26\*log10(predator mass) vs. log10(prey mass) = 0.71\*log10(predator mass), p-value < 0.001).

MORE FIGURES:

SI Table: Prey families

57959504c8a8bcc91008178decf79d8af2639fb3