Running Head:

Title: Predator-prey body mass scaling is mediated by species identity revealed by individual predator-prey interactions from diet DNA metabarcoding

Author list: Ana Miller-ter Kuile, Austen Apigo, Bart DiFiore, An Bui, Daniel Preston, Elizabeth Forbes, Carina Motta, Michelle Lee, John McLaughlin, Kevin Lafferty, Hillary Young

To invite: Jasmine Childress, Katie Plummer, Rodolfo Dirzo, David Weber, Maggie Klope, Cora Johnston, Taylor Bogar, Magalay Espinoza

Lab?: Ronny Young, Chelsea Steele, Marisa Morse, Vahan Tertarian, who else???

Needs to be 20 MS pages from start to finish

**Abstract (200 words)**

**Keywords (6-12)**

**Introduction**

Baby: Food webs and general patterns based on species traits, particularly body size.

Werewolf: While building food webs based on predator-prey body size scaling are promising for many datasets in which interactions can be well observed (e.g. via gut dissections), datasets for small-bodied species have historically been based on inference rather than observation due to small body sizes and particular feeding habits of some invertebrate predators (e.g. spider feeding habits). Therefore, whether body size scaling allometries apply for these species or whether these allometries have represented relics of the body size rules used to infer interactions in the first place is imperative for the development of food webs in these systems and the generalization of food web patterns across systems (e.g. Brose citations). Furthermore, food web models across body size ranges often approach food webs as species-based, and so intraspecific variation and whether predator species identity (as a proxy for functional traits) mediates the body size scaling allometry are not often incorporated into these allometries. This may be particularly important for small-bodied predator species, which can range multiple orders of magnitude in size from juvenile to adult life stages (e.g. the predator *Heteropoda venatoria* in this current study ranges roughly three orders of magnitude from 1 - 1000 mg throughout its lifespan). Given recent evidence that predator identity or functional traits mediate cross-species body size scaling (Brose et al. 2019, Rudolf et al. 2014) and that predator-prey body size relationships are an integral part to a growing number of food web models (e.g. Gravel et al. 2013, Pomeranz et al. 2019), verifying predator-prey size allometries for small-bodied predators along with how species identity mediates these allometries is imperative.

Silver Bullet: In this study, we employ novel diet DNA metabarcoding data from \_\_ individuals comprising a set of nine 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 species identified in their diets, we examine the predator-prey body size relationships for these predators across and within species. Specifically, we ask: do larger individuals eat 1) larger prey and 2) a wider range of prey sizes, and is this mediated by predator species identity? Understanding whether individuals across species display body size scaling allometries will be key to building predictive models across systems for these species (give example). Furthermore, understanding whether larger individuals eat a larger range of prey sizes is imperative to parameterizing some of the most promising models of food webs that capture size based rules within and across species (e.g. extensions of the niche model: Gravel et al. 2013, Pomeranz et al. 2019; stage-structured food webs: Rudolf and Lafferty 2011). 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 on the atoll, including different forest types and microhabitats, including understory vegetation, canopy vegetation, and soil habitats (supplementary summary table of 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). We identified all predators to morphospecies in a laboratory environment and the predator species used in this study represent the most common predator species found in each habitat type and a representative distribution of body size distributions for each predator in each habitat. All individuals were stored in 95% EtOH a -20ºC before DNA extraction.

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

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). Following methods in Krehenwinkel et al. (2017), we standardized concentrations of 40uL of each sample to 20ng/ul and used Ampure XP (CITE COMPANY) beads to remove higher molecular weight predator DNA prior to PCR steps. We then amplified the CO1 gene, which is well-represented in online databases (Porter and Hajibabaei 2018) with general metazoan primers (mlCOIintF/Fol-degen-rev; Krehenwinkel et al. 2017, Leray et al. 2013, Yu et al. 2012). We ran total reaction volumes per sample of 25μL, with 9μL nuclease free water, 12.5μL GoTaq Green Master Mix (Promega Corp.), 1.25μL of each primer (at 10mM), and 1μL of DNA template (at 10ng/μL) and ran a duplicate for each sample. We followed a PCR protocol as followed: 3 minutes at 95ºC, 35 cycles of: 95ºC for 30 seconds, 46ºC for 30 seconds, 72ºC for one minute; ending with 72ºC for five minutes. We removed reaction dimer with Ampure XP beads (CITE company) at 0.8x bead-to-DNA ratio. We then attached Illumina index primers (Nextera XT Index Kit v2) with 5μL of PCR product per reaction and the recommended PCR protocol for these primers (CITE). We combined and cleaned successfully amplified duplicate samples using Ampure XP beads (0.7x beads-to-DNA) and diluted each sample to 5nM in 10mM TRIS, using 1uL of each sample for sequencing.

Because of the sample size and the need for a large number of sequences per predator in order to detect rarer prey DNA (Krehenwinkel, miller-ter Kuile, supplementary), we ran samples for this study across four separate sequencing runs. All individuals within a predator species were sequenced on the same run and each run contained one to five predator species (Supplementary TABLE). We ran 19 samples of one predator species (*H. venatoria*) across all runs to quantify run-to-run variation in sequencing (Supplement repeated measures ANOVA). For each run, we multiplexed all samples along with one negative control and two PCR4-TOPO TA vectors (Invitrogen, Carlsbad, CA, USA) containing the internal transcribed spacer 1 region from two fungal species as positive controls (GenBank accession numbers: MG840195 and MG840196; Apigo & Oono, 2018; Clark et al., 2016; Toju et al., 2012). 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. Our full protocol from DNA extraction through submission for Illumina sequencing can be found in Miller-ter Kuile et al..

We merged, filtered (max ee = 1.0), and denoised (clustered) our sequences around amplicon sequence variants (ASVs) using the DADA2 algorithm in R (dada2 package version 1.1.14.0; Callahan et al., 2016). Prior to denoising with DADA2, we used cutadapt (version 1.18, Martin, 2011) to remove primers from each sequence. We compared results to a similar protocol using the UNOISE3 algorithm (unoise3 function in unoise CITE EDGAR), but found that DADA2 gave more high-read abundance ASVs (Supplementary Figure). We ran DADA2 on sequences from all sequencing runs combined but verified that this was appropriate by first ensuring that error rates per run were similar, following recommendations from the algorithm developers (Appendix Figure CITE). We removed samples from analysis that had not been sequenced to sufficient depth using iNEXT (CITE) and a lower quantile cutoff (Supplement). 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*

From the output of the DADA2 algorithm, we created a list of unique ASVs which we matched to taxonomies both in the GenBank and BOLD databases. For GenBank, we used BLAST (version 2.7.1) with the blastn command for taxonomic assignment of each ASV using the computing cluster at UC Santa Barbara, comparing against the GenBank nucleotide database with an evalue of 0.01 (downloaded on November 20, 2019). We visualized and exported taxonomic alignment 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). For BOLD taxonomic assignment, we used the BOLD IDEngine of the CO1 gene with Species Level Barcode Records (accessed May 21, 2020; 4,070,029 Sequences, 225,114 Species, and 104,607 Interim Species in database) to match each ASV list to taxonomies. We combined taxonomic assignments from both programs and discarded taxonomic assignments that were mismatched at the family level or higher (Elbrecht et al., 2017). Most (\_\_%) ASVs received a family-level taxonomic assignment, so 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 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*). We determined a by-species scaling relationship by fitting a mixed effect model predicting the log10 transformed predator mass from log10 transformed predator length with a random slope term for log10 predator length and a random intercept by predator species (e.g. letting the mass-length relationship vary by species; Supplementary Information). We then used these scaling relationships to predict each predator individual mass based on length measurements. Prey masses were taken as the average mass for individuals across species within each family. We determined prey mass range in each predator individual by determining the range of largest to smallest prey items for predator individuals for which we documented more than one prey item n=\_\_/222).

*Predator-prey body size scaling*

To determine whether individual predator size, species, or both predicted 1) prey size and 2) prey size range, we used model selection based on AICc to compare a set of nested linear models with each response variable (prey mass and prey mass range). The full model for prey size included the interaction between log10 predator mass (in mg) and species identity, with a random effect of predator individual. The full model for prey size range included the interaction between log10 predator mass (in mg) and species identity. We log10 transformed each response variable for each model (log10 prey mass and log10 prey mass range, both in 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).

**Results**

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

Of a total of 280 samples, we successfully extracted DNA from 99% of samples (n = 278 of 280). Amplification success across all samples was 80%, with 222 of 278 initially extracted samples successfully amplified and sequenced. The Illumina MiSeq runs yielded 3.9, 3.8, 3.8, and 3.3 x 107 unpaired reads that passed filtering and had Q30 quality scores of 87.82, 85.46, 79.34, and 78.02%, respectively. After quality filtering and denoising with DADA2, 3.02 x 107 paired-end reads corresponded to 1,738 ASVs. Following filtering and clustering, we determined that 40 samples had too-low sequencing depths (less than 11,211 reads total) and so they were removed from further analyses.

*ASV taxonomic assignment with BLAST and BOLD*

Sixty-seven percent (1,167 of 1,738) of ASVs matched to a taxonomic assignment at Class level or lower. Thirty percent (n = 524) of these taxonomic assignments corresponded to prey items at the family level or lower and so were used in analyses. There were two conflicting taxonomic assignments at the family level or higher between the BOLD and BLAST assignments which were removed from the final total above.

*Quantifying cross-run variation*

There was significant cross-run variation, with significant differences (pair-wise differences between runs with p-value ≤ 0.05) between run 1-4, 2-3, and 2-4. On average, samples had: A: 2.26 ± 0.15, B: 2.33 ± 0.24, C: 1.72 ± 0.23, and D: 1.44 ± 0.15 species in each sample. Because each species was run on a sequencing run with all other individuals from that species and because we did not compare species richness as a response variable across predator species in this study, we report this as the variation across sequencing runs but do not correct for it in future analyses (Supplementary Information).

*Predator and prey size determination*

To build predator mass-length relationships, we compiled data from 2,362 individual predators (13 – 1652 per predator species). These data had strong species-specific mass-length scaling relationships (R2m = 0.62, R2c = 0.95) which were used to predict individual predator masses from lengths. We compiled data from 3,458 individual prey (1 – 1721 individuals per family), with 59% (2,028 out of 3,458) coming from individuals from Palmyra Atoll.

*Predator-prey body size scaling*

**Discussion**

Do larger individuals eat larger prey, and is this mediated by predator species?

Larger individuals eat larger prey with a power relationship.

This is mediated by species, suggesting that some life history traits dispose predators to being able to target different sizes of prey (e.g. gape limitation vs. scavenging, vs. tools such as webs)

Do larger individuals eat a larger range of prey, and, again, is this mediated by species? This is the foundation of the niche food web model – larger preds eat larger range of prey sizes, however, could be mediated by optimal foraging and or species traits (eg new citations added to Mendeley Dec 11) such that a prescriptive size-based size range approach is not sufficient to capture this variation.

Find that, again, larger individuals do eat larger prey with a power relationship AND again, that species identity is really important here, and this could be related to foraging traits (e.g. gape limitation etc).

**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 \_\_\_\_\_\_\_ 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**

**Figures**