Review - Miller-ter Kuile et al Miller-ter Kuile and others have conducted a worthwhile study on the impact of surface sterilization on prey item detection in spiders from the natural environment as well as from mesocosms. They have done an excellent job providing detailed, well-documented laboratory and analytical methods. The supplementary material is a great asset to the paper.

There are two main issues with the manuscript. First, I found the written section about the sterilization procedure(s) confusing. Perhaps I have misinterpreted, but it seems as though bleach is only used as a sterilization method for the spiders from the mesocosm experiment. I would recommend re-writing that section to clarify which sterilization methods were used for which spiders.

*We agree that this section was confusing as previously written. We have attempted to clarify this aspect of the study. Specifically, all consumers were sterilized using the same bleach-wash method, however, the laboratory environment in which this occurred differed. For feeding trial consumers, we performed this process immediately after they had been killed because we were going to be storing them in 95% EtOH. For natural environment consumers, because we had been able to keep these frozen at -80ºC, we were able to perform this process in the sterilized environment of a laminar flow hood in a university laboratory setting. We have attempted to clarify this in the abstract (lines 25-28), in the introduction (lines 120-122), and especially in the methods section (lines 218-247) highlighting this methodology.*

Most importantly, the biggest issue concerns the interpretation of the mesocosm results. There is no such thing as “marginally significant,” and suggesting that the results are marginally significant (p = 0.07) is a misinterpretation of the results. A substantial re-write will be required to address this misinterpretation in the abstract, results, and the discussion. I have discussed this issue in more detail below (under lines 409-410)

*Thank you for pointing this misinterpretation out to us. We have chosen to re-phrase these model results in light of your suggestion, reporting the comparison between this model and the null model in the abstract and highlighting the term’s lack of significance (lines 30-33). We have also re-written the results section (lines 537-542) and the discussion (lines 625-652) to contextualize this result more accurately.*

Please find more detailed comments below:

Abstract lines 23-25 Were the spiders from natural environments sterilized in bleach or via a laminar hood (presumably via UV)? This needs to be clarified in the methods and, if necessary, clarified in the abstract.

*We have clarified this as part of the comment above but see lines 25-28 of the abstract for clarified phrasing.*

Lines 27-28 There is no such thing as “marginally significant” (p = 0.07). A more detailed discussion of this point follows below (lines 409-410).

*We have re-framed this finding here, as well as throughout (see previous comment and lines specified).*

Introduction line 52 “form the base of most food webs” Please clarify \*terrestrial\* food webs

*We have added “terrestrial” to this sentence (line 65)*

line 66 Please provide citations.

*We have added citations (lines 79-80)*

Lines 93-101 This argument about the potential of contamination to “inflate” or “depress” estimates of consumption is a bit confusing, For instance, if you are considering alpha diversity, contaminants that are not potential diet actually inflate estimates of consumption. Please state in other terms for clarity.

*We removed the entire paragraph for clarity and brevity (suggestion of another reviewer) and do not believe that it altered the interpretation of our study.*

Line 104 To clarify beyond “used in analyses” perhaps it would be better to write “used for diet DNA metabarcoding”

*Changed to your suggestion, “used for diet DNA metabarcoding (lines 117-118).*

Lines 104-105 It is awkward to state “high throughput sequencing results of the COI gene…” Rather, you could state “Targeting the COI gene region, we produced high throughput sequencing results….” or “We performed COI DNA metabarcoding with the full….”

*We changed the order of this sentence per your suggestion (line 118)*

Line 107 No comma necessary

*Removed comma*

Line 114 By abundance, do you mean relative abundance?

*Abundance is measured as a rarefied measure (i.e. we standardized the sequencing depth across all samples and then used this abundance). We have changed all instances of “abundance” to “rarefied abundance” and hope this clarifies this point (e.g. line129)*

Lines 115-118 To reduce verbosity and over-interpretation in the intro, remove “suggesting that contaminants either hide or inflate diet consumption amount” and “suggesting that surface contamination could alter ecological interpretations of community-scale species interactions”

*We have taken your suggestion to change this wording (lines 129-148).*

Line 130 For the non-arachnologists, what kind of prey do these spiders typically target? Based on the Handler et al (2007) reference, do they mostly consume arthropods? It is probably worth explicitly mentioning what they typically ingest.

*We have inserted more natural history in this section for the non-arachnologists. Specifically, we have included a list of potential diet items as well as feeding behaviors for this consumer (lines 154-162).*

Line 131-134 It seems out of order to explain the collections in 2017 before the collections in 2015.

*We had originally chosen to present the study this way to provide methods for the mesocosm (now “feeding trial”) study prior to the field study. However, we have switched the order of these two parts of the study throughout and this section now reads chronologically. (Lines 165-169).*

Lines 142-143 According to Macias-Hernandez (Figure S1), in the opithosoma of the woodlouse hunter spider, woodlice could be detected for over 100 hours after feeding. At 50 hours after feeding, there was a nearly 100% positive detection rate. In comparison, waiting only 12 hours in the present study to allow for the digestion of previously consumed prey items seems like a very short time period. What is the reasoning behind waiting only 12 hours in light of the cited reference?

*Thank you for pointing out this discrepancy between our methods and findings from this study. This study (Macias-Hernandez et al.) was not the initial methodological justification for this approach. We had a limited time in which to complete this project at the field site in 2017 (~2 weeks) and so chose to provide this 12-hour period as an adjustment period given the time constraints of running trials with limited time to complete them. We have changed the wording of this section so that we do not justify it based on this citation (Lines 204-215).*

Lines 169-171 2 minutes is a very long time to submerge an small-bodied organism in 10% bleach. It is my understanding that surface sterilization with bleach is generally effective within a much shorter time period. I wonder if the bleach can seep into the exoskeleton of the spider and damage the internal DNA after 2 minutes? Can you speak to the permeability of an exoskeleton?

*We have addressed this comment in our study by providing two studies which washed similarly-permeable invertebrate consumers with bleach for equal or longer times and with equal or greater concentrations of bleach with no evidence of substantial degradation (lines 229-232). We agree that the time and concentration of bleach should be a consideration in future studies.*

Lines 155-182 The whole surface sterilization section is confusing to follow – it would be helpful to clarify when and what was done to each set of samples (natural vs mesocosm). Were the natural environment consumers sterilized in bleach before being frozen at -80C? It sounds like they were immediately frozen (lines 157-159), then they were sterilized in a sterilized laminar flow hood before extraction? If I am not misinterpreting, then it sounds like only the mesocosm spiders were surface sterilized in bleach, while the natural spiders were sterilized in a laminar flow (UV)? Using different methods of sterilization could pose significant problems for comparability, although this isn’t necessarily the point of the study, so it isn’t a fatal flaw. From the abstract, I was under the impression that bleach surface sterilization was the only method of sterilization used in this study. I’m hoping this is a misinterpretation on my part…

*We have attempted to clarify this section per your clarifying question (lines 218-247).*

Line 189 “….lower molecular weight consumer or diet DNA…” Is the “or” a typo?

*Rephrased to “to isolate a proportion of lower molecular weight DNA prior to PCR steps…” to increase clarity. (lines 254-255).*

Lines 190 “DNA prior to PCR steps with Ampure XP beads” It sounds like Ampure XP beads are part of the PCR steps. For clarity, please re-word as “…DNA with Ampure XP beads prior to PCR”

*We have changed this wording as per your suggestion (line 255).*

PCR amplification, library preparation, and sequencing + bioinformatics sections Great - nicely detailed, easy to read.

*Thank you!*

Lines 291-294 (similar issue to lines 93-101) This concept is confusing: “…increasing detection because of “false” diet detection or by decreasing detection because of abundance of non-diet DNA…” Although I understand the authors’ intentions, this could be misleading since you could equally say that including non-diet DNA items equals increased detection, depending on how “detection” is defined. This can either be reworded or “detection” can be more clearly defined.

*We have removed this phrasing because we agree that it is confusing (current lines 360-361).*

Line 311 Should use “that” instead of “which”

*We have changed to “that” instead of ”which”.*

Line 317 Should use “that” instead of “which”

*We have changed to “that” instead of ”which”.*

Lines 319-323 The GLM descriptions were incomplete here, but then I saw that much more detail is provided later in the methods. I would recommend putting all the GLM information in the same section (probably the latter section) to prevent confusion.

*We agree that having these descriptions in all sections is confusing. We have moved all of these descriptions to a single section (Statistical Analyses, lines 455-488).*

Lines 327-330 “This is especially important…” This does not belong in the methods. This could be relevant information to include in the intro. However, it’s unclear what the other ~92% of DNA represents in that study: host tissue, likely contamination, or other? I would think that host tissue, not contamination, would be responsible for swamping a sequencing run to that extent, which would make this argument obsolete in the context of this paper.

*We agree that this is unnecessary here, so we have removed the sentence at now line 404-408.*

Lines 397-399 Perhaps I have overlooked the definition of “potential diet items,” but if 23% of the taxonomicallyassigned ASVs correspond to potential diet items and 8% of them correspond to consumer DNA, what do the remaining 69% represent? Could there possibly be more diet items unknown to this spider before this study? Or is all of this DNA considered secondary predation (prey items of the prey items)?

*We have included more description of ASV taxonomic assignments in this section to clarify this point, as we agree that reporting these percentages hides the biological diversity and interpretation of these ASVs. (lines 517-520). We also report the percentages of each type of ASV to highlight that although non-diet ASVs were diverse, they were a minority of the total sequencing abundance compared to consumer and potential diet DNA.*

Lines 409-410 There is no such thing as “marginally significant.” If you choose to use frequentist stats, which is perfectly acceptable, then you must also accept the significance level of 0.05. Frequentist statistics is based on the idea that the null hypothesis will be rejected 5% of the time even if it’s true. Otherwise, you are interpreting your results with post-hoc bias. Naming p-values between 0.06 and 0.10 as “marginally significant” is a misinterpretation of the data, and it needs to be corrected throughout the manuscript. This includes the language surrounding the results as well as the interpretation of the data in the discussion and abstract. It’s okay to have non-significant results. I also wonder whether the long bleaching period led to the decreased detection of offered prey items. How permeable is the exoskeleton of a spider?

*We have updated this section to clarify the interpretation of these results (see response to your comment above).*

Line 422 Should use “that” instead of “which”

*Changed from “which” to “that”.*

Lines 438-447 This is a misinterpretation of the results. This section needs to be re-written to account for the results: surface sterilization before metabarcoding does not change our perception of diet in natural environments or mesocosms. It may also be worthwhile mentioning that two different sterilization techniques were used (if that is indeed the case), so it is difficult to directly compare the natural vs mesocosm results.

*See comments above, we have clarified the interpretation of this section and spent more time discussing the interpretation of the model in question (lines 625-657).*

Lines 471-474 This needs to be tweaked. It is okay to discuss why mesocosms may be more prone to contamination; however, the language should not say that the study provides “some evidence” that mesocosms are more prone to contamination.

*As per above comments, we have removed this misinterpretation of the model results.*

Line 489 Should be “further” not “farther” because it is an expression of time, not distance

*Changed to “further” from “farther”.*

Lines 500-509 This recommendation cannot be made in light of non-significant results.

*We have updated to remove this recommendation. (included in edits from comments above).*

Supplementary Information – Thank you for providing such thorough, detailed information.

Comments to the Author  
Miller-ter Kuile and colleagues present a study on the effects of body surface sterilization on the detection of prey DNA in spiders. Surface sterilization may be desirable in molecular gut content analyses when a consumer has come into external contact with potential diet taxa; however, not much is yet definitively known about the best methods for sterilization, their effectiveness or their potential side effects.  
  
This manuscript speaks to a timely and important issue: external contamination could cause major problems for data interpretation in gut content studies. However, I worry that the results are not sufficient to address the central question of the study, and I disagree with many of the authors’ decisions regarding data collection and analysis. At a minimum I recommend extensive reanalysis of the sequence data; but better yet would be a repeat of the feeding experiment with higher sample sizes and visual ascertainment of whether the spiders actually feed.  
  
I think this project has potential to make an important contribution to the field of diet metabarcoding, but substantial work is needed before that can happen. I encourage the authors to build upon what they have done so far to make it a truly robust and reliable study.  
  
My comments on specific aspects of the study, as well as the manuscript itself, are detailed below.  
  
  
-Laboratory experiment:  
  
The purpose of this experiment was to see how sterilization affects detection of prey consumption. However, if the authors do not actually know which spiders have eaten and which ones have not (lines 149-150), then how can they reliably test for the effects of surface sterilization on detection of prey in the spiders’ guts? What is known about feeding habits of H. venatoria – do they macerate their prey, or simply suck out the digested tissues? It seems like there should be a way to tell from looking at the grasshopper remains. Otherwise, I think this experiment should be repeated and someone should observe the spiders to see if/when they feed.  
  
I am also concerned about the sample size, which is insufficient to draw reliable conclusions from the data (n=8 for sterilized, n=11 for unsterilized spiders).

*We agree that there are limitations to the methods used to conduct our feeding trial study. Many of these limitations are due to the remote nature of the field work and the limited amount of time available for researchers to spend at the field site to conduct multiple concurrent field and laboratory studies. At this time, we do not have funding to either return to the field site or repeat laboratory procedures (an additional funding source of $30,000+ would be needed for this combined effort), nor do we have laboratory access to conduct further molecular work due to the COVID-19 pandemic.*

*H. venatoria likely suck out digested tissue from prey items, thus, visually determining ingestion is challenging even when closely observed. We do not believe that repeating this trial would aid in detection given that ingestion may often be partial. Indeed, it may be particularly important to assess surface contamination when feeding is difficult to detect (when feeding rates may be low and thus more easily confused with surface contamination results). Certainly, even imperfect information confirming when ingestion occurs would be helpful in interpreting these results, and we highlight that future studies, especially in taxa where feeding is more obvious, should include such a design.*

*We also agree that the sample size for this study is low due to the limitations we have highlighted. We have performed a power analysis with the values from our GLM model, which suggest that a sample size of n = 36 would be sufficient for a follow-up study (at a power of 0.8). We have also performed a simulation-based sensitivity analysis which suggests that at our current sample size, 86% of the time, our p-value would be above an α = 0.05 threshold based on 1000 simulated draws of our presence-absence detection model, suggesting that the outcome we interpret from our dataset (no effect of surface sterilization) is more common than the alternative hypothesis. However, we agree that this is a trial of a method that needs to be repeated to draw complete conclusions across consumers and contexts. Simulated p-value distribution:*

**

*We address these two concerns (feeding trial methods and sample size) in an updated portion of the discussion (lines 625-657) highlighting that these results suggest next directions and a consideration of several alternative hypotheses. Furthermore, we have chosen to switch the order in which data are presented in the manuscript, specifically presenting the field results first followed by the feeding trial results to highlight our more robust findings prior to the feeding trial findings.*  
One other potential issue is that these spiders were stored in 80% ethanol without refrigeration (lines 172-173). At 80% at room temperature, the DNA of both the spider and its prey may degrade, especially when the spider is large and the water inside its body further dilutes the ethanol. The size selection method of Krehenwinkel et al. (2017) works best when the predator’s own DNA has not had a chance to degrade – e.g. when kept frozen and/or in 100% ETOH – otherwise some degraded predator DNA may be enriched along with the prey DNA. This could artificially depress the apparent prey abundance or prey detection rate in the spiders’ guts.

*We have updated this concentration based on an error we made in reporting these concentrations in the previous draft. The concentration was 95% EtOH (the highest concentration available at the field site). Furthermore, although no deep-freezing options were available at the time in the field site, we were able to keep samples stored at -20ºC until DNA extraction. We have updated these methods in lines 238-242. We agree that the use of sterilization fluids as opposed to freezing samples is worth considering, however, for potential users of diet DNA metabarcoding in similar remote field conditions, preservation fluids without freezer options may be a real constraint. We highlight that our data are promising to users in these contexts (e.g. many real-world field research projects) despite the limitation of freezer storage (lines ).*   
  
  
-Field experiment:  
  
Sample size is reasonable and experimental methods seem sound. However, I do not agree with the decision to collapse prey sequences to family level. Doing so probably obscures important variation, as a single family could represent any number of species. I therefore do not trust results addressing diversity of recovered prey DNA at such a coarse level. There could in fact be a significant difference between sterilized and unsterilized individuals that simply is not detectable at the family level.  
  
Why not forget about prey taxonomy and simply work with OTUs clustered by sequence similarity? Or even work at the ASV level (though this will probably inflate diversity estimates)?

*We agree that there are limitations to analyzing at the family level, which might have effects on results. We have chosen to collapse prey diversity at the family level because this makes these data comparable to other studies in these types of systems (e.g. Brose et al. 2019 citation in paper). We agree that this may hide patterns at the ASV level, so we have updated our methods to examine both ASV-level and family-level richness and have found the results to be comparable. We present both to maximize depth and comparability across studies (lines 481-487, 563-573, Figure 3).*  
  
A bit more detail is needed on how the authors processed sequence data. How did the authors deal with NUMTs? How exactly did they determine whether an ASV matched H. venatoria or a prey item (was there a minimum BLAST % match)?

*We have now included details throughout the manuscript that we hope will clarify this important point. Specifically, we assigned H. venatoria as any sequence that matched the family Sparassidae, as this is the only species in that family on the atoll (lines 162-163, lines 382). We have also reported more of the parameters from MEGAN highlighting the coverage of taxonomic assignments (lines 345-346, lines 523-527) to clarify how ASVs were assigned through the program. We also report the percentage rarefied reads of ASVs in each category (consumer, potential diet, non-diet lines 519-520). Specific to NUMTs, because MEGAN assigned family-level taxonomic* *assignments with 100% coverage (lines 523-527) we take this as evidence for a low representation of NUMTs in this dataset. Additionally, we consider combining taxonomic units at the family level also provides a conservative richness estimate that is less susceptible to the over-estimation of diversity that arises when NUMTs are present.*   
  
-Molecular work:  
  
The reported success rate for PCR amplification is low at 78% (lines 391-392). Did the authors attempt to optimize PCR or extractions for those that did not work the first time? PCR inhibitors can be a big problem in spiders, especially in larger individuals. It may be worth rerunning PCR using diluted DNA template, and if this does not work, then the authors could try performing a repurification of the DNA extracts.

*We optimized our PCR and extraction conditions compared to those reported in the study used to derive these results (Krehenwinkel et al. 2017). Specifically, we optimized the PCR conditions (varied number of cycles, multiple annealing temperatures, varied ratios of Ampure XP beads for all bead cleaning steps), over the period of several months as well as adjusted the concentrations of DNA template used in PCRs and tested PCRs with two different master mixes. We did re-run samples that did not work through PCR two times. We agree that this success rate seems to suggest PCR inhibitors common to spiders. We did attempt to repurify DNA extracts for ~8 samples during this optimization period using a Zymo PCR inhibitor kit and found PCR amplification to be similar across these. We have provided our full DNA extraction and PCR amplification protocol as a supplemental document to this manuscript in an effort for these results to be fully reproduced or altered by future researchers. We have used this set of primers and this protocol (without sterilization procedure) across multiple other consumer species (n = 10, some other spiders, other insect and arachnid predators) and have found that amplification varies from 75% - 100% per species. Overall this primer set has a high amplification success rate of 97% (± 1.0% SD) across n = 262 additional samples using this protocol.*

Alternatively, the MCO/Fol-degen-rev fragment may be too long for some of the most heavily digested/degraded prey DNA. The authors could try the primers from Zeale et al. 2011 (Mol Ecol Resour), which amplify a fragment of just 211 bp. Another option would be the primers of Krehenwinkel et al. 2019 (Methods in Ecol & Evol), which also amplify short fragments, but furthermore suppress amplification of spiders and therefore give much higher yields of prey DNA.

*We agree that considering multiple molecular approaches is a key point worth further consideration in this type of study. We chose to use the longer fragment length from Krehenwinkel et al. 2017 because it was explicitly a broader metazoan primer set with higher amplification success and which detected the most diet orders in that methods study. The consumer in our study is known to feed on insects, spiders, and vertebrates on the island (two gecko species in the genus Lepidodactylus) (line 160). Hence, we chose not to use the arthropod primers from Zeale et al. 2011 or others because of the broad diet of the consumer in this study and the evidence of success for broad diet with the chosen primer set. This method from Krehenwinkel et al. 2019 was not available when we started this study but seems very promising as a next step for this type of work. Again, because of the COVID-19 pandemic and because of the cost-prohibitive nature of the field and laboratory work to re-do these procedures, we are unable to repeat or alter these methods at this time. We attempt to address some of these important points in the current draft of the discussion and hope this is a fruitful area of future research in the field (lines 625-657).*

-Writing:  
  
The manuscript generally reads smoothly and has apparently been well proofread for grammar and spelling. However, there is too much text and many sections are repetitive. Removal of repetitive text, e.g. in lines 162-173, 217-232, 298-307 and others, would probably bring the manuscript well within the word limit. Some details could be removed from the descriptions of molecular and statistical methods.

*We have removed a section of the introduction as well as several sections of the discussion that we agree were redundant (lines 98-114 and 659-677). We have replaced the discussion sections with the important caveats, findings, and next directions pointed out by your comments above (lines 468-492).*   
  
There are some serious issues with word choice that need to be corrected in future versions:  
  
“mesocosm”: This does not seem like the appropriate term for the laboratory housing of the spiders. Based on the authors’ descriptions – essentially just empty containers – I would call these “enclosures” or “containers.”

*We have chosen to change this wording to “feeding trial” throughout the manuscript and agree mesocosm is misleading in this instance.*   
  
“concatenate”: To concatenate is to string together multiple elements into a series. This is not at all the same as collapsing ASV data to the family level (e.g. line 316).

*We have changed this word to “combined”*

“taxonomy”: The authors use “taxonomy”/”taxonomies” when they mean “taxonomic level” (e.g. lines 311, 314, 317) or “taxonomic identity” (line 397).

*We agree that this is confusing and have attempted to clarify when we mean to refer to “taxonomic level” versus what we are now calling “family-level taxonomic units”.*   
  
“population consumption frequency”: This is not a well established term as far as I know. The authors should not use it in the Abstract. If they wish to use it elsewhere in the manuscript, then they should briefly define it after the first usage.

*We have removed this term.*  
  
The authors often neglect to include words like “perceived,” “apparent,” “detected,” etc. and thus imply that their experimental protocols are influencing the \*actual\* diet of the spiders rather than the results obtained from metabarcoding. Some examples are in lines 23 (“We compared diet from individuals...”) and 114-117. Future versions should be careful to make this distinction clear.

*We have included “potential” or anther such qualifier at every instance of “diet” to point out this important distinction that this DNA is perceived diet as opposed to observed.*