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