**Title**

**Effects of surface sterilization on diet DNA metabarcoding data of invertebrate consumers in mesocosms and natural environments**

**Authors:** Ana Miller-ter Kuile, Austen Apigo, Hillary Young

**Abstract**

DNA metabarcoding is an emerging tool to quantify diet in environments and consumer groups where traditional approaches are unviable, including in small body size taxa. However, metabarcoding for small taxa often requires DNA extraction from full body parts, and it is unclear if surface contamination from the body parts alters diet metrics. Here we examine different measures of diet (consumption frequency, abundance, diversity, and composition) in both mesocosms (where aconsumers offered a known diet item) and natural systems for a group of small invertebrate consumers (spider *Heteropoda venatoria*) using DNA metabarcoding of full body parts (opisthosomas). We compared individuals surface sterilized in 10% commercial bleach solution followed by deionized water with a matched set of unsterilized individuals. We found that surface sterilization did not significantly alter any measure of diet (consumption, abundance, diversity, or composition) for consumers collected in either environment. However, in a mesocosm environment, there was a large (reduction from 91% detection to 50%). But marginally significant reduction (p-value = 0.07) in consumption detected in surface sterilized consumers, suggesting that without surface sterilization consumption frequency may be inflated due to contaminant DNA from sharing space with an offered diet item. While surface sterilization thus does not seem to be a critical need for DNA metabarcoding of full body parts of invertebrate consumers in natural systems, it is probably advisable in contexts where environmental conditions increase the likelihood of surface contamination.

**Keywords**

consumptive interactions, arthropods, contamination, food web, predator prey interactions

**Introduction**

Biological communities and ecosystem function are shaped by the interactions between organisms (Brown et al. 2001, Hooper et al. 2005, Delmas et al. 2019, Schleuning et al. 2015). Among the many interaction types, consumptive interactions (including herbivory, predation, and parasitism) can shape the stability of biologically diverse communities (Ings et al. 2009, Delmas et al. 2019). Until recently, these consumptive interactions were most often measured by visual observations of feeding or by gut dissection or inspection of fecal contents (Hylsop 1980, Nielson et al. 2018, Baker et al. 2014, Duffy and Jackson 1986), which made it challenging or impossible to conduct diet analyses for many consumer groups.

Specifically, accurate diet analyses was typically not possible for consumers that are too small for dissection and food identification, have feeding habits which render food items unidentifiable, have food items that contain few components which pass through digestion in recognizable form, or have cryptic habits that prevent them from being observed in large enough numbers (Sheppard and Harwood 2005). This composes an enormous part of most foodwebs… SOME INFO HERE> For these consumer groups the use of high-throughput sequencing methods for determining gut contents is one of the most promising emerging approaches. High-throughput sequencing (hereafter referred to as “diet DNA metabarcoding”) can identify a suite of diet species at once and provides a comprehensive and efficient method for determining intraspecific, intra-population, and interspecific diets (Pompanon et al. 2012, Lucas et al. 2018, Soininen et al. 2015, Quemere et al. 2013). These methods have already illuminated new interactions and ecological trends in a variety of environments (e.g. host-parasitoid: Wirta et al. 2014; plant-herbivore: Kartzinel; host-parasite: Schnell et al. 2012, predator-prey: Toju et al. 2018).

As diet DNA metabarcoding methods have continued to advance, however, they need to be validated so that the ecological inference made from them is robust. Focusing on the challenges of small organisms where small body size has limited other diet analysis methods), DNA diet analyses has gained traction as an approach to survey diet but little methods validation has been conducted. In contrast to large animals where gut contents can generally be isolated prior to analyses, in these small animals diet data is often extracted from full organisms or body parts without dissection, and so the possibility of surface contamination altering the detection and species composition of presumed diet items is a major concern, wth the potential to dramatically alter results. Surface sterilization is systematically used in other fields to reduce the risk of similar contamination in other DNA metabarcoding contexts (e.g. fungal endophyte research; Zimmerman and Vitousek 2012, Burgdorf et al. 2014), but has not been systematically used in diet metabarcoding studies. While some fields have developed informed protocols based on decades of research into best practices and study-specific considerations (Brown et al. 2018, Hallman et al. 1997), Unfortunately, ue to its relative infancy, the field of diet-based DNA metabarcoding has not developed a systematic approach to surface sterilization with many studies not sterilization, and others using a range of approaches (e.g. ethanol washes in Doña et al. 2019, bleach washes in Anslan et al. 2016, no sterilization in Wirta et al. or Jacobsen et al. 2018). The lack of systemic surface sterilization in diet metabarcoding when using full individuals may be due to the desire to avoid destruction of DNA in relatively permeable animal cells (Greenstone et al. 2012). A crirtical first step in evaluating the utility of surface sterilization is to test if it changes results of dietary analyses.

Importantly he value of surface sterilization in diet DNA metabarcoding may vary strongly by context. There are environmental, ecological, and methodological factors that may contribute to the value of surface sterilizion of consumers prior to DNA metabarcoding for diet analsyes. These factors are related to the risk of contamination by environmental DNA, and could be regulated by abiotic or biotic conditions that alter DNA persistence (Strickler et al. 2015, Barnes et al. 2014, Neilson et al. 2007, Carini et al. 2016, Collins et al. 2018), DNA abundance (Elbrecht et al. ), or the likelihood that consumers come in contact with diet items in the environment (Greenstone et al.). In any environment, the ecological interpretation of diet data with surface contamination could alter the interpretation of these data regardless of the diet measure in question (e.g. frequency of consumption versus diversity of interactions). If diet data are being used to ask questions about the species-, population-, or community-level rate of consumption of a diet item or items (Kaunisto et al. 2020), then surface contaminants could either inflate (i.e. contaminants are potential diet) or depress (i.e. contaminants are not potential diet) estimates of consumption (Greenstone et al. 2011/12). Similarly, if diet data are being used to ask questions about a range of interactions between groups of organisms (e.g., Kartzinel et al.), then surface contaminants could introduce false diet diversity or hide real diet diversity, especially if this diet diversity is low abundance because of low biomass or infrequent consumptive interactions (Macias-Hernandez et al, Elbrecht et al. 2017).

In this study, we look at the effects of surface sterilization on our understanding of consumer diets where the DNA of full body parts (no dissection) is used in analyses. We use high throughput sequencing results of the CO1 gene from the full body parts (opisthosomas) of invertebrate consumers (the spider, *Heteropoda venatoria*) from two environments – a “contained” mesocosm environment in which we offered consumers a potential diet item, and a natural environment in which consumers could feed on naturally-occurring diet items. In each environment, we surface sterilized half of the consumers prior to DNA extraction using a series of washes in a 1:10 dilution of bleach (10% commercial bleach) and deionized water; we left the other half of consumers unsterilized. Specifically, we ask: Does surface sterilization alter 1) detection of offered (mesocosm environment) or all potential (natural environment) diet items or 2) the abundance of offered or all potential diet DNA, suggesting that contaminants either hide or inflate diet consumption amount? Last, 3) Does surface sterilization alter the richness or composition of potential diet items in a natural environment, suggesting that surface contamination could alter ecological interpretations of community-scale species interactions? Examining how surface sterilization alters estimated diet metrics for studies of consumptive studies in invertebrates is an important step for verifying and standardizing these methods for future studies.

**Materials and Methods**

*Field site and collections*

We conducted this work on Palmyra Atoll National Wildlife Refuge, Northern Line Islands (5º53’ N, 162º05’W). We targeted a generalist, active hunting spider species (*Heteropoda venatoria*) as the consumer for this project because 1) it occurs in high abundance on the atoll and is therefore easy to collect and observe and 2) it is a generalist species with no viable non-genetic methods of diet analysis, making it an ideal target species for DNA metabarcoding of diet analyses. Palmyra Atoll has a well-characterized species list, and like many atolls, is relatively species poor, allowing for relatively easy characterization of potential diet items (Handler et al. 2007, Food webs if published). We collected consumer individuals during two summers. In 2015, we collected individuals in natural habitats across the atoll. In 2017, we collected consumer individuals which we kept in mesocosm environments in the lab (explained below). All individuals were collected individually in sterilized collection containers to avoid contamination (Greenstone et al. 2011).

*Mesocosm consumer set-up and feeding*

In 2017, we conducted mesocosm trials (n = 26) in order to test whether DNA metabarcoding would detect DNA from a diet item a consumer was offered in a common “contained” environment (mesocosm) often used in studies of feeding interactions (e.g. Gao 2017, Rudolf et al. 2014, Srivistava et al. 2004). We created feeding mesocosms out of one-liter yogurt containers with holes for air transfer. We placed an individual *H. venatoria* in each of these mesocosms. After a 12-hour period alone in the mesocosms, all *H. venatoria* individuals were offered one individual of a large grasshopper species (*Oxya japonica*), which is a common introduced species on the island and a likely diet item (Handler et al. 2007). While many mesocosm experiments include natural “habitats” for consumers and diet to hide, we did not include anything in mesocosms to avoid contamination from DNA by these items. We left all mesocosms for 24 hours, after which we immediately froze at (-20°C) *H. venatoria* individuals which had killed an *O. japonica* individual (n = 25 of 26 trials); consumption of killed individuals was not easily detectable and thus not considered in analyses. All mesocosms were cleaned between each trial with a 10% bleach solution and kept closed to avoid contact of other organisms with the inside of the mesocosms.

*Natural environment consumer collection*

In 2015, we collected consumers (n = 47) in natural environments in order to test whether DNA metabarcoding would detect diet DNA from consumers occurring naturally as they fed on available diet items and came into contact with natural environmental surfaces. We collected and froze all individuals in separate containers at -80°C immediately following collection until surface sterilization and DNA extraction in 2019.

*Both mesocosm and natural environment consumers: surface sterilization*

Because we planned to extract DNA from entire body parts (opisthosoma) of consumer individuals (following methods from Krehenwinkel et al. 2016 and Macias-Hernandez et al. 2018), we wanted to determine whether surface sterilization of *H. venatoria* consumer individuals altered 1) offered (mesocosm) or potential diet (natural environment) DNA detection, and 2) diet item richness and composition (natural environmental context only). We used a surface sterilization treatment to remove possible contaminants from some consumer individuals while leaving some individuals unsterilized. We used surface sterilization techniques common in other fields of molecular ecology (i.e. plant endophytes Schulz et al. 1993, Burgdorf et al. 2014) by submerging and stirring each sample in 10% commercial bleach by volume (0.5% sodium hypochlorite) for 2 minutes and then washing each sample by submerging and stirring in deionized water for 2 minutes. We surface sterilized mesocosm consumers (2017) in the lab on the atoll following freezing at -20°C and then stored each sample in individual vials of 80% ethanol because no long-term refrigeration was available on the atoll at the time (n = 10 surface sterilized; n = 14 not surface sterilized). Natural environment consumers (2015) had been frozen at -80ºC since collection; these consumers were surface sterilized in a sterilized laminar flow hood in 2019 just before DNA extraction (n = 22 surface sterilized, n = 25 not surface sterilized; Table). Prior to DNA extraction, all samples from both 2015 and 2017 were allowed to dry for 1-3 hours in a sterilized laminar flow hood. Finally, the opisthosoma was removed from every consumer individual for DNA extraction using a sterilized scalpel. For all sterilization steps forceps, scalpels, and laboratory surfaces were sterilized between handling each individual.

*DNA extraction and removal of consumer DNA with Ampure XP beads*

We extracted DNA from each *H. venatoria* consumer opisthosoma following a modified CTAB extraction protocol (Fulton et al. 1995). At least twenty-four hours following extraction, we quantified DNA using a Qubit (Invitrogen) fluorometer and the high sensitivity DNA quantification kit with 1L of DNA per reaction. We used methods developed by Krehenwinkel et al. (2016) to remove a proportion of lower molecular weight consumer or diet DNA prior to PCR steps. To do this, we diluted each DNA sample to 20ng/l (a total of 40L per sample), bead cleaned each sample using Ampure XP beads (0.75x bead ratio), keeping the supernatant from the bead cleaning step. Because Ampure XP beads preferentially bind to heavier molecules, during this step, more intact consumer DNA binds to the beads, leaving semi-digested diet DNA in the supernatant (Supp figure). Thus, by keeping the supernatant, we aimed to work with a sample that had a larger proportion of lower molecular weight diet DNA after removing some consumer DNA that bound to beads (Krehenwinkel et al. 2016). With the supernatant, we repeated the CTAB protocol steps for precipitating DNA pellets with isopropanol and 5M potassium acetate and cleaning DNA pellets with ethanol washes. Waiting at least another twenty-four hours, we quantified DNA again using a Qubit fluorometer and high sensitivity kit with 1L of DNA per reaction tube and diluted all samples to 10ng/L prior to PCR steps. All DNA pellets were stored in and diluted with TE buffer.

*PCR amplification, library preparation, and sequencing*

The CO1 gene is well-represented in the GenBank sequencing database (Porter et al. 2018), so we chose to amplify this gene using general metazoan primers (Krehenwinkel et al. 2016, Yu et al. 2012, Leray et al. 2013). We performed all PCR amplification steps in a UV-sterilized biosafety cabinet. We used a standard desalted primer set tested by Krehenwinkel et al. (2016) for use in diet analyses of invertebrate predatory consumers (Table). Our primers included overhang adapters compatible with the Illumina indexing PCR (Illumina 2009).

We PCR amplified the CO1 gene in our samples with a 25L reaction volume, including 9L nuclease free water, 12.5L GoTaq Green Master Mix (Promega Corp.), 1.25 L of each of the primers, and 1 L of DNA template (at 10ng/L). When DNA concentrations were lower than 10ng/L, we added more DNA to the sample to equal 10ng of total template and reduced the amount of water added. Each sample was run in duplicate until after the Illumina indexing PCR, and we ran a duplicated negative sample each PCR run. We ran each reaction with an initialization step at 95°C for 3 minutes, and then 35 cycles of a denaturation step at 95°C for 30 seconds, an annealing step at 46°C for 30 seconds, an elongation step at 72°C for one minute. We ended each run with a final elongation step at 72°C for 5 minutes and then held samples at 4°C until placed in a 4°C refrigerator until attaching Illumina indices via an additional PCR step. To remove reaction dimer, we performed a bead cleaning step with Ampure XP beads at a 0.8x bead ratio prior to the Illumina index PCR. Samples were re-suspended from beads using a 10mM TRIS resuspension buffer.

We then attached Illumina index primers with an additional PCR step (Nextera XT Index Kit v2). Each total reaction volume was again 25L, with 5L of nuclease free water, 12.5L GoTaq Green Master Mix, 1.25L of each primer, and 5L of PCR product. These were run in a standard PCR protocol for these primers: an initialization step at 95°C for 3 minutes, followed by 10 cycles of a denaturation step at 95°C for 30 seconds, an annealing step at 55°C for 30 seconds, and an elongation step at 72°C for 30 seconds. We ended each run with a final elongation step at 72°C for 5 minutes and then held samples at 4°C until placed in a 4°C refrigerator.

We verified PCR amplification by visualizing 3-4 L of each PCR product in a 1.5% agarose gel using GelRed (Biotium) at 100V and 170mA for 30-40 minutes. Gels were visualized with a Bio-Rad Gel Doc XR+ imager using Image Lab 5.0. We kept samples for which both duplicates successfully amplified during the PCR steps and only kept samples on PCR runs in which both negative control duplicates resulted in no product detection. For successful samples, we combined duplicates and bead cleaned with an Ampure XP bead ratio of 0.7x. We determined the average length of the gene region using an Agilent TapeStation with a D1000 ScreenTape System following the standard protocol from the quick start guide. We then quantified these final PCR products using a Qubit fluorometer and a high sensitivity kit with 1L of sample per reaction tube and diluted each sample in 10mM TRIS to a final concentration of 5nM.

We multiplexed all samples along with one negative control and two positive controls (cloned fungal species of the ITS gene region; GenBank accession numbers: MG840195 and MG840196, Apigo and Oono 2018; 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.

*Sequence merging, filtering, and clustering with UNOISE3*

We merged, filtered, and denoised (clustered) our sequences around exact amplicon sequence variants (ASVs) using the UNOISE3 algorithm (unoise3 command in the open-source USEARCH 32-bit version 11.0.667; Edgar 2016). This ASV denoising approach incorporates sequence abundance, quality, and error rates to cluster reads in high throughput sequencing data into a smaller subset of real biological units (Supp Figure). We also repeated analyses with the DADA2 algorithm run through R (dada2 package version 1.1.14.0; Callahan et al. 2016) and with a data cleaning step run through BBSplit (Bushnell 2018) to remove consumer DNA prior to ASV assignment (since ASV assignment is abundance-sensitive); however, UNOISE3 assigned more sequence reads to positive controls than DADA2 (on average, 3x as many per positive control) and the cleaning step did not increase diet DNA detection, so we chose to continue analyses from the UNOISE3 algorithm only (summary and comparisons in Supplement).

Prior to denoising, we used cutadapt (version 1.18, Martin 2011) to remove primers from each sequence. With trimmed sequences, we ran UNOISE3 in USEARCH with a maximum error rate of 1 in the filtering step. From the output, we created a list of unique ASVs and a matrix of ASV abundances across samples. We matched ASVs 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), selecting the subtree with all possible diet items for this species (Kingdom: Animalia, Clade: Bilateria). For taxonomies which were not assigned below the order level (n =24), we submitted each ASV individually to the BLAST Basic Local Alignment Search Tool and assigned them a family based on the best sequence match in the database, given that the top database matches were from the same family. For BOLD taxonomic assignment, we used the BOLD IDEngine of the CO1 gene with Species Level Barcode Records (accessed February 5-16, 2020; 3,825,490 Sequences, 216,704 Species, and 95,537 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).

*Detection of potential diet items*

For consumers from both mesocosm and natural environments, we wanted to know whether surface sterilization altered the detection of potential diet items for each consumer (either by increasing detection because of “false” diet detection or by decreasing detection because of abundance of non-diet DNA). For mesocosm consumers, we focused our detection analysis on the offered diet item we provided the consumers in the mesocosm environment (*O. japonica*, which all consumers were observed to have killed, but not necessarily ingested). For natural environment consumers, we examined all potential diet items (which could represent either diet or surface contaminants). Samples were rarefied (McKnight et al. 2018) because sequencing depth (total number of DNA sequences assigned) can vary considerably across samples in high throughput sequencing runs. We did this using the rrarefy() function in the vegan (version 2.5.6) package in R and we rarefied based on the sample with the lowest sequencing depth (but which had been sequenced with 95%+ sampling completeness based on iNEXT (version 2.0.20) interpolation and extrapolation methods (Hsieh and Chao 2017). We rarefied to 55,205 reads per sample for the mesocosm and 16,004 reads per sample for the natural environment consumers. We rarefied these separately because samples had been preserved in different ways and for different times, which can have large effects on DNA extraction outcomes (Murphy et al. 2002).

Following rarefying, we selected all ASVs which matched to the offered diet item for the mesocosm consumers (species: *Oxya japonica*, genus: *Oxya*, family: Acrididae), and all potential diet items for the natural environment consumers (Kingdom: Animalia; Clade: Bilateria, excluding consumer DNA). In addition, for all consumers, because BLAST and BOLD matched multiple ASVs to the same taxonomies (e.g. at species, genus, or family level), we concatenated all ASVs based on shared taxonomic assignment. All ASVs received a family-level taxonomic assignment, and family-level analyses are common in DNA metabarcoding studies (e.g. Kartzinel et al.), so we chose to concatenate at the family level. We did this by combining ASVs which matched at the family level into one combined taxonomy with cumulative read abundance (i.e. all ASVs matched to *diet family A* were combined into one *diet family A* taxonomy with cumulative read abundance). For both mesocosm and natural environment consumers, we assessed per sample detection of offered diet (mesocosm; *O. japonica*) or all potential diet (natural environment) using generalized linear models with offered (mesocosm) or all potential (natural environment) diet item detection (presence-absence per sample) as the response variable, surface sterilization treatment as a fixed effect, and a binomial distribution.

*Abundance of potential diet DNA*

Because potential diet DNA can represent a rare subset of total sequence abundance in DNA metabarcoding studies (e.g. 0.03 – 8.43 percent of all sequencing reads in one study; Krehenwinkel et al. 2016), we also wanted to determine whether surface sterilization altered this proportion of reads assigned to possible diet. Again, because contaminants can represent “false” diet or can be non-diet items, surface sterilization could lead to either an increase or decrease in the proportion of reads representing potential diet. To test whether surface sterilization altered the proportion of DNA representing offered (mesocosm) or all potential (natural environment) diet items, we assessed per sample offered or potential diet DNA proportion for both sets of consumers (mesocosm and natural environment) separately. For this analysis, we subset only consumer individuals for which we detected offered or potential diet DNA (n = 14 out of 19 for mesocosm; 33 of 37 for natural environment), since we were interested in whether contaminants alter diet abundance only when potential diet DNA is present. We assessed diet proportional abundance in consumers using generalized linear models with the number of offered (mesocosm; *O. japonica*) or all potential (natural environment) diet DNA reads per sample as the response variable, surface sterilization treatment as a fixed effect, total read abundance of the sample (constant across all) as an offset term, and a Poisson or negative binomial distribution (to correct for overdispersion when needed). We also examined the proportions of other potential diet items for the mesocosm consumers as well as DNA which was sequenced but was not diet (e.g. fungi and potential endoparasites) with results in the Supplement.

*Potential diet richness and composition in natural environment consumers*

In addition to allowing detection of diet items, DNA metabarcoding also enables the analysis of diet communities, allowing explorations of individual-, population-, and species-level diet richness and composition. If surface contaminants alter these metrics, ecological interpretation of these community-level data could be misleading, either by increasing the number of consumptive interactions attributed to a consumer, or by hiding interactions that occur more rarely or further back in time (e.g. Marcias-Hernandez et al. 2018, MacKenzie et al. 2002). We assessed whether surface sterilization altered richness and composition of potential diet items in our natural environment consumers. For per sample potential diet richness, we used taxonomies concatenated at the family level to represent diet richness and assessed differences in per sample potential diet richness among sterilization treatments using generalized linear models with the number of potential diet items per sample as the response variable, surface sterilization treatment as the fixed effect and a Poisson or negative binomial distribution (to correct for overdispersion when needed). We assessed differences in potential diet item composition between surface sterilized and unsterilized consumers using a presence-absence PERMANOVA model fit with a binomial mixed effects model with surface sterilization treatment as a fixed effect, a random intercept term for potential diet item, and a random slope term for surface sterilization treatment. Incorporating a random intercept term for potential diet item combined with a random slope term for surface sterilization treatment allows the effect of surface sterilization treatment to vary by potential diet item, such that some potential diet items may increase in presence with surface sterilization (i.e. hidden by contaminants), while others may decrease in presence (i.e. potential diet item is a contaminant; Zurr). We repeated the natural environment consumer potential diet item PERMANOVA with abundance data (Poisson distribution), conducted both presence-absence and abundance based PERMANOVA analyses on all potential diet items (including offered item) for mesocosm consumers, and repeated each analysis using the adonis() function from the vegan package (version 2.5.6) in R (dist = “jaccard” with binary = TRUE for presence/absence and dist = “bray” for abundance; Supplement).

*Model selection*

For all generalized linear models, we performed model selection by comparing the full model (including the fixed effect of surface sterilization treatment) to a null model without this effect. All models were called in the glmmTMB package (version 1.0.0, Brooks et al. 2017) in R (version 3.6.1) We chose the best fitting model based on size corrected AIC values (MuMIn package version 1.43.15). For responses for which the best model included the surface sterilization treatment term, we examined the model summary to determine the standardized coefficients () and p-value of the significance between marginal means of the levels of the surface sterilization fixed effect. We assessed model fit using diagnostics in the DHARMa package (version 0.2.7), including tests for heteroskedasticity, and for count models (Poisson or negative binomial), zero inflation and overdispersion (Zurr, Bolker citations here). All raw data, data cleaning, and data analyses can be found in the data (CITE data here).

**Results**

*PCR success, sequence merging, filtering, and clustering with UNOISE3 and DADA2*

We successfully extracted DNA from 100% of samples (n = 72). Amplification success across all samples was 78%, with 56 of 72 initially extracted samples successfully amplified and sequenced (mesocosm: n = 8 surface sterilized, n = 11 unsterilized; natural environment: n = 18 surface sterilized, n = 19 unsterilized). The Illumina MiSeq run yielded 33,332,804 unpaired reads and had a Q30 quality score of 78.03%. After quality filtering and denoising with UNOISE3, we had a set of 8,029,959 paired-end reads that corresponded to 176 ASVs. Seventy-three percent (128 of 176) of the ASVs matched to a taxonomic assignment. Twenty-three percent of these taxonomies corresponded to potential diet items (41 of 176) and eight percent (14 of 176) corresponded to consumer DNA. Eighty-five percent of the potential diet ASVs received a species-level taxonomic assignment (35 of 41) from either the BLAST or BOLD taxonomic assignments, and every potential diet species received a family-level and order-level taxonomic assignment. There were no conflicting taxonomic assignments at the family level or higher between the BOLD and BLAST assignments.

*Detection of potential diet items*

We detected offered diet (*O. japonica*) in 74% (14 of 19) of mesocosm consumers and potential diet in 89% (33 of 37) of natural environment consumers. For mesocosm consumers, one ASV matched to the offered diet (*O. japonica*), and the best model for diet detection included the fixed effect of surface sterilization treatment with a marginally significant difference between marginal means (( = -2.3; p-value = 0.07). Based on this model, offered diet detection decreased with surface sterilization treatment from a detection in 91% (10 of 11) of all consumers when unsterilized to 50% (4 of 8) of all consumers when surface sterilized. For natural environment consumers, concatenated ASVs corresponded to 20 families of potential diet items. The best model for diet detection was the null model that did not include surface sterilization treatment as a fixed effect (Figure). (Supplement).

*Proportion of potential diet DNA*

Offered diet DNA sequence reads (*O. japonica*) represented 0.8% (± 0.7% SE) of total per-sample DNA sequence abundance for mesocosm consumers; all potential diet DNA sequence reads represented 2.0% (± 1.0 %) of total per-sample DNA sequence abundance for natural environment consumers (Figure). For both the mesocosm and natural environment consumers, the null models which did not include surface sterilization treatment as a fixed effect were the best models of diet DNA read abundance.

*Potential diet richness and composition in natural environment consumers*

For natural environment consumers, the best model for per sample diet richness was the null model which did not include surface sterilization treatment as a fixed effect. Diet richness per consumer was an average 2.08 (± 0.26) diet families per individual sample, with a maximum of 5 diet families in one consumer diet (Figure). The best model for potential diet composition also did not include surface sterilization treatment as a fixed effect. Diet composition consisted of twenty families from ten arthropod orders (*Class Insecta*: Dermaptera, Diptera, Blattodea, Lepidoptera, Orthoptera, Odonata, Hymenoptera; *Class Arachnida*: Araneae, Scorpiones; *Class Chilopoda*: Geophilomorpha; Figure; Supplement).

**Discussion**

In natural environments, surface sterilization does not appear to change diet measures in diet DNA metabarcoding data for the predatory consumer *H. venatoria*, suggesting that surface contamination is not a necessary step. In this environment, all measures of diet, including diet detection, abundance, richness, and composition, were not altered by surface sterilizing consumers prior to DNA metabarcoding, suggesting the potential diet DNA extracted from these individuals represents consumed diet items encompassing 20 families from ten arthropod orders. In a contained mesocosm environment, we found potential evidence of the value of surface sterilization to reduce contamination. In this context the best model included surface sterilization, with a marginally significant reduction in detection of an offered diet item following surface sterilization (a reduction of offered diet item from 91% of the unsterilized to 50% of the surface sterilized population, p-value = 0.07). This outcome, while inconclusive, suggests that surface sterilization may reduce false detections in some high-contact environments prone to surface contamination; correcting for this surface contamination could alter the ecological interpretations of diet metabarcoding data. This evidence suggests that surface sterilization may be an appropriate validation step prior to conducting a diet DNA metabarcoding study in contained environments where there is a high likelihood of consumers coming into contact with diet items. Overall, we observed high rates of diet DNA (74% of consumers with an offered diet item, 86% with natural-fed diet items), suggesting that DNA diet analyses are an effective method in both mesocosm and natural environments for capturing consumptive interactions.

The field of diet DNA metabarcoding uses highly variable surface sterilization practices, in particular for studies including DNA extraction of full organisms of organism body parts without dissection (e.g. Wirta et al., Jacobsen et al. 2018). This is the first study to explicity test for a benefit of this practice. We demonstrate that surface sterilization is not necessary in most or all of these types of studies to capture consumed diet items. The evident lack of surface contaminants, including both “false” diet items and non-diet DNA, (Supplement figure) that hide or inflate diet detection or diversity in DNA metabarcoding studies contrast with obvious surface contaminants that alter ecological interpretations in other fields using high-throughput sequencing to determine community diversity (e.g. fungal endophytes, Burgdorf et al. 2014). One reason for this difference may be that fungal spores are widespread on and in the surfaces of most environments and organisms (Despres et al. 2012, Philippot et al., Colston et al. 2016) and so likely to contaminate studies targeting specific subgroups of these communities. Indeed, even in our dataset, many sequences matched to fungal taxonomies (Data). The fact that these non-target sequences did not alter our DNA metabarcoding data by hiding target diet DNA, even with the relative rarity of diet DNA compared to consumer DNA (0.006 – 26% of each sample, similar to other studies; Krehenwinkel et al. 2016), is likely due to the differences in biomass of these sources of DNA in our samples and the specificity of our PCR amplification protocol (Elbrecht et al. 2017). Our results are promising both in validating the robustness of findings from past diet DNA studies that have not implemented surface sterilization treatments, but also highlight that diet DNA metabarcoding is an effective tool even when DNA sequence data contain potential environmental contaminants (Supplementary Figure).

While we saw no widespread evidence of the necessity for surface sterilization in our study, we did observe some evidence that more contained environments (here, mesocosms) may be more prone to contamination than open terrestrial environments. This outcome highlights that the decision to surface sterilize prior to DNA metabarcoding may matter more in some environments and experiments than others. Mesocosms, either built or natural, are a widely used type of environment in studies of consumptive interactions (e.g. built mesocosms: Gao 2017, Rudolf et al. 2014; ponds, lakes, and natural microcosms: Srivistava et al. 2004, De Meester et al. 2005) and DNA metabarcoding in any of these environments may benefit from surface sterilization. While our results highlight potential benefit of surface sterilization in one environmental context (mesocosms), abiotic and biotic factors of any environment can alter the risk of contamination as well, and thus be good candidates for surface sterilization. For example, any aspect of an environment that alters the persistence of DNA is likely to alter the risk of environmental contamination (e.g. high pH and low salinity increase persistence; Collins et al. 2018, Strickler et al. 2019, microbial growth decreases DNA persistence: Nielson et al; Strickler et al. 2014). Ecological factors of an environment or consumer could shape contamination risk as well, including diet density, consumer-diet population ratios, consumer-diet body size ratios, interaction frequency, and hunting mode, which can alter contact probability and handling times as well as the likelihood that high-biomass contaminants will hide diet items that were consumed farther back in time (Greenstone 2011 and 2012, Scharf et al. 1998, Jeschke et al. 2002, Samu and Biro 1993, Marcias-Hernandez et al., Abrams and Ginzberg 2000). Considering these aspects of any environment prior to conducting a diet DNA metabarcoding study is key in confidently assigning consumptive interactions and could help build predictive frameworks of when surface sterilization may be necessary.

Future work on surface sterilization in diet DNA metabarcoding (particularly when extracting DNA from full body parts) should 1) determine mechanisms and factors influencing contamination risk, such as factors contributing to contaminant persistence (Collins et al. 2018, Strickler et al. 2019, Machler et al. 2018, Pilliod et al 2014) or ecological or methodological factors contributing to contamination risk and abundance (Greenstone, Elbrecht) and 2) test for any negative effects of surface sterilization on data quality .

Given our results and the low time and effort involved in pursuing this approach (~5 minutes per sample), our current recommendations is 1) that surface sterilization is not necessary in most natural systems with open habitats, and 2) that in any system where it seems likely that there may be prolonged contact with potential diet items surface sterilization should be utilized (if extensive pre-trial of benefits of surface sterilization is not feasible) use surface sterilization (e.g. when small habitats are shared and in mesocosms) or seeking to detect small changes in focal diet items.

Diet DNA metabarcoding is providing the first glimpse at comprehensive diet for a suite of consumers important to the field of food web ecology and to the maintenance of biodiversity on the planet (Nielson et al. 2018). Being able to determine consumptive interactions for many species and environments for the first time will continue to build a better picture of the complex structure of nature, and how species interactions can scale up to create unique ecosystems and ecosystem functions and how these systems will change with anthropogenic disturbance (Pilosof et al. 2017, Ives et al. 2005, Rudolf and Lafferty 2011, Brophy et al. 2017, Dunne et al., Tylianakis et al. 2008, Harvey et al. 2017). Like any method for determining consumptive interactions in nature, DNA metabarcoding continues to be refined, especially as new tools and data are continually emerging (Zinger et al. 2019, Gomez-Rodriguez et al. 2017, Kyist 2013, Jain et al. 2015). This study builds on past efforts to refine the field of diet DNA metabarcoding by using surface sterilization to pinpoint potential sources of error in diet DNA data, Here we find that, on the whole, surface sterilization seems to be unnecessary in most contexts. However, continual context-specific refinements of surface sterilization protocols, along with refinement of other steps in diet DNA metabarcoding studies, is still needed. Cumulatively these methods developments will increase the validity and widespread utility of diet DNA metabarcoding across consumer groups and environments.

**Acknowledgements**

Field work for 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 Carina Motta and Michelle Lee for help collecting samples for this project and laboratory technicians Emily Lutz and Tessa Chou for helping prepare genetic samples. 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 D. Orr, E. Forbes, H. Lowman, D. Preston, D. Trovillion, E. Crone, E. Sauer, L. Falke, B. DiFiore, 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.

**References**

Will be putting in after some rounds of edits.

**Data Accessibility**

Sequence data are available on GenBank (reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA639981?reviewer=2k2u1qmrtehqbsmc05vdqivkor) and will be made publicly available following the acceptance of the manuscript.

Data and analyses are currently available in a GitHub repository (https://github.com/anamtk/DNA\_Diet\_Methods.git) and will be uploaded to Dryad after acceptance of the manuscript.

**Author Contributions**

AM-tK, AA, and HY conceived the idea for this study. AM-tK collected field samples and conducted mesocosm study. AM-tK and AA designed laboratory analyses for this study. AM-tK performed all lab processing and data analyses for the study. AA and HY provided feedback on data analysis methods. AM-tK led the writing of the manuscript. All authors contributed to editing of the manuscript.

**Figures and Tables**

Table 1: Sample sizes for successfully extracted and PCR amplified samples of surface sterilized and unsterilized *H. venatoria* individuals in the mesocosm and natural environments. Bold numbers indicate final sample sizes for statistical analyses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Surface sterilized* |  | *Unsterilized* |  |
| *Environment* | Extracted | Amplified | Extracted | Amplified |
| Mesocosm | 10 | **8** | 14 | **11** |
| Field | 22 | **18** | 25 | **19** |

Table 2: Primers with Illumina overhang adapters (in bold) used to amplify the CO1 region in this study.

|  |  |  |
| --- | --- | --- |
| Primer | Sequence (5’ – 3’) | Source |
| mICOIintF | **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGWACWGGWTGAACWGTWTAYCCYCC | Yu et al. 2012 |
| Fol-degen-rev | **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TANACYTCNGGRTGNCCRAARAAYCA | Leray et al. 2013 |



Figure 1: a) Detection of offered diet (*Oxya japonica*) DNA in mesocosm consumers that were and were not surface sterilized. The decrease from 91% to 50% detection is marginally significant (p-value = 0.07). B) Detection of all potential diet DNA in natural-environment consumers that were and were not surface sterilized. Detection of diet DNA did not change with sterilization treatment.

****

Figure 2: Neither the a) proportion of offered diet item DNA in mesocosm consumers or b) proportion of total potential diet DNA in natural environment consumers significantly changed with surface sterilization treatment.

****

Figure 3: Surface sterilization did not alter per sample diet richness (concatenated at the family level) for natural environment consumers..

****

Figure 4: Surface sterilization did not alter the composition (either with a presence-absence of abundance model) of potential diet items (at the family level) for natural environment consumers. In this figure, presence is indicated by a colored box and abundance is indicated by color depth (divided by quartiles due to wide variation in DNA sequence abundance).

**Supplementary Information:**

Supplement A: Comparisons between UNOISE3 and DADA2 with and without BBSplit.

Supplement B: BBSplit methods.

Supplement C: Model outputs for GLMMs.

Supplement D: Full DNA extraction and PCR protocol

Supplement E: Model outputs from supplementary data analyses.

Supplement F: Supplementary figures on DNA extraction and amplification protocols, as well as additional figures of diet composition and diversity for mesocosm and natural-environment consumers.