COMPARING ENVIRONMENTAL DNA AND MICROSCOPE IDENTIFICATION FOR ASSESSMENTS OF MACROINVERTEBRATE BIODIVERSITY IN NORTHERN ARIZONA

By Yazhmin A. Dozal-Young

A Thesis

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Science

in Environmental Science and Policy

Northern Arizona University

May 2023

Approved:

Rebecca Best, Ph.D., Chair

Rachel Mitchell, Ph.D.

Joseph Mihaljevic, Ph.D.

ABSTRACT

COMPARING ENVIRONMENTAL DNA AND MICROSCOPE IDENTIFICATION FOR ASSESSMENTS OF MACROINVERTEBRATE BIODIVERSITY IN NORTHERN ARIZONA

YAZHMIN A. DOZAL-YOUNG

In the arid southwestern region of the United States, increased temperatures and more variable precipitation associated with climate change will change the biodiversity and productivity of small water bodies such as ponds and wetlands. Ponds are isolated and closed systems, the lack of water flow allows DNA to accumulate within these systems, which makes ponds an excellent system to use environmental DNA (eDNA). This study compared the frequency and richness of invertebrates found in eight stock ponds in 2019 by eDNA and community sampling followed by microscope identifications. To target kingdom Animalia and phylum Arthropoda, eDNA was sequenced for the cytochrome oxidase I (COI) mitochondrial marker using BR5 and F230 primers. Environmental DNA metabarcoding for ponds resulted in higher richness than microscope identification (p-value=0.0024). However, when eDNA sequences with low resolution taxonomic identifications (e.g., only to phylum) were removed from the dataset, microscope richness was higher (p-value= 0.03). Further development of sampling and data processing workflows, and comprehensive reference sequence databases for regional species would help the utility of eDNA as an identification method in this region. Coupling eDNA with microscope identification could provide the information to determine stock pond biodiversity the most accurately, which can function as a proxy for pond secondary productivity. By integrating microscope and eDNA methods, managers can more efficiently identify which ponds are priority to maintain with their limited resources.

ACKNOWLEDGEMENTS

I would like to thank Kaitlin Haase for collecting macroinvertebrate samples and working with undergraduate students in the Aquatic Biodiversity and Conservation lab on the microscope identifications. I would like to thank Dylan Chandler for collecting and filtering water samples and extracting DNA, and Jill Cocking at the NAU Genetics Core Facility for sequencing. Thank you to the Research Initiative for Scientific Enhancement (RISE) for funding my thesis projects, and the valuable feedback and support from my RISE cohort and mentors throughout the two years of my program. Thank you to the Lab of Conservation Ecology meeting members for providing support and additional feedback throughout the development of my thesis. I'm incredibly grateful to my friends, family, and husband that have provided endless support throughout these last two years. Finally, I'd like to thank Dr. Rebecca Best for her endless support, mentorship, and dedication as a graduate advisor.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

The effects of climate change are reducing biodiversity and ecosystem functionality in small water bodies, like ponds and wetlands, in the southwestern United States (Reid et al., 2019). Small ponds are one of the most frequently found water bodies on Earth, which makes ponds critical for migrating species and for the maintenance of aquatic biodiversity (Cereghino et al., 2014; Biggs et al., 2017). Biodiversity in any environmental system can increase biomass productivity and resources use, along with providing essential ecosystem services valuable to humanity (Duffy 2009). Therefore, the biodiversity in these lentic systems is critical to maintain ecosystem functioning and provide ecosystem services in natural and urban landscapes. Functioning pond ecosystems provide ecosystem services of small-scale water storage, purifying pollutants, nutrient retention, and carbon-cycling (Thornhill et al., 2018; Cereghino et al., 2014; Biggs et al., 2017). However, biodiversity within these isolated environments is vulnerable to invasive species, habitat loss, and climate change. Although threats are human induced, efforts for mitigation can be taken. Effective management requires data about the distribution of biodiversity across the landscape, and how it is changing over time. However, biodiversity data can be challenging to obtain in an efficient timeframe in response to increasing threats.

Drought due to climate change along with water withdrawals and land management are impacting where and when aquatic habitats are available in the network of small ponds in Arizona (Parsons and Mathew, 1990). Numerous states are pulling from groundwater and rivers to meet increasing population demand for water in growing metropolitan areas, such as Phoenix, Arizona (Parsons and Mathew, 1990). By allocating groundwater resources for human populations, there is less available water for naturally occurring and anthropogenic ponds. Stock ponds, which are ponds smaller than 15-acre feet and used for watering livestock and/or wildlife,

provide ecosystem services and are essential for water connectivity, and wildlife and cattle survival in arid landscapes (Arizona Legislature, n.d.). However, with a more variable climate and precipitation patterns, ponds can become more ephemeral, with shorter and less predictable wetted periods (i.e., shorter hydroperiods). Many amphibians and invertebrates may be unable to successfully migrate northward without a network of ponds, whether natural or artificial (Cereghino et al., 2014), which could pose a problem with predicted hotter and drier climates in Arizona. Pond management can entail sedimentation removal and vegetation buffer maintenance, invasive species control, and pollution or fertilizer cleaning. Ponds can also act as catchment systems for water and prevent water flow downhill to streams to support fish or other stream species (Lockwood et al., 2022). Therefore, Arizona pond managers may need to evaluate the importance of ponds and remove some of them in drought conditions. To determine which ecosystems are of most importance for ecosystem functioning and conservation, aquatic ecologists and managers need the most effective tools to measure biodiversity to prioritize and maintain the most productive and diverse stock pond ecosystems.

One option for monitoring aquatic biodiversity is environmental DNA (eDNA), which can be used to identify a wide variety of organisms ranging from chordates to phytoplankton. The eDNA found in aquatic environments can originate from saliva, skin cells, urine, and feces of animals living within the aquatic environment or passing animals, like cows or birds (Rees et al., 2014). If eDNA sequences can be matched to species, biodiversity assessments can be made efficiently from water samples, rather than time consuming microscope identifications which require specialized taxonomic expertise. Previous studies in many diverse types of aquatic ecosystems have compared eDNA identification to observation-based or microscope identification and its accuracy when measuring biodiversity. In coral reef systems, depending on

abiotic factors of where the samples were collected, like salinity, temperature, and depth, eDNA can result in similar or higher rates of species identification than diver observations, especially when identifying gastropods, hydrozoans, crabs, bryozoans, barnacles, or polychaetas (Staehr et al., 2022). When using different forms of sequencing and bioinformatics, eDNA consistently identified more taxa compared to microscopy identification in a New Zealand freshwater river environment (Dowle et al., 2016). In the New Zealand study, cryptic species that are difficult to identify morphologically below the taxonomic level of family were successfully identified through eDNA, (Dowle et al., 2016).

Several factors can influence the success of species identifications from eDNA. Environmental factors that degrade DNA, such as temperature, or inhibit its amplification can lead to low data quality and poor identifications (Harper et al., 2019). In addition, researchers need to use appropriate primers for the targeted species and need to have or build an established DNA reference library for the identification of new DNA sequences (Yang et al., 2021). However, if sampling, sequencing, and analysis workflows are optimized, eDNA can be a useful tool when working with cryptic species because eDNA can repeatedly detect the exact species based on genetic sequencing and is a cost and time effective alternative for taxonomic identification (Kim et al., 2019). Using eDNA metabarcoding has its limitations, but based on these studies, could provide valuable information, especially when coupled with other forms of morphological identification.

Depending on the targeted organisms, different genetic markers can be used to maximize the resolution and accuracy of sequence identifications. To assess animal biodiversity, cytochrome oxidase I (COI) is the best marker currently available due to its widespread use and reference database (Shokralla et al., 2012). Within the COI gene, multiple primer pairs can be

used to target regions that are particularly useful for identifying specific groups of animals. For this study, we used primer pairs BR5 and F230 due to their ability to identify a broad range of arthropod orders (Gibson et al., 2015). There are many primers available for use in aquatic ecosystems, other common arthropod forward and reverse primers are LepF1/LepR1 and LCO1490_tl/HCO2198_tl primers, but they are generally used when DNA can be extracted directly from arthropod tissues, enabling longer, less degraded sequences (Herbert et al., 2004; Hajibabaei et al., 2011). By using different identification methods, like eDNA and microscope identification, researchers can identify different species depending on their goal, whether they are aiming for a broad array of taxa to quantify biodiversity or to test for specific species present in an aquatic ecosystem (Harper et al., 2019; Dowle et al., 2016).

To compare the value of eDNA and microscope identification methods as biodiversity monitoring tools in small waterbodies in the Southwest, we conducted a study across an elevation gradient of 1366 to 2255 meters in eight different ponds in northern Arizona. We aimed to compare the richness and frequency of taxa found in samples collected for microscope identification in summer and fall of 2019 to those identified from eDNA water samples collected at the same sites in November of 2019. We collected microscope samples twice, once before the monsoon season, and one after, because the timing for the community sweep net samples determined the aquatic community present in the pond. However, eDNA can remain detectable up to 8 weeks after shedding from an organism (Dejean et al., 2011; Pilliod et al., 2014; Buxton et al., 2018). Therefore, we only sampled once in November of 2019 for eDNA because it could capture DNA in the pond shed in summer and fall.

RESEARCH QUESTIONS AND HYPOTHESES

The overarching research question we seek to answer is whether microscope identification or eDNA is a better method for assessing aquatic invertebrate biodiversity in southwestern ponds. Microscope analysis allows users to both identify and count the abundance of organisms, whereas eDNA cannot provide abundance data because the DNA from different organisms cannot be assumed to amplify equally (Harper et al., 2019; Takahasi et al., 2023). However, microscope identification is time intensive and depends on the taxonomic expertise of the user. Environmental DNA data can provide a large amount of data in a brief time span, however due to limitations in data quality and/or the representation of local taxa in sequence databases, accurate identification may not be possible.

To compare these two methods, we ask the following specific questions:

1. Do we detect more taxa from multiple primers, compared to using a single primer?

Hypothesis: Animal species are typically identified using the COI (cytochrome oxidase I) genetic marker, but there are multiple primer sets that capture different sections of this gene. We predict that using multiple primer pairs can increase the number of taxa identified.

2. Does eDNA identify more taxa than microscope data collected in the same year?

Hypothesis: We predict that eDNA from water samples will capture more taxa and species than the microscope identification, which is often only to the level of order and limited to those species collected in a single 5 meter transect per pond.

3. Do we detect more widespread taxa using eDNA or microscope identification methods?

Hypothesis: Because community sampling is limited to a single transect per pond, it may miss rare species. Therefore, we predict that community samples are more likely to find widespread species present in many ponds, whereas eDNA might also identify rare species found only in one or two ponds.

4. Does DNA degradation at higher water temperatures limit the effectiveness of eDNA methods at lower elevations?

Hypothesis: Because higher water temperature speeds the degradation of DNA, we predict that we will detect declining biodiversity at lower elevations using eDNA. This decline should be steeper than any decline in microscope-identified biodiversity.

METHODS

Field Sites

We selected eight study ponds in northern Arizona, which we sampled for both aquatic macroinvertebrates for microscope identification and for water samples for eDNA analysis. We collected community samples for microscope identification in summer and fall of 2019, and then sampled each pond again in November 2019 for eDNA (Figure 1).

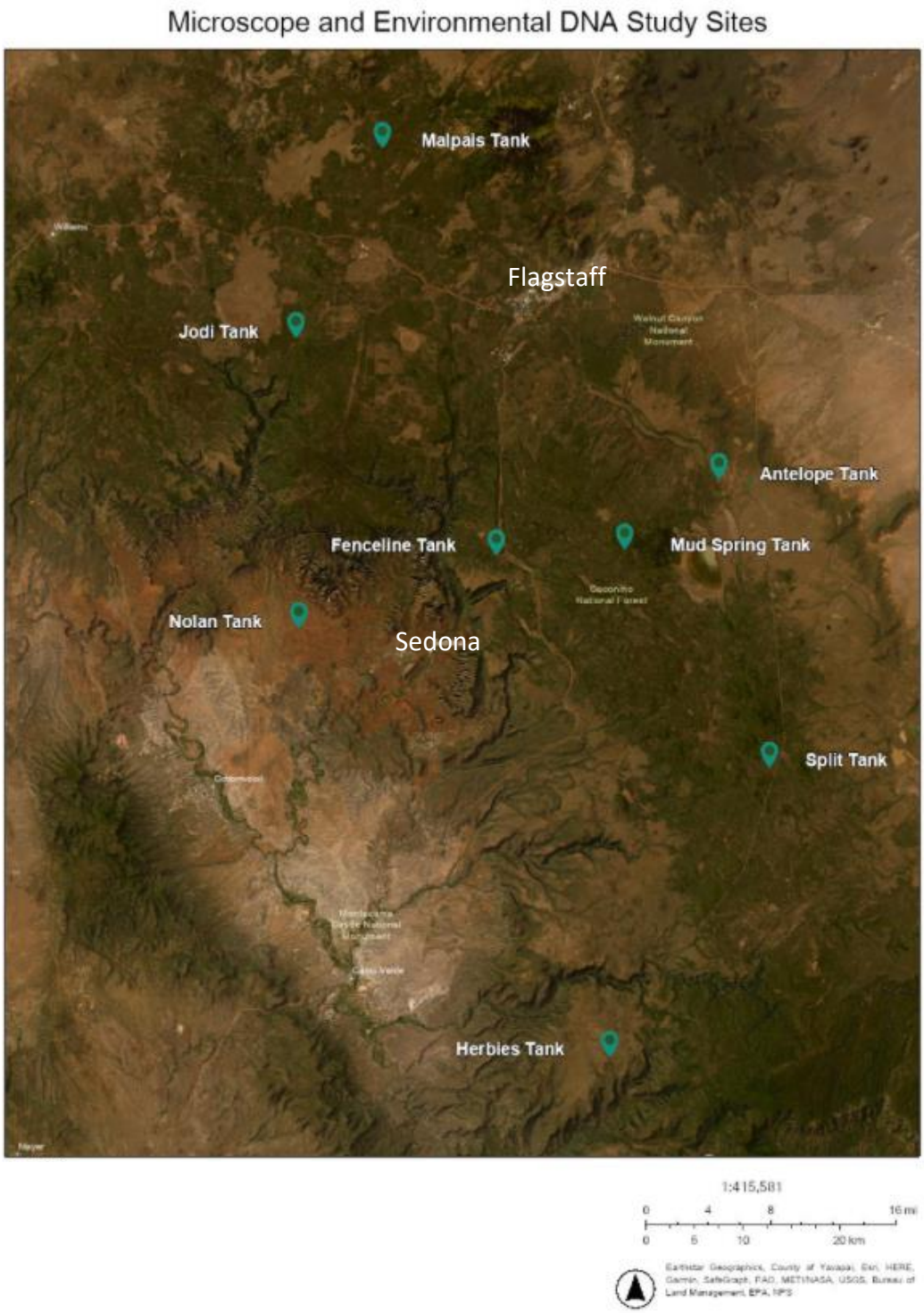

Figure 1: Location of each pond in northern Arizona. Map was made using ArcGIS Pro Online spatial imagery data.

Field Sampling

In each of the eight ponds, we sampled invertebrate communities using a single five meter transect swept with a 12x10 D-net with a 500-micron screen. The collected invertebrates were placed in a specimen collection cup with ethanol and taken to the lab. The samples were stored in the refrigerator until we were able to sort and count invertebrates to the finest taxonomic level consistently possible for a team of undergraduate researchers under the microscope (usually only to Family or Order).

The same eight stock ponds were sampled for eDNA in November of 2019, when all ponds were still in their wetted phase (only one of the ponds later dried over the following winter; all others are permanently filled). Water samples were collected in one liter plastic bottles sterilized with bleach and then rinsed with pond water before being filled from four separate locations at each stock pond. This resulted in one mixed liter of water per pond, which was transported to the lab and stored in an -80^{o}C freezer until filtration. In January 2020, the samples were thawed and then filtered through cellulose nitrate membrane filters with 0.45 um pores. All equipment was sterilized with bleach between samples. Because pond water was visibly turbid and clogged filters quickly, we were only able to filter 100-300 mL of water per filter. To maximize DNA capture, we used two filters per pond, resulting in two DNA subsamples.

Analysis and identification of eDNA sequences

We extracted DNA from all filters using DNeasy Blood and Tissue kits (Qiagen) and a protocol shared by the Goldberg Lab at Washington State University (https://labs.wsu.edu/goldberglab/edna-assays-and-protocols/), and quantified DNA

concentrations using a NanoDrop (range of concentrations 40-140 ng/uL) and PicoGreen (range of concentrations 0.5 to 22 ng/uL). After extraction and quantification, all PCR and sequencing reactions were completed by the NAU Genetics Core Facility. To optimize identification of arthropods (insects and crustaceans) we selected the BR5 and F230 primer sets for amplification of the Cytochrome Oxidase I (COI) mitochondrial gene. To capture the 229 bp F230 region we used the forward primer LC01490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') from Folmer et al. (1994) and the reverse primer 230R (5'-CTT ATR TTR TTT ATI CGI GGR AAI GC-3') from Gibson et al. (2015). To capture the 310bp BR5 region we used the forward primer B (5'-CCI GAY ATR GCI TTY CCI CG-3') from Hajibabaei et al. (2012) and the reverse primer ArR5 (5'-GTR ATI GCI CCI GCI ARI ACI GG-3') from Gibson et al. (2014). Possibly due to freshwater inhibitors not being fully removed by DNeasy kits, DNA required substantial dilution to achieve successful amplification. PCR reactions were 25 uL reactions with 0.5 uL of Taq (at 5U/uL), 2.5 uL of 10x reaction buffer, 1uL of MgCl2 (at 50 uM), 0.5 uL of dNTPs (at 10 mM), 0.5 uL of each primer (at 10 uM), and 1 uL of template DNA diluted to 5 ng/uL. The PCR program was 95°C for 5min; 35 cycles of 94°C for 40s, 46°C for 1min, and 72°C for 30s; and a final extension at 72°C for 5min.

PCR products for each sample paired with each primer were sequenced on an Illumina MiSeq. Sequencer files were demultiplexed into Casava format forward and reverse sequences and checked for library quality control using FastQC. We used DADA2 implemented in Qiime2 to filter, and de-noise reads. We classified sequences using the MIDORI web server (http://www.reference-midori.info/server.php), using the naive Bayesian RDPClassifier (Wang et al. 2007) and the unique read database for COI, which includes all unique reads for all Eukaryote species present in GenBank. To explore all possible identifications, we set a confidence cutoff of

0.5 to output all identifications to any taxonomic level from species to kingdom over 50%. We then used Qiime2 to integrate sample metadata, sequence data, and taxonomy identifications, producing a community presence absence matrix for all taxa. Again, to allow maximal exploration of this preliminary dataset, we included all taxa identified to phylum or below at >50% confidence, treating each identified sequence as a unique taxon. Finally, to accurately compare richness detected using eDNA and microscope methods, we excluded organisms detected by eDNA that could not possibly have been captured in our microscope community samples. This included zooplankton and rotifers, which were too small for our macroinvertebrate mesh size to accurately capture; chordates there were not captured in microscope community samples due to large size, low abundance, active avoidance, or only temporary visitation to ponds; bryozoans; and cnidarians that are either sessile or unlikely to be accurate identifications under microscope identification. By excluding these organisms from eDNA, we were able to accurately compare the richness and frequency of taxonomic species identified under both eDNA and microscope.

Statistical Analysis

We tested all hypotheses using statistical analyses in R (R Core Team, 2023). First, we tested if we detected more taxa from multiple primer pairs than a single primer pair by comparing the total number of taxa identified (richness) for BR5, F230, and then both primers combined. When combining both primers, we made sure that each taxon identified for BR5 and/or F230 was only counted once. Then, we tested whether eDNA identified more taxa than microscope data collected in the same year by conducting paired t-tests. Furthermore, we explored whether eDNA or microscope identification methods could identify more widespread

vs. rare species by graphing distributions of taxon frequencies for each method. Additionally, we tested if the two methods were equally able to detect richness along the elevation gradient in water temperatures using a linear mixed model. With richness as the response variable, we used elevation, method (microscope vs. eDNA) and their interaction as fixed effects and set pond as the random effect. We tested whether the difference between methods depended on elevation by evaluating the method x elevation interaction. We checked the residuals of the model for normal distribution and equal variance using residuals plots. We implemented the mixed model using the *lme4* package and tested for significance of fixed effects using the Kenward-Roger approximation implemented in the *car* package.

RESULTS

To determine which primer detects more taxa, or if they detect the same taxa, we graphed the taxonomic richness detected with each primer, along with a combined value of both primers (Figure 2). The total value was determined by the combination of both primers, but if both primers detected the same taxa, it would only be counted once. There does not seem to be one primer that consistently identifies more taxa than the other. Therefore, by using the combination of the two primers, BR5 and F230, users can detect higher overall richness (Figure 2).

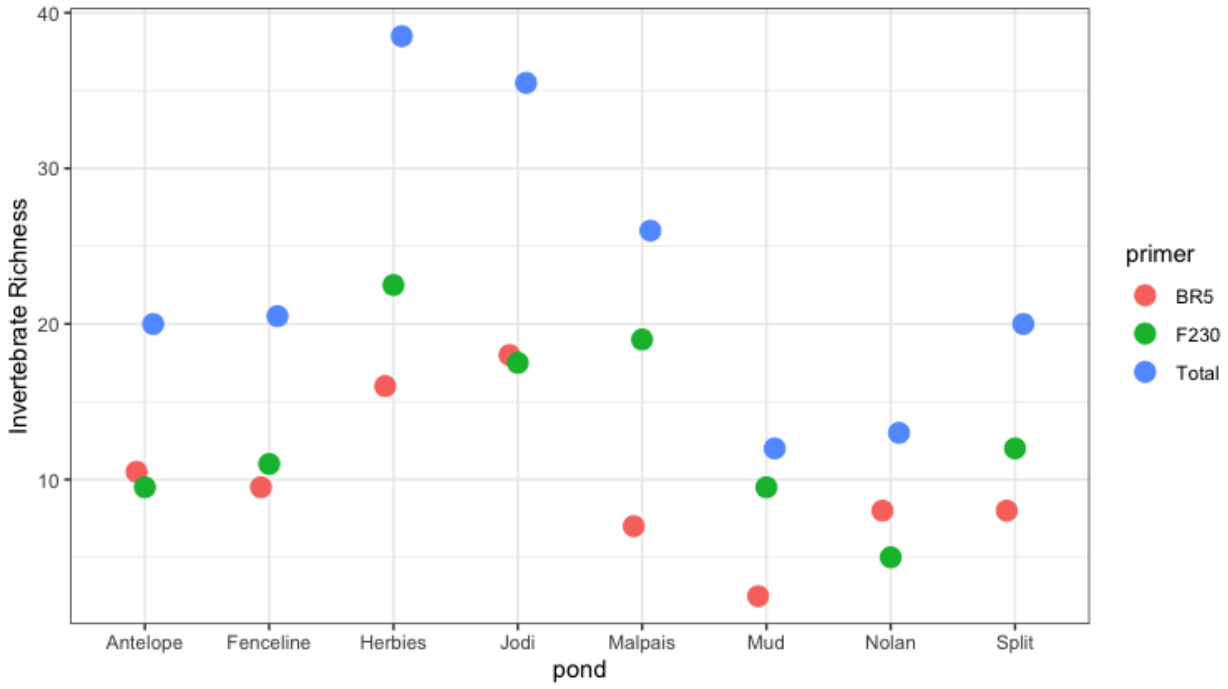

Figure 2: Taxonomic richness by primer in eDNA data.

Regardless of the season, the collected samples using microscope identification detected between seven to thirteen taxa in each pond. Environmental DNA identified taxa at an even lower taxonomic resolution, but indicated from one to 25 taxa per pond (Figure 3). However, the taxa identified using eDNA and the MIDORI database were mostly >50% confident only to the level of phylum Arthropoda or class Insecta, and we are not confident whether they are distinct species or variation within the same species. There were many taxa identified in eDNA, while the taxonomic identification identified fewer separate taxa, but on a higher taxonomic resolution in the microscope data. Richness was significantly higher using the eDNA method (paired t-test p-value: 0.0024), but the specificity of the identifications from the eDNA data is low.

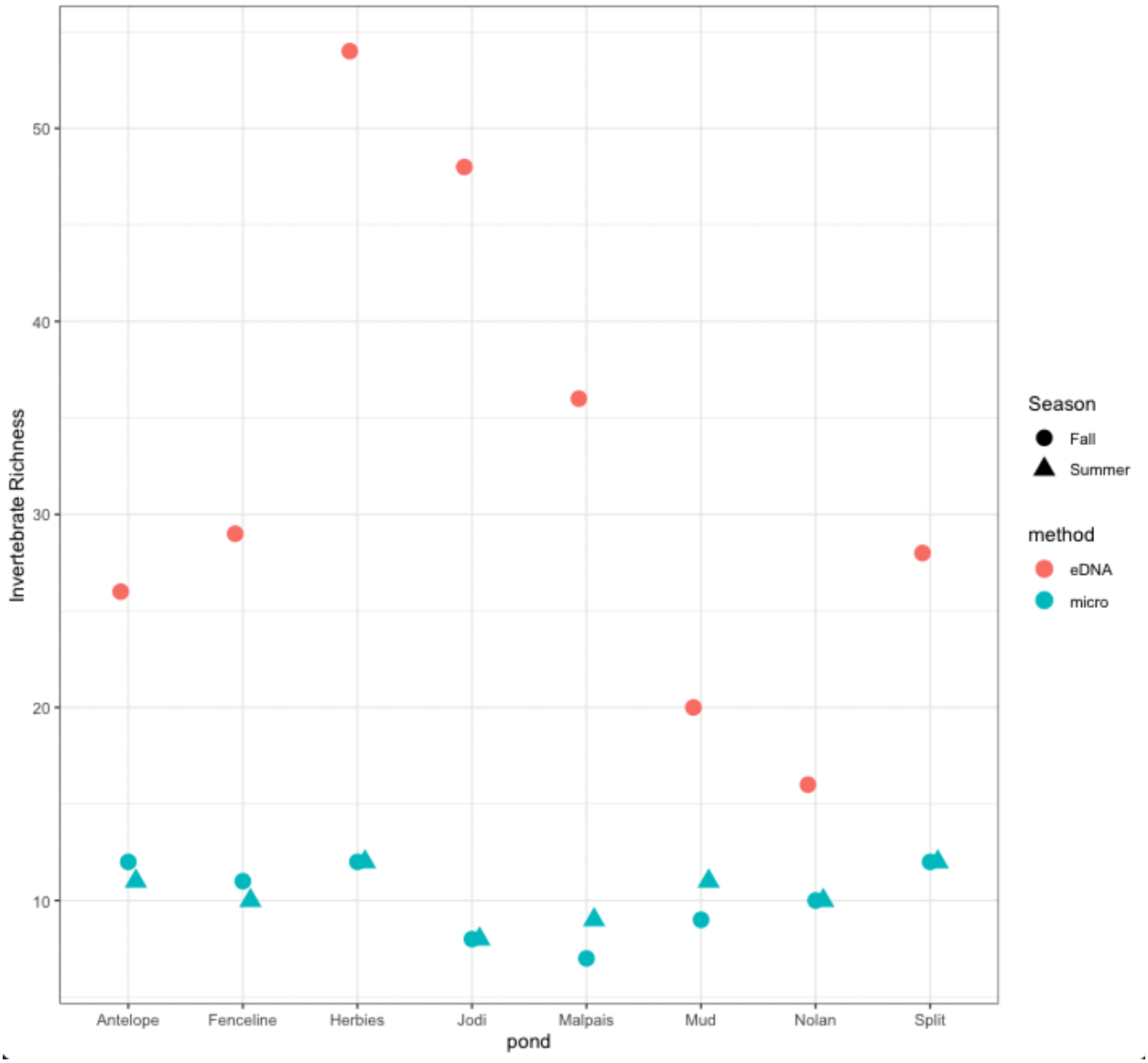

Figure 3: Total species richness between eDNA and microscope identification methods for all taxa.

Since we are not confident in the identification of unknown Arthropoda and Insecta taxa found in the eDNA, we conducted a paired t-test for the eDNA and microscope data without these low-resolution Arthropoda or Insecta taxa to see if there was a significant difference. All the eDNA data for richness are identified as more specific taxa than kingdom or phylum (Figure

4). Richness was significantly higher in the microscope data (paired t-test value t=2.69 and p-value = 0.03).

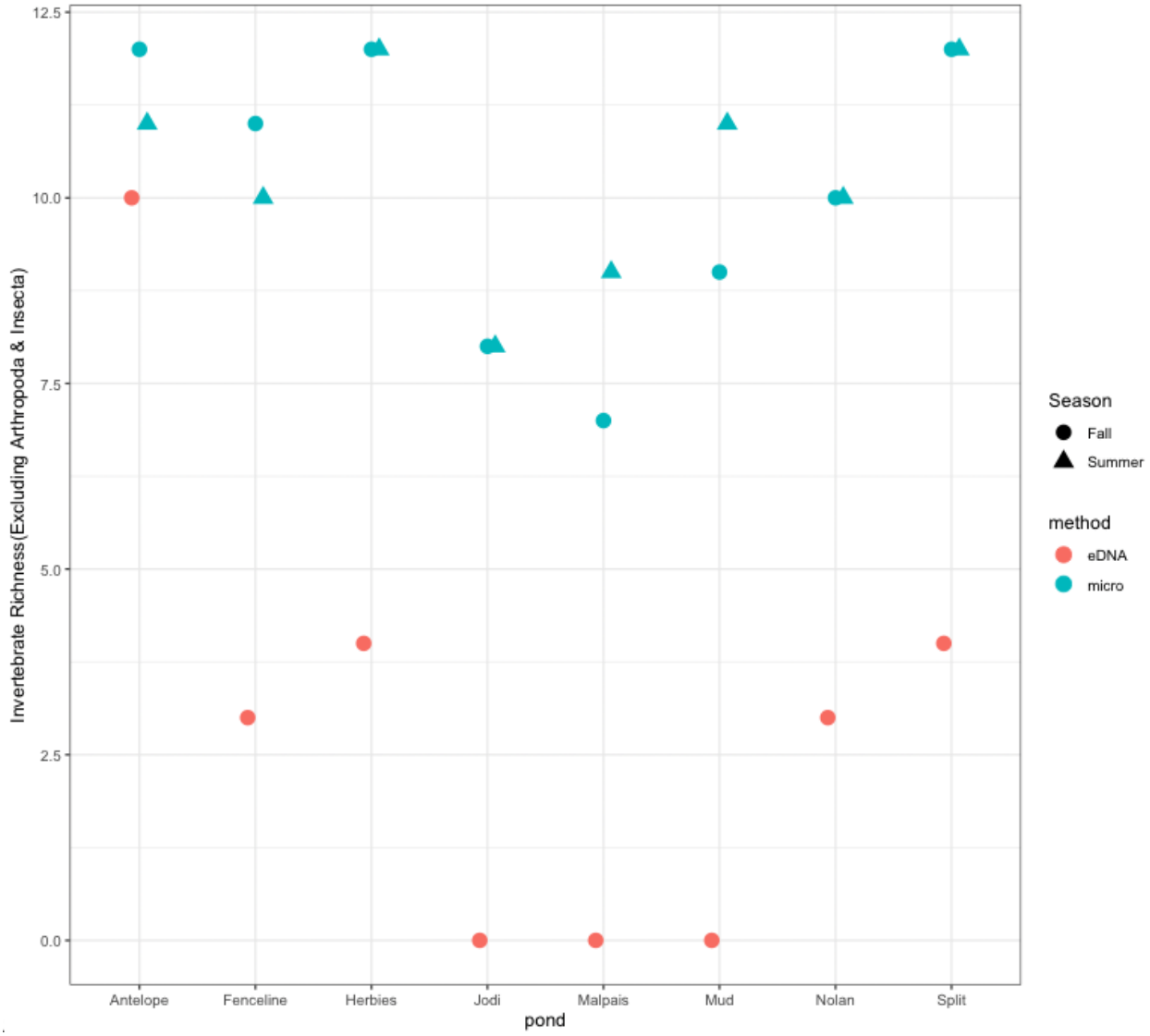

Figure 4: Total species richness between eDNA and microscope identification method for all taxa except taxa identified only as Arthropoda or Insecta.

We explored whether either method was more useful for identifying widespread taxa vs. rare taxa found only in single locations. We found that microscope-identified taxa were present in multiple ponds, with four taxa found in 7 ponds out of 8 ponds and 2 taxa found in all 8 ponds (Figure 5). In the microscope dataset, the taxonomic identification was coarse (Family or Order), so there could have been more variability in distribution at the level of species. However, most eDNA taxa were unique to the one pond where they were found (Figure 5). This extreme result is also unlikely. We expect that all taxa identified were not different species, but instead indicated genetic variation within species, or low-confidence identifications due to low quality or quantity of animal sequences.

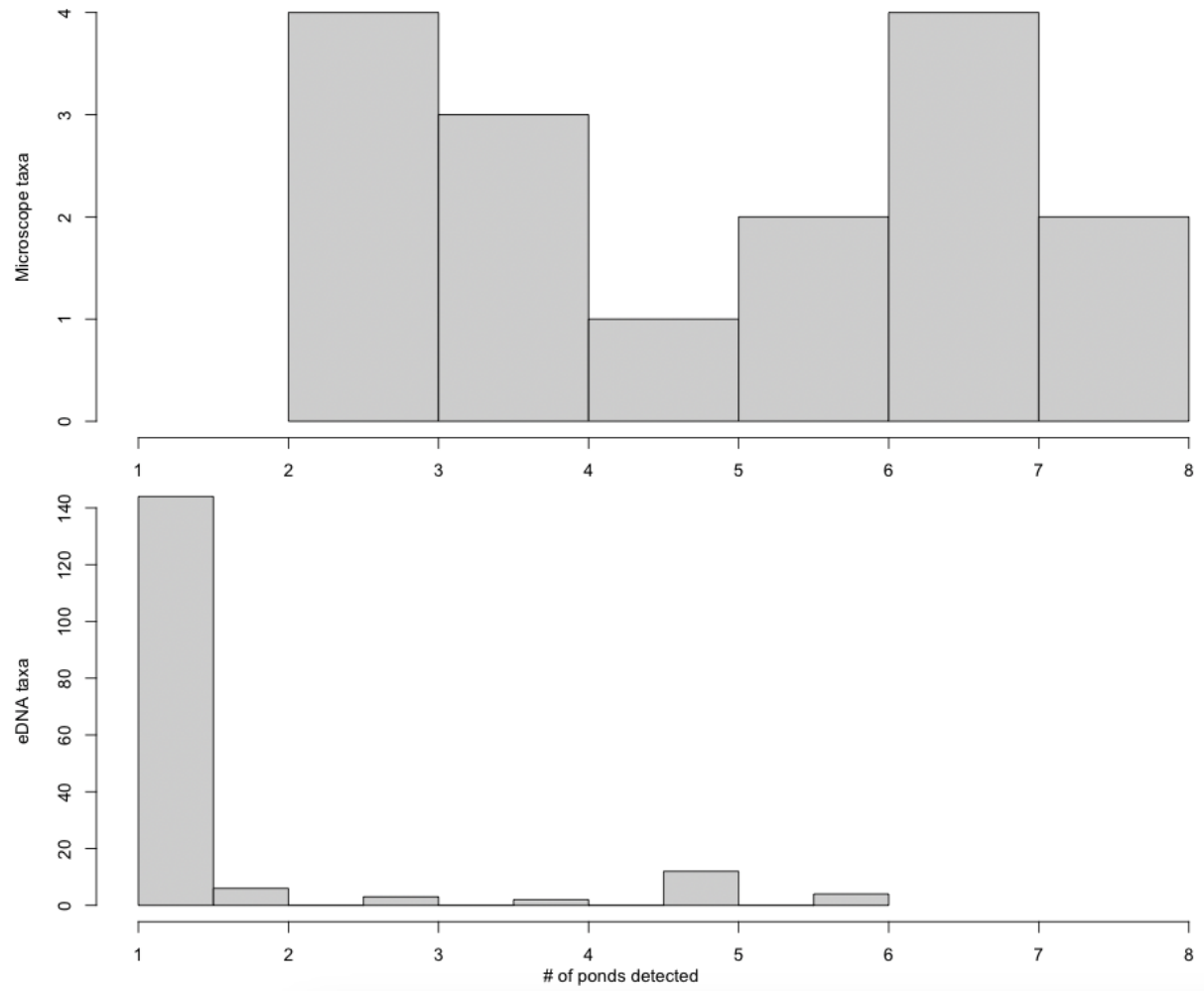

Figure 5: Taxa found throughout eight ponds sampled by eDNA and microscope identification.

To test the effectiveness of eDNA and microscope identification with increasing water temperature at lower elevations, we used a linear mixed model. We found that elevation was not a good predictor of richness detected using either method (Figure 6). As for previous analyses comparing richness across the two methods, we excluded all taxa identified as only Arthropoda or Insecta, which resulted in much lower richness values for eDNA (Figure 7). However, there does not appear to be a consistent pattern from lower to higher elevation in either method of identification, regardless of whether the low resolution eDNA taxa are included (Figure 6 & Figure 7). Both identification methods identified approximately the same number of taxa throughout the elevational gradient. For the complete eDNA data, there was no significant interaction between identification method and the effect of elevation on richness (p-value=0.26), or an overall relationship between elevation and richness (p-value=0.98) (Figure 6). There was also no significant random effect of pond identity (p-value= 1). For the data with sequences more specific than order, there was no significant interaction between identification method and the effect of elevation on richness (p-value= 0.37), or an overall relationship between elevation and richness (p-value= 0.61) (Figure 7). There was also no significant random effect of pond identity (p-value= 1).

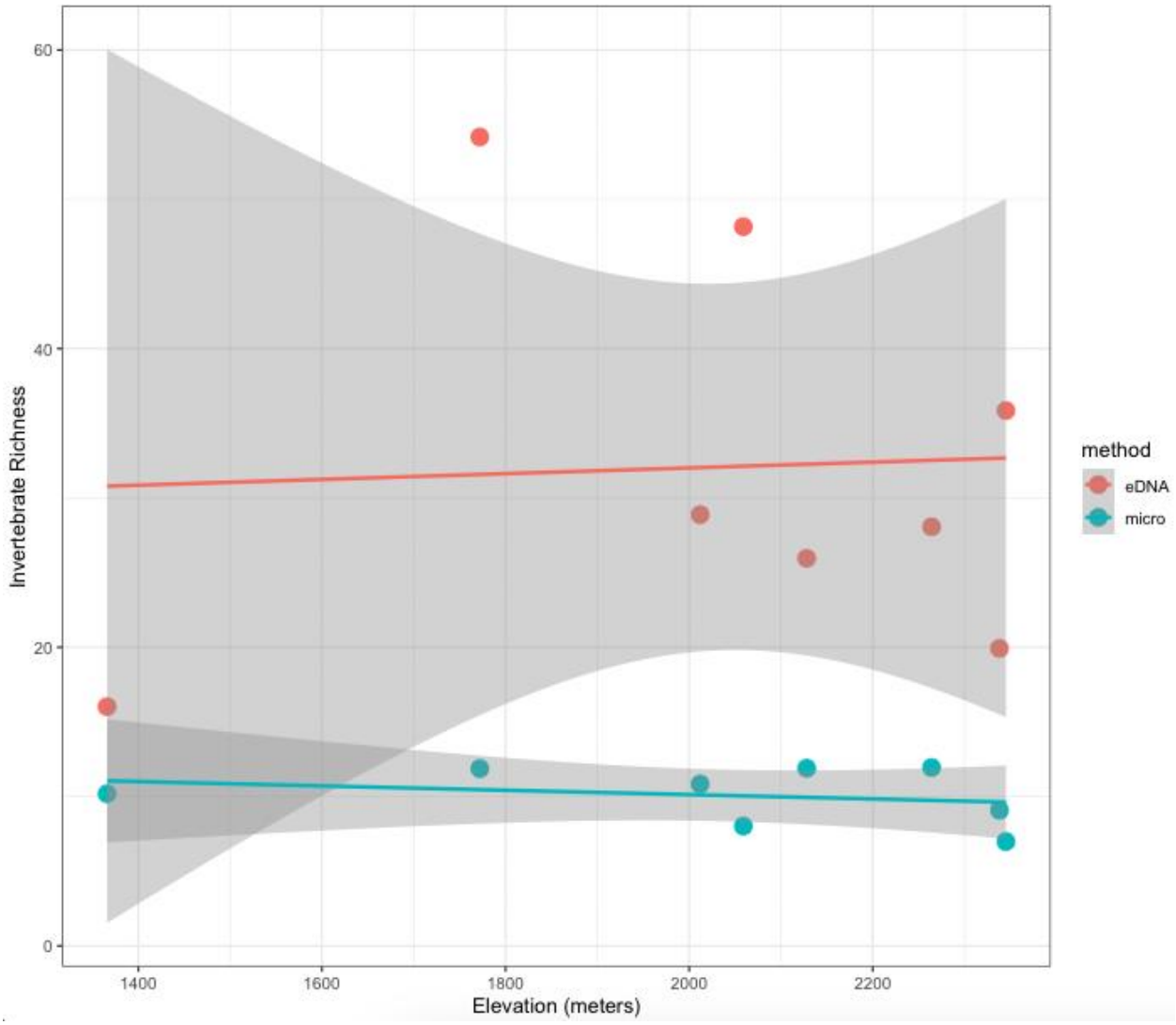

Figure 6: Richness of identification methods by elevation location for all taxa.

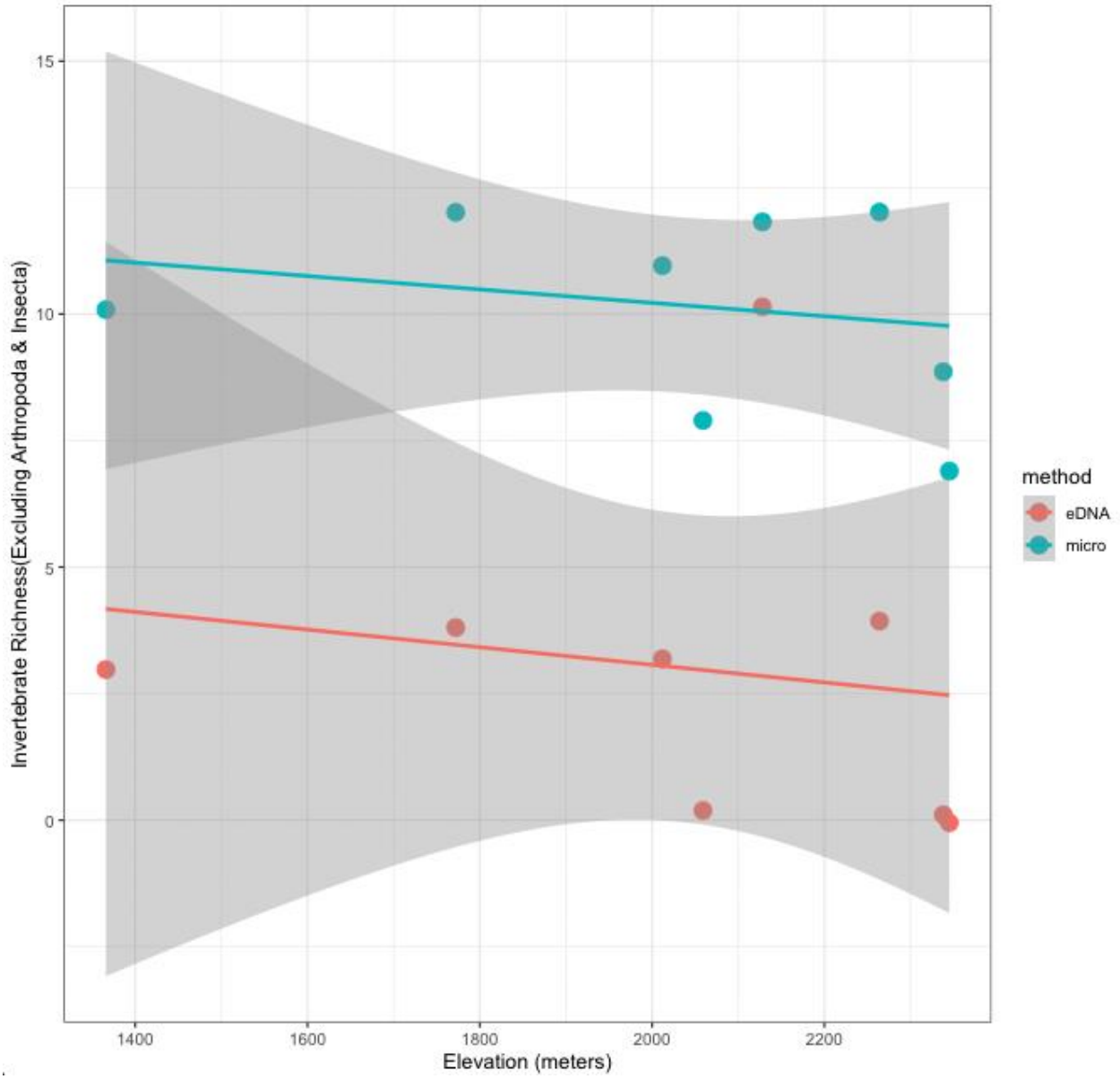

Figure 7: Richness of identification methods by elevation location for all taxa except for taxa identified only as Arthropoda or Insecta.

DISCUSSION

To conserve aquatic biodiversity in small bodies of water we need reliable and effective monitoring methods, especially with future climatic changes in the southwestern United States.

By analyzing richness and frequency of macroinvertebrates, we can test whether eDNA or microscope taxonomic identification would be the best suited as a reliable biodiversity measurement. Taxa identified by microscope were widespread across multiple ponds, whereas taxa identified using eDNA were localized to a single pond. Additionally, richness found in the entire eDNA dataset was significantly higher than the microscope data (p-value= 0.0024). However, when we excluded all genetic sequences identified as only Arthropoda or Insecta, we instead found higher taxonomic richness using microscope identification (p-value= 0.03). We did find that combining two primer pairs, BR5 and F230, maximized the richness identified from eDNA. Finally, we found that pond elevation (as a proxy for water temperature) did not affect the taxonomic richness detected with either method. Future prospects for the use of eDNA as a monitoring tool for aquatic biodiversity in this region will require further research on several key steps in the collection and sequencing pipeline, including obtaining sufficient DNA quantity and quality, understanding the persistence of DNA in different environments, and building a regional reference database.

Sampling sufficient DNA quantity

One of the most important determinants of success in biodiversity monitoring using eDNA is optimizing sampling techniques to obtain enough DNA, and this can be heavily influenced by water characteristics such as turbidity. The lack of water flow and small water volumes in ponds helps eDNA to accumulate over time to concentrations that are not found in other water bodies (Harper et al., 2019). However, due to the stagnant water and soil run-off, ponds also support high levels of algae build-up, which leads to high levels of natural turbidity (Harper et al., 2019). In addition, the ponds we studied are filled with fine sediment easily

disturbed by cattle, wildlife, and collection activity. Thus, filters can quickly become clogged with sediment or algae, reducing the volume of water filtered and blocking extraction spin columns, which inhibits DNA recovery (Harper et al., 2019; Takahasi et al., 2023). The larger the volume of water sampled, the more local biodiversity will be detected because of the higher probability of shed DNA captured. However, sampling a sufficient volume of turbid water can be logistically and mechanically challenging (Takahashi et al., 2023).

Ensuring sufficient DNA quality

Environmental DNA results can also be easily influenced by field collection methods and environmental inhibitors. Field collection of eDNA samples could influence sequencing identification within a collected sample. It is paramount to thoroughly disinfect materials and, when possible, to use single-use sample containers (Goldberg et al., 2016). Decontamination of all equipment used, which includes boots and other field gear, is vital to the integrity of the eDNA sample. In addition, many freshwater systems contain chemical compounds which inhibit sequence amplification during PCR (Hunter et al., 2019). There are many competing strategies for removing these inhibitors to maximize the generation of high-quality DNA sequences for identification. One strategy is DNA dilution during PCR (Hunter et al., 2019), which was necessary for our samples. However, additional strategies might improve sequencing outcomes and improve yield of sequences that can be identified to a higher taxonomic resolution than we found in our study.

Choice of DNA marker and primers

One of the most important choices to make when using environmental DNA is which marker and specific primers to use. If the wrong primer or marker is used the eDNA identification can be severely limited. There have been instances where only about 43% to 64% of the species were identified in the sample, while the rest of the DNA was broadly identified to kingdom, according to an eDNA review (Deagle et al., 2014). For animals, which were the target of our study, COI is clearly the marker with the highest representation in sequence databases globally (Deagle et al., 2014). When deciding on an amplification primer for COI, the ideal primer pair should be able to identify known taxa even with a short amplification length, across a wide array of taxonomic coverage, at high taxonomic resolution, and have no biased amplification of a specific organism (Shu et al., 2021). However, this requires careful tests and verification of primer combinations. Potential primers for COI are limited, but testing can determine which primers are most suitable for the dataset (Deagle et al., 2014; Takahashi et al., 2023). We chose amplification primers BR5 and F230, as they are what is used to amplify Arthropoda genetic sequences (Compson et al., 2019). There are other primers that can be used to amplify Arthropoda sequences, but other primers need higher quantities of DNA sequences (Herbert et al., 2004; Hajibabaei et al., 2011.

Reference libraries

High-quality sequences should be identifiable to coarse taxonomic levels such as order or family, even if regional species are not present in a database. However, improving the identification of DNA sequences in new geographic regions, including for southwestern macroinvertebrates, will also likely require significant investment in building an adequate

reference library. In a freshwater study comparing microscopy and eDNA identification, researchers used metabarcoding with mitochondrial gene marker COI and gene enrichment using MYbaits, which targets nine different genes including COI (Dowle et al., 2016). Researchers built their sequence reference library from the microscope samples collected (Dowle et al., 2016). By using the gene-enrichment method, which targeted nine different genes including COI, researchers were able to detect the highest number of taxa, the highest number of sequences reads, abundance, and biomass in 67% of the samples in the eDNA (Dowle et al., 2016). Increasing local representation will be necessary to obtain species identifications and would need to be constructed for more accurate eDNA identification.

Understanding the persistence of DNA in the environment

The detection of eDNA is based on the presumption that the organism is present where and when the DNA is found, but the presence of an organism's DNA sequence in the environment does not mean that the organism currently resides in that environment. Ponds are considered closed systems, but can receive inflow, land surface run-off, and mobile species visitation (Harper et al., 2019). Thus, DNA deposited from visiting animals via feces or tissue can possibly be found within pond closed systems (Merkes et al., 2014). Additionally, DNA from rivers can also invade closed pond ecosystems during flooding (Harper et al., 2019). Therefore, the detection of DNA sequences within pond ecosystems does not necessarily mean the sequences originate from these systems.

In addition to movement of DNA from outside the system, environmental DNA can also persist in an environment for a certain amount of time before it is degraded to a point that it can no longer be amplified and identified. Environmental DNA starts to decay as soon as the DNA is

separated from the organism and continues to decay even after a sample is collected (Goldberg et al., 2016). DNA bound to sediments can remain identifiable far longer than DNA in water because DNA is preserved longer in undisturbed sedimentation (Goldberg et al., 2016), so that disturbance of sediments could introduce buried DNA from eradicated or seasonally absent species in the aquatic ecosystem (Turner et al., 2015). Therefore, samples collected in sediment heavy environments, like some ponds sampled, may include identified sequences from an uncertain time. Additionally, environmental abiotic factors such as light, oxygen, pH, salinity, microbial communities, UV, and the abundance and composition of substrates can also influence eDNA persistence within an aquatic system (Barnes & Turner, 2016; Shu et al., 2021; Huang et al., 2022). Once eDNA has been collected, the DNA will continue to degrade by microbial activity, exogenous enzymes, and by spontaneous chemical reactions (Takahashi et al., 2023). Ideally, on-site filtering is preferred to avoid further DNA degradation, but that may not be possible in all cases. Therefore, practices such as chilling, freezing, or adding a preservative to the sample, like ethanol, take place in efforts to avoid further DNA degradation (Takahashi et al., 2023). Studies recommend that water samples should either be filtered on-site or within 24 hours, regardless of DNA preservation practices (Takahashi et al., 2023). Temperature can accelerate DNA degradation and inhibit eDNA detection in water sampling (Goldberg et al., 2016). Environmental DNA degradation can limit the detectability of targeted DNA to between 1 day to 8 weeks (Dejean et al., 2011; Pilliod et al., 2014; Buxton et al., 2018). With significant DNA degradation, a delay in time between organism presence and sample collection, could result in a false absence of that organism in the aquatic ecosystem.

Taxonomic resolution

The eDNA method produced higher richness values than the traditional microscope identification. However, species identified under microscopy were identified at least to order whereas the lowest level of identification included in this study for eDNA is phylum Arthropoda. For COI to determine sequences as different species, sequences need to differ 2% of base pairs between closely related species (Ji et al., 2013). The identification confidence level for eDNA was set at 50% or higher. We treated each sequence identified as Arthropoda or Insecta as a distinct species, as we are uncertain whether the sequences are different species or genetic variation within the same species that the eDNA was unable to identify as a species. To gain more insight on the richness difference between microscope and eDNA data, we excluded all sequences identified as only Arthropoda and Insecta to determine if the taxa identified at a more specific taxonomic level had a significantly different richness than the taxa found in the microscope data. As a result, we drastically changed the number of sequences identified by eDNA. After excluding Arthropoda and Insecta, microscope identification had much higher richness (p-value= 0.03). Many of the taxonomic identifications using eDNA are low resolution with low confidence levels, whereas most of microscope identifications are at least to Order or Family.

Next steps for environmental DNA identification

The environmental DNA identification method resulted in many poor resolution identifications with a low identification confidence of 50% or higher, resulting in lack of confidence and trust in the identifications of the eDNA method. Even though there are freshwater inhibitors present within these ponds and possible DNA degradation in aquatic

environments, improvements in DNA collection, preservation, and extraction techniques can decrease the number of false interpretations. Environmental DNA researchers can increase the volumes of the sample collected increase the number of filters to capture more DNA, and test for, and potentially eliminate, freshwater PCR inhibitors (Jerde & Mahon, 2015). In addition to improving collection methods, multiple steps can be taken to improve the levels of trust in the eDNA that is identified in the sequencing methods used. With more high-quality sequences, researchers should be able to set the confidence threshold for identification to 80% or higher (cite something that suggests this) and still identify a high percentage of sequences. Increasing the MIDORI refence library database for the Arizona region could also increase the number of sequences that can be identified to the level of species and genus. Improvements in the collection, DNA extraction, and the sequencing of eDNA, would help eDNA develop into a more accurate method of taxonomic identification.

Conclusion

There is a significant difference in richness identification between eDNA and microscope, however we do not know enough about the identified taxa to be confident in the eDNA results. Richness data detected by eDNA is not completely reliable, as we do not know if all the specified Arthropoda or Insecta are different species or are just genetic variation within species. By further developing eDNA collection, DNA extraction, sequencing of eDNA, and increasing the confidence level of identification, we can improve the accuracy and resolution of eDNA identification. Additionally, the more eDNA is used in the same aquatic systems and regions, the reference libraries will improve in detecting sequences, and eDNA can improve as an identification method. While microscope data can be limited due to the community sampled

by field collection and time needed to identify the sampled organisms, we found that it currently provides the most reliable taxonomic identification. Ideally, using a combination of the two methods in ecosystems without an existing eDNA library would provide the most comprehensive information on the macroinvertebrates within the ecosystem.

Being able to identify biodiversity is a valuable tool for pond and aquatic management in the southwestern United States. Monitoring biodiversity could help managers and policy makers identify the critical tanks or aquatic systems to upkeep to maintain the matrix of aquatic connectivity or discover a decline in biodiversity and ecosystem services over time in the face of climate change. By having an effective method of measuring biodiversity, managers can monitor how aquatic diversity recovers after an extreme drying event, which is predicted to become more common in the arid southwestern United States (Biggs et al., 2016). Studies from other systems indicate that Environmental DNA, when properly prepared and processed, can be used to assess biodiversity for many samples over many time points, or as an early detection tool for invasive species, or a non-invasive method to record rare or endangered species for these small aquatic ecosystems (Lockwood et al., 2022). This suggests that with further development, eDNA can become a valuable tool with more development for measuring biodiversity in aquatic ecosystems in future climatic changes for Arizona.

REFERENCES

Arizona Legislature . 45-271 - definition of stockpond. (n.d.). Retrieved January 18, 2023, from https://www.azleg.gov/ars/45/00271.htm.

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation genetics*, *17*(1), 1-17.

Biggs, J., von Fumetti, S., & Kelly-Quinn, M. (2016). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia*, *793*(1), 3–39. https://doi.org/10.1007/s10750-016-3007-0.

Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic environmental DNA influenced by substrate type?. *PLoS One*, *12*(8), e0183371.

Céréghino, R., Boix, D., Cauchie, H. M., Martens, K., & Oertli, B. (2013). The ecological role of ponds in a changing world. *Hydrobiologia*, *723*(1), 1–6. https://doi.org/10.1007/s10750-013-1719-y.

Compson, Z. G., Monk, W. A., Hayden, B., Bush, A., O'Malley, Z., Hajibabaei, M., ... & Baird, D. J. (2019). Network-based biomonitoring: exploring freshwater food webs with stable isotope analysis and DNA metabarcoding. *Frontiers in Ecology and Evolution*, *7*, 395.

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: not a perfect match. *Biology Letters*, *10*(9), 20140562. https://doi.org/10.1098/rsbl.2014.0562.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., & Miaud, C. (2011). Persistence of environmental DNA in freshwater ecosystems. *PloS one*, *6*(8), e23398.

Dowle, E. J., Pochon, X., C. Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene

enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, *16*(5), 1240-1254.

Duffy, J. E. (2009). Why biodiversity is important to the functioning of real-world ecosystems. *Frontiers in Ecology and the Environment*, *7*(8), 437–444. https://doi.org/10.1890/070195.

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., ... & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, *111*(22), 8007-8012.

Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PloS one*, *10*(10), e0138432.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., Spear, S.F., McKee, A., Oyler-McCance, S.J., Cornman, R.S., Laramie, M.B., Mahon, A.R., Lance, R.F., Pilliod, D.S., Strickler, K.M., Waits, L.P., Fremier, A.K., Takahara, T., Herder, J.E. and Taberlet, P. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods Ecol Evol. 2016;7: 1299–1307. doi:10.1111/2041-210X.12595.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS one*, *6*(4), e17497.

Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., ... & Hänfling,

B. (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, *826*, 25-41.

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences*, *101*(41), 14812-14817.

Huang, S., Yoshitake, K., Watabe, S., & Asakawa, S. (2022). Environmental DNA study on aquatic ecosystem monitoring and management: Recent advances and prospects. *Journal of Environmental Management*, *323*, 116310.

Hunter, M. E., Ferrante, J. A., Meigs-Friend, G., & Ulmer, A. (2019). Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. *Scientific Reports*, *9*(1), 5259.

Jerde, C. L., & Mahon, A. R. (2015). Improving confidence in environmental DNA species detection.

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... & Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters*, *16*(10), 1245-1257.

Kim, D. K., Park, K., Jo, H., & Kwak, I. S. (2019). Comparison of water sampling between environmental DNA metabarcoding and conventional microscopic identification: a case study in Gwangyang Bay, South Korea. *Applied Sciences*, *9*(16), 3272.

Lemmens, P., Mergeay, J., De Bie, T., Van Wichelen, J., De Meester, L., & Declerck, S. A. J.

(2013). How to Maximally Support Local and Regional Biodiversity in Applied Conservation? Insights from Pond Management. *PLoS ONE, 8*(8), e72538. https://doi.org/10.1371/journal.pone.0072538.

Lockwood, T., Freer, J., Michaelides, K., Brazier, R. E., & Coxon, G. (2022). Assessing the efficacy of offline water storage ponds for natural flood management. *Hydrological Processes, 36*(6), e14618.

Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS One, 9*(11), e113346.

Parsons, W., & Mathews, D. (1990). The Californiazation of Arizona Water Politics. *Natural Resources Journal*, *30*(2), 341–359. http://www.jstor.org/stable/24883576.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2014). Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular ecology resources*, *14*(1), 109-116.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., ... & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849-873.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular ecology*, *21*(8), 1794-1805.

Shu, L., Ludwig, A., & Peng, Z. (2021). Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. *Ecology and Evolution, 11*(12), 8281-8294.

Staehr, P. A., Dahl, K., Buur, H., Göke, C., Sapkota, R., Winding, A., ... & Sundberg, P. (2022). Environmental DNA monitoring of biodiversity hotspots in Danish marine waters. *Frontiers in Marine Science*, *8*, 1993.

Takahashi M, Saccò M, Kestel JH, Nester G, Campbell MA, van der Heyde M, et al. Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. Sci Total Environ. 2023;873: 162322. doi:10.1016/j.scitotenv.2023.162322.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological conservation*, *183*, 4-18.

Thornhill, I. A., Biggs, J., Hill, M. J., Briers, R., Gledhill, D., Wood, P. J., Gee, J. H. R., Ledger, M., & Hassall, C. (2018). The functional response and resilience in small waterbodies along land-use and environmental gradients. *Global Change Biology*, *24*(7), 3079–3092. https://doi.org/10.1111/gcb.14149.

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation, 183*, 93-102.

Wang, J., Soininen, J., & Heino, J. (2021). Ecological indicators for aquatic biodiversity, ecosystem functions, human activities and climate change. *Ecological Indicators, 132*, 108250. https://doi.org/10.1016/j.ecolind.2021.108250.

Yang, J., Zhang, X., Jin, X., Seymour, M., Richter, C., Logares, R., ... & Klymus, K. (2021).

Recent advances in environmental DNA-based biodiversity assessment and conservation. *Diversity and Distributions*, *27*(10), 1876-1879.