2013

DNA Barcode Examination of North American Mayflies Across Their Natural Distribution Reveals Cryptic Species Complexes

Robert J. George
University of Pennsylvania

DNA Barcode Examination of North American Mayflies Across Their Natural Distribution Reveals Cryptic Species Complexes

Abstract
The application of DNA barcoding to distinguish between two or more closely related taxa has been used more frequently in recent years. The typical approach has been to isolate, amplify and sequence cytochrome c oxidase subunit 1 (COI), a mitochondrial DNA (mtDNA) encoded subunit of respiratory Complex IV. COI has proven useful as a marker for identification purposes because the frequency of mutations for this gene is relatively high. Thus, conserved mutations and variability in the COI sequence can be used to determine relatedness of individuals. Many valid criticisms about the practice have arisen, not the least of which is that intraspecific COI variability has not been examined and compared across the entire range of a given species. Rather, most studies employing DNA barcoding focus on relatively few individuals and even fewer sites. Together, this may underestimate intraspecific variance, confounding efforts to distinguish between intraspecific and interspecific differences. Therefore, criteria for delimiting species may need revision. To test whether intraspecific differences at COI are influenced by geographical scale, a COI-barcode library was constructed for three species of North American mayflies (Ephemeroptera) across their natural distribution. This order is important for water quality monitoring of streams and rivers and hence species level identifications have the potential for great application. These three species were chosen because they had relatively wide distributions (i.e., throughout eastern North America) yet were presently considered single species based on morphological characters. Sampling sites included in the study were widespread and represented a range of geographical diversity. Two of three species examined (i.e., *Eurylophella funeralis* and *Leptophlebia cupida*) exhibited genetic differences between individuals that frequently exceeded 2% base pair deviation at the COI locus. There were three or more distinct barcode clusters within each of these two species. Our data suggests that these two species may represent species complexes that are morphologically cryptic. In contrast, genetic differences between individuals for species (*Siphloplecton basale*) did not greatly exceed the 2% base pair deviation at the COI locus. The presence of morphologically cryptic species within *Eurylophella funeralis* and *Leptophlebia cupida* illustrates the need for a robust library of barcodes with morphological vouchers for North American mayflies to resolve the phylogenetics of this group so their contributions in water quality assessments can be maximized.

Disciplines
Environmental Sciences | Physical Sciences and Mathematics

DNA BARCODE EXAMINATION OF NORTH AMERICAN MAYFLIES ACROSS THEIR NATURAL DISTRIBUTION REVEALS CRYPTIC SPECIES COMPLEXES

Robert J. George

Fall 2013

University of Pennsylvania

John K. Jackson, PhD, Stroud Water Research Center
Sarah A. Willig, PhD, University of Pennsylvania

Dedication and Acknowledgements

Dedicated to a true original, the original Robert George.

The Greg and Susan Walker Endowment funded this thesis. I would like to thank: Dr. John Jackson for his guidance and assistance, not the least of which included his honest and constructive introduction to technical and scientific writing; Sally Willig for her counsel and infectious enthusiasm; Dave Funk for his scrupulous curation of aquatic insect specimens that allowed this project to be possible; Jan Battle for helping me solve many data analysis problems; Dr. Teresa Reyes for acknowledging my potential and promoting an atmosphere conducive to maximal personal and professional development; my family for their support, Alyssa for the encouragement and stability necessary to climb the mountain; and last, but certainly not least, my dad and grandmother, who make the sharing of this accomplishment worth the effort.

ABSTRACT

DNA BARCODE EXAMINATION OF NORTH AMERICAN MAYFLIES ACROSS THEIR NATURAL DISTRIBUTION REVEALS CRYPTIC SPECIES COMPLEXES

Robert J. George

The application of DNA barcoding to distinguish between two or more closely related taxa has been used more frequently in recent years. The typical approach has been to isolate, amplify and sequence cytochrome c oxidase subunit 1 (COI), a mitochondrial DNA (mtDNA) encoded subunit of respiratory Complex IV. COI has proven useful as a marker for identification purposes because the frequency of mutations for this gene is relatively high. Thus, conserved mutations and variability in the COI sequence can be used to determine relatedness of individuals. Many valid criticisms about the practice have arisen, not the least of which is that intraspecific COI variability has not been examined and compared across the entire range of a given species. Rather, most studies employing DNA barcoding focus on relatively few individuals and even fewer sites. Together, this may underestimate intraspecific variance, confounding efforts to distinguish between intraspecific and interspecific differences. Therefore, criteria for delimiting species may need revision. To test whether intraspecific differences at COI are influenced by geographical scale, a COI-barcode library was constructed for three species of North American mayflies (Ephemeroptera) across their natural distribution. This order is important for water quality monitoring of streams and rivers and hence species level identifications have the potential for great application. These three species were chosen because they had relatively wide distributions (i.e., throughout eastern North America) yet were presently considered single species based on morphological characters. Sampling sites included in the study were widespread and represented a range of geographical diversity. Two of three species examined (i.e., *Eurylophella funeralis* and *Leptophlebia cupida*) exhibited genetic differences between individuals that frequently exceeded 2% base pair deviation at the COI locus. There were three or more distinct barcode clusters within each of these two species. Our data suggests that these two species may represent species complexes that are morphologically cryptic. In contrast, genetic differences between individuals for species (*Siphloplecton basale*) did not greatly exceed the 2% base pair deviation at the COI locus. The presence of morphologically cryptic species within *Eurylophella funeralis* and *Leptophlebia cupida* illustrates the need for a robust library of barcodes with morphological

vouchers for North American mayflies to resolve the phylogenetics of this group so their contributions in water quality assessments can be maximized.

Table of Contents

List of Tables

List of Figures

Introduction

Genetic identification techniques, such as DNA barcoding have the potential to increase the accuracy of species identifications over traditional, morphological-based identification methods (Sweeney, Battle, Jackson, & Dapkey, 2011). Another benefit of DNA barcoding is its use to discover new and cryptic species (Hajibabaei, Janzen, Burns, Hallwachs, & Hebert, 2006; Pauls, Blahnik, Zhou, Wardwell, & Holzenthal, 2010). The approach is founded upon the theoretical concept that sequence variability within a standardized section of the genome can provide consistent identification at the species level. The thorough analysis of the mitochondrial gene cytochrome b across major vertebrate taxonomic classes led Johns & Avise (1998) to conclude that sister species, congeneric species, and confamilial genera regularly show >2% deviation in base pair sequence at this locus. This finding was further advanced by research that found differences in a 658-bp region in the cytochrome c oxidase subunit 1 (COI) gene, a mitochondrial DNA (mtDNA) encoded subunit of respiratory Complex IV, to be highly effective at distinguishing closely related species (Hebert, Ratnasingham, & deWaard, 2003). Since then, this 658-bp COI locus has been widely accepted as the standardized barcode sequence. With sequences from a variety of organisms accumulating (Janzen et al., 2009), a centralized effort to

catalogue this data has culminated in the International Barcode of Life Project (http://www.barcodeoflife.org).

One area where DNA barcoding has great potential is for its use to identify aquatic insects accurately for water quality assessments (Dapkey, 2008). In a river or stream, benthic macroinvertebrates (e.g., aquatic insects, crayfish, worms) commonly serve as bioindicators that are sensitive to environmental fluctuations and stressors. Therefore, water quality can be assessed indirectly by monitoring these populations (Bauernfeind & Moog, 2000; Cain, Luoma, Carter, & Fend, 1992). Multiple studies have validated the efficacy of using aquatic macroinvertebrates, especially Ephemeroptera (mayfly), Plecoptera (stonefly), and Trichoptera (caddisfly), to qualitatively examine conditions in stream and river ecosystems. One factor that can limit the overall contribution and comparability of these assessments is the accuracy of identifications, especially at the species level. Species identifications are difficult because: (i) morphological traits for distinguishing species (and taxonomic keys to those traits) are often known only for adult (non-larval) stages of most aquatic insect orders, (ii) for some species, including *Eurylophella funeralis*, identifiable traits are only or best expressed by adults, and (iii) even when taxonomic keys for discerning species based on larvae are available, it is often difficult to identify closely related, damaged, or immature specimens (Funk & Sweeney, 1994; Gresens, Belt, Tang, Gwinn, & Banks, 2007). Because each species exhibits varying sensitivity to

environmental stressors, imprecise or inaccurate identifications can compromise the reliability of this technique. As a result, many aquatic insect surveys identify organisms only to the family or order level and this limits their ability to detect change (Jackson et al., 2014). DNA barcoding may provide an ideal solution to the difficulty of identifying larval aquatic insects to the species level.

Concerns have been raised questioning the intraspecific uniformity of the COI locus across spatial distribution (Bergsten et al., 2012). The criticism has focused on the unknown pressure that natural selection may place on the COI locus: is COI conserved independent of natural selection; or does selection strongly influence COI? Further, environmental history, such as periodic glaciation of North America, may result in genetic bottlenecks, founder effect and haplotype refugia (Rich, Light, Hudson, & Ayala, 1998). In addition, studies of insect barcodes have found that the threshold of 2% differences often used to discriminate species may not be a universal cutoff, and some have argued that recent widespread integration of DNA barcoding in molecular taxonomy may be premature (Meier, Shiyang, Vaidya, & Ng, 2006; Meier, Zhang, & Ali, 2008; Meyer & Paulay, 2005).

The purpose of this study was to (i) measure variation in the genetic structure for three mayfly species with wide geographic ranges (i.e., *Eurylophella funeralis, Leptophelbia cupida, Siploplecton basale*), and (ii) test the efficacy of current DNA

barcoding methods by comparing intraspecific uniformity within these species against the 2% genetic distance threshold commonly used to define species with COI. Past work on mayflies using allozyme electrophoresis to better understand the relationship between geographic variation and population genetics have yielded intriguing results. Notably, the data suggested the presence of morphologically cryptic complexes for certain North American ephemerellid species (Funk, Sweeney & Vannote, 1988). In addition, congeneric species of Ephemerellidae have been resolved using older molecular techniques (Sweeney, Funk & Vannote, 1987). Consequently, novel DNA barcoding techniques have the potential to elucidate further the genetic structure of this group while examining the consistency of barcode sequences over a wide geographic range.

Methods & Materials

Study species

To measure genetic variation of aquatic insects using COI barcodes, I selected three univoltine mayfly species that have a broad distribution throughout North America. The three species examined are from three families: Ephemerellidae (*Eurylophella funeralis*), Leptophlebiidae (*Leptophlebia cupida*), and Metretopodidae (*Siphloplecton basale*).

Eurylophella funeralis: occur across eastern North America and can be found in small woodland spring seeps and low-order streams (Sweeney, & Vannote, 1987). Sixty-seven individuals were collected from 9 field locations throughout the species range (Trapper Cabin Creek, Neversink River, Wyalusing Creek, White Clay Creek, Fourpole Creek, Hamrick Run, Slate River, and Panther Creek)(Table 2). To supplement the barcode library generated from field collections, 25 published barcodes (Table 1) were sourced from the BOLD public portal. BOLD automatically assembles submitted barcode sequences that closely resemble species groupings into OTUs or barcode index numbers (BINs). A search for "*Eurylophella funeralis*" yielded one BIN, AAC0324, which contained 25

sequences >350 bp. These published sequences were collected from individuals at 7 additional sites (NS1, NB1, Green River, Crabtree Creek, Monacacy River, Potomac River, and Overflow Creek).

Leptophlebia cupida: are also found in a broad area across eastern North America. Typically, larvae occur in leaf-litter accumulations in streams with slow to intermediate flow (Sweeney, Jackson, Newbold & Funk, 1992). Seventy-eight individuals were collected from 9 field locations (Beaver Creek, Swan River, Tomah Stream, White Clay Creek, Spring Creek, Swan Point Creek, Station Spring Creek, Eno River, and Indian Creek)(Table 3). 48 barcodes were sourced from BOLD. The sequences were organized in 2 BINs: AAB1563 and AAA7018. BIN AAB1563 contained 28 sequences from 3 additional sites (Gravel Run, Potomac River, and Nassawango Creek). BIN AAA7018 contained 20 sequences from 3 additional sites (MB 1, Saskatchewan River, and St. John River).

Siphloplecton basale: are widely distributed across eastern North America and can be found in a variety of benthic habitats, including submerged root mats in slow to fast moving freshwater. Larvae are strong swimmers that consume detritus and hydrophytes (Clifford, 1976). Fifty individuals were collected from 8 sites (Beaver Creek, Penobscot River, Battenkill River, Nescopeck Creek, Blackbird Creek, Sheep Pen Ditch, Meherrin River, and Eno River)(Table 4). There were not any publicly available barcodes on BOLD to supplement our field collections.

Sample collection, preparation, and barcoding

Larvae were collected from 34 streams in North America using kick nets. Using a 12x magnification microscope, an expert entomologist identified each individual using morphological characters. The larvae were kept alive and brought to Stroud Water Research Center, where they were reared to the winged adult stage (Fig. 1). Adult specimens were curated and stored individually at -80°C until sequencing. A leg from each mayfly was removed and placed in a 96-well plate. Each well was filled with 30 µL of ETOH for preservation of nucleic acids. The plates were sent to the Canadian Centre for DNA Barcoding at the University of Guelph, where mtDNA was extracted and the COI locus was amplified and sequenced using a widely accepted protocol (Ivanova, Dewaard, & Hebert, 2006). The sequences were posted on the Barcode of Life Data Systems' (BOLD) workbench along with information about the specimens, including life stage and the collection site.

Data interpretation and analysis

All qualifying barcodes with sequence lengths >350 bp were exported from the BOLD workbench as a FASTA file and imported into MEGA 5.2 (Tamura, et al., 2011). In the MEGA program, the sequences were aligned using the ClustalW operation with default parameters. To examine whether significant OTUs exist for each population, pairwise distances were computed and phylogenetic trees were constructed using the neighbor-joining (NJ) model with pairwise deletion and Kimura-2 parameter distance (K2P) (Kimura, 1980). Bootstrap values were based on 500 replications and were interpreted in accordance with an established methodology (Efron, Halloran, & Holmes, 1996). Branches that surpassed 2% variance were organized into OTUs. *Eurylophella verisimilis* was used as an outgroup while analyzing *Eurylophella funeralis* because of the relatedness and broad distribution of the two species. *Leptophlebia intermedia*, a morphologically distinct congener, was used as an outgroup for *Leptophlebia cupida. Siphloplecton costalense* was used as an outgroup while analyzing *Siphloplecton basale*. Pairwise distances were compared and analyzed to illustrate relatedness of clusters. In order to determine haplotypes present in the data, full sequences (658 bp) were plotted using a NJ tree. Individuals that shared identical sequences were organized into distinct haplotypes. Haplotype abundance was spatially represented using a bubble graph. All graphs produced in this paper were assembled using Graph Pad Prism 5.

Results

DNA barcode analysis

COI sequences >350 bp were obtained for 268 individuals, 195 were from new sequences from individuals collected at the field sites and 73 were mined from previous studies (Table 1). *Leptophlebia cupida* had the greatest number of specimens (126 from 15 sites), followed by *Eurylophella funeralis* (92 from 16 sites), and *Siploplecton basale* (50 from 8 sites) (Tables 2, 3, and 4). Similarities and differences among sequences within each species were examined using K2P pairwise comparisons, phylogenetic trees, and spatial mapping. A gap in the frequency of genetic differences was observed beginning at approximately 2% in two of three species (*Eurylophella funeralis* and *Leptophlebia cupida*)(Fig. 3). The <2% versus >2% difference was used as a genetic threshold to demarcate putative species in the interpretation of the results. Based on the degree of genetic differentiation within and between clusters, barcode sequences were grouped into Operational Taxonomic Units (e.g., OTU 1, OTU 2, OTU 3) and treated as putative species if the genetic difference was >2%. The data retrieved

from BOLD consistently fit into the cluster groupings delimited by new barcodes generated as part of this study. There were not any clusters composed entirely of publicly mined barcodes. Thus, the genetic relationships observed among individuals within a species do not appear to be a function of the source of the genetic data (new versus retrieved).

Eurylophella funeralis

Barcodes from 92 individuals collected from 16 sites that included 12 states and provinces (Table 2) were used to measure genetic variation across the natural distribution of *Eurylophella funeralis*. K2P distance comparing pairs of individuals varied from 0.000-0.169, with a mean of 0.070 and a median of 0.019 (Fig. 2). Further examination of pairwise distances exposed a gap between 2.0 and 10.8% (Fig. 3a). The great distance observed for some pairwise comparisons - up to 16.9% - far exceeded the 2% difference expected for intraspecific variation. A neighbor-joining (NJ) phylogenetic tree was created to determine (i) how many haplotype clusters were represented, (ii) how many individuals were in each cluster, (iii) relatedness of each cluster, (iv) and to analyze the inference that a cryptic species complex had been discovered. Four haplotype clusters were resolved based on the distribution of individuals in the NJ tree (Figs. 4a, 5a). Based on the genetic differences among the clusters, I treated each cluster as a putative species or Operational Taxonomic Unit (OTU). The majority of individuals aggregated into *E. funeralis* OTU 1 (65 individuals, 12 distinct

haplotypes), with the remaining sequences were classified as *E. funeralis* OTU 2 (9 individuals, 3 distinct haplotypes), *E. funeralis* OTU 3 (8 individuals, 3 distinct haplotypes), and *E. funeralis* OTU 4 (10 individuals, 5 distinct haplotypes). Members of *E. funeralis* OTU 1 ranged from Nova Scotia to Georgia (Fig. 6a), which is similar to the present range currently attributed to *E. funeralis*. The remaining three clusters are represented by individuals from single sites - *E. funeralis* OTU 2 was collected only at Station Spring Creek in Virginia, *E. funeralis* OTU 3 was collected only at Hamrick Run in West Virgina, and *E. funeralis* OTU 4 was collected only at Fourpole Creek in West Virginia. Thus, in combination with *E. funeralis* OTU 1 collected at Slate River in Virginia, these results show that all four putative species currently identified as *E. funeralis* were collected in the geographic region represented by West Virginia and Virginia. To identify distinct haplotypes across all field sites, full COI sequences (53 out of 92 were 658 bp in length, 57.6%) were used to construct a NJ tree (Fig. 5a). 11 distinct haplotypes were present in *E. funeralis* OTU 1, while the remaining 11 haplotypes, located in the geographic region represented by West Virginia and Virginia, were present for three putative species (*E. funeralis* OTU 2- 3 distinct haplotypes, *E. funeralis* OTU 3- 3 distinct haplotypes, and *E. funeralis* OTU 4- 5 distinct haplotypes). To approximate the temporal span between these putative species, the NJ trees were subject to molecular clock analysis. Nucelotide substitution rates for insects have been estimated to range between 1.5% (Farrell, 2001) and 3.54% (Papadopoulou, Anastasiou & Vogler, 2010) million years $^{-1}$.

Both approximations were included in the analysis. The model suggests that the four clusters diverged around 2.2-5.0 million years ago, with more recent divergences between the putative species occurring around 1.7-4 million years ago.

Leptophlebia cupida

Barcodes from 126 individuals collected across 15 sites that included 12 states and provinces were used to measure genetic variation across the natural distribution of currently identified as *Leptophlebia cupida* sensu lato (Table 3). K2P distance comparing pairs of individuals varied from 0.000-0.117, with a mean of 0.046 and a median of 0.057 (2b). The distribution of pairwise distances showed a gap between 1.7 and 4.6% (Fig. 3b), with the majority of distances >2% falling between 4.6-6.4%. The NJ tree distinguished 3 distinct clusters with >2% deviance relative to other clusters (Fig. 4b). The majority of individuals aggregated into in *L. cupida* OTU 1 (70 individuals, 7 distinct haplotypes), but *L. cupida* OTU 2 (39 individuals, 5 distinct haplotypes) and *L. cupida* OTU 3 (17 individuals, 4 distinct haplotypes) were both represented by numerous individuals. Unlike *Eurylophella funeralis*, each cluster was represented by individuals from more than one site. There were not any sites with >1 OTU present. Individuals clustering together as *L. cupida* OTU 1 were collected at 8 sites distributed east of the Appalachian Mountains from southeastern

Pennsylvania to South Carolina. Individuals clustering together as *L. cupida* OTU 2 were most closely related to OTU 1, and were from five northern sites ranging from the St. John River in Nova Scotia to Swan River in Montana. Individuals clustering together as to *L. cupida* OTU 3 were collected at two sites: Station Spring Creek in Virginia and Beaver Creek in Quebec. Thus, two of the three putative species currently identified as *L. cupida* appear geographically isolated, with *L. cupida* OTU 1 having a more southern distribution than *L. cupida* OTU 2. It is more difficult to characterize the distribution of *L. cupida* OTU 3 with data from only 2 sites. To identify distinct haplotypes across all field sites, full COI sequences (81 out of 126 were 658 bp in length, 64.3%) were used to construct a NJ tree (Fig. 5b). Calibration of the molecular clock placed the divergence between *L. cupida* OTU 1 and *L. cupida* OTU 2 at 0.9-2 million years, while *L. cupida* OTU 3 diverged much earlier (1.5-3.7 million years ago).

Siphloplecton basale

Barcodes from 50 individuals collected across 8 sites that included 7 states and provinces (Table 4) were used to identify genetically distinct OTUs across the natural distribution of *Siphloplecton basale*. The number of individuals per site ranged from 2-9, with a mean of 6.25 individuals per site (Table 4). All of these data are new as there were not any publicly available barcodes for *S. basale*. K2P distance comparing pairs of individuals varied from 0.000-0.036, with a mean of

0.011 and a median of 0.005 (Fig. 2). Unlike *Eurylophella funeralis* and *Leptophelbia cupida*, there was no clear gap in the frequency of pairwise distances. The Beaver Creek population (n=9) displayed the greatest amount of genetic divergence in the dataset. For example, all of the pairwise comparisons that were greater than the 2% threshold consisted entirely of comparisons of individuals from Beaver Creek with individuals from other sites (Fig. 3c); these differences ranged from (0.020-0.036). This population did not yield any full sequences (658 bp in length). The remaining pairwise comparisons had less than <2% deviation. The NJ tree analysis did not indicate that Beaver Creek population should be excluded or split from the main S*iphloplecton basale* cluster. Thus, I concluded that all individuals of *S. basale* examined should be considered members of a single species (Fig 4c) that is relatively consistent across a wide geographic range (Fig. 5c). Unlike *Eurylophella funeralis* and *Leptophelbia cupida*, the data for *S. basale* do not suggest the presence of morphologically cryptic species.

Morphologically cryptic species revealed by barcoding

Two of the of North American mayfly species (i.e., *Eurylophella funeralis* and *Leptophelbia cupida*) that I studied had high intraspecific differences among sites across the wide spatial distribution sampled. Pairwise distance comparisons for *Eurylophella funeralis* and *Leptophlebia cupida* had a mean distance of 7.0 and 4.6%, respectively. Every specimen for each species, regardless of source (e.g. sampled

by our group vs. data mined from BOLD), shared conserved sequence differences that allowed for accurate appointment to an OTU on our NJ trees. Sequence divergence within each cluster was <2%, with the exception of *Siphloplecton basale*, which did contain one population that produced genetic differences up to 3.6%. There were not any outliers in the entire barcode library and bootstrap values for divergences were consistently high. Thus, the data are consistent and conforms to past findings supporting the use of barcoding as an effective species identification technique. Adopting this conclusion, the data can be considered evidence for the discovery of a cryptic species assemblage for both *Eurylophella funeralis* and *Leptophlebia cupida. Leptophlebia cupida* sensu lato is represented by three genetically distinct, but morphologically similar, populations. One *L. cupida* OTU was more common than the others, but all three were represented by individuals collected at more than one site. Molecular clock analysis suggests that these putative species have been genetically isolated for >900,000 years. Similarly, *E. funeralis* sensu lato is represented by four genetically distinct populations that are morphologically similar. One *E. funeralis* OTU was much more common than the others, with three of the four clusters represented by individuals from only one site. Molecular clock analysis suggests that these putative species have been genetically isolated for >2.2 million years. *Siphloplecton basale* collected across a range of watersheds similar to those sampled for *E. funeralis* and *L. cupida* did not exhibit evidence for

morphologically cryptic species similar to that observed for *E. funeralis* and *L. cupida*.

Discussion

The species problem and utility of DNA barcoding

Efforts to organize the natural world by establishing hierarchical orders culminating in the discipline of taxonomy have long struggled with the inherently dynamic behavior of life. Species are no exception. The most successful attempt to address the issue of defining species began when John Ray proposed the core idea of what is now know as the Biological Species Concept. He formulated that, "animals likewise that differ specifically preserve their distinct species permanently; one species never springs from the seed of another nor vice versa" (Ray, 1686). This description evolved and was refined by Ernst Mayr into the most widely accepted definition of a species- "groups of actually or potentially interbreeding populations, which are reproductively isolated from other such groups" (Mayr, 1942). Isolation in this context can be further divided into four categories: (i) spatial, (ii) temporal, (iii) behavioral, or (iv) anatomical. Past studies, using two species of grasshopper, have manipulated behavioral

isolation resulting in production of a highly viable line of hybrids (Gottsberger, 2007). This violation confounds the simplicity of the Biological Species Concept. Further, because morphological change is not always associated with speciation (e.g. cryptic species complexes), it can be difficult to estimate species diversity. Herein lies the appeal of genetic analysis, objective and quantifiable results may serve as a grand solution to a dilemma that has bemused biologists and taxonomists. A core concept of my study was to examine usefulness and reliability of COI barcoding to identify mayfly species. My results are consistent with past findings that support the efficacy of barcoding to accurately identify species accurately. All OTUs consistently displayed <2% divergence. High bootstrap values associated with the NJ trees support the clustering patterns that lead to these general conclusions. Further, in the entire library of 268 specimens, there were not any singletons (i.e., a genetically unique cluster represented by a single individual that could appear as an outlier); the lowest number of individuals in an OTU was 7. Lastly, there were not any sites that supported individuals from more than 1 cluster in that species (i.e., the OTUs were never sympatric). Therefore, the examples we have presented here support the use of DNA barcoding as a valid means to explore the genetic structure of North American mayflies.

Discovery of cryptic species complexes within North American mayfly populations

Many insect species remain undescribed worldwide. One challenge in the effort to quantify species diversity is the presence of morphologically cryptic species complexes that represent a group of closely related species that cannot be easily resolved using traditional morphological characters. The introduction of barcoding has given researchers a new set of characters to use to detect and identify these organisms. In the last decade, many cryptic species complexes have been discovered or resolved for insect populations considered a single, polymorphic species or possibly a problematic species (Bickford et al., 2007; Hebert, Penton, Burns, Janzen, & Hallwachs, 2004). Recent analysis of diversity for aquatic insect communities demonstrates more diversity than previously thought (Jackson et al., 2014). The results presented here support the conclusion that *Eurylophella funeralis* and *Leptophlebia cupida* sensu lato are both a species complex. The levels of COI diversity observed suggest the presence of four distinct species for *Eurylophella funeralis* and three distinct species for *Leptophlebia cupida*.

Moving forward

Because whole genome sequencing is cost prohibitive, it is understandable why sequencing a small, stable portion of the genome to resolve species has become widespread. A limitation of this approach is the inability to parse out details of gene flow between isolated populations. Still, it is useful in identifying novel

populations that have undergone speciation but have not yet developed distinct morphological characters (or we have not yet recognized derived traits as species specific). Thus, our data show the potential that COI barcoding has for addressing issues relating to North American mayfly biology, ecology, and phylogenetics. It is likely that whole-genome sequencing will soon be cheap and accessible. Whether it will be a superior tool for examining genetic structure remains to be examined.

If the approach that all populations exhibiting >2% deviation are to be universally declared as new species is to be widely adopted, this has the potential to generate enormous challenge for taxonomists worldwide. Each newly discovered species will need to be formally described morphologically (or at least genetically), which is not a trivial undertaking given the staggering diversity that insects already display under current taxonomic regulations. In order to aid in solving this challenge, a more robust library of morphological and genetic vouchers for North American mayflies must be assembled and curated.

Lastly, to assist with the issues that this rapidly growing field faces, a uniform and objective approach should be universally adopted. Recent analysis of four common species delimiting techniques found inconsistencies and conflicts among the differing techniques (White, Pilgrim, Boykin, Stien & Mazor, 2014). More data such as were generated in this study are needed to improve objectivity

in the interpretation of COI barcodes, especially for closely related species that can be a challenge morphologically and genetically.

Tables

Table 1. Sources of mayfly specimens used to construct the COI barcode library.

Species	Collected	Mined	n
Eurylophella funeralis	67	25	92
Leptophlebia cupida	78	48	126
Siphloplecton basale	50	0	50

Table 2. Collection sites of *Eurylophella funeralis* samples included in COI library.

Site	State/Province	Latitude	Longitude	(n)	658 bp	Haplotypes
Trapper Cabin Creek [a]	QC	50.3208	-65.9606	10	8	101
NS1 [b]	NS	46.713	-60.383	2	2	101,111
NB1 [b]	NB	45.9184	-66.6405	5	2	102
Green River [b]	VT	43.119	-73.208	2	0	NA
Neversink River [a]	NY	41.902	-74.581	5	0	NA
Wyalusing Creek [a]	PA	41.8197	-75.9333	8	7	108,112
White Clay Creek [a]	PA	39.8631	-75.7853	14	9	103-106
Crabtree Creek [b]	MD	39.495	-79.166	1	0	NA
Monacacy River [b]	MD	39.201	-77.401	2	0	NA
Potomac River [b]	MD	38.436	-77.252	8	0	NA
HF Fourpole Creek [a]	WV	38.3883	-82.4356	9	6	401-404
Hamrick Run [a]	WV	38.2281	-80.4011	9	7	301-303
Slate River [a]	VA	37.474	-78.658	2	1	107
Station Spring Creek [a]	VA	37.0872	-81.4022	9	7	201-203
EF Overflow Creek [b]	NC	35.0182	-83.2445	5	3	109,110
Panther Creek [a]	GA	34.673	-83.355	1	1	111

[a] From specimens collected by the Stroud Water Research Center

[b] Sourced from BOLD BIN: AAC0324

Table 3. Collection sites of *Leptophlebia cupida* samples included in COI library.

Site	State/Province	Latitude	Longitude	(n)	658 bp	Haplotypes
MB1 [b]	MB	58.663	-94.167	1	1	24
Saskatchewan River [b]	SK	53.017	-105.578	5	5	24,25
Beaver Creek [a]	QC	50.3061	-65.95483	10	5	31
Swan River [a]	MT	48.0801	-114.0197	10	10	21,25
St. John River [b]	NB	45.976	-66.719	14	12	22,24
Tomah Stream [a]	ME	45.4718	-67.5928	9	8	23,25
White Clay Creek [a]	PA	39.8631	-75.7853	11	9	16
Gravel Run [c]	MD	39.039	-76.047	2	0	NA
Spring Creek [a]	DE	39.0103	-75.5297	9	9	13,16
Swan Point Creek [a]	MD	38.759	-76.708	2	0	NA
Potomac River [c]	MD	38.43	-77.251	16	0	NA
Nassawango Creek [c]	MD	38.263	-75.462	10	0	NA
Station Spring Creek [a]	VA	37.095	-81.3811	7	6	32
WF Eno River [a]	NC	36.1392	-79.1703	8	8	11-13,15,16
Indian Creek [a]	SC	34.2979	-79.8878	12	8	11,12,14,15

[a] From specimens collected by the Stroud Water Research Center

[b] Sourced from BOLD BIN: AAB1563

[c] Sourced from BOLD BIN: AAA7018

Table 4. Collection sites of *Siphloplecton basale* samples included in COI library. All specimens were from the collections of the Stroud Water Research Center.

Site	State/Province	Latitude	Longitude	Samples (n)	658 bp	Haplotypes
Beaver Creek	QC	50.3061	-65.9548	9	0	NA
Penobscot River	ME	45.7789	-69.0126	2	0	NA
Battenkill River	VT	43.0978	-73.1419	3	0	NA

Nescopeck Creek	PA	41.0583	-75.9311	8	2	104
Blackbird Creek	DE	39.3425	-75.6843	6	1	102
Sheep Pen Ditch	DE	38.6032	-75.3208	5	4	101,102
Meherrin River	VA	36.7028	-77.5864	9	5	101,105,106
WF Eno River	NC	36.1392	-79.1703	8	3	101,103

Figures

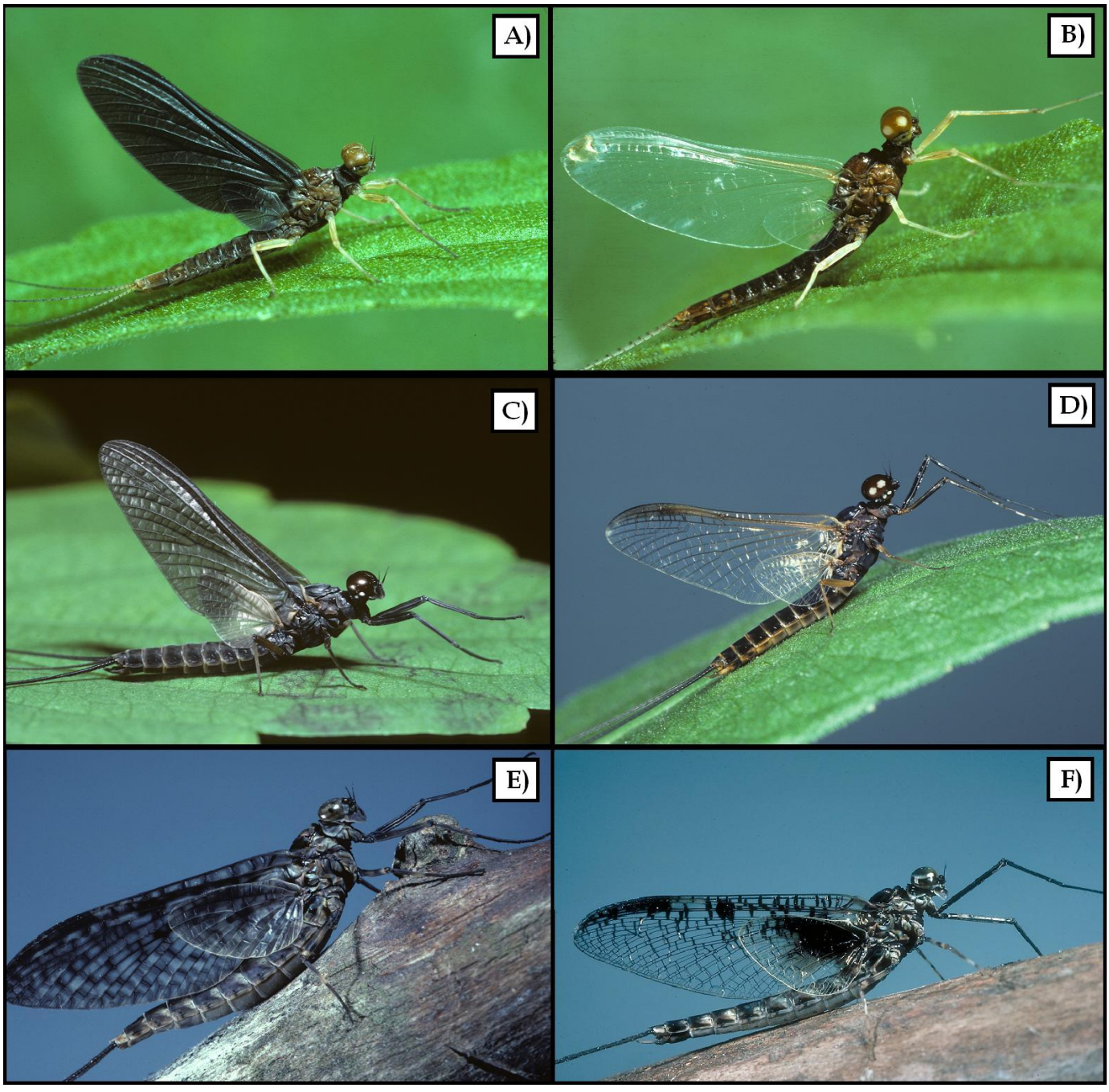

Figure 1. Pictures of 6 specimens submitted for COI sequencing. (A) Subimago of *Eurylophella funeralis* and (B) imago of *Eurylophella funeralis*. (C) Subimago of *Leptophlebia cupida* and (D) imago of *Leptophlebia cupida*. (E) Subimago of *Siphloplecton basale* and (F) imago of *Siphloplecton basale*. Pictures courtesy of D. H. Funk.

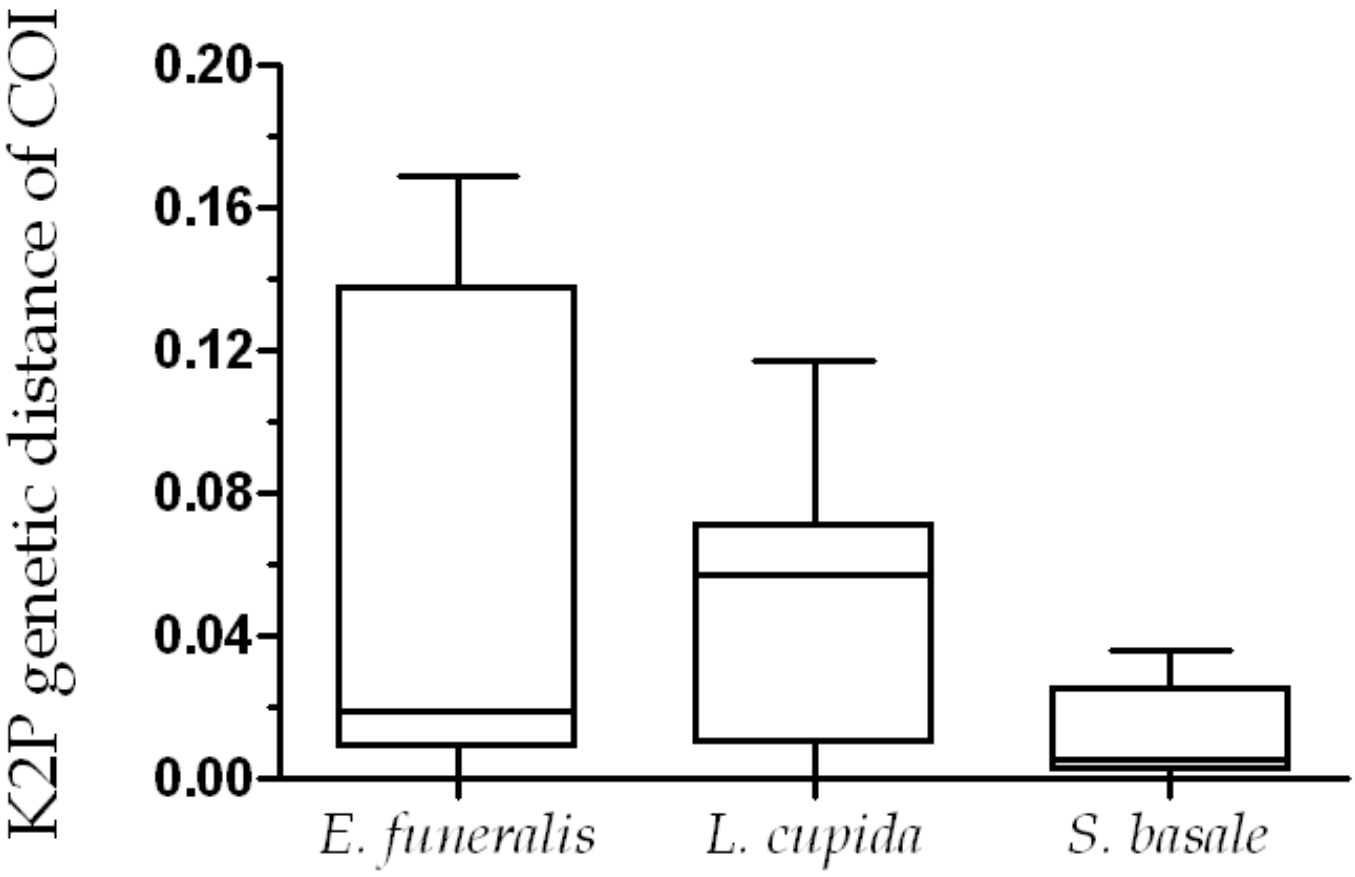

Figure 2. Box plots depicting differences between the minimum interspecific variation and the maximum interspecific variation using the K2P distance model.

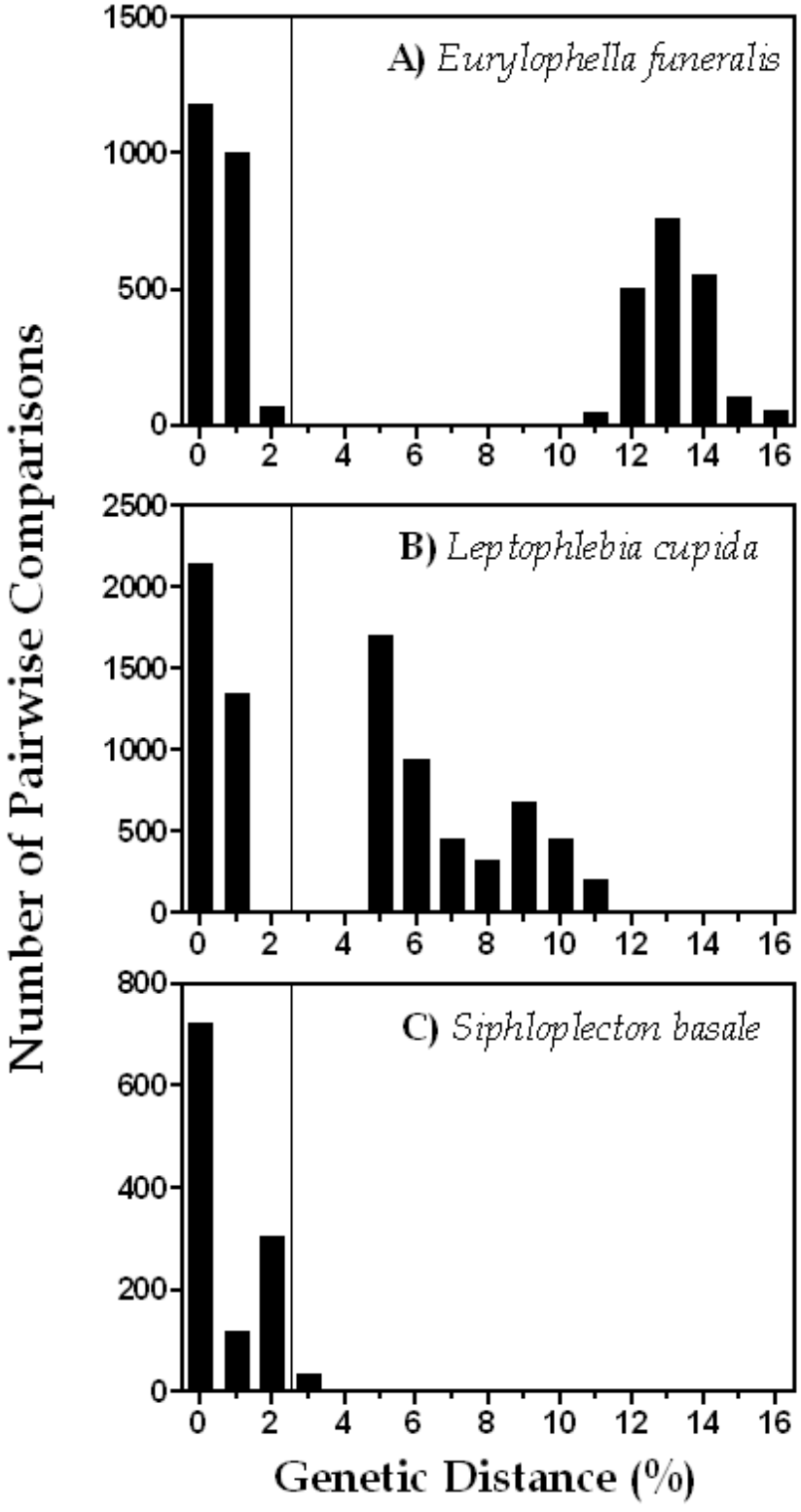
A) Eurylophella funeralis
B) Leptophlebia cupida
C) Siphloplecton basale
Number of Pairwise Comparisons
Genetic Distance (%)

Figure 3. Number of pairwise comparisons versus % genetic distance for A) *Eurylophella funeralis*, B) *Leptophlebia cupida,* and C) *Siphloplecton basale*. Solid black line at 2% genetic divergence represents adopted intraspecific threshold.

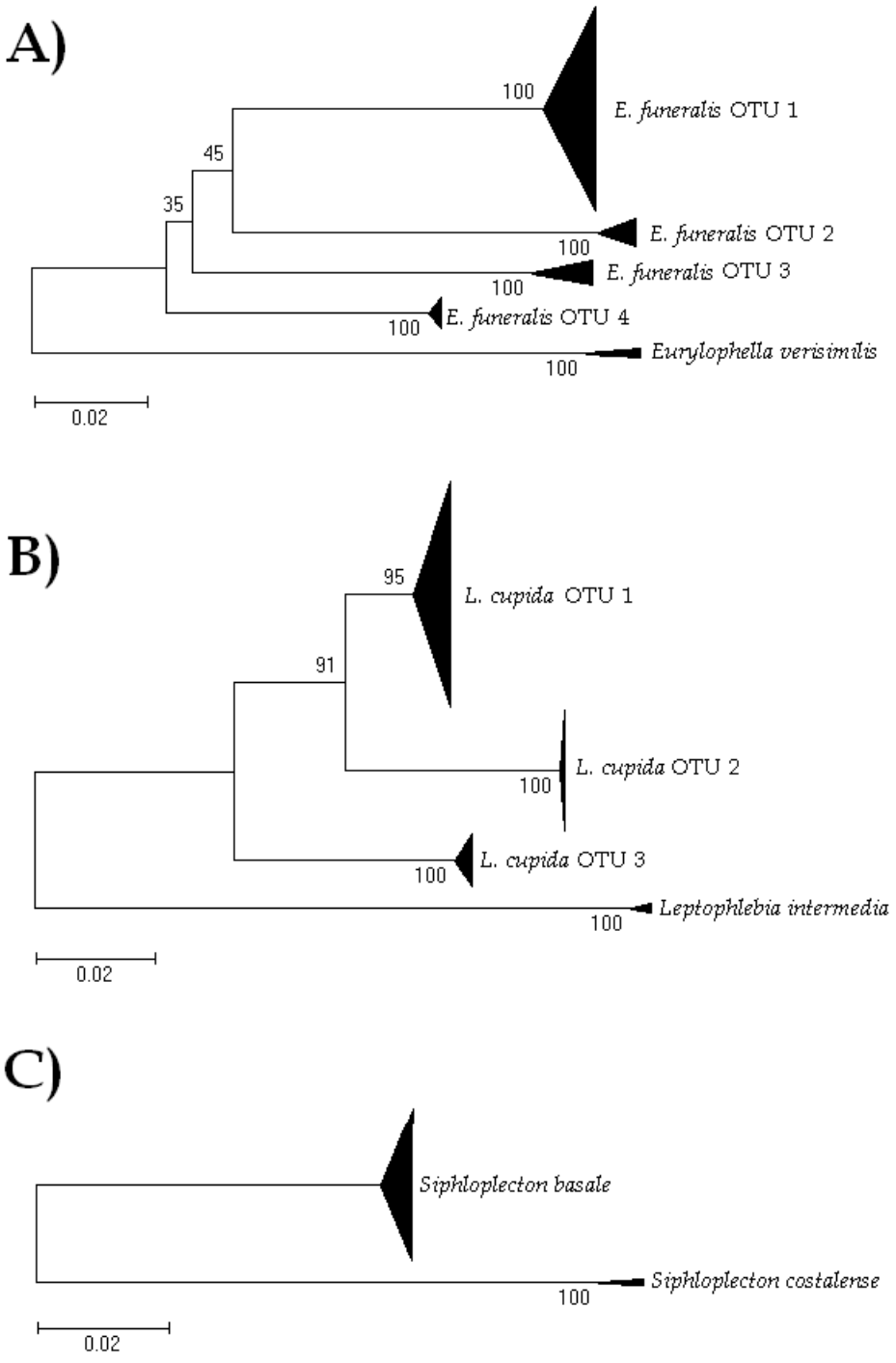

Figure 4. Neighbor-joining tree constructed using barcodes (>350 bp) for three species of North American mayfly. (A) Four distinct OTUs or clusters (>2% genetic divergence) were discovered for *Eurylophella funeralis* (A) and Three distinct OTUs were discovered for *Leptophlebia cupida*. All of our *Siphloplecton basale* (C) barcodes were grouped into one cluster. Bootstrap values and genetic distance (K2P) are listed. Bold triangles represent amount of divergence within branch along x-axis and proportion of individuals on y-axis.

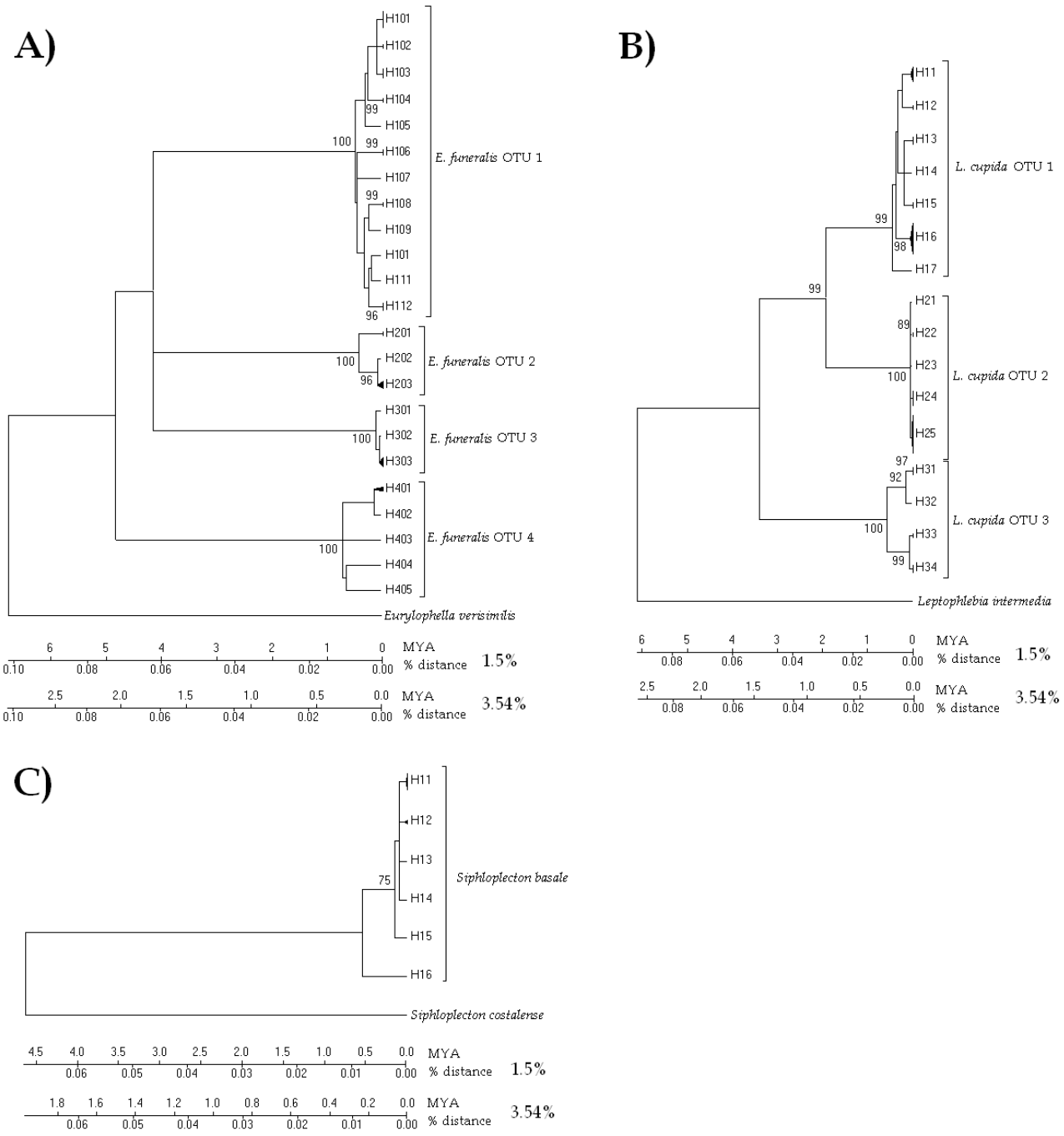

Figure 5. Detailed neighbor-joining tree constructed using full COI barcodes (658 bp) for three species of North American mayfly. Haplotypes, determined by aggregations of identical sequences, are listed for each species (first number indicates OTU classification). Time to most recent common ancestor (TMRCA) is estimated using two, 1.5% (Farrel, 2001) and 3.54% (Papadopolou, Anastasiou, & Vogler, 2010), mtDNA substitution rates per million years.

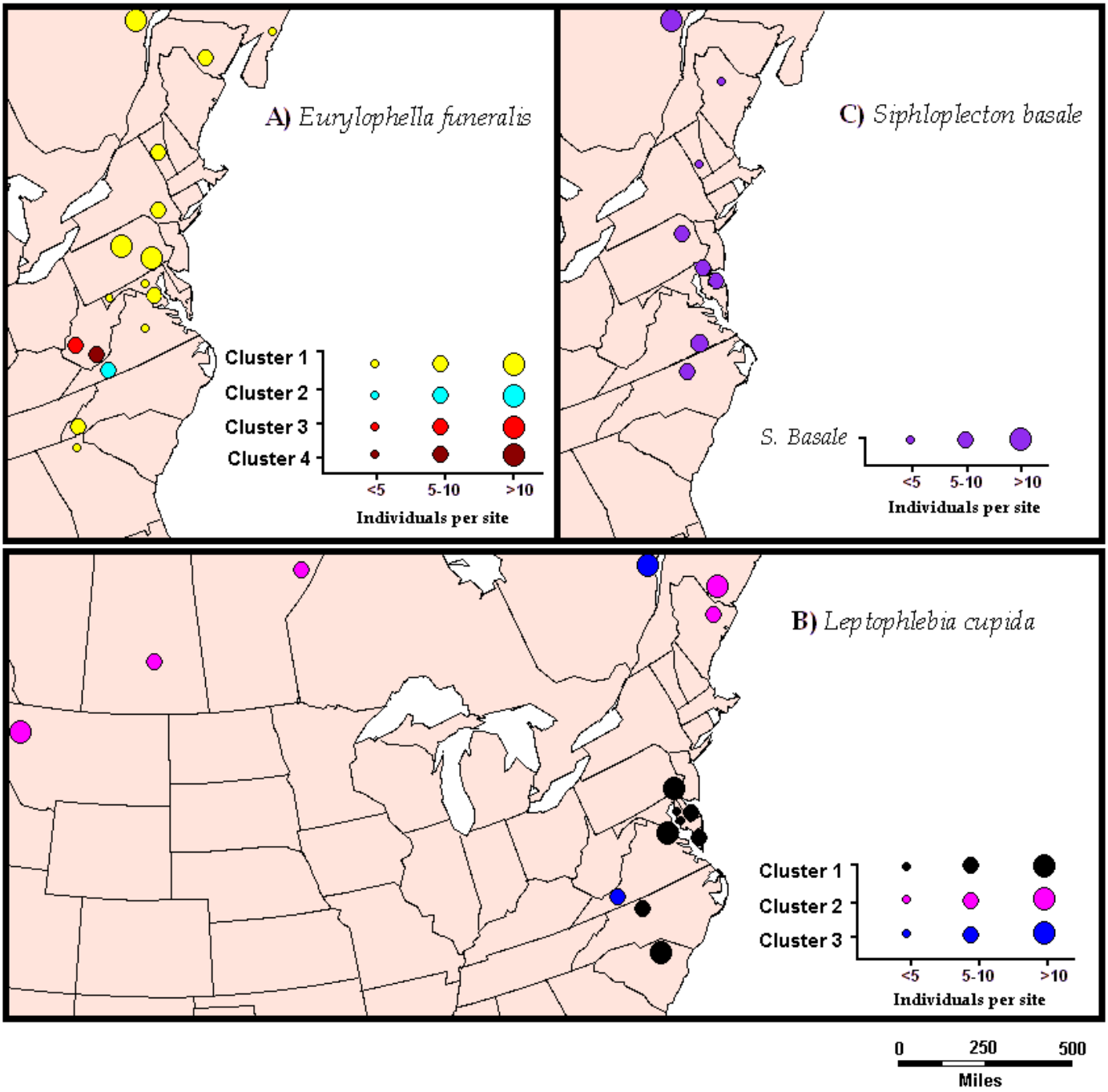

Figure 6. Bubble graph representing the location and abundance of barcodes for populations of (A) *Eurylophella funeralis*, (B) *Leptophlebia cupida*, and (C) *Siphloplecton basale.* The size of each bubble is equivalent to the (n) at each site. Cluster designation is color-coded. All three maps are equal in scale.

Literature Cited

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications, 4*, 1892. doi:10.1038/ncomms2881

Bauernfeind, E., & Moog, O. (2000). Mayflies (Insecta : Ephemeroptera) and the assessment of ecological integrity: A methodological approach. *Hydrobiologia, 422*, 71-83. doi:10.1023/A:1017090504518

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., . . . Vogler, A. P. (2012). The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology, 61*(5), 851-869. doi:10.1093/sysbio/sys037

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., . . . Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution, 22*(3), 148-155. doi:10.1016/j.tree.2006.11.004

Cain, D., Luoma, S., Carter, J., & Fend, S. (1992). Aquatic insects as bioindicators of trace-element contamination in cobble-bottom rivers and streams. *Canadian Journal of Fisheries and Aquatic Sciences, 49*(10), 2141-2154.

Clifford, H. F. (1976). Observations on the life cycle of *Siphloplecton basale* (Walker)(Ephemeroptera: Metretopodidae)[insects, Alberta]. *Pan-Pacific Entomologist*, *52*(4), 265-271.

Dapkey, T. (2008). Combining DNA barcoding and macroinvertebrate sampling to assess water quality. Master of Environmental Studies Capstone Projects, Retrieved from http://repository.upenn.edu/mes_capstones/37/

Efron, B., Halloran, E., & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees (vol 93, pg 7085, 1996). *Proceedings of the National Academy of Sciences of the United States of America, 93*(23), 13429-13434. doi:10.1073/pnas.93.23.13429

Farrell, B. D. (2001). Evolutionary assembly of the milkweed fauna: Cytochrome oxidase i and the age of Tetraopes beetles. *Molecular Phylogenetics and Evolution*, *18*(3), 467-478. doi: doi:10.1006/mpev.2000.0888

Funk, D. H., Sweeney, B. W., & Vannote, R. L. (1988). Electrophoretic study of eastern North American *Eurylophella* with the discovery of morphologically cryptic species. *Entomological Society of America*, *81*(2), 174-186.

Funk, D., & Sweeney, B. (1994). The larvae of eastern North-American Eurylophella-tiensuu (Ephemeroptera, Ephemerellidae). *Transactions of the American Entomological Society*, 120(3), 209-286.

Gottsberger, B. (2007). Interspecific hybridization between the grasshoppers *Chorthippus biguttulus* and *C. brunneus* (Acrididae; Gomphocerinae). (Doctoral dissertation).

Gresens, S. E., Belt, K. T., Tang, J. A., Gwinn, D. C., & Banks, P. A. (2007). Temporal and spatial responses of Chironomidae (diptera) and other benthic invertebrates to urban stormwater runoff. *Hydrobiologia, 575*, 173-190. doi:10.1007/s10750-006-0366-y

Hajibabaei, M., Janzen, D., Burns, J., Hallwachs, W., & Hebert, P. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America, 103*(4), 968-971. doi:10.1073/pnas.0510466103

Hebert, P., Penton, E., Burns, J., Janzen, D., & 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 of the United States of America, 101*(41), 14812-14817. doi:10.1073/pnas.0406166101

Hebert, P., Ratnasingham, S., & deWaard, J. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B-Biological Sciences, 270*, S96-S99. doi:10.1098/rsbl.2003.0025

Ivanova, N. V., Dewaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes, 6*(4), 998-1002. doi:10.1111/j.1471-8286.2006.01428.x

Jackson, J., Battle, J., White, B., Pilgrim, E., Miller, P., & Sweeney, B. (2014). Cryptic biodiversity in streams – a comparison of macroinvertebrate communities based on morphological and DNA barcode identifications. *Freshwater Science*.

Janzen, D. H., Hallwachs, W., Blandin, P., Burns, J. M., Cadiou, J., Chacon, I., . . . Wilson, J. J. (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources, 9*, 1-26. doi:10.1111/j.1755-0998.2009.02628.x

Janzen, D. H., Hallwachs, W., Harvey, D. J., Darrow, K., Rougerie, R., Hajibabaei, M., . . . Hebert, P. D. N. (2012). What happens to the traditional taxonomy when a well-known tropical Saturniid moth fauna is DNA barcoded? *Invertebrate Systematics, 26*(5-6), 478-505. doi:10.1071/IS12038

Johns, G., & Avise, J. (1998). A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution, 15*(11), 1481-1490.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution, 16*(2), 111-120. doi:10.1007/BF01731581

Mayr, E. (1942). Systematics and the origin of species, from the viewpoint of a zoologist. Cambridge: Harvard University Press.

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology, 55*(5), 715-728. doi:10.1080/10635150600969864

Meier, R., Zhang, G., & Ali, F. (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology, 57*(5), 809-813. doi:10.1080/10635150802406343

Meyer, C., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLOS Biology, 3*(12), 2229-2238. doi:10.1371/journal.pbio.0030422

Papadopoulou, A., Anastasiou, I., & Vogler, A. (2010). Revisiting the insect mitochondrial molecular clock: The Mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27*(7), 1659-1672. doi: doi:10.1093/molbev/msq051

Pauls, S. U., Blahnik, R. J., Zhou, X., Wardwell, C. T., & Holzenthal, R. W. (2010). DNA barcode data confirm new species and reveal cryptic diversity in

Chilean Smicridea (smicridea) (trichoptera:Hydropsychidae). *Journal of the North American Benthological Society, 29*(3), 1058-1074. doi:10.1899/09-108.1

Ray, J. (1686). *Historia Plantarum*.

Rich, S., Light, M., Hudson, R., & Ayala, F. (1998). Malaria's eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America, 95*(8), 4425-4430. doi: 10.1073/pnas.95.8.4425

Smith, M. A., Fernandez-Triana, J. L., Eveleigh, E., Gomez, J., Guclu, C., Hallwachs, W., . . . Zaldivar-Riveron, A. (2013). DNA barcoding and the taxonomy of Microgastrinae wasps (Hymenoptera, Braconidae): Impacts after 8 years and nearly 20,000 sequences. *Molecular Ecology Resources, 13*(2), 168-176. doi:10.1111/1755-0998.12038

Sweeney, B. W., Battle, J. M., Jackson, J. K., & Dapkey, T. (2011). Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? *Journal of the North American Benthological Society, 30*(1), 195-216. doi:10.1899/10-016.1

Sweeney, B., Jackson, J., Newbold, J., & Funk, D. (1992). Climate change and the life histories and biogeography of aquatic insects in eastern North America. 143-176.

Sweeney, B. W., & Vannote, R. L. (1987). Geographic parthenogenesis in the stream mayfly *Eurylophella funerals* in eastern North America. *Holarctic Ecology*, *10*(1), 52-59.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*(10), 2731-2739. doi:10.1093/molbev/msr1

White, B., Pilgrim, E., Boykin, L., Stien, E., & Mazor, R. (2014). Comparing four species delimitation methods applied to a DNA barcode data set of insect larvae for use in routine bioassessment. *Freshwater Science*.