Gene Sequencing as a Tool for Identifying Native and Nonindigenous Sphaeriid Clams in Lake Erie

Final Report

Submitted in completion of

Lake Erie Protection Fund Grant SG 499-2015

Submitted by

Kylee L. Spencer[1], Ph.D, Co-Principal Investigator

Jakob A. Boehler[2], B.S., Co-Principal Investigator

Kenneth A. Krieger[2], Ph.D., Project Director

[1]Department of Biological and Environmental Sciences and

[2]National Center for Water Quality Research

Heidelberg University

Tiffin, OH 44883

to

Sandra Kosek-Sills, Ph.D., Project Officer

Ohio Lake Erie Commission

347 North Dunbridge Road

Bowling Green, Ohio 43402

28 July 2016

[blank page for 2-sided printing]

Gene Sequencing as a Tool for Identifying Native and Nonindigenous Sphaeriid Clams in Lake Erie

Executive Summary

The health of the Lake Erie ecosystem, and therefore the quality of the many services the lake can provide to Ohio citizens, is strongly reflected in the health and diversity of the biological communities in the lake. For this reason, the Ohio EPA and Ohio DNR, along with other states and the province of Ontario conduct surveys of the fish, invertebrate and algal communities of Lake Erie. Knowledge of the many species that make up those communities provides crucial information about the amount and quality of food resources for the sport and commercial fisheries, the health of the fisheries themselves, and changes in the status of exotic species. Aquatic invertebrate communities, which are near the bottom of the fisheries food chain, are very useful for monitoring lake conditions, and in fact, certain invertebrates such as native burrowing mayflies and the exotic invasive zebra and quagga mussels, play an outsized role in the Lake Erie food web and as well have directly influenced human activities on the lake.

Non-native species, whether aquatic invertebrates or fishes, can have rapid and deleterious impacts on the lake ecosystem, as demonstrated by the invasion of the zebra mussel in the 1980s followed soon thereafter by the quagga mussel. Many other aquatic invertebrates have invaded Lake Erie over the past few decades, including various crustaceans living in the plankton and on the bottom as well as bottom-dwelling worms, snails and clams. Among the clams is the family Sphaeriidae, consisting of tiny "pill clams" and "fingernail clams", most of which grow no larger than one-eighth to one-quarter inch. Sphaeriid clams are common, occasionally abundant, in the bottom mud of Lake Erie and also in its coastal wetlands and tributary streams. Along with their much larger relatives, the freshwater mussels, they filter microscopic food particles from the water, thereby filtering and helping to clarify the water and also serving as a link in the food chain of a variety of Lake Erie fishes.

Despite their nearly ubiquitous presence and their role in the Lake Erie food web, the role of individual kinds of sphaeriid clams is not well studied. This is in part because of their very small size and in part because of the difficulty of identifying which of the various species are present in a particular locality in the lake. Several non-native (nonindigenous) pill and fingernail clams have invaded Lake Erie (Mackie et al. 1980), but the extent to which they have spread throughout the lake and how and to what extent each species has impacted the food web or the bottom habitat are poorly studied. Furthermore, additional non-native pill clams and fingernail clams have the potential to invade the lake with unknown consequences.

Because of the tedium and difficulty involved in the identification of sphaeriid clams, most Ohio surveys of bottom-dwelling (benthic) macroinvertebrates in Lake Erie and rivers and wetlands throughout the state over the past four decades or more have only identified pill and fingernail clams to the level of family, or at most, genus. The few available published taxonomic keys for identification of sphaeriid clams in the Great Lakes (Burch 1975, Mackie et al. 1980) are outdated and somewhat ambiguous, and

the keys cannot be used to identify the very small immature individuals that often make up most or all of Lake Erie collections. However, the advent and increasing application of gene sequencing technology to a growing array of aquatic invertebrates provides a new approach to species identification. Its application to the identification of sphaeriid clams in Lake Erie and its watershed holds promise to make their identification quicker and more accurate, while the costs of gene sequencing may be offset by the time saved by eliminating traditional tedious morphological identifications. The primary purpose of this study was to investigate the potential of gene sequencing for providing a fast and reliable method of identification as an alternative to traditional morphological taxonomy for determining the distribution and abundance of both native and exotic sphaeriid clams in Lake Erie. We predicted that gene sequencing would prove to be the more economical and accurate of the two approaches. This project addressed the strategic objective of the Lake Erie Protection and Restoration Plan (LEPR 13) to "develop an early detection strategy for invasive species" within Priority Area: Invasive Species (p. 12.).

Clam specimens used in the study were collected mostly from western Lake Erie and several marshes and tributaries of the western basin in 2015 and 2016. Additional specimens were collected in 2015 from the central basin of Lake Erie and from sites in lakes Huron, Ontario and Superior. A few specimens were derived from archived samples from earlier projects in the Sandusky River and Olentangy (Ohio River Basin) watersheds. In all, 160 specimens were examined both morphologically and genetically.

Soft tissues were removed from each specimen for genetic identification and the shell was retained for morphological identification using the best available taxonomic keys. For genetic identification, DNA was extracted from the soft tissues using a molluscan DNA kit. The 16S, 18S and 28S rRNA genes were next amplified by polymerase chain reaction (PCR) and were purified with a PCR clean-up kit. Purified samples were shipped to GeneWiz for sequencing. Only one or two genes were successfully sequenced for some specimens. We compared resulting sequences with known sequences available in GenBank. Throughout the project, we recorded the time accrued for specimen processing and identification by both the morphological and genetic methods.

Ten sphaeriid species of 24 known in Lake Erie were positively identified using taxonomic keys, and an eleventh species was tentatively identified. Available keys are outdated because they do not include all clam species now know to be in the Great Lakes and North America, and the keys contain a number of inconsistencies that results in some ambiguous identifications. Of 160 specimens examined morphologically, we identified 137 to species; 22 specimens could not be identified because of their small size, and one shell was too damaged for identification. All 23 specimens lacking morphological identification were successfully identified by gene sequencing.

We attempted to sequence 175 individual clam specimens and were successful in making genetic identifications for 152 of them. Of those 152 samples, 58 were identified as a single species, 62 were narrowed down to two possible species, 24 had 3 possibilities, and for 8 samples four or more species were equal possibilities. We were able to obtain DNA sequences for all three genes for 69 specimens, at least two genes for 53, and at least one gene for 30 specimens. Gene sequencing indicated the possible presence of 25 species. However, several of the possible species were identified along with one or more additional potential species for the same individual specimen, and six of those tentative species have not been reported from North America. In all six cases, additional species known from Lake Erie matched those gene sequences equally well; that is, gene sequencing was unable to distinguish those

species from each other, and it is unlikely that our samples contained species that have not previously been reported in the Great Lakes using traditional morphological identification. A global electronic repository for gene sequences, called GenBank, provides gene sequences that researchers find for individual species, and it is possible that some species are not accurately identified in GenBank, resulting in the possibility of incorrect or ambiguous identifications. Further, because of inherent natural variability in DNA sequences within species, 100% matches of samples to published gene sequences is not always expected. Gene sequencing did unambiguously identify numerous fragile clams from other Great Lakes (*P. moitessierianum*) and very small immature clams that could not be identified morphologically. Two challenges to identifying sphaeriid clams by genetic means include the need for (1) development of truly universal primers in clams and (2) determining the reliability of GenBank records. The library of DNA sequences that we generated through this project greatly extends our knowledge of the clam genome. Of the 129 specimens identified both morphologically and genetically, genetic identifications matched the morphological identifications for 28%, were ambiguous for 60%, and did not match the morphological identifications for 12%.

The time required for processing and identifying specimens for both morphological identification and genetic identification was inflated in our study by several factors, including a learning curve for students learning the genetic techniques, increasing familiarity over time with the taxonomic keys, and procurement later in the study of instrumentation that increased the efficiency of sample preparation for gene sequencing. In our project, the average time for morphological identification of a specimen was 12.1 minutes, whereas the average time per specimen for gene sequencing was 51 minutes (4.2 times longer). With extensive experience and greater automation of sample preparations, we expect that the gene-sequencing time per specimen would be greatly reduced. However, given the added costs for gene sequencing of sample reagents and preparation kits as well as analytical costs for commercial gene sequencing (approximately $39 more per specimen), it seems unlikely that gene sequencing would be affordable by a typical biological survey laboratory for routine clam identification and would be reserved for studies in which it is essential to know the species of immature and damaged specimens.

Several recommendations follow from our results:

- Further research should be pursued to identify gene sequences accompanied by refinement of the gene sequencing technology that will provide unambiguous identifications of the sphaeriid clam species known in the Great Lakes and those suspected of potential colonization of the Great Lakes Basin.
- For the present time, gene sequencing can be used as a tool to augment traditional morphological identification of sphaeriid clams in the Great Lakes. Because gene sequencing yielded ambiguous species identifications for many specimens, it should be used primarily to aid identification of small and particularly fragile specimens.
- In about 22% of cases, gene sequencing methods applied to sphaeriid clams yielded ambiguous identifications in our study. Until a definitive taxonomy of those species already known to live in Lake Erie and those nonindigenous species capable of establishing populations in Lake Erie is developed, the choice to use e-DNA sampling for assessment of the present distributions and future range expansions of invasive non-native sphaeriid clams in Lake Erie is probably impractical.

This project has broad management implications in that it indicated the potential efficiency and affordability of gene sequencing and DNA barcoding compared to traditional morphological identifications. Despite additional costs, it is likely that gene sequencing will soon make it practical for government agencies and environmental consulting firms to routinely identify the species of some difficult taxonomic groups and to identify the species of heretofore unidentifiable immature specimens of most invertebrate groups. Eventual application of e-DNA techniques will improve early detection of new invasive species, the accuracy of biodiversity estimates, calculations of trophic condition indexes, and resulting interpretations of ecosystem health.

Introduction

It has long been recognized that the environmental health of a lake or stream can be understood to a large extent through the characterization of its biological inhabitants (Barbour et al. 1999, Davis and Simon 1995, Ohio EPA 1987a, Rosenberg and Resh 1992). As a result, several approaches enabling the assessment of the quality of Lake Erie and the other Laurentian Great Lakes have been developed over the past century based on the composition and abundance of different kinds of aquatic animals that occur as predictable natural assemblages called communities in healthy (unpolluted and physically unmodified) freshwater systems (e.g., Reynoldson et al. 1989). Biotic indexes, which consist of assortments of measurable community traits, have long been applied and continue to be refined for specific groups of aquatic organisms in the Great Lakes, such as the fish communities (for example, Ohio's Lake Index of Biotic Integrity, L-IBI) and major components of the invertebrate communities, including midges, aquatic oligochaete worms, and other groups. Individual kinds of invertebrates that play an outsized role in the functioning of the lake ecosystem also can be excellent indicators of lake quality. For example, the Ohio Lake Erie Quality Index applies the trend over time in the distribution and abundance of larvae of burrowing mayflies (genus *Hexagenia*, locally known in Ohio as "Canadian soldiers") in western Lake Erie as one of the three metrics of its Biological Indicator, the other two being the abundance of bald eagles and walleyes (Ohio Lake Erie Commission 2004).

The relatively tiny fingernail and pill clams (Family Sphaeriidae, also called Pisidiidae; **Figure 1**) are a common group of mollusks found in most sedimentary benthic (bottom) habitats of Lake Erie and its coastal wetlands and tributary rivers and creeks (Mackie et al. 1980, Krieger and Ross 1993). They are generally considered moderately tolerant of summer oxygen depletion (Pennak 1989) and play an important role in a balanced ecological system, including being a component in the diets of numerous Lake Erie fishes. Though not yet incorporated in biotic indexes for Lake Erie, the sphaeriid clams do contribute to the scoring of metric 8 (Percent Other Diptera and Non-Insects) of Ohio's Index of Biotic Integrity as applied to rivers and creeks of the state (Ohio EPA 1989b).

However, even though commonly encountered, sphaeriid clams ("sphaeriids") are difficult to identify to species, and sometimes even to genus, because they present several difficulties to the taxonomist: (1) Most of the morphological features needed for identification are on the inside of the shell. These include the shape and arrangement of teeth, length of the shell hinge relative to total shell length, and other features (Burch 1975, Mackie et al. 1980). Sphaeriids can be enticed to open their shells if placed

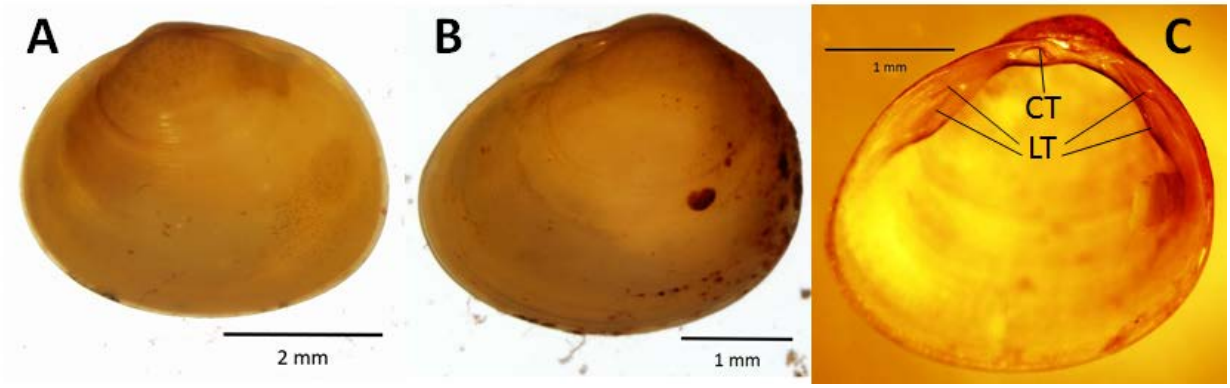

Figure 1. Three specimens typical of adult sphaeriid clams. (A) External view of *Sphaerium*. (B) External view of *Pisidium*. (C) Inside of *Pisidium* shell (right side), showing one cardinal tooth (CT) and four lateral teeth (LT), which are some of the shell parts important for identifying species; soft parts have been removed. (Photos by J. Boehler)

in warm water immediately prior to placing them in preservative, thus enabling the taxonomist to pry the shells open further to see the minute internal structures. In most cases, however, sphaeriids are collected as part of a general benthic invertebrate collection and are preserved in the field along with the rest of the collection without special separation and processing to ensure the shells will be open. As a result, the fragile shells need to be opened in the laboratory prior to identification to the species level (and sometime to genus), often yielding broken shells and the lost ability for identification. (2) Considerable time is required to open shells for identification, and additional time is required to measure dimensions of internal shell features. (3) Young clams, which often comprise the majority of individuals in samples, are even more difficult to identify because they are extremely fragile and their diagnostic features are not fully developed. (Diagnostic keys for most invertebrate groups are published only for "mature" individuals with the recognition that immature specimens will not have fully-developed diagnostic features.) Very young sphaeriids of *Musculium* and *Sphaerium* often cannot reliably be separated.

As a result of the taxonomic challenges, most Lake Erie and Ohio agency biologists identify sphaeriids only to the level of family (Sphaeriidae) or genus (*Musculium, Pisidium*, or *Sphaerium*), both to avoid dubious species identifications and to save substantial time and costs. However, by doing this, important information such as the discovery of exotic invasive species at new locations in Lake Erie and its wetlands and tributaries is lost. Records indicate the presence in Lake Erie of 19 native sphaeriid species and 5 non-native species (*Pisidium amnicum, P. henslowanum, P. moitessierianum, P. supinum,* and *Sphaerium corneum*) (see **Table 2**), and potential future increases in mean annual water temperatures of the lake as the result of global warming may make the benthic habitat more suitable for additional exotic species as well as permit increases in the abundance of those exotics that are already established (e.g., Rahel and Olden 2003). Deleterious impacts of the spread of non-native sphaeriids on the Lake Erie ecosystem have not been identified, but non-natives potentially may have been superior competitors with the native species and as a result might have reduced the abundance of the natives, but impacts such as this have not been documented. To the contrary, the spread of non-native

sphaeriids in Lake Erie may have increased the availability of appropriately-sized food items for various fish species, but again this appears to be undocumented.

Because of the potential value of sphaeriid clams in elucidating the past, present and future environmental quality of Lake Erie in conjunction with other members of the benthic invertebrate community, the primary purpose of this study was to investigate the potential of gene sequencing technology for providing a fast and reliable method of identification as an alternative to traditional morphological taxonomy for determining the distribution and abundance of both native and exotic sphaeriid clams. We predicted that gene sequencing would prove to be the more economical and accurate of the two approaches. This project addressed the strategic objective of the Lake Erie Protection and Restoration Plan (LEPR 13) (Ohio Lake Erie Commission 2013) to "develop an early detection strategy for invasive species" within Priority Area: Invasive Species (p. 12).

From the standpoint of future management implications and broader impacts, gene sequencing of aquatic organisms of all types, including bacteria, harmful cyanobacteria, algae, macroinvertebrates and fishes, and the establishment of DNA barcodes (unique gene sequences) at the species and subspecies levels is now well established as a modern tool to help researchers know the true biodiversity of Lake Erie and lake and river ecosystems throughout Ohio. We expect that DNA barcoding will soon make it practical for the first time to routinely identify the species of some taxonomic groups and to identify immature specimens of most groups, such as the sphaeriid clams, oligochaete worms, and early instars of aquatic insects, which cannot be identified beyond the level of family or genus using traditional morphological taxonomy. The approach we explored through this project has broad management implications by comparing the efficacy of the traditional taxonomic approach to that of DNA barcoding for potential application to many groups of aquatic organisms by government agencies (e.g., OEPA, ODNR) in routine monitoring and TMDL (total maximum daily load) studies as well as surveys to detect and monitor invasive species. Use of DNA barcoding will ultimately improve the accuracy of biodiversity estimates, calculations of trophic condition indices (such as those based on the species composition of oligochaete worms and chironomid midges), and resulting interpretations of ecosystem health.

Study Area

Specimens of sphaeriid clams were collected by personnel aboard the U.S. EPA research vessel *R/V Lake Guardian* at stations where they were collecting other samples in western Lake Superior near Duluth, MN, northern and southern Lake Huron, western Lake Ontario, and the central and western basins of Lake Erie (**Figure 2A**). Additional specimens were collected by J. Boehler and K. Krieger from the Old Woman Creek coastal wetland, Metzger Marsh, Muddy Creek Bay and Rock Creek of the Sandusky River watershed, and the Little Portage River of the Portage River watershed (**Figure 2B**). Several preserved specimens from earlier studies in the Sandusky, Portage and Olentangy River watersheds (**Figure 2B**) were also included in an effort to increase the taxonomic diversity included in the study. The collection date, a brief description, and the geographic coordinates of each collection site are included in the **Appendix**.

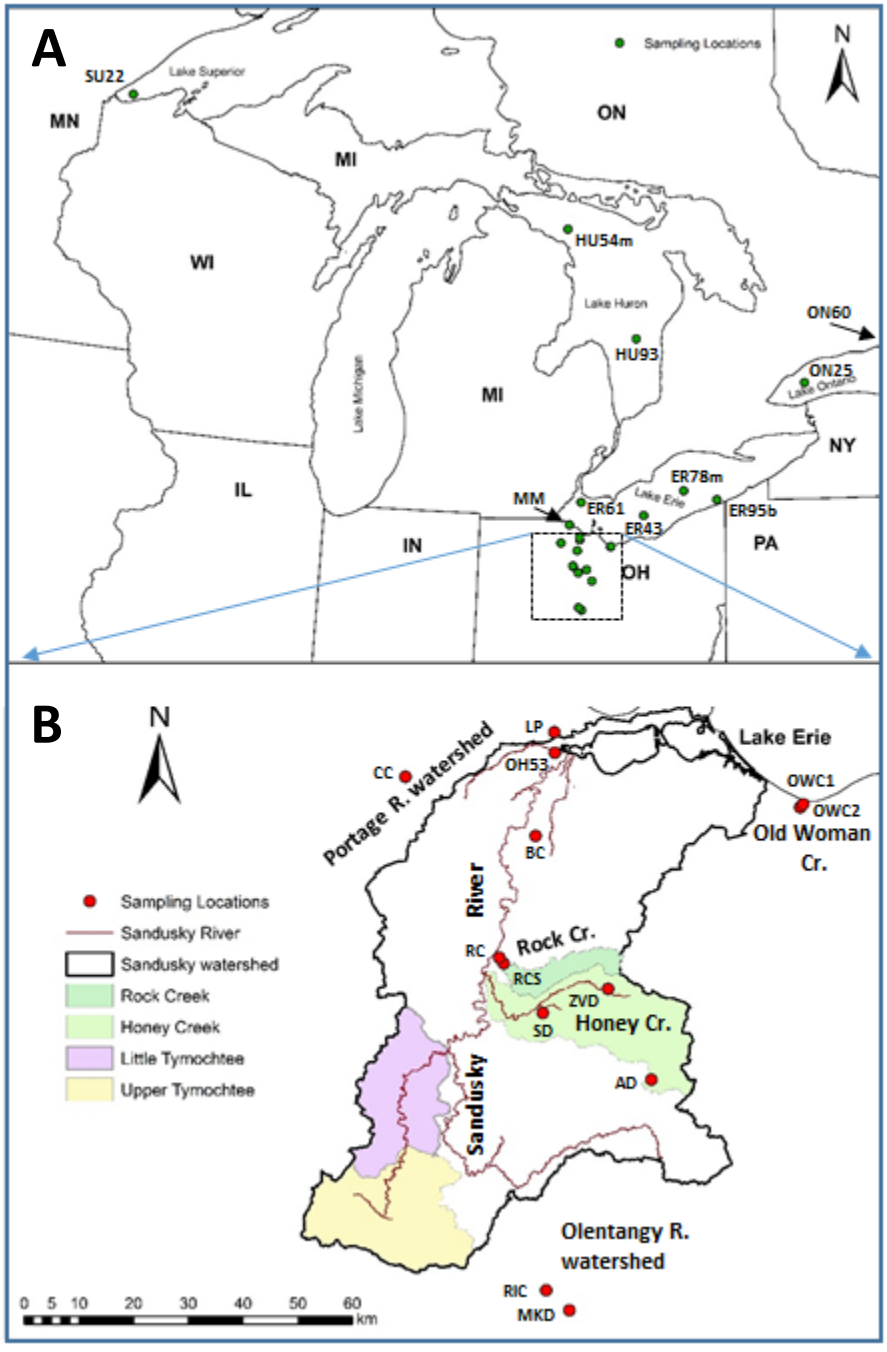

Figure 2. Locations of the collection sites in lakes Superior, Huron, Ontario and Erie and in several watersheds of western Lake Erie and headwaters of the Olentangy River of the Ohio River Basin. Site coordinates are provided in the Appendix.

Methods

Sample Collections

The clams were collected from lake, wetland, and riverine systems, with each type of system requiring different sampling protocols. The lake samples [ER(43, 61, 78m, 95b), HU(54m, 93), ON(25, 60), and SU22] were all collected by technicians from Buffalo State College aboard the USEPA's R/V Lake Guardian using a ponar sampler which is designed to grab sediment samples from the lake bottom. Upon bringing the sediment sample aboard the vessel it was rinsed over a #35 sieve (500 µm screen) to remove excess material. The remaining sediment was then picked through by technicians under a dissecting microscope to remove the clams and they placed each clam into a labeled site specific vial. The vial from each site was then placed in the freezer and hand delivered to Heidelberg University at a later date.

Specimens collected from streams required the use of either a D-framed dip net (500 µm mesh), or a standard kick seine (500 µm mesh). The kick seine was used for the stream sampling riffle areas of the larger streams with flowing water following standard kick seining protocol. In streams where riffles and flow were not always present (i.e. agricultural ditches), then the D-framed dip nets were used following standard sampling protocols. Specimens collected in these types of habitats were either preserved in 100% ethanol (sites: AD, BC, CC, MKD, RIC, SD, and ZVD) or kept alive and then opened in the lab using a warm water bath method (sites: LP, OH53, and RC). All wetland collections (sites: MM, OWC.1, OWC.2 and RCS) were also collected using the D-framed dip nets, kept alive and then opened in the lab using the warm water bath method.

In order to dissect the soft tissue needed for DNA extractions and expose the diagnostic characteristics of the shells needed for morphological identification it was necessary to open each clam prior to identifying the specimens. Specimens that were field preserved using ethanol were opened using a dissecting microscope and fine tipped dissecting tools. When field preserved, the clam's muscles constrict making them difficult to open without damaging the shells beyond use for morphological identifications. Due to the difficulties associated with opening preserved clams, a warm water bath method was developed to open live clams prior to preservation. Reverse Osmosis (RO) water was heated in a 150 mL beaker using a hot plate to a temperature of approximately 50° C. Each live clam was placed in a shallow watch glass and then the warm water was poured over the clam. When in the warm water, the clam's muscles would relax and the shell opened slightly. Fine tipped forceps were then used to pry the shell open completely and the specimen was placed directly into a labeled vial containing a 95% ethanol solution. Upon preserving the specimen the soft tissue was then removed from the shell of each specimen and placed into a separate labeled vial also containing a 95% ethanol solution. The time spent opening each clam was recorded in order to provide an accurate measure of the time necessary to prepare the specimens for morphologically identification.

Morphological Identifications

For this project, the morphological identification of specimens was performed by J. Boehler (research assistant; NCWQR) who had some previous experience identifying fingernail and pill clams to both genus and species prior to this project. All morphological identifications were based on the characteristics of each clams shell and not on the clam's soft anatomy. The taxonomic keys used for identification included Mackie, White and Zdeba (1980) and Burch (1975). Both keys were used to make the identifications for each specimen, helping to eliminate identification errors, with Mackie, White and Zdeba's key being the primary source, since it includes detailed pictures of the cardinal teeth. Burch's key was then used to confirm the identifications made using the primary key. The time spent identifying

each clam was recorded and combined with time spent opening each clam to determine a total processing time for the morphological identification of each individual specimen. When possible, one species name was attributed to each specimen with the exception of specimens that were not distinguishable between two similar species. When this occurred, both names were listed for that specimen. Any specimens that were either too small to identify, or the shells were too damaged to identify were not identified morphologically in order to prevent misidentifications.

Gene Sequencing and Specimen Identification

The E.Z.N.A. Mollusc DNA Kit (Omega Bio-tek, Inc., Norcross, GA) was used to extract the DNA from the soft tissues of the clam samples according to the manufacturer's protocol. During the final elution steps of the protocol, 75µL of Elution Buffer was added to column, and the extracted DNA was collected in a 1.5mL microcentrifuge tube and stored at -20°C.

Polymerase chain reaction (PCR) was then used to amplify the 16S, 18S, and 28S rRNA genes for sequencing. In each 0.2 mL tube, 4µL DNA, 25µL TopTaq Mastermix (Qiagen, Valencia, CA), 19µL nuclease free water, 1µL forward primer, and 1µL reverse primer were added. The primers were synthesized by Eurofins MWG Operon, LLC. (Huntsville, AL) and the sequences for each are presented in **Table 1**. The thermal cycler was programed with the temperature at 94°C for 3 minutes, and then 30 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute. After the temperature cycling, there was a final extension at 72 °C for 10 minutes, and then the samples were held at 4 °C. After PCR, the samples were purified using the ExoSAP-IT PCR Product Cleanup Kit (Affymetrix, Santa Clara, CA), according to the manufacturer's protocol and sent to Genewiz, LLC (North Plainfield, NJ) for Sanger sequencing. The resulting chromatograms were analyzed using Finch TV version 1.4 (GeoSpiza, Inc., Seattle, WA) and the forward and reverse sequences were compared using SeqTrace (Stucky 2012). Each sequence was then compared with the known sequences available in GenBank using BLAST (Basic Local Alignment Search Tool) from the National Center for Biotechnology Information (NCBI) and the top match or matches based on the percent identity and length of the DNA sequences were recorded for each gene. We then compared the results across all 3 genes to make a final genetic identification. If the same species was a top match for each of the 3 genes, then we identified the sample as that species, even if there were multiple top hits for any one of the three genes. For example, for sample ER43.1, the best match in BLAST for both 16S and 18S was *Sphaerium corneum*. For the 28S gene, that sample matched sequences from *S. corneum* and *S. striatinum* equally well. Putting the data together from all 3 genes, we gave this sample a genetic ID of *S. corneum*, and this matched the morphological determination. When multiple species matched equally well across multiple genes, we listed all those species as possibilities for the genetic identification. If DNA sequence data was missing for a gene for a given sample we used only the data from the other two genes to make the genetic identification. In a few cases, DNA sequence data was missing for two genes, and we used only the sequence data from one gene to make the identification.

Time Calculations for Morphological Identifications

Factors that could influence the time spent on morphological identifications include the species diversity of the sample, the complexities of identifying one species over another, and the expertise level of the person performing the identifications. For the time analysis of morphological identifications, the time

Table 1. Primers used for the amplification of the 16S, 18S, and 28S genes

Sequence Name	Sequence (5' to 3')	Source
16Sar_F	CGCCTGTTTATCAAAAACAT	Kessing 2000
16Sbr_R	CCGGTCTGAACTCAGATCACGT	
18SLab_F	GGCGAGGGTTTAAAGAGTGG	Designed by our lab
18SLab_R	ATTTCGTTTGACGGGTTGTC	
D23Forward_28S	CCAGCTATCCTGAGGGAAACTTCG	Park and Foighil 2000
D6Reverse_28S	GAGAGTTCAAGAGTACGTG	

spent identifying each specimen was recorded and the average time for all 160 identifications was calculated. (These data are available electronically by request.) As is the case with identifying any group or organisms using morphological characteristics, there is a learning curve associated with familiarizing oneself with the key features and characteristics to identify the specimens. To account for the learning curve associated with the identification process, we included all specimen identifications in the average. For example, it may have taken 20 minutes per specimen to identify the first specimen for each taxon, but as more specimens of the same taxon were identified it may have only taken 5 minutes per specimen to identify each of the remaining clams. Another reason to include the learning curve in the average time spent on identifications is that the researcher does not know how diverse the sample truly is until the identifications have been completed. A sample could contain only 1 species, or it could have several species, making it difficult to know how much time will be spent on identifications beforehand. For example, sample OWC.2 had 20 specimens that were all the same taxon, and it took 159 minutes to identify all of these specimens (8 minutes per specimen). By contrast, sample ER78m had 33 specimens consisting of 6 individual species and one indeterminate species pair that took 477 minutes (14.5 minutes per specimen) to identify. This shows that one can expect it to take longer to identify a species-rich sample as compared to a less diverse sample.

Results

Species Identified by Morphological Identification

The species identified in our study, as well as additional species reported to be present in Lake Erie, are listed in **Table 2**. There were 10 confirmed and one possible species identified morphologically, for a total of 11 different species, using traditional taxonomic identification keys (**Table 3**). The single unconfirmed species was *Pisidium supinum*, which is very similar to *Pisidium henslowanum*, and the shells of the three specimens were too damaged to distinguish between the two species. Undamaged specimens from both sites (ER43 and ER78) where these three specimens occurred were identified as *Pisidium henslowanum*, but we did not want to speculate that the damaged specimens were also *P. henslowanum*, as it could be possible to have both species at a single collection site.

Table 2. Sphaeriid species identified by morphological characteristics in this study and other species also reported from Lake Erie.

Species identified in this study	Additional species reported from Lake Erie*
Musculium partumeium (Say, 1822)	*Musculium lacustre (Müller, 1774)*
M. transversum (Say, 1829)	*Pisidium amnicum (Müller, 1774)*
Pisidium adamsi Stimpson, 1851	*P. conventus Clessin, 1877*
P. casertanum (Poli, 1791)	*P. dubium (Say, 1817)*
P. compressum Prime, 1852	*P. fallax Sterki, 1896*
P. henslowanum (Sheppard 1825)	*P. furrugineum Prime, 1852*
P. lilljeborgi (Clessin, 1886)	*P. moitesserianum Paladilhe 1866*
P. subtruncatum Malm, 1855	*P. nitidum Jenyns, 1832*
Sphaerium corneum (Linnaeus, 1758)	*P. punctatum Sterki, 1895*
S. striatinum (Lamarck, 1818)	*P. supinum Schmidt, 1850*
	P. variable Prime, 1852
	P. ventricosum Prime, 1851
	P. walkeri Sterki, 1895
	Sphaerium occidentale (J. Lewis, 1856)

*Mackie et al. (1980); U.S. EPA GLERL: www.glerl.noaa.gov/seagrant/GLWL/Benthos/Mollusca/Bivalves/Sphaeriidae.html (accessed 25 July 2016); GLANSIS: www.glerl.noaa.gov/res/Programs/glansis/glansis.html (accessed 25 July 2016)

Species Identified by Gene Sequencing

Twenty-five species were identified as possible hits in BLAST searches. Of these, 8 species were identified as the single most likely species across all 3 genes for at least one sample. All 8 of these were species that had been previously identified in North America and likely represent correct identifications. Six species identified by BLAST have never been observed in North America. In all 6 of these cases, additional species known to be found in Lake Erie matched the unknown sequence equally well, so it is likely that these are spurious hits.

Difficulties Encountered in Identification by Each Method

Morphological Identifications. Unfortunately, the morphological identification of clam specimens is hindered by the fact that two clam specimens, even of the same species, can show great variability in the shape of the key diagnostic features. Several of the diagnostic features, such as the shape of the cardinal teeth, can be somewhat variable depending on the age of the clam, or even the environmental living conditions. When identifying clams it is sometimes necessary to make assumptions and provide the most likely species name, even if the taxonomist still has some level of doubt. This uncertainty and doubt can lead to more time being spent looking at difficult specimens, increasing the time and cost of identifications. Adding to this uncertainty are issues with the keys used such as disagreements on certain features of specimens between the two taxonomic keys. For example, in one key the cardinal teeth of *Pisidium casertanum* are described as anterior to the center of the hinge, whereas in the other key they are described as central. Disagreements such as these can also add to both the time and cost of identifying specimens using morphological characteristics. The uncertainty associated with

Table 3. Comparison of identifications of *Musculium* (*M*.), *Pisidium* (*P*.) and *Sphaerium* (*S*.) using morphological and genetic methods, and locations where specimens were collected.

Species	Identified by morpho-logical or genetic method	Lake Erie specimens found in L. Erie (L), wetland (W) or tributary (T)	Specimens found in lakes Huron (H), Ontario (O) or Superior (S)	Specimens identified morpholo-gically	Specimens identified genetically*	Specimens identified genetically to one species	Specimens genetically ambiguous for 2 species	Specimens genetically ambiguous for 3 or more species	Complete/ partial/ or non-matches
M. lacustre	Genetic	W, T	-	0	17	6	8	3	0/0/17
M. partumeium	Both[a]	W, T	-	17	7	0	7	0	0/7/6
M. transversum	Both	T	-	3	3	1	2	0	1/2/0
P. adamsi	Both	T	-	1	2	0	0	2	0/1/1
P. casertanum	Both[b]	L, W, T	-	64	74	26	33	15	19/50/4
P. compressum	Both	L, W	-	17	19	1	6	12	1/18/0
P. fallax	Genetic	L, W, T	-	0	42	0	28	14	0/0/42
P. henslowanum	Both[c]	L	-	13	15	0	1	14	0/14/1
P. lilljeborgi	Both[d]	L, W	H, O, S	4	34	7	15	12	4/11/2
P. moitesserianum	Genetic[e]	-	H, O, S	0	15	1	14	0	0/0/15
P. subtruncatum	Both[f]	L	-	11	17	12	0	5	7/5/1
P. supinum	Both[g]	L	-	3	15	0	1	14	0/14/1
P. ventricosum	Genetic	L	-	0	1	0	1	0	0/0/1
S. corneum	Both	L	-	3	3	3	0	0	3/0/0
S. occidentale	Genetic[h]	W	-	0	1	0	0	1	0/1/0
S. striatinum	Both	T	-	4	4	0	1	3	0/4/0
		Total specimens identified		140**					

See footnotes on following page.

Table 3 Footnotes:

*The following species were all listed as one of many possible species on ambiguous genetic identifications but are not considered likely identifications because none of these species have been reported in the Great Lakes and/or North America: *P. edlaueri*, *P. globulare*, *P. hallae*, *P. hibernicum*, *P. obtusale*, *S. nucleus*
**20 additional specimens were examined but could not be identified because they were too small
[a]4 specimens were identified morphologically, but were not identifiable genetically due to failed PCR
[b]1 specimen was not identifiable morphologically, but genetic analysis yielded 2 ambiguous identifications (*P. casertanum* or *P. subtruncatum*)
[c]3 ambiguous morphological identifications (*P. henslowanum* or *P. supinum*)
[d]17 genetic identifications were not assigned morphological identifications because the specimens were too small to identify
[e]15 specimens were too small to identify morphologically
[f]2 ambiguous morphological identifications (*P. casertanum* or *P. subtruncatum*); 4 genetic identifications were not assigned morphological identifications because the specimens were too small to identify
[g]3 ambiguous morphological identifications (*P. henslowanum* or *P. supinum*)
[h]Listed as a possible species on a confirmed *Pisidium* specimen

morphological identifications shows one reason why the genetic identification of sphaeriid clams potentially could be used to aid morphological identifications.

Genetic Identifications. For the genetic identifications we relied on reference sequences published in GenBank to make comparisons. However, given the known difficulty in morphologically identifying these species, it is possible that some samples in GenBank are not accurately identified, which would cause challenges for our genetic identifications as well. There is also a notable lack of data available for many of the *Pisidium* species in GenBank for the 28S gene. Another challenge is designing universal primers for the 3 genes. We chose the 16S, 18S, and 28S genes for sequencing because universal primers for them have been previously published for some species. However, those universal primers were not always successful in clams. Lastly, we would not always expect 100% matches to the reference sequence because even within a single species there will be some individual variability in the DNA sequences.

Extent of Agreement between Morphological and Genetic Identifications

We attempted to sequence 175 individual clam specimens, and were able to make genetic identifications for 152 of these. All 14 samples that had previously been stored in formaldehyde failed, suggesting that our DNA extraction method does not work for these samples. The remaining 9 failed samples were all processed in the first batch by a student new to the lab, and likely represent experimental error. For each sample, we attempted to sequence 3 genes, but we were not successful in all cases. We were able to obtain DNA sequence for all 3 genes for 69 samples, for at least 2 genes for 53, and at least one gene for 30 samples. Of the 152 samples that were genetically identified, 58 were identified as a single species, 62 were narrowed down to two possible species, 24 had three possibilities, and for 8 samples four or more species were equal possibilities (**Figure 3)**.

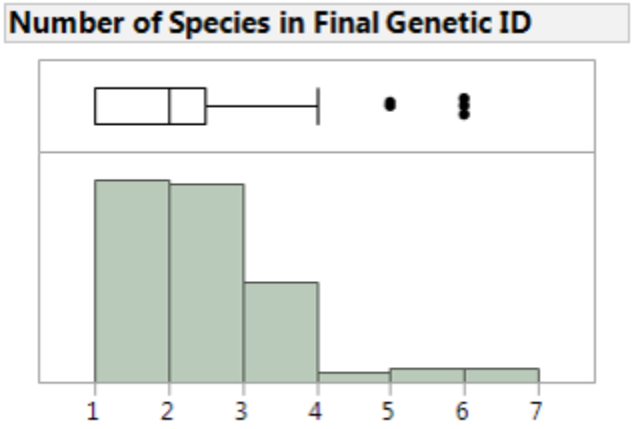

Figure 3. Histogram and box-and-whiskers plot of number of species in the final genetic identification for all samples. In the box-and-whiskers plot the vertical line within the box indicates the median. The ends of the box represent the 25th and 75th percentiles (P25 and P75). The whiskers extend from the ends of the box to the outermost data point that falls within P25-1.5*(interquartile range) to P75+1.5*(interquartile range). Data points outside of this range are represented as black dots.

For the morphological analysis, 160 samples were examined and 137 samples were identified to the species level. In 22 cases, an identification could not be made due to the small size of the sample, and in one case due to damage to the shell. Notably, all 23 samples that were not identifiable morphologically were identifiable using genetic means.

Of all the samples tested, 129 had both genetic and morphological identifications. When comparing the genetic and morphological identifications, we considered the identifications be a “match” if the sample was identified as one single species both morphologically and genetically, and the identifications were concordant. We considered an identification to be a “partial match” if there were multiple possibilities for either the genetic or morphological identifications, and there was overlap between the two. We considered an identification to be a “non-match” if the genetic and morphological results were discordant. Based on this terminology, 36 samples matched between the morphological and genetic identifications, 77 were partial matches, and 16 were non-matches.

In the group of partial matches, we noticed a few species that consistently could not be distinguished from each other, leading us to suspect that perhaps these different species were, in fact, only a single species (**Table 4**). To investigate this further, we aligned 16S rRNA gene sequences from GenBank for each of the conflicting species. An example alignment for 5 *P. casertanum* and 5 *P. fallax* is shown in **Table 5**. Each row in the alignment represents the DNA sequence for this gene in a particular sample. The alignment shows a striking similarity between the two species for this gene. Of the 437 nucleotides that are present in all 10 samples, 385 (88.1%) are identical across all 10 samples. Looking at each of the 52 differences, only one seems to correlate with species status, the nucleotide highlighted in yellow at position 322 using the numbering for the first sequence. At this position, the “A” allele is present in all *P. casertanum* samples and the “G” allele is present in all *P. fallax* samples. The other 51 differences do not

Table 4. Species that are difficult to distinguish using genetic identification.

Species	Number of specimens that were identified as matching the given species equally well
P. casertanum and *P. fallax*	27
P. henslowanum, P. lilljeborgi, and *P. supinum*	14[a]
M. lacustre and *M. partumeium*	7
S. striatinum and *S. simile*	2[b]

[a] In one case, the identification included only *P. henslowanum* and *P. supinum*.

[b] In one case, the identification also included *S. nucleus*.

show species-specific patterns. This suggests that it is possible the categories *P. fallax* and *P. casertanum* could be collapsed into a single species. The alignments for the other ambiguous species in **Table 4** showed similar results.

If the ambiguous species in **Table 4** actually are single species, it would greatly impact our results. Reclassifying the data under this assumption, the genetic and morphological identifications would result in 84 matches, 29 partial matches, and 16 non-matches. Overall, this increases our match percentage from 27.9% to 65.1% and decreases our partial match percentage from 59.7% to 22.5%. The rate of non-matches (12.4%) remains the same under either classification scenario.

Some patterns emerged upon further examination of the 16 non-matches. In 13 of the 16 non-matches, a genetic identification was made when sequence data was missing for at least one of the genes, suggesting that caution should be used when attempting to identify species without having sequence data available for all 3 genes. Even in the case of non-matches at the species level, we did find good agreement between the genetic and morphological identifications at the genus level. For all 16 non-matches, the genetic and morphological approaches agreed on the genus.

Geographic Distribution of Species as Described by this Data Set

Of the ten species confidently identified morphologically in our limited collections, *Pisidium casertanum* and *P. compressum,* were widely distributed at the Lake Erie Basin sites, including all four lake sites and most wetland and tributary sites, and they were absent from the collections from the other Great Lakes (**Table 6**). Most of the remaining species were found only at two or three sites in Lake Erie or its southwestern watershed and not the other Great Lakes. Only two sphaeriids, *Pisidium lilljeborgi* and *Pisidium moitesserianum* (a non-native species) were found in our Lake Huron, Lake Superior and Lake Ontario collections. The non-native species *P. henslowanum*, *Sphaerium corneum*, and possibly *P. supinum* were found only in our collections from Lake Erie but not in its watershed or the collections from the other Great Lakes. The samples included in this project were very limited in geographic coverage and the numbers of individuals collected and by no means indicate the geographic ranges of these species. Our samples only indicate where the species are present and do not indicate where they are absent.

Table 5. Alignment of five *Pisidium casertanum* and five *P. fallax* 16S rRNA gene sequences. The five *P. casertanum* samples are listed first with the initials "PC" at the beginning of the identifier, and are followed by the *P. fallax* samples with the initials "PF" in the identifier. The gi and gb numbers indicate GenBank accession and version numbers. Dashes indicate gaps in the sequence. The numbers at the end of each row represent the number of nucleotides present in the sequence. The gene sequence for each sample continues on nine rows.

```
PC|gi|575058781|gb|KF483316.1|      GCCCCTTGTTTTTTTTTTATGAGGGGTTGGACCTGCCCGGTGATTAGTCAAATTTTGAAG    60
PC|gi|575058758|gb|KF483293.1|      CGCCCCTTGTTTTTTTTTATGAGGGGTTGGACCTGCCCGGTGATTAGTCAAATTTTGAAG    60
PC|gi|28379992|gb|AY093557.1|       ----------------------GGGGTCGGACCTGCCCGGTGATTAGTCAAATTTCGAAG    38
PC|gi|28379994|gb|AY093559.1|       ----------------------GGGGTTGGACCTGCCCGGTGATTAGTCAAATTTTGAAG    38
PC|gi|28379993|gb|AY093558.1|       ----------------------GGGGTTGGACCTGCCCGGTGATTAGTCAAATTTTGAAG    38
PF|gi|62079835|gb|AY957868.1|       -------------------------------CTGCCCGGTGATTAGTCAAATTTTGAAG    28
PF|gi|62079780|gb|AY957813.1|       -------------------------------CTGCCCGGTGATTAGTCAAATTTTGAAG    28
PF|gi|62079791|gb|AY957824.1|       -------------------------------CTGNCCGGTGATTAGTCAAATTTTGAAG    28
PF|gi|62079783|gb|AY957816.1|       -------------------------------CTGCCCGGTGATTAGTCAAATTTTGAAG    28
PF|gi|28379996|gb|AY093561.1|       ----------------------GGGGTTGGACCTGCCCGGTGATTAGTCAAATTTTGAAG    38
                                                            *** ******************** ****

PC|gi|575058781|gb|KF483316.1|      TTGTTAATATAATAAGAAATTTATTTAAACGGCTGCAATTAATAGTTGTGCTAAGGTAGC    120
PC|gi|575058758|gb|KF483293.1|      TTTTTAATATAATAAGAAATTTATTTAAACGGCTGCAATTAATAGTTGTGCTAAGGTAGC    120
PC|gi|28379992|gb|AY093557.1|       TTATTAAATTAATTAGAAATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    98
PC|gi|28379994|gb|AY093559.1|       TTATTAAATTAATTAGAGATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    98
PC|gi|28379993|gb|AY093558.1|       TTATTAAATTAATTAGAGATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    98
PF|gi|62079835|gb|AY957868.1|       TTATTAAATTAATAAGAGGTTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    88
PF|gi|62079780|gb|AY957813.1|       TTATTAAATTAATAAGAGGTTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    88
PF|gi|62079791|gb|AY957824.1|       TTATGAAGGTAACAAGGAATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    88
PF|gi|62079783|gb|AY957816.1|       TTAT-GAGGTAACAAGAGATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    87
PF|gi|28379996|gb|AY093561.1|       TTAT-GAGGTAACAAGAGATTTTTTTAAACGGCTGCAATTAGTAGTTGTGCTAAGGTAGC    97
                                    ** *  *  ***  **   *** ****************** ******************

PC|gi|575058781|gb|KF483316.1|      GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    180
PC|gi|575058758|gb|KF483293.1|      GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    180
PC|gi|28379992|gb|AY093557.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAGAACTGTC    158
PC|gi|28379994|gb|AY093559.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    158
PC|gi|28379993|gb|AY093558.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    158
PF|gi|62079835|gb|AY957868.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    148
PF|gi|62079780|gb|AY957813.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    148
PF|gi|62079791|gb|AY957824.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    148
PF|gi|62079783|gb|AY957816.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    147
PF|gi|28379996|gb|AY093561.1|       GTAATCATTTGCCCCCTAATTAGGGGAAGGTATGAATGGTTTGACGTGGGAAAAGCTGTC    157
                                    ********************************************** ** * *****

PC|gi|575058781|gb|KF483316.1|      TCCTTTATATAAAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATTTATATAAAAGACG    240
PC|gi|575058758|gb|KF483293.1|      TCCTTTATATGAAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATTTATATAAAAGACG    240
PC|gi|28379992|gb|AY093557.1|       TCTTTTATATATAAGGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTTTAAAAGACG    218
PC|gi|28379994|gb|AY093559.1|       TCTTTTATATTTAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTATAAAAGACG    218
PC|gi|28379993|gb|AY093558.1|       TCTTTTATATTTAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTATAAAAGACG    218
PF|gi|62079835|gb|AY957868.1|       TCTTTTATATAAAAAGAAGTTTATTTTTGAGTGAAAAAGCTTGGATGTTTTTAAAAGACG    208
PF|gi|62079780|gb|AY957813.1|       TCTTTTATATAAAAAGAAGTTTATTTTTGAGTGAAAAAGCTTGGATGTTTTTAAAAGACG    208
PF|gi|62079791|gb|AY957824.1|       TCTTTTATATAGAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTGTAAAAGACG    208
PF|gi|62079783|gb|AY957816.1|       TCTTTTATATAGAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTGTAAAAGACG    207
PF|gi|28379996|gb|AY093561.1|       TCTTTTATATAGAAAGAAGTTTATTTTTGAGTGAAAAAGCTTAGATGTTTGTAAAAGACG    217
                                    ** *******  ** ************************** *** * * *********

PC|gi|575058781|gb|KF483316.1|      AGAAGACCCTATCGAACTTGAATTGTGTGTTAAAATTTTTTTTATGCAGAAAGTTTAGTT    300
PC|gi|575058758|gb|KF483293.1|      AGAAGACCCTATCGAACTTGAATTGTGTGTTAAAAATTTTTTGATACGGAAAGTTTAGTT    300
PC|gi|28379992|gb|AY093557.1|       AGAAGACCCTATCGAACTTGAATTATGAATTATAAAT-TTATAATACAGAAAGTTTAGTT    277
PC|gi|28379994|gb|AY093559.1|       AGAAGACCCTATCGAACTTGAATTATGTATTGTAGGT-TTATAATGCAGAAAGTTTAGTT    277
PC|gi|28379993|gb|AY093558.1|       AGAAGACCCTATCGAACTTGAATTATGTATTGTAGAT-TTATAGTGCAGAAAGTTTAGTT    277
PF|gi|62079835|gb|AY957868.1|       AGAAGACCCTATCGAACTTGAATTGTGTGTTTTAGTT-TTGGGATACAGAAAGTTTAGTT    267
PF|gi|62079780|gb|AY957813.1|       AGAAGACCCTATCGAACTTGAATTGTGTGTTTTAGTT-TTGGGATACAGAAAGTTTAGTT    267
PF|gi|62079791|gb|AY957824.1|       AGAAGACCCTATCGAACTTGAATTGTGTATTTTAGTT-TTGGAATACAGAAAGTTTAGTT    267
PF|gi|62079783|gb|AY957816.1|       AGAAGACCCTATCGAACTTGAATTGTGTATTTTAGTT-TTGGAATACAGAAAGTTTAGTT    266
PF|gi|28379996|gb|AY093561.1|       AGAAGACCCTATCGAACTTGAATTGTGTATTTTAGTT-TTGGAATACAGAAAGTTTAGTT    276
                                    ********************** **  **  *  * **    * * ************

PC|gi|575058781|gb|KF483316.1|      GGGGAAACTTAAAGTTAAAAGAAACGCTTTTTTGTTATAAAGTGATCCTGTATTATGGAA    360
PC|gi|575058758|gb|KF483293.1|      GGGGAAACTTAAAGTTAAAAGAAACGCTTTTTTGTTATAAAATGATCCTGTATTATGGAA    360
PC|gi|28379992|gb|AY093557.1|       GGGGAAACTTAAAGTTAAGAAAAACTCTTTTTTGTTGTAAAATGATCCTGTATTACAGAA    337
PC|gi|28379994|gb|AY093559.1|       GGGGAAACTTAAAGTTAAGAAAAACGCTTTTTTGTTGTAAAATGATCCTGTATTATAGAA    337
```

```
PC|gi|28379993|gb|AY093558.1|       GGGGAAACTTAAAGTTAAGAAAAACGCTTTTTTGTTGTAAAATGATCCTGTATTATAGAA    337
PF|gi|62079835|gb|AY957868.1|       GGGGAAACTTAAAGTTAAGAAGAACGCTTTTTTGTTGAAAAATGATCCTATATTATGGAA    327
PF|gi|62079780|gb|AY957813.1|       GGGGAAACTTAAAGTTAAGAAGAACGCTTTTTTGTTGAAAAATGATCCTATATTATGGAA    327
PF|gi|62079791|gb|AY957824.1|       GGGGAAACTTAAAGTTAAGAAGAACGCTTTTTTGTTGTAAAATGATCCTGTATTATGGAA    327
PF|gi|62079783|gb|AY957816.1|       GGGGAAACTTAAAGTTAAGAAGAACGCTTTTTTGTTGTAAAATGATCCTGTATTATGGAA    326
PF|gi|28379996|gb|AY093561.1|       GGGGAAACTTAAAGTTAAGAAGAACGCTTTTTTGTTGTAAAATGATCCTGTATTATGGAA    336
                                    ****************** *  *** **********  *** ******* *****  ***

PC|gi|575058781|gb|KF483316.1|      AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATTAAAGA    420
PC|gi|575058758|gb|KF483293.1|      AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATTAAAGA    420
PC|gi|28379992|gb|AY093557.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    397
PC|gi|28379994|gb|AY093559.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    397
PC|gi|28379993|gb|AY093558.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    397
PF|gi|62079835|gb|AY957868.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    387
PF|gi|62079780|gb|AY957813.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    387
PF|gi|62079791|gb|AY957824.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    387
PF|gi|62079783|gb|AY957816.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    386
PF|gi|28379996|gb|AY093561.1|       AAATGAAAAAGTTACCGTAGGGATAACAGCGCTTTCTTCTCTGAGAGGACTAATCAAAGA    396
                                    ************************************************** *****

PC|gi|575058781|gb|KF483316.1|      GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    480
PC|gi|575058758|gb|KF483293.1|      GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    480
PC|gi|28379992|gb|AY093557.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    457
PC|gi|28379994|gb|AY093559.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    457
PC|gi|28379993|gb|AY093558.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    457
PF|gi|62079835|gb|AY957868.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAA------    441
PF|gi|62079780|gb|AY957813.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAA------    441
PF|gi|62079791|gb|AY957824.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAA------    441
PF|gi|62079783|gb|AY957816.1|       GTTGGTTGCGACCTNNATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAG-----------    435
PF|gi|28379996|gb|AY093561.1|       GTTGGTTGCGACCTCGATGTTGGATTAAGGTTTCTTTTTGGGGTAGGAGTTAAAATAGTA    456
                                    **************  *********************************

PC|gi|575058781|gb|KF483316.1|      AAACTGTTCGTTTTTTAAAACCTT     504
PC|gi|575058758|gb|KF483293.1|      AAACTGTTCGTTTTTTAAAACCTT     504
PC|gi|28379992|gb|AY093557.1|       AAACTGTTCGTTTT----------     471
PC|gi|28379994|gb|AY093559.1|       AAACTGTTCGTTTT----------     471
PC|gi|28379993|gb|AY093558.1|       AAACTGTTCGTTTT----------     471
PF|gi|62079835|gb|AY957868.1|       ------------------------     441
PF|gi|62079780|gb|AY957813.1|       ------------------------     441
PF|gi|62079791|gb|AY957824.1|       ------------------------     441
PF|gi|62079783|gb|AY957816.1|       ------------------------     435
PF|gi|28379996|gb|AY093561.1|       AAACTGTTCGTTTT----------     470
```

Time and Cost Efficiency of Morphological versus Genetic Approach

Based on the average for all 160 identifications, it took just over 12 minutes to open and morphologically identify a single clam (**Table 7**). The final average time per specimen for genetic identification was 0.84 hours (~51 minutes) (**Table 8**). However, several factors must be considered in interpreting this result. First, this time estimate only includes “hands on” time by the student. It does not include long intermediate steps that do not require someone to be present, for example, the overnight incubation of samples at 37 degrees during DNA extraction. Second, it is not a simple matter to extrapolate how much time would be required to process larger batches of samples. Though students most often handled batches of 8, they sometimes handled 16 or 24 samples. We found that while doubling or tripling the number of samples did result in some increase in processing time, it was much less than double or triple the time needed for 8 samples. For example, while it takes about 22 minutes to set up PCR for 8 samples, it takes only about 30 minutes to set up PCR for 16 samples. The major exception to this was for the data analysis step, where there was a fairly linear increase between time and number of samples processed. The time for data analysis per sample also varied widely (from a few seconds to 10 minutes) depending on the speed of BLAST, which was likely due to variability in how many users around the world were running BLAST at any given time.

Table 6. Geographic distributions of species identified morphologically in this project. A site in bold type indicates a confirmed identification using both identification methods, whereas plain type indicates that the genetic identification did not completely or at all match the morphological identification.

Species	Locations (see Figure 2.)
Musculium partumeium	LP, OH53, AD
Musculium transversum	RIC, **BC**
Pisidium adamsi	MKD, OWC.2
Pisidium casertanum	AD, BC, **ER43**, **ER61**, **ER78m**, ER95b, MKD, MM, OH53, **OWC.1**, **OWC.2**, RCS, ZVD
Pisidium compressum	ER43, **ER61**,ER78m, ER95b, MM, OH53, RCS, MKD, OWC.1, OWC.2
*Pisidium henslowanum**	ER43, ER78m
Pisidium lilljeborgi	**ER78m**, HU54m, HU93, ON25, ON60, SU22b
Pisidium moitesserianum†*	HU54m, HU93, ON25, ON60, SU22b
Pisidium subtruncatum	**ER43**, **ER78m**
*Pisidium supinum**	ER43, ER78m
*Sphaerium corneum**	**ER43**, **ER78m**
Sphaerium striatinum	SD, CC

* non-native species; all 3 *P. supinum* specimens may have been *P. henslowanum*

† This species was only found using the genetic method

Furthermore, for the molecular genetics techniques in particular, the time to complete them greatly depends on the technology available in the lab. For example, students A and B were able to use a multi-channel pipette to transfer 8 samples at a time to plates for packaging, whereas student C had only a single-channel pipette. Extending this further, a lab equipped with a liquid handling robot would see a drastic decrease in the time needed to process samples, perhaps completing the project in just a few weeks. Given the advancements in next-generation sequencing, it is not unrealistic to speculate that a lab with that capability could process all the samples in this project in as little as two days.

We compared the approximate cost of the two identification methods by calculating the average time required per specimen (sample) and cost of reagents and services. We arbitrarily established salary and fringe benefits for a technician performing either method at $25/hour. Therefore, average time cost for morphological identification was $5.00/specimen, and for genetic identification it was $21.25. We estimated that reagents and storage vials for morphological identifications cost about $0.50/specimen, whereas those costs for genetic identification, based on invoices, came to $9.07/specimen. The cost for GeneWiz analysis was $30.38/specimen. Therefore, in total it cost about $61/specimen (11 times more) for genetic analysis compared to $5.50 for morphological analysis. One of a number of factors that would lower the cost is that the price per sample charged by GeneWiz declines as the number of samples submitted in one batch increases.

Table 7. Time analysis for morphological identification (n=160).

	Time (minutes) with opening of shells included
Average Time per Specimen	12.1
Standard Deviation	9.8
Range for all Taxa	49
Median	10
Mode	10

Table 8. Time analysis in minutes for genetic identification (n=175).

Student	DNA extraction	PCR & ExoSap	Packaging	Analysis	Total Time for 8 Samples
A	135	45	60	135	375
B	147	42	65	134	388
C	167	36	87	154	444
Overall Average	149.7	41.0	70.7	141.0	402.3
Average ID time per specimen = 0.84 hours					

Discussion

Adequacy of Available Morphological Taxonomic Keys for Sphaeriid Clams

For at least two centuries, traditional taxonomic identification has relied on the availability of authoritative identification keys based on morphological and anatomical characteristics. An authoritative key for a particular group of organisms is developed by one or more experts who have specialized in studying that group. In the case of the sphaeriid clams, there are very few taxonomic experts and only two available, dated keys (Burch 1975, Mackie et al. 1980) for identifying Great Lakes species. As noted previously, an updated and revised North American key, particularly for the Great Lakes, is needed that will incorporate North American species that might have colonized the lakes since the 1970s and 1980s and species that are not indigenous to North America (e.g., *Pisidium moitessierianum*; Grigorovich et al. 2000) that now are present in the Great Lakes but are not included in the existing keys. Ambiguities within each key and inconsistencies between the keys, some of which we have noted above, also need to be corrected. Despite the shortcomings of the present keys, a technician who is sufficiently familiar with them and is experienced in identifying sphaeriid species can identify most mature specimens with some degree of confidence. However, those keys are less useful in identifying badly damaged specimens and inadequate for identifying young specimens, and it is for those specimens, which often comprise the majority of individuals in Lake Erie collections, that gene sequencing has special potential as an important tool for identifying the species.

Need for Further Development of Gene Sequencing Methods for Sphaeriid Clams

We identified and overcame several challenges over the course of this project. For example, we found that it is critical to use a DNA extraction process that has been optimized for molluscs. The typical phenol-chloroform extraction method used in mammals is not efficient in clams. We also overcame some difficulties in designing PCR primers. In all, we tested 9 different sets of primers across 5 genes (16S rRNA, 18S rRNA, 28S rRNA, COI, and PGD) before arriving on the final 3 pairs that we used for all samples.

Two major areas continue to present ongoing challenges to identifying sphaeriid clams by genetic means: 1) development of truly universal primers in clams, and 2) determining the reliability of GenBank results. Though we were able to develop primers for 3 genes that worked reasonably well, there is still a possibility that other primer pairs would work more efficiently. We chose the genes in our study because previous sources have used universal primers for those genes in some species, but we quickly found that some of these primers were not universal for clams. Ultimately, we designed our own primers for the 18S rRNA gene. The library of DNA sequences that we have generated greatly extends our knowledge of the clam genome. Using this information, we may be able to identify better primer pairs for the 16S, 18S, and 28S rRNA genes for the most common species in our dataset.

As mentioned above, there is some question regarding how accurate each species identification in GenBank is. Going forward, it would be extremely useful to have several representative members of each species known to be in the Great Lakes morphologically identified by an expert taxonomist and sequenced for at least 3 genes to use as a gold standard for comparing unknown sequences. This would remove the uncertainty that arises from using GenBank and also fill in the gaps in the data where some species do not have entries in GenBank for some genes. It would also give us a better gauge for determining whether DNA differences in our sample are due to individual-level (intraspecific) variation or are likely to be true distinguishing differences between species.

Choice of Methods Based on Time and Cost Efficiency

We experienced increases in efficiency in accomplishing both the morphological identifications and the genetic sample processing and species identifications. As noted above, there was a considerable learning curve in all areas of this project. Therefore, our estimates of both time and cost efficiencies are undoubtedly quite high compared to estimates that would be calculated if one were to begin with the level of experience we now have. Although the time per specimen to reach an identification was around 4.2 times greater with the gene sequencing approach, much of that time, as explained above, resulted from the students learning the sample processing and gene amplification techniques. Based on prior experience in a medical genetics setting, all of our 175 samples probably could have been processed in a single batch by an experienced laboratory technician in two to three days with the latest equipment and using next-gen sequencing.

Choice of Methods Based on Quality of Identifications

This project confirmed that the two approaches to sphaeriid clam identification are complementary. Where gene sequencing pointed to a single species, it usually but not always confirmed the morphological identification. Where the same species was not indicated by both methods, we re-keyed those specimens to determine whether the morphological identification was mistaken. In a few instances, it was agreed that the morphological identification should be changed to match the genetic

identification; however, in most instances the morphological identification definitely seemed correct. This comparative process was beneficial in that it raised questions as to the validity of paired species that the identification keys indicate are closely related and that possess overlapping characteristics (such as in tooth shape and position) within populations. Further research might show some of those paired species to in fact be single variable species. Questionable pairings based on our results include *Musculium lacustre* and *M. partumeium, Pisidium casertanum* and *P. fallax*, *P. henslowanum* and *P. supinum*, and *Sphaerium striatinum* and *S. simile*.

As presented earlier, gene sequences suggested two or more species instead of a single species for a majority of the specimens so that a definitive identification was not possible with that approach. Because of that ambiguity, it seems that gene sequencing cannot yet stand alone as a means for identifying sphaeriid clams, and the traditional morphological approach, with its own set of drawbacks with currently available keys, should remain as the most reliable identification approach. Gene sequencing was successful, however, in clearly identifying small clams from the upper Great Lakes (*P. moitessierianum*), immature clams, and several specimens in poor condition that could not be successfully identified using the keys. Therefore, in studies where it is critical to know decisively all of the species present, gene sequencing could be called upon as an important identification tool.

Potential Application of DNA Barcoding of Sphaeriid Clams to Environmental DNA (e-DNA) Analysis

Given the challenges discussed above of unambiguously identifying sphaeriid species through gene sequencing, caution is warranted in using DNA barcoding on eDNA samples. While it is possible to identify some samples to the species level using genetics alone (~38% of the samples in our study), in many cases the genetic analysis provided 2 (~41%) or 3 (~16%) possible matches. Without a shell to examine morphologically, as would be the case with an eDNA sample, it would be impossible to further narrow down the list of possibilities. At the very least, it would be essential for gene sequencing to be able to unequivocally identify the presence of individual nonindigenous species in order to detect their spread if already established or their new establishment in a Great Lake.

Recommendations

- We recommend that further research be pursued to identify gene sequences accompanied by refinement of the gene sequencing technology that will provide unambiguous identifications of the sphaeriid clam species known in the Great Lakes and those suspected of potential colonization of the Great Lakes Basin.
- The results of this project indicate that, for the present time, gene sequencing can be used as a tool to augment traditional morphological identification of sphaeriid clams in the Great Lakes. Because gene sequencing yielded ambiguous species identifications for many specimens, it should be used primarily for the tentative identification of small and particularly fragile specimens that cannot be identified morphologically, and it should not serve as the sole method of species identification of adult specimens.
- In about 22% of cases, gene sequencing methods applied to sphaeriid clams yielded ambiguous identifications in our study. The application of e-DNA sampling to assess the present distributions and future range expansions of invasive non-native sphaeriid clams in Lake Erie

should wait until a definitive taxonomy of those species already known to live in Lake Erie and those nonindigenous species capable of establishing populations in the lake is developed.

Summary of Project Deliverables

We listed four project deliverables in our proposal, and they are summarized here.

Deliverable 1 consists of this technical report.

Deliverable 2 was to be the submission of a manuscript to a professional journal. Within a few weeks following acceptance of our final project report, we plan to condense our findings into a suitable length and format for a professional journal article. We have yet to decide on which journal we will submit to; however, we most likely will try to publish the paper in *Freshwater Science*, the *Journal of Great Lakes Research*, or one of the specialized journals devoted to molluscan studies, such as *The Nautilus* or *Malacologia***.**

Deliverable 3 was the presentation of the project results at one or more conferences. Three Heidelberg University students made presentations on campus, as listed in Table 9, and we will also present our results at a scientific conference in 2017, most likely at the annual conference of the International Association for Great Lakes Research in Detroit in May or the Society for Freshwater Science (formerly North American Benthological Society) in Raleigh, NC, in June.

Deliverable 4, along with Deliverable 1.(c), was to recommend whether or not the results of the gene sequencing, or "DNA barcoding", approach as applied in this project can be applied to environmental (water and sediment) samples to detect the presence of the individual species making up entire biological communities through the collection and amplification of e-DNA (environmental DNA). We have discussed this topic in both the Discussion and Recommendations sections of this report.

Additional Outreach

Several student research projects were associated with this project (**Table 9**), including initial method development by Brittan Labry in the fall semester of 2013. This final report is being submitted to Mr. Jeff DeShon, our agency advisor for this project and also manager of the Ecological Assessment Section of Ohio EPA, for his feedback and for potential application to their work in Lake Erie and rivers and creeks throughout Ohio. Also, we will forward this project report, once approved by OLEC, to Mr. Ron Maichle, a senior investigator at the Northeast Ohio Regional Sewer District. Ron is one of the foremost aquatic invertebrate taxonomists in Ohio with respect to the bottom-dwelling (benthic) invertebrates of Lake Erie and its tributaries. He holds Level 3 QDC and is a certified Level 3 trainer for benthic macroinvertebrate biology. Lastly, for samples we have identified with a high degree of certainty, we will submit the associated DNA sequences to GenBank, so that they are publicly available to all interested researchers.

Project Evaluation

In our proposal we stated that we would evaluate the success of this project on the basis of (1) determining the extent of congruence of taxonomic identifications by gene sequencing and traditional morphological taxonomic keys, and (2) providing a definitive answer as to which identification approach is more accurate and cost effective. We accomplished both of these objectives but discovered the

existence of greater complexity in both taxonomic approaches, as discussed earlier in this report. We also are asking our agency advisor, Mr. Jeff DeShon, to provide an independent evaluation of the success of the project.

Table 9. Research awards and honors associated with this project.

Person	Award or Event	Title	Date
Brittany Labry	Martha and Ernest Hammel Student Research Grant, $600	NA	10/11/2013
Brittany Labry	Honors Presentation	Genetic Comparison and Identification of Different Pill Clam Species within the Family Sphaeriidae	4/23/2014
Emily Glor	Martha and Ernest Hammel Student Research Grant, $412	NA	4/10/2015
Emily Glor	Pepsi Fund Grant for $188	NA	4/10/2015
Emily Glor	Honors Presentation	Gene Sequencing as a New Tool for Identifying Native and Invasive Sphaeriid Clams in the Great Lakes	12/1/2015
Emily Glor	Minds At Work Student Research Conference	Gene Sequencing as a New Tool for Identifying Native and Invasive Sphaeriid Clams in the Great Lakes	2/23/2016
Dr. Kylee Spencer	Distinguished Scholar Award	NA	2/26/2016
Emily Glor	Martha and Ernest Hammel Research Award--Best Paper	Gene Sequencing as a New Tool for Identifying Native and Invasive Sphaeriid Clams in the Great Lakes	4/22/2016
Ethan Rodgers	Honors Presentation	The Use of DNA Sequencing to Aid Clam Identification	4/26/2016
Alex Thompson	Research Internship	NA	5/23/2016 to 7/29/2016

Acknowledgements

"This project was funded in part through the Lake Erie Protection Fund, administered by the Ohio Lake Erie Commission (www.lakeerie.ohio.gov). The LEPF is supported by donations or through the contributions of Ohioans who purchase Lake Erie license plates, featuring the Marblehead Lighthouse or Lake Erie life preserver." Undergraduate student participation in the project was partly funded by Hammel and Pepsi student research grants through Heidelberg University. Additional matching funds were provided by the National Center for Water Quality Research. Figure 2 was supplied by Dr. Remegio Confesor, NCWQR, Heidelberg University. Dr. Glenn Warren at U.S. EPA's Great Lakes National Program Office gave permission for the special collection of sphaeriid clams from samples of mud from the bottoms of lakes Superior, Michigan, Huron and Erie from the *R/V Lake Guardian*. Susan Daniel, Katie

Hastings and Sonya Bayba of Buffalo State College (SUNY)'s Great Lakes Institute sieved those samples for the fingernail and pill clams that were subsequently used in this project. Heidelberg University students Brittany Labry, Emily Glor, Ethan Rodgers, and Alexander Thompson made substantial contributions to this project through their research projects and an internship guided by Dr. Kylee Spencer. Dr. Katy Klymus in the laboratory of Dr. Carol Stepien at the University of Toledo's Lake Erie Center provided helpful advice during the project.

Literature Cited

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999**.** *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish*. 2nd Ed. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, DC.

Burch, J.B. 1975. *Freshwater sphaeriacean clams (Mollusca: Pelecypoda) of North America*. Malacological Publications, Hamburg, Michigan. 96 pp.

Cooley, L.R, and D. Ó Foighill. 2000. Phylogenetic analysis of the Sphaeriidae (Mollusca: Bivalvia) based on partial mitochondrial 16s rDNA gene sequences. *Invertebrate Biology* 119:299-308.

Davis, W.S., and T.P. Simon. (eds.) 1995. *Biological assessment and criteria, tools for water resource planning and decision making*. Lewis Publ., Boca Raton. 415 pp.

Deiner, K., R.A. Knapp, D.M. Boianos, and B. May. 2013. Increased accuracy of species lists developed for alpine lakes using morphology and cytochrome oxidase I for identification of specimens. *Molecular Ecology Resources* 13:820–831.

Grigorovich, I., A. Korniushin, and H. MacIsaac. 2000. Moitessier's pea clam *Pisidium moitessierianum* (Bivalvia, Sphaeriidae): a cryptogenic mollusc in the Great Lakes. *Hydrobiologia 435*:153-165.

Kessing B., H. Croom, A. Martin, C. McIntosh, W.O. Mcmillan, S. Palumbi 2000. *The simple fool's guide to PCR*. University of Hawaii, Honolulu, HI.

Krieger, K.A., and L.S. Ross. 1993. Changes in the benthic macroinvertebrate community of the Cleveland Harbor area of Lake Erie from 1978 to 1989. *Journal of Great Lakes Res*earch 19:237-249.

Mackie, G.L., D.S. White, and T.W. Zdeba. 1980. *A guide to freshwater mollusks of the Laurentian Great Lakes with special emphasis on the genus* Pisidium. U.S. Environmental Protection Agency. EPA-600/3-80-068. 144 pp.

Ohio EPA. 1987a. *Biological criteria for the protection of aquatic life: Vol. I: The role of biological data in water quality assessment*. (Updated Feb. 15, 1988). Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio. 44 pp.

Ohio EPA. 1987b. *Biological criteria for the protection of aquatic life: Vol. II: Users manual for biological field assessment of Ohio surface waters.* (Updated Jan. 1, 1988). Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus, Ohio.

Ohio Lake Erie Commission. 2004. *State of the Lake Report 2004, Lake Erie Quality Index*. Ohio Lake Erie Commission, Toledo, OH. 80 pp.

Ohio Lake Erie Commission. 2013. *Lake Erie Protection and Restoration Plan*. LEPR 13. Ohio Lake Erie Commission, Sandusky, OH. 70 pp.

Park, J.K, and D. Ó Foighill. 2000. Sphaeriid and corbiculid clams represent separate heterodont bivalve radiations into freshwater environments. *Molecular Phylogenetics and Evolution* 14:75-88.

Pennak, R.W. 1989. *Fresh-water invertebrates of the United States, Protozoa to Mollusca*. 3rd. Ed. John Wiley & Sons, Inc., NY. 628 pp.

Rahel, F.J., and J.D. Olden. 2008. Assessing the effects of climate change on aquatic invasive species. *Conservation Biology* 22:521–533.

Reynoldson, T.B., D.W. Schloesser, and B.A. Manny. 1989. Development of a benthic invertebrate objective for mesotrophic Great Lakes waters. *J. Great Lakes Res*. 15:669-686.

Rosenberg, D.M., and V.H. Resh. (eds.) 1992. *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman & Hall, NY. 488 pp.

Shokralla, S.I., J.L. Spall, J.F. Gibson and M. Hajibabaei. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology 21:1794-1805.

Stucky, B.J. 2012. *SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms.* Department of Ecology and Evolutionary Biology, University of Colorado, Boulder CO.

Appendix

Site locations of sphaeriid clam specimens used in this study

Site Identification	Sample Date	Descriptive Location (Watershed)	County & State, or Province	Latitude	Longitude
Lake Erie -- open water					
ER43	20150811	Lake Erie central basin north of Westlake, OH	Cuyahoga, OH	41.78833	-81.94500
ER61	20150810	Lake Erie western basin southeast of Detroit River mouth	Ontario, Canada	41.94667	-83.04500
ER78m	20150811	Lake Erie central basin north of Painesville	Lake, OH	42.11667	-81.25000
ER95b	20150811	Lake Erie central basin northwest of Conneaut	Ashtabula, OH	42.00000	-80.66639
Lake Erie -- watershed					
McKibben Ditch downstream section A (MKD)	20060707	McKibben Ditch at Ohio Hwy 529 (Olentangy R.)	Marion, OH	40.53523	-82.99460
Riffle Creek (RIC)	20070710	Riffle Creek at Ohio Hwy 98 (Olentangy R.)	Marion, OH	40.56730	-83.04543
Ackerman Ditch (AD)	20100629, 20110606, 20110630	Ackerman Ditch at Johnston Road (Honey Cr. of Sandusky R.)	Crawford, OH	40.91769	-82.82797
Slee Ditch (SD)	20100608	Slee Ditch at County Road 6 (Honey Cr. of Sandusky R.)	Seneca, OH	41.02353	-83.06697
Zurbach Vogel Ditch (ZVD)	20100601	Zurbach-Vogel Ditch at County Road 56 (Honey Cr. of Sandusky R.)	Seneca, OH	41.06572	-82.92614
Rock Creek slough (RCS)	20160520	Rock Creek slough, behind Hedges-Boyer Park softball field, Tiffin, OH (Sandusky R.)	Seneca, OH	41.10367	-83.15469
Rock Creek (RC)	20160520	Rock Creek, riffle at football stadium,	Seneca, OH	41.11311	-83.16444

(specimens not included in report)		Heidelberg University campus (Sandusky R.)			
Bark Creek (BC)	20080821, 20100526	Bark Creek at Smith Rd. (Sandusky R.)	Sandusky, OH	41.31492	-83.09178
Old Woman Creek @ bend in channel south of railroad bridge (OWC1)	20151021	Old Woman Creek Wetland near railroad bridge (Old Woman Cr.)	Erie, OH	41.37093	-82.51522
Old Woman Creek @ cove east of observation deck (OWC2)	20151021	Old Woman Creek Wetland near visitor center (Old Woman Cr.)	Erie, OH	41.37639	-82.50901
Coon Creek (CC)	20110601	Coon Creek at Wendler Rd. (Portage R.)	Sandusky, OH	41.40678	-83.37917
Muddy Creek @ OH 53 bridge (OH53)	20160518	Muddy Creek Bay at OH Hwy 53 (Sandusky R.)	Sandusky, OH	41.45257	-83.05432
Little Portage @ West Oak Harbor SE Rd. 4 (LP)	20160518	Little Portage R. @ W. Oak Harbor SE Rd. (Portage R.)	Ottawa, OH	41.48633	-83.05612
Metzger Marsh (MM)	20151015	Metzger Marsh (south shore of western L. Erie)	Lucas, OH	41.64946	-83.24118
Lake Huron					
HU54m	20150807	northern Lake Huron north of Presque Isle, MI	MI	45.51667	-83.41667
HU93	20150808	southern Lake Huron west of Ripley, Ontario	Ontario, Canada	44.10000	-82.11667
Lake Ontario					
ON25	20150814	western Lake Ontario south of Pickering	Ontario, Canada	43.516667	-79.08
ON60	20150814	eastern Lake Ontario north of Marion, NY	NY	43.58	-77.20
Lake Superior					
SU22	20150829	western Lake Superior east of Duluth, MN	WI	46.80	-91.75