Investigating the species diversity of *Planorbella* and *Helisoma* (Gastropoda: Hygrophila: Planorbidae) in Manitoba

By: Emma Rempel

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba,

In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba, Winnipeg, MB.

Abstract

Species-level resolution is critical to understanding the ecological role and evolutionary history of taxa. However, factors like cryptic species, phenotypic variation, and subjective interpretations of morphology have confounded efforts to characterize species diversity, leading to species misidentifications and poor taxonomic resolution. To test whether these factors have resulted in inaccurate estimates of species-level diversity in freshwater snails, I used integrative taxonomy to characterize the diversity of five nominal planorbid species (four *Planorbella* and one *Helisoma* species). These species commonly occur in North America and have been subject to taxonomic controversy. In Chapter 1, I explored whether shell morphology was sufficient to reliably distinguish five nominal species by comparing morphology-based identifications made using descriptive taxonomic keys with phylogenetic and geometric morphometric analyses of shell shape. Snails were collected from 18 field sites (n = 257) ranging across two watersheds in central North America, as well as from two laboratory strains: Manitoba (n = 6) and Oklahoma (n = 1). Specimens were digitally imaged and a subset (n = 56) were sequenced at the mitochondrial cytochrome c oxidase I (*COI*) gene (586 bp). Phylogenetic analysis of 48 unique sequences (including 29 unique sequences from field collections (n = 27) and lab strains (n = 2), and 19 from GenBank) revealed five clades that were at least 5% different. However, multiple nominal species were assigned to three of the five clades according to both GenBank identifications and my putative shell-based identifications. Additionally, maximum within-group p-distances for three of the clades approached or surpassed the genetic benchmark of 2% that has been used to suggest the presence of cryptic species. Geometric morphometric analysis of 12 landmarks also suggested that these snail species are cryptic as only two of the five clades were clustered separately, while the remaining three species had overlapping clusters. By employing integrative taxonomy to these

species, I have shown that the genetic and morphological differences among two nominal species are congruent and that shell-based identification of these species is reliable. In contrast, the best method to discriminate the remaining three species is DNA barcoding, as morphological analyses did not reveal clear and reliable delimiting traits. These findings suggest that solely using shell-based traits for identification is not a reliable approach to distinguishing species of *Planorbella* and *Helisoma*.

To further explore the validity of the *COI* clades for use in distinguishing species, Chapter 2 used genome skimming to obtain the mitogenomes and nuclear rRNA repeat genes (*18S*, *5.8S*, *ITSI*, *ITSII*, and *28S*) for each of the five clades found in Chapter 1. The mitogenome phylogeny was comprised of specimens identified as *Helisoma anceps, Planorbella campanulata, Planorbella duryi, Planorbella pilsbryi,* and *Planorbella trivolvis* (n = 1 individual per clade and 2 for *P. trivolvis* clade), and an additional 18 unique sequences in superorder Hygrophila from GenBank (12,907 bp). I compared the results of the barcode phylogeny to those of the mitogenome phylogeny to determine if there were any differences in the topology and support for the monophyly of clades and to more confidently delimit species boundaries. Further, I compared the mitochondrial gene trees to nuclear gene phylogenies for *18S* (20 GenBank sequences, 1,685 bp) and *28S* (19 GenBank sequences, 1,479 bp) to determine if there was any evidence for introgression. I found that the mitochondrial *COI* tree and the mitochondrial genome tree had congruent topologies. I found topological comparisons between mitochondrial and nuclear gene trees to be difficult owing to the lack of branching in nuclear trees, thereby providing no evidence for introgression. Therefore, phylogenetic reconstructions using the entire ribosomal RNA repeat sequence were generated for a subset of taxa that included members from the five clades uncovered in the *COI* analysis (n = 7). These were compared to reconstructions of the *COI*, mitogenome, and

concatenated *18S/28S* rRNA sequences using identical taxa. These comparisons revealed incongruities between mitochondrial and nuclear sequence trees, which could be interpreted as past introgression events among some of the five focal species.

In order to improve taxon sampling for *Planorbella* and *Helisoma*, I have generated the first complete mitochondrial genome sequences, as well as the first complete nuclear rRNA repeat sequences for five planorbid species. To better understand how morphology corresponds to genetic differentiation, I recommend more sampling across the entire geographic range of the focal species, as well as experimentation with different ecological parameters like water flow, to test the range in phenotypic variation in these taxa. The broader significance of this work is to show that problems with the identification of these taxa exist, and that an integrative approach to species delimitation will be essential for freshwater gastropods given the similarity and overlap in their shell morphology, ecology, and geographic distributions. These species are commonly used to reconstruct food webs, explore the evolutionary ecology of host-parasite interactions, and assess ecosystem health. Therefore, this work will improve understanding of parasite-host relationships and promote better conservation of molluscs by grounding *Planorbella* and *Helisoma* in a clearer taxonomic framework.

Acknowledgements

First, I would like to express sincere gratitude to my advisor Dr. Jillian Detwiler for her continuous support, encouragement, and guidance during the course of our professional relationship. Dr. Detwiler has always had the innate ability to recognize when I needed a push or when I needed a dose of confidence. I am very grateful to have had this opportunity to collaborate with her and have learned so much from her example. I would also like to thank Dr. Olivia Wilkins and Dr. Jeffrey Marcus for acting as committee members and providing valuable feedback throughout this process, with special thanks to Dr. Marcus for his collaboration in genomic sequencing.

Second, I acknowledge that my fieldwork took place in the lands of the Anishinaabeg, Cree, Oji-cree, Dakota, and Dene peoples, and on the homeland of the Métis Nation in Manitoba. I acknowledge and respect the Treaties made on these territories.

Field collections throughout Canada and the US were carried out with the help of many people. I want to thank Dr. Matthew Bolek and Scott Malotka for providing snails from Oklahoma and Minnesota, respectively. I thank the many people who have helped me with collections: Dr. Olwyn Friesen, Ashley Pidwerbesky, Ali Kandil, Donald Rempel, Paula Rempel, and Megan Rempel. Thank you to Carly Gair for assistance in DNA sequencing. I want to extend a special thanks to my lab mates for their help and support.

My research was funded by several scholarships and grants, including NSERC Discovery Grants and Fieldwork support grants from the University of Manitoba awarded to Drs. Jillian Detwiler and Jeffrey Marcus.

Finally, I express profound gratitude for my family and friends, for their continued support and encouragement over the course of my academic career. This accomplishment would not have been possible without them.

Table of Contents

List of Tables

List of Figures

Introduction

Taxonomic resolution for the majority of North American gastropods is poorly understood, with both higher and lower level classifications being regularly reassessed (Albrecht et al. 2007; Johnson et al. 2013). Resolution at the family level has been subject to recent changes, while species-level resolution is lacking for even the most widely studied and charismatic genera (Albrecht et al. 2007; Johnson et al. 2013; Saadi et al. 2020). As taxonomic inventories largely rely on surveys and species descriptions conducted more than 30 years ago, resolving these issues is difficult (Johnson et al. 2013). Further, inferring current geographical ranges of taxa from historical surveys may not be accurate due to rapid environmental changes and anthropogenic threats to freshwater habitats that have even led to endemic extinctions (Johnson et al. 2013). Habitat loss, in conjunction with outdated information and endemic extinctions, can lead to identification errors and potentially overestimated range sizes for gastropod species, as similar species can be mistaken for each other outside of their known range (Johnson et al. 2013). Updated information on extant species diversity and their distributions are needed for accurate estimates of gastropod biodiversity. This information should be accompanied by voucher specimens for both molecular and morphological characteristics. When such an approach has been used, our knowledge of gastropod biodiversity has improved with the re-discovery of species previously believed to be extinct, as well as more current information on the conservation status of several endangered species (Johnson et al. 2013). This last outcome is of particular importance for gastropods, which have a current extinction rate that is 9,539 times greater than background rate (Johnson et al. 2013).

Species-level resolution is critical to understanding the ecological role and evolutionary history of taxa (Bolek et al. 2019). However, confounding factors like cryptic species, phenotypic

variation, and subjective interpretations of morphology have led to species misidentifications, poor taxonomic resolution in evolutionary ecology studies, and uncertain estimates of species diversity within gastropod taxa (Gustafson and Bolek 2015; Bolek et al. 2019). Cryptic species (i.e. species that are morphologically similar but genetically distinct), may lead to underestimates of species diversity (Bickford et al. 2007). The likelihood of cryptic species is predicted to increase as geographic distance between populations increases since the reduced gene flow would result in morphologically similar populations that are genetically dissimilar (Bickford et al. 2007). In addition to geographic distance, habitat differences may influence the rate of differentiation of cryptic species (Pérez-Ponce de León and Poulin 2016). For example, freshwater habitats are discontinuous with many barriers to dispersal, which may promote speciation and result in faster rates of diversification in freshwater animals compared to marine animals (Wiens 2015; Pérez-Ponce de León and Poulin 2016; Poulin and Pérez-Ponce de León 2017). Therefore, freshwater gastropods may exhibit faster rates of diversification than their marine counterparts.

As the number of studies in search of cryptic species has increased, there is heterogeneity among higher taxa for the discovery of cryptic species (Poulin and Pérez-Ponce de León 2017). Based on relative species richness and study unit effort, cryptic species are under-reported in many mollusc groups (Pérez-Ponce de León and Poulin 2016). Moreover, there is a lack of studies testing for crypsis in freshwater gastropods, despite some evidence that crypsis has resulted in uncertainty at the species level (Bolek et al. 2019). Additionally, there is a wide range in the rigor applied to analyses that support whether or not cryptic species are present in a population (Struck et al. 2018). In a meta-analysis, nearly half (47%) of the studies that proposed cryptic species included no morphological data, and relied solely on genetic differences (Struck et al. 2018). Though approximately half of studies on cryptic species did incorporate morphological characters, few

(15.4%) presented quantitative analysis of morphology or phenotype (Struck et al. 2018). This omission is problematic because there is no way to assess the degree of morphological similarity of the two proposed cryptic species (Struck et al. 2018). Further, 36% of the genetics-based studies assessed genetic divergence at only a single locus, which may be insufficient to accurately describe cryptic species for several reasons (Struck et al. 2018). For example, mitochondrial cytochrome c oxidase I (*COI*) gene, the uniparental marker commonly used in metazoan DNA barcoding, is vulnerable to horizontal gene transfer (i.e. through hybridization) (Stegemann et al. 2012). In addition, this locus reaches saturation (i.e. multiple substitutions occurring at a single site in a locus) quickly, which masks the actual sequence divergence (Blouin et al. 1998; Xia et al. 2003; Kane et al. 2008). Both of these characteristics potentially hamper the effectiveness of using *COI*, as a single locus, to estimate genetic divergence and assume the estimate reflects genome-wide divergence (Good et al. 2008; Stegemann et al. 2012; Marcus 2018; Struck et al. 2018).

The identification of cryptic species within nominal taxa provides insights into the underlying causes of speciation (Bickford et al. 2007; Poulin and Pérez-Ponce de León 2017). The occurrence of cryptic species could be the result of several scenarios, but two of the most common are: recent speciation or morphological stasis owing to similar stabilizing ecological conditions and selective pressures (Poulin and Pérez-Ponce de León 2017; Struck et al. 2018). A common assumption of speciation is that selection primarily acts on physiological, behavioural, and reproductive traits, while observable changes in morphology take more time to accumulate (Struck et al. 2018). In order to accurately observe speciation, it is essential to combine analyses of phenotypic disparity and genetic divergence in order to confidently assign cryptic species (Struck et al. 2018).

In addition to cryptic species, within-species morphological variation may contribute to the confusion surrounding gastropod identification (Gustafson et al. 2014; Gustafson and Bolek 2015). Historically, snail shell dimensions have been used to identify species and are considered to be phylogenetically informative (Baker 1945; Hubendick 1955; Albrecht et al. 2007). However, experts have also used shell characteristics of fossils to infer details about the snail's habitat, indicating the extent to which the environment influences an individual's development (McKillop 1996). Likewise, in extant populations, some shell traits, such as the size of the aperture, are affected by differences in stream size and flow (Gustafson et al. 2014). Further, predation stressors can affect shell thickness, length, and aperture size (Gustafson et al. 2014). If intraspecific variation is mistakenly interpreted as interspecific variation, then species will be misidentified and species diversity will be overestimated (Gustafson et al. 2014). Shell phenotypes can be affected at all stages of gastropod development including the adult stage. Certain freshwater snails (e.g. *Planorbella trivolvis*) maintain the capacity to alter their shell phenotype past sexual maturity (Hoverman and Relyea 2007). This phenotypic variation within a species, both among mature adults and at various developmental stages, must be accounted for when attempting to identify members of morphologically similar species. To determine the phenotypic limits of variation within a species, many specimens must be collected from a range of environments and locations.

Consideration of phenotypic variance is crucial to accurately identifying snails as shell traits are most often used by non-mollusc specialists to identify species. Ecologists require accurate identifications of snail taxa to estimate their role in food webs (Morgan et al. 2002; Albrecht et al. 2007; Johnson et al. 2013). Likewise, identification to species is essential for parasite evolutionary ecologists who are interested in the role of snails in parasite life cycles (Bolek et al. 2019; Gagnon and Detwiler, 2019). Without species-level identification of the hosts, host-parasite interactions

cannot be described or studied with any certainty or specificity (Bickford et al. 2007). Snails serve as intermediate hosts for many disease-causing digenean parasites, like echinostomes and schistosomes (Detwiler et al. 2010; Gordy and Hanington 2019; McPhail et al. 2021). Without accurate species identifications of their snail hosts, our understanding of fundamental biological traits of parasites (i.e. host specificity, distributions, and life cycles) will remain in doubt (Adamson and Caira 1994; Lockyer et al. 2004; Sultana 2018; Bolek et al. 2019).

Prior to the advent of genetic analysis, snails were traditionally identified by their external shell morphology and classified using a combination of internal and external anatomy (Baker 1945; Hubendick 1955; Albrecht et al. 2007). The use of internal anatomical traits (e.g., reproductive organs) as traditional means of identification and classification have resulted in disagreements and controversy about how to divide the higher taxa (sub-families and tribes) (Morgan et al. 2002; Albrecht et al. 2007). For example, when Hubendick (1978) proposed that two families within the Hygrophila (freshwater panpulmonate snails), Planorbidae and Ancyclidae, be unified into a single family based on the similarity of gonad and prostate characters, the change was not accepted for many years (Burch 1989; Brown 1994) until it was confirmed with molecular systematics (Morgan et al. 2002; Jørgensen et al. 2004; Albrecht et al. 2007; Saadi et al. 2020). Frequently, when traditional morphology-based delimitations have been tested against phylogenetic analysis, intergeneric relationships have been extensively revised (Agapow et al. 2004; Albrecht et al. 2007).

Additionally, traditional anatomical characters used by snail taxonomists for identification are not uniformly informative at the species level. For example, two closely related species of the freshwater snail genus *Planorbella* are not distinguishable by the traditional anatomical analysis of radula or penial complex (Baker 1928). Instead, Baker identified the two species by subtle

differences in their external shell morphology and habitat type (i.e. standing versus moving water), the latter of which has come under question since the two species commonly co-occur (Baker 1928; Pip 1987). Among morphologically similar species, a purely morphological approach to classification may be insufficient for delimitation, or in some cases to discern the limits of phenotypic variation (Dayrat 2005). Therefore, the limits of morphological species delimitation require that these proposed species be tested with additional types of data (Dayrat 2005; Becker et al. 2016; Horsáková et al. 2019).

When delimiting species boundaries, it is crucial to use a set of criteria that can allow for practical identifications (Samadi and Barberousse 2006). For gastropods, the limitations of morphological species delimitation suggest that the Morphological Species Concept (MSC) may be difficult to apply (Agapow et al. 2004). An alternative is the Biological Species Concept (BSC, Mayr 1942), in which species are primarily defined as reproductively isolated populations (Agapow et al. 2004; Baker and Bradley 2006). However, this concept fails to adequately address hybridization between species (Baker and Bradley 2006; Samadi and Barberousse 2006). Hybridization has been documented both in wild and experimental populations of closely related species of gastropods and can be traced in mitochondrial lineages long after one of the species has been locally extirpated from the hybrid zone (Mello-Silva et al. 1998; Shimizu and Ueshima 2000; DeJong et al. 2001). Any species concept to be applied to a gastropod study system must be able to sufficiently address the potential for hybridizing populations (Horsáková et al. 2019).

The Genetic Species Concept (GSC) is better suited to delimiting taxa that have the potential to hybridize (Baker and Bradley 2006; Horsáková et al. 2019). Here, species are defined as genetically isolated, wherein lineages diverge to the point where they no longer share an evolutionary fate (Baker and Bradley 2006). While there is considerable overlap in the Genetic

and Biological Species Concepts, the former is more expansive in its definition and allows for easier recognition of hybrids, which allows for a greater understanding of speciation (Baker and Bradley 2006). Further, the GSC posits that reproductive isolation can be viewed as an 'endpoint' in the divergence of two lineages, rather than the initial breaking off point of divergence (Zink 1996; Baker and Bradley 2006; Samadi and Barberousse 2006). Species boundaries are likely not distinct but rather indefinite, and the species concept that best addresses this vague boundary is the GSC (Agapow et al. 2004; Baker and Bradley 2006; Horsáková et al. 2019).

However, it is important to consider the caveats of the GSC: the degree of taxonomic resolution uncovered is highly dependent on the sample size of taxa and loci (Agapow et al. 2004). To increase taxonomic resolution, sampling efforts should focus on including a sufficient breadth of focal species (Agapow et al. 2004). Taxon sampling may be limited for several reasons, including the limited availability of desired species in museum collections and the feasibility and cost of conducting field collections. The number of independent loci in an analysis also affects the resolution for the GSC (Avise 2000). Additionally, it has been argued that the GSC records phylogenies of genes, rather than phylogenies of species, given that events like gene duplication, gene loss, horizontal transfer, and incomplete lineage sorting may result in incongruities between the lineage for a gene and the lineage for a species as a whole (Agapow et al. 2004). Adoption of the GSC tends to result in over-splitting of taxa, which can have adverse effects on conservation efforts, since over-splitting could create more species with smaller ranges and decreased abundance. If the GSC is widely adopted, it could have the unintended consequence of spreading conservation efforts too thin (Agapow et al. 2004).

In contrast to the GSC, the BSC has a tendency to lump species together. (Agapow et al. 2004). When metazoan species that had been described using either BSC or traditional

morphological identification were reevaluated with quantitative morphological and/or molecular analyses, the average number of species per group increased significantly (Agapow et al. 2004). However, molluscan taxa were an exception, with a notable 50% decrease in the number of species following reevaluation (Agapow et al. 2004). This may be indicative of a tendency in amateur molluscan taxonomists to identify "new" species erroneously based on phenotypic variation or other variable features such as geography, habitat, and anatomy (Agapow et al. 2004). These problems highlight the need for the development and use of a methodology that can integrate morphological and genetic variation to better understand the taxonomy of freshwater gastropods.

I chose to integrate morphological and molecular analyses of members of the Family Planorbidae, the largest group of freshwater panpulmonate snails, containing genera of medical, economical, and ecological importance (Morgan et al. 2002; Jørgensen et al. 2004; Albrecht et al. 2007; Saadi et al. 2020). Members of this family are ubiquitous and diverse, inhabiting different water bodies like lakes, creeks, and streams (Morgan et al. 2002; Albrecht et al. 2007). Phylogenetic relationships within Planorbidae remain confused at higher and lower taxonomic levels. In particular, there is controversy about whether internal anatomical traits (e.g., male copulatory organs, radula) can be reliably used to identify and classify higher taxa (e.g. families), as well as species (Baker 1928; Morgan et al. 2002; Albrecht et al. 2007). Additionally, a majority of the molecular phylogenies focus on resolving higher levels of taxonomy, while largely ignoring species-level resolution (Albrecht et al. 2007; Juqueira Cunha and Giribet 2019). Due to this confusion, there are no precise estimates of species richness for the Planorbidae (Albrecht et al. 2007).

In Canada, the range boundaries of 14 planorbid species converge in Manitoba, a region of rich geological diversity (Pip 1987; Johnson et al. 2013). Of these species, I selected five

Planorbella species and one *Helisoma* species to examine because they have been subject to taxonomic controversy, with disagreements on whether to split or to lump taxa (Pip 1987; Johnson et al. 2013). At present, the following species are recognized as nominal taxa in the most recent reassessment in North America: *Planorbella campanulata, Planorbella corpulenta, Planorbella pilsbryi, Planorbella subcrenata, Planorbella trivolvis,* and *Helisoma anceps* (Johnson et al. 2013). An example of some of the taxonomic controversy is the elevation of *P. subcrenata* as a nominal species. Formerly a subspecies of *P. trivolvis,* it was later elevated to species status due to differences in the axial height of the shell (Burch 1989). However, this character trait was found to be highly variable and may be leading to misidentifications (Pip 1987; Johnson et al. 2013). Additionally, some consider *P. pilsbryi* to be a genetic variant of *P. trivolvis*, or possibly an eco-phenotype that arose due to dimorphism (Pip 1987). Delimitation between these two species is not made any easier by other aspects of their biology since they frequently co-occur and share ecological and chemical habitat preferences. These examples from some of the focal species in this study indicate the depth of ignorance around intra- and intergeneric relationships in *Planorbella* and *Helisoma* and highlight the need for further investigation to address these questions.

My thesis consists of two chapters: the first tested whether cryptic species and phenotypic variation have led to inaccurate estimates of freshwater snails by using integrative taxonomy. I employed an integrative approach that compared identifications of focal species based on shell morphology, phylogenetic clades, and geometric morphometric analysis. I compared the accuracy of identifications based on descriptions of shell morphology using dichotomous keys to molecular identifications using a single mitochondrial locus (*COI*). I also tested if geometric morphometric analysis of shell shape is a more reliable method of comparative morphological species

delimitation than descriptive shell-based identifications. I predicted that geometric morphometric analysis would be more objective and therefore more reliable than descriptive shell-based identifications. The second chapter used additional genomic markers, including mitogenome and nuclear ribosomal RNA repeat sequences, to reconstruct the evolutionary relationships of these taxa. I aligned and annotated mitogenomes for a minimum of one specimen per clade based on the *COI* phylogeny. I predicted that phylogenies reconstructed from the *COI* barcode region and the entire mitogenome would share congruent topologies, given that all mitochondrial DNA loci are linked and there is no recombination. Additionally, I compared the nuclear rRNA repeat phylogeny to the mitochondrial phylogenies to test whether there was evidence of introgression. It was hypothesized by Pip (1987) that *P. trivolvis* may have hybridized with *P. corpulenta*, given their frequent co-occurrence in past studies. However, given the limited evidence of successful hybridization within Planorbidae, I predicted no introgression in the focal species. Overall, this work will clarify understanding of the diversity of planorbids in North America and will help to elucidate evolutionary relationships within these taxa, which will improve our knowledge of snail host-parasite relationships.

Literature Cited

Adamson, M.L., and Caira, J.N. 1994. Evolutionary factors influencing the nature of parasite specificity. Parasitology **109**(S1): S85–S95.

Agapow, P.M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., and Purvis, A. 2004. The impact of species concept on biodiversity studies. Q. Rev. Biol. **79**(2): 161–179.

Albrecht, C., Kuhn, K., and Streit, B. 2007. A molecular phylogeny of Planorboidea (Gastropoda, Pulmonata): Insights from enhanced taxon sampling. Zool. Scr. **36**(1): 27–39.

Avise, J.C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusetts.

Baker, F.C. 1928. The freshwater Mollusca of Wisconsin. Wisconsin Academy of Sciences, Arts, and Letters.

Baker, F.C. 1945. The molluscan family Planorbidae. University of Illinois Press, Urbana, IL.

Baker, R.J., and Bradley, R.D. 2006. Speciation in mammals and the genetic species concept. J. Mammal. **87**(4): 643–662.

Becker, M., Zielske, S., and Haase, M. 2016. Conflict of mitochondrial phylogeny and morphology based classification in a pair of freshwater gastropods (Caenogastropoda, Truncatelloidea, Tateidae) from New Caledonia. Zookeys **2016**(603): 17–32.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., and Das, I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. **22**(3): 148–155.

Blouin, M.S., Yowell, C.A., Courtney, C.H., and Dame, J.B. 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Mol. Biol. Evol. **15**(12): 1719–

1727.

Bolek, M.G., Detwiler, J.T., and Stigge, H.A. 2019. Selected wildlife trematodes. *In* Digenetic Trematodes, 2nd edition. Springer Nature.

Brown, D. 1994. Freshwater snails of Africa and their medical importance. *In* 2nd edition. Taylor & Francis, Bristol, PA.

Burch, J.B. 1989. North American freshwater snails. Malacological Publications, Hamburg, MI.

Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. **85**(3): 407–415.

DeJong, R.J., Morgan, J.A.T., Lobato Paraense, W., Pointier, J.P., Amarista, M., Ayeh-Kumi, P.F.K., Babiker, A., Barbosa, C.S., Brémond, P., Canese, A.P., De Souza, C.P., Dominguez, C., File, S., Gutierrez, A., Nino Incani, R., Kawano, T., Kazibwe, F., Kpikpi, J., Lwambo, N.J.S., Mimpfoundi, R., Njiokou, F., Poda, J.N., Sene, M., Velásquez, L.E., Yong, M., Adema, C.M., Hofkin, B. V., Mkoji, G.M., and Loker, E.S. 2001. Evolutionary relationships and biogeography of *Biomphalaria* (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. Mol. Biol. Evol. **18**(12): 2225–2239.

Detwiler, J.T., Bos, D.H., and Minchella, D.J. 2010. Revealing the secret lives of cryptic species: Examining the phylogenetic relationships of echinostome parasites in North America. Mol. Phylogenet. Evol. **55**(2): 611–620.

Good, J.M., Hird, S., Reid, N., Demboski, J.R., Steppan, S.J., Martin-Nims, T.R., and Sullivan, J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol. Ecol. **17**(5): 1313–1327.

Gordy, M.A., and Hanington, P.C. 2019. A fine-scale phylogenetic assessment of digenean trematodes in central Alberta reveals we have yet to uncover their total diversity. Ecol. Evol.

9(6): 3153–3238.

Gustafson, K.D., and Bolek, M.G. 2015. Effects of trematode parasitism on the shell morphology of snails from flow and nonflow environments. J. Morphol. **277**(3): 316–325.

Gustafson, K.D., Kensinger, B.J., Bolek, M.G., and Luttbeg, B. 2014. Distinct snail (*Physa*) morphotypes from different habitats converge in shell shape and size under common garden conditions. Evol. Ecol. Res. **16**(1): 77–89.

Horsáková, V., Nekola, J.C., and Horsák, M. 2019. When is a "cryptic" species not a cryptic species: A consideration from the Holarctic micro-landsnail genus *Euconulus* (Gastropoda: Stylommatophora). Mol. Phylogenet. Evol. **132**: 307–320.

Hoverman, J.T., and Relyea, R.A. 2007. How flexible is phenotypic plasticity? Developmental windows for trait induction and reversal. Ecology **88**(3): 693–705.

Hubendick, B. 1955. Phylogeny in the Planorbidae. Trans. Zool. Soc. London **28**(1): 453–542.

Hubendick, B. 1978. Systematics and comparative morphology of the Basommatophora. *In* Pulmonates. Academic Press, New York, NY.

Johnson, P.D., Bogan, A.E., Brown, K.M., Burkhead, N.M., Cordeiro, J.R., Garner, J.T., Hartfield, P.D., Lepitzki, D.A.W., Mackie, G.L., Pip, E., Tarpley, T.A., Tiemann, J.S., Whelan, N. V., and Strong, E.E. 2013. Conservation status of freshwater gastropods of Canada and the United States. Fisheries **38**(6): 247–282.

Jørgensen, A., Kristensen, T.K., and Stothard, J.R. 2004. An investigation of the "Ancyloplanorbidae" (Gastropoda, Pulmonata, Hygrophila): Preliminary evidence from DNA sequence data. Mol. Phylogenet. Evol. **32**(3): 778–787.

Juqueira Cunha, T., and Giribet, G. 2019. A congruent topology for deep gastropod relationships. Proc. R. Soc. B Biol. Sci. **286**(1898).

Kane, R.A., Stothard, J.R., Emery, A.M., and Rollinson, D. 2008. Molecular characterization of freshwater snails in the genus *Bulinus*: A role for barcodes? Parasites and Vectors **1**(1): 1–15.

Lockyer, A.E., Jones, C.S., Noble, L.R., and Rollinson, D. 2004. Trematodes and snails: An intimate association. Can. J. Zool. **82**(1): 251–269.

Marcus, J.M. 2018. Our love-hate relationship with DNA barcodes, the Y2K problem, and the search for next generation barcodes. AIMS Genet. **5**(1): 1–23.

McKillop, B. 1996. Geographic and environmental distribution of freshwater gastropods in Manitoba, Canada. Manitoba Museum Man Nat. Occas. Ser. **1**: 1–34.

McPhail, B.A., Rudko, S.P., Turnbull, A., Gordy, M.A., Reimink, R.L., Clyde, D., Froelich, K., Brant, S. V., and Hanington, P.C. 2021. Evidence of a putative novel species of avian schistosome infecting *Planorbella trivolvis*. J. Parasitol. **107**(1): 89–97.

Mello-Silva, C.C., Grault, C.E., Da Costa, V.A., and Barbosa, F.S. 1998. Possible hybridization of Brazilian planorbid snails and its importance in population dynamics. Mem. Inst. Oswaldo Cruz **93**(SUPPL. I): 227–232.

Morgan, J.A.T., DeJong, R.J., Jung, Y., Khallaayoune, K., Kock, S., Mkoji, G.M., and Loker, E.S. 2002. A phylogeny of planorbid snails, with implications for the evolution of *Schistosoma* parasites. Mol. Phylogenet. Evol. **25**(3): 477–488.

Pérez-Ponce de León, G., and Poulin, R. 2016. Taxonomic distribution of cryptic diversity among metazoans: Not so homogeneous after all. Biol. Lett. **12**(8): 1–5.

Pip, E. 1987. Ecological differentiation within genus *Helisoma* in Central Canada. Nautilus (Philadelphia). **101**(1): 33–44.

Poulin, R., and Pérez-Ponce de León, G. 2017. Global analysis reveals that cryptic diversity is

linked with habitat but not mode of life. J. Evol. Biol. **30**(3): 641–649.

Saadi, A.J., Davison, A., and Wade, C.M. 2020. Molecular phylogeny of freshwater snails and limpets (Panpulmonata: Hygrophila). Zool. J. Linn. Soc. **190**(2): 518–531.

Samadi, S., and Barberousse, A. 2006. The tree, the network, and the species. Biol. J. Linn. Soc. **89**(3): 509–521.

Shimizu, Y., and Ueshima, R. 2000. Historical biogeography and interspecific mtDNA introgression in *Euhadra peliomphala* (the Japanese land snail). Heredity (Edinb). **85**(1): 84–96.

Stegemann, S., Keuthe, M., Greiner, S., and Bock, R. 2012. Horizontal transfer of chloroplast genomes between plant species. Proc. Natl. Acad. Sci. U. S. A. **109**(7): 2434–2438.

Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.-H., Liow, L.H., Nowak, M.D., Stedje, B., Bachmann, L., and Dimitrov, D. 2018. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. **33**(3): 153–163.

Sultana, A. 2018. Integrating morphology and genetics to estimate species diversity, host specificity and life cycles of echinostome trematodes. University of Manitoba.

Wiens, J.J. 2015. Faster diversification on land than sea helps explain global biodiversity patterns among habitats and animal phyla. Ecol. Lett. **18**(11): 1234–1241.

Zink, R.M. 1996. Species concepts, speciation, and sexual selection. J. Avian Biol. **27**(1): 1–6.

Chapter 1: An Integrative Approach to Untangle a Taxonomic Quagmire in Freshwater Snails (Family: Planorbidae).

1.1 Introduction

Freshwater snails are essential components of the ecosystems they inhabit, playing key roles as principal grazers, prey for vertebrate and invertebrate species alike, and contributing to nutrient cycling (McKillop 1996; Johnson et al. 2013). Snails are also crucial to parasitic life cycles because most parasites require snails as hosts at some stage in their life cycle (Adamson and Caira 1994; Lockyer et al. 2004; Bolek et al. 2019). Thus, consideration of snail host diversity is critical to parasite evolutionary ecology (Bickford et al. 2007; Bolek et al. 2019). In addition to host-parasite interactions, a thorough understanding of snail species diversity is critical for the conservation of molluscs and the ecological interactions that they are involved in (McKillop 1996; Agapow et al. 2004; Johnson et al. 2013; Bolek et al. 2019). For instance, assessing gastropod community composition (e.g. species richness) in freshwater sites provides information on the success of natural resource management plans, as snails can be indicators of nutrient content, chemical composition (including pollution from agricultural run-off), and water quality (McKillop 1996).

Although snails are ecologically important and indicators of ecosystem health, one major barrier that prevents them from being the focus of more studies is the current state of their taxonomy. Historically and currently, shell characteristics are most commonly used to identify snails and classify them into different taxonomic groups (Clarke 1981; Burch 1989). In addition, some internal features are also examined, such as the patterning of the radula as well as penial characteristics (Baker 1945; Hubendick 1955). However, the use of the shell and internal anatomical traits as the primary means of classification have resulted in disagreements and controversy about how to divide the higher taxa of freshwater pulmonates (sub-families and tribes)

(Morgan et al. 2002; Albrecht et al. 2007). More recently, molecular phylogenies using one or more independent molecular markers (*COI*, *16S*, and *18S*), focused on resolving groupings at higher taxonomic levels (i.e. family and subfamily), while ignoring species-level diversity (Jørgensen et al. 2004; Albrecht et al. 2007). However, species-level resolution is most important for studying fine scale interactions and relationships for ecology and conservation (Bolek et al. 2019). At present, there is a lack of taxon sampling within many pulmonate groups to make comparisons at the species level meaningful. Within freshwater pulmonates, the superorder Hygrophila comprises some of the most medically and economically important snails worldwide (Saadi et al. 2020). Notable unifying anatomical characters of this group include the presence of a lung, the absence of an operculum, and the placement of eyes at the base of the tentacles (Johnson et al. 2013; Saadi et al. 2020). Snails in this group serve as intermediate hosts for a range of parasites of both medical and veterinary importance (Saadi et al. 2020). Of the families within the Hygrophila, Planorbidae is the most diverse and species-rich, with over 250 species (Saadi et al. 2020). Most previous molecular studies within the Planorbidae have focused on the most medically important genera, while less attention has been paid to others (Saadi et al. 2020).

The taxonomic history of Planorbidae is complicated and has undergone many revisions, as technologies have improved and systematic methods have been reevaluated (Albrecht et al. 2007; Saadi et al. 2020). Revisions have largely altered classifications that were based on comparative morphology. For instance, Hubendick (1955) lumped two previously distinct families, Planorbidae and Ancyclidae, into one family based on gonad and prostate characters (Saadi et al. 2020). Some researchers chose not to accept this reclassification (Burch 1989), while others accepted it (Albrecht 2007), with little explanation provided in either case (Saadi et al. 2020). Among the molecular phylogenies of Planorbidae, few have focused on the family as a

whole; rather most have focused on phylogenetic relationships of medically important genera (Saadi et al. 2020). The currently accepted taxonomy includes 52 species of Planorbidae within 16 genera in North America, with 25 species endemic to Canada (Johnson et al. 2013). Of these, ten species (19%) are presumed extinct, with an imperilment rate of 44% owing to the highly restricted distributions of certain taxa (Johnson et al. 2013).

The traditional anatomical characters used to divide the Planorbidae at higher levels (subfamilies and tribes), such as male copulatory organ structure or radulae, have not proven informative for distinguishing *Planorbella* spp. (Baker 1945; Albrecht et al. 2007). Even Baker (1928; 1945), whose work on describing the family is considered the most thorough and foundational, did not use anatomical traits to divide certain species within the genus *Planorbella,* as they were considered uninformative and insufficiently distinct. Examinations of the Planorbidae like that of Baker (1945) and Hubendick (1955) attempted to define diversity past the genus level, though the distinctions made between closely related species were often in continuous shell character traits and are therefore of limited utility for species delimitation. For example, Baker (1945) found the closely related species *Planorbella pilsbryi* and *Planorbella trivolvis* to be indistinguishable based on their anatomy, including radula and genitalia, so he distinguished these two species using several shell characteristics such as whorl height, aperture shape, and umbilicus shape. The problem is that all of these shell traits are continuous, and Baker noted that there were intermediate forms for these two species which were endemic to specific localities (1945). In this and other *Planorbella* spp., morphological comparisons between taxa can be complicated by interspecific morphological similarity.

Additionally, phenotypic plasticity in diagnostic characters can lead to misidentifications. Snail shells, which are the predominant feature used for identification, often demonstrate

"perplexing variation" which can befuddle experts in taxonomy (Baker 1945; Hubendick 1955; Morgan et al. 2002; Albrecht et al. 2007). The typical planorbid shell is planospiral and can vary widely in its size (5-25mm) (Johnson et al. 2013). The morphological keys typically used for identification often only describe a particular life stage (e.g. adults), making identifications of juveniles difficult or impossible. Also, these morphological keys often employ technical language with relative comparisons across taxa, requiring that the user possess a high level of expertise. Failure to meet this expertise commonly results in misdiagnoses (Hebert et al. 2003).

Another limitation of a purely morphology-based identification system is the inability to account for cryptic taxa. Cryptic species, or species that are morphologically similar but genetically distinct, may lead to underestimates of species diversity (Bickford et al. 2007). There are very few reports of cryptic species in the Planorbidae, but this may largely be due to a lack of investigative effort devoted to testing for cryptic species. A search of Web of Science (July 9, 2021) with the query terms "Planorbidae" and "cryptic species" returned 17 entries, of which only three examined planorbid groups for crypsis (Tuan and dos Santos 2007; Macher et al. 2016; Hobbs et al. 2021) while the remainder discussed invasive species ("cryptic invaders") or host-parasite relationships, in which the parasite is part of a cryptic species complex. These three studies were united in their use of molecular systematics to identify cryptic species, while they differed in the loci they selected for their respective study systems: Macher et al. (2016) used *COI* in *Ancyclus* spp. in Spain; Tuan and dos Santos (2007) used *ITS2* for *Biomphalaria* spp. in Brazil; and Hobbs et al. (2021) used both *ITS* and *COI* regions in *Segmentina* spp. in Europe. Additionally, only Hobbs et al. (2021) determined the morphological similarity of their taxa using detailed analysis (geometric morphometric analysis), while the others used only cursory analysis and qualitative

features. The results of the Web of Science search suggest that no studies have investigated crypsis in North American planorbid taxa such as *Planorbella* or *Helisoma*.

One approach that is increasingly being used to genetically delimit species is genetic barcoding. DNA barcoding is a system of rapid species identification that uses a short, standardized gene that acts as the species' unique 'barcode', or identifier, that is compared across a vast library of sequences (Hebert et al. 2003; Hebert and Gregory 2005). Various loci, commonly the mitochondrial *COI* gene and ribosomal *16S* gene, have proven effective barcodes for identifications at higher-level and species-level taxa (Hebert et al. 2003; Hebert and Gregory 2005). These DNA barcodes may reveal diversity among morphospecies that was not captured by comparative morphology and lead to the discovery of novel taxa (Kress and Erickson 2008; Layton et al. 2014). However, barcoding also has its limitations, as its utility is constrained to taxa that are represented in databases with sufficiently diagnostic sequences (Dayrat 2005; Albrecht et al. 2007). For example, some genetically distinct clades of *Planorbella trivolvis* were recently found, but no detailed assessment of shell morphology or internal anatomy was included, making it difficult to biologically interpret the genetic differences and compare to specimens involved in other studies (Martin et al. 2020). Additionally, some sequences can be of poor quality and species identifications assigned to them can be misdiagnoses. Although having the DNA sequences for snails is helpful for species delimitation, a link to a phenotype is critical for comparisons to past and future studies.

One approach that links genotype and phenotype in species delimitation is integrative taxonomy (Dayrat 2005; Perez and Minton 2008; Schlick-Steiner et al. 2010). This method of combining multidisciplinary findings to come to a consensus in the classification of species provides a more robust strategy to delimit species boundaries and identify diagnostic genetic and

morphological differences between species (Dayrat 2005; Perez and Minton 2008; Schlick-Steiner et al. 2010). Integrative taxonomy can also incorporate more objective approaches to quantifying intraspecific and/or interspecific differences such as geometric morphometric analysis (Rohlf and Marcus 1993; Webster and Sheets 2010). This type of landmark analysis can be used to identify and quantify minute morphological differences in the shape of a given character (Rohlf and Marcus 1993; Webster and Sheets 2010). This method has frequently been used in gastropods as a means of analyzing the changes in shell shape of conspecifics caused by external factors (e.g. stream flow, parasitism) (Gustafson and Bolek 2015; Parra and Liria 2017). Fewer studies have explored geometric morphometric analysis as a potential means of species identification and none have applied it to *Planorbella* and *Helisoma*. This study will examine the utility of one method of landmark application and subsequent geometric morphometric analyses as a means of species identification for the focal species.

Here, an integrative taxonomic approach was implemented by combining phylogenetic analyses, morphological analyses, and geographic information to assess the species status of six nominal species of Planorbidae that occur in Manitoba. This approach has been used successfully in many metazoan taxa, including gastropods. An integrative approach is helpful because differences in geographic range cannot be used to help discriminate among species. The geographic range boundaries of 5 *Planorbella* and 1 *Helisoma* species overlap in central Canada and around the Great Lakes, chiefly in Saskatchewan, Manitoba, and Ontario (Pip 1987; Johnson et al. 2013). The focal species for this study are *Planorbella campanulata*, *Planorbella corpulenta, Planorbella pilsbryi, Planorbella subcrenata, Planorbella trivolvis,* and *Helisoma anceps.* Of these species, *P. trivolvis, H. anceps,* and *P. campanulata* are the most common and widely distributed in Canada (Boerger 1975). Ecological surveys suggested that these species coexist by

minimizing competition and favoring habitats with slightly different physical and chemical compositions, a tendency which may be important in divergent evolution (Pip 1987). Co-occurrences are common, however, and two or occasionally all three of these species can inhabit the same lake or pond. Some species co-occur more frequently; for instance the occurrence of *H. anceps* with either of the other two species was more commonly observed than a *P. trivolvis*-*P. campanulata* co-occurrence (Boerger 1975).

Planorbella trivolvis has the widest tolerance for water body types, bottom substrates, and water chemistry of all the six focal species and this wider ecological tolerance has enabled *P. trivolvis* to colonize habitats over a vast spatial distribution (Boerger 1975). *Planorbella trivolvis* previously included two sub-species: *P. t. trivolvis*, found east of Manitoba, and *P. t. subcrenata*, found in Manitoba and westward (Clarke 1981; Pip 1987). These two subspecies also co-occur regionally, as both were recorded in southern Manitoba, with *P. t. subcrenata* being the more common of the two (Pip 1987). *Planorbella subcrenata* is currently considered a valid species on the Integrated Taxonomic Information System (ITIS), however some taxonomic experts argue that *P. subcrenata* is not valid and should have remained a subspecies (Dillon 2019). These two species have historically been distinguished by differences in the axial height of the shell, but the wide range in this continuous character trait and the occurrence of intergradations of both morphotypes suggest that further study is needed to assess the status of these two forms (Clarke 1981; Pip 1987; McKillop 1996).

Similar to *P. trivolvis*, *H. anceps* is common, widespread, and has two formerly recognized subspecies, *H. a. anceps* and *H. a. royalense,* which have subsequently been deemed invalid by ITIS (Boerger 1975; McKillop 1996). It was theorized by McKillop (1996) that a single population of *H. a. royalense* existed in Manitoba, at Little Limestone Provincial Park, but recent surveys at

this site did not recover any members of *H. anceps* (Pip 2000, E. Rempel, unpublished). As mentioned above, *H. anceps* commonly co-occurs with *P. trivolvis* and *P. campanulata* (Boerger 1975). A combination of different life history traits may enable coexistence while minimizing competition: *H. anceps* lay their eggs later in the spring (April-May) season compared to both *Planorbella* species, and juveniles have slower growth rates compared to both of the other species (Boerger 1975). This temporal separation of reproduction and maturity during the crucial early months of the life cycle when snail density is highest may be an important mechanism to lessen the pressures of competition for space and resources (vegetation) (Boerger 1975).

Unlike the other species, *P. campanulata* has a relatively short life cycle; *H. anceps* and *P. trivolvis* have two-year life cycles, while *P. campanulata* has a one-year cycle (Boerger 1975). It is also less flexible in its habitat choice, which renders the species particularly susceptible to environmental changes (Boerger 1975). *Planorbella campanulata* also has two subspecies that have been deemed invalid per ITIS: *P. c. campanulata* was considered widespread across its range, while *P. c. collinsi* was considered rare with a constrained range (McKillop 1996).

The remaining two species, *P. corpulentum* and *P. pilsbryi*, are reportedly rare, with infrequent sightings in their current range (Boerger 1975; Pip 1987). Little is known about *P. corpulentum*, given its scarcity; Whitefish Lake near the Saskatchewan-Manitoba border is its only western location outside of Winnipeg and its current conservation status is 'vulnerable' (Baker 1928; Pip 1987). Similarly, *P. pilsbryi* has been the focus of few studies, though that may be because it has been misidentified as *P. trivolvis*. *Planorbella pilsbryi* was previously considered a subspecies of *P. trivolvis* (Baker 1945) and has since been elevated to the species rank based on the axis of its shell coiling (Burch 1989; McKillop 1996; Turgeon and Quinn 1998; Johnson et al. 2013).

Given the geographic overlap in distribution of the six focal species, morphological differences have been used to delimit species. However, as alluded to above, a recurring theme found in past studies is that variations in morphology can be difficult to interpret. For example, the morphological variation associated with shell characteristics makes it difficult to distinguish between *P. pilsbryi* and *P. trivolvis* (Baker 1928, 1945; Clarke 1973; Pip 1987). Several hypotheses explain the close relationship between these two species. First, Clarke (1973) posited that *P. pilsbryi* was a hybrid of *P. trivolvis* and *P. corpulenta*; however, Pip (1987) noted that *P. corpulenta* was far less commonly observed in the study area indicating that hybridization was unlikely.

Alternatively, a second hypothesis is that the two species are distinct but hybridize when they co-occur (Pip 1987). Intermediate phenotypes between the two parent species have been observed in water bodies where the two species co-occur, which could be evidence of hybridization (Baker 1928; Pip 1987). Hybridization between other species in Family Planorbidae has been proposed based on morphology, but these hypotheses have not been confirmed with genetic evidence (Mello-Silva et al. 1998; Lotfy et al. 2005). Most studies have focused on *Biomphalaria* spp., with successful experimental crosses between two Brazilian species, *Biomphalaria glabrata* and *Biomphalaria tenagophila* (Mello-Silva et al. 1998). To my knowledge, no experimental or observational studies, incorporating morphological and/or molecular analyses, have investigated hybridization among *Planorbella* or *Helisoma* spp.

A third possibility is that *P. pilsbryi* is a genetic variant of *P. trivolvis*; with dimorphism exhibited in populations where the two forms appear distinct (Pip 1987). Finally, *P. pilsbryi* may be an ecophenotype of *P. trivolvis*; this hypothesis is supported by the well-defined ecological parameters within which *P. pilsbryi* has previously been identified (Pip 1987). In the presence of

these ecological criteria, including sand substrates, low water chemistry parameters (e.g. nitrate, chloride), and high macrophyte diversity, the ecophenotype is manifested if the associated genotypes are present in the population (Pip 1987). Whether any of these proposed explanations account for the association between the two taxa, it is clear that this relationship requires a closer critical examination using both morphological and molecular analyses.

In this study, I used integrative taxonomy to characterize species diversity in the genera *Planorbella* and *Helisoma* in Manitoba. I field-collected snails from several watersheds in an attempt to collect 6 species of Planorbidae that had formerly been identified in Manitoba (*Planorbella campanulata, Planorbella corpulenta, Planorbella pilsbryi, Planorbella trivolvis, Planorbella subcrenata,* and *Helisoma anceps*). I then compared the numbers of distinct species based on three different approaches: traditional shell-based identification, genetic clades from phylogenetic analysis of the partial *COI* gene, and geometric morphometric analysis. I hypothesized that the use of a purely shell-based identification system would not match the results of the genetic and geometric morphometric analyses because of the known issues with this approach. I also hypothesized that the use of a purely shell-based identification system would not uncover occurrences of cryptic species. I predicted that diversity was underestimated and that the use of molecular data would uncover cryptic species of *Planorbella* and *Helisoma*.

1.2 Methodology

1.2.1 Specimen collection

Specimens were hand collected at 16 locations throughout Manitoba, Canada, and one location each from Minnesota and Oklahoma, USA (Table 1.1). Snails originating as two separate laboratory strains (Manitoba, Oklahoma) were also vouchered and analyzed as they have been part of host-parasite interaction studies (Eliuk et al. 2020) (Table 1.1). To ensure that the phenotypic

variation measured was not influenced by parasitism, all snails were screened for trematode parasites and only snails with no evidence of patent infection (i.e. emerging cercariae, the free-swimming larvae of trematodes) were used. Briefly, snails were assessed for parasites by placing them in individual well plates containing non-chlorinated water that were then placed under lamps to artificially stimulate the emergence of cercariae (Eliuk et al. 2020). All snails were preserved in 100% ethanol and stored at -20 ˚C.

1.2.2 Morphological vouchering

Prior to DNA extraction, a series of voucher images were captured for the shells of each specimen using a Canon EOS 90D camera with a 100 mm f/2.8L Macro IS USM lens. Images were taken of the apical and basal views of the shell, as well as a view of the aperture (Nantarat et al. 2019). Several meristic traits were recorded for each specimen: shell width, shell height, aperture width, aperture height, and the number of whorls visible from both the apical and basal sides of the shell (Madjos and Demayo 2018). Measurements were performed using ImageJ software (Schneider et al. 2012).

1.2.3 Geometric morphometric analyses

To apply 12 landmarks, I superimposed a fan on each image using the software MakeFan 8.0 (Parra and Liria 2017). Fans originated from the centroid of three distinct anatomical loci around the aperture: the apex of the aperture, the basal intersection of the aperture and the body whorl, and the apical intersection of the aperture and the body whorl (Parra and Liria 2017). The placement of the three markers on these loci coincided with the subsequent placement of three landmarks: 1, 8, and 12, respectively (Figure 1.1). Each fan consisted of 10 equidistant rungs and was used to place the remaining eight sliding semi-landmarks along the curve of the shell (Parra and Liria 2017). The tps software suite was used to capture and digitize the 12 landmarks (Rohlf

2015). Landmarks were aligned using tpsRelw software, which uses generalized Procrustes analysis to extract configurations and centroid size (Rohlf 2015; Parra and Liria 2017). The final digitized landmark configurations were comprised of a mix of both traditional type I landmarks (landmarks 1, 8, 11, and 12) and type II sliding semi-landmarks (2-7, 9, and 10) for a total of 12 landmarks (see Figure 1.1) (Parra and Liria 2017). Any specimens missing one or more of the landmarks were excluded from the analysis.

Geometric morphometric analyses were conducted using the Integrated Morphometrics Package (IMP) v.8 software suite (Sheets 2014). Procrustes superimposition isolated the variation in shell shape and removed signal from the size and orientation of the shells from which I generated partial warp scores using CoordGen v.8 (Sheets 2014). The mean form of all samples was computed and the variation in shell shape for each specimen was quantified as the average partial Procrustes distance from the mean form (Webster and Sheets 2010). Bootstrap resampling (1600 replicates) provided the basis for upper and lower confidence limits (Webster and Sheets 2010). A principal component analysis (PCA) of partial warp scores was conducted with the package "geomorph" in RStudio v.1.2.1335 to determine if specimens clustered according to their putative identifications (Adams et al. 2021). Measurement error was assessed for one specimen (specimen ID: Plsp63), which was cleaned, mounted, photographed, and dismounted a total of ten times. Ten landmark configurations were digitized, one from each replicate image, to provide an estimate of measurement error, as recommended by Webster and Sheets (2010).

1.2.4 Molecular sequencing

After capturing a voucher image of each specimen, two individuals per phenotype per location were selected for barcoding at the *COI* gene. This region is commonly used to identify invertebrates and more specifically, gastropods, to species (Kress and Erickson 2008; Bolek et al.

2019; deWaard et al. 2019). Owing to the more rapid mutation rate of mitochondrial DNA versus nuclear, intrageneric differences are more apparent (Yang et al. 2014). A 30 mg piece of tissue was removed from the head and mantle region, while entire specimens were used in DNA extractions when the body mass was less than 30 mg. Preserved tissue samples were rehydrated in MilliQ water, pulverized in liquid nitrogen, and incubated in 350 µL ML1 buffer solution with 25 µL proteinase K at 60 ˚C for a minimum of three hours. DNA extractions were performed using the E.Z.N.A. Mollusc kit; following incubation, DNA samples were vortexed and eluted in 25 µL of MilliQ water. Eluted samples were stored at -20 ˚C until ready to perform polymerase chain reaction (PCR).

Prior to performing PCR, DNA was quantified using the Take3 micro-volume (2 µL) plate absorbance detection application for the Bio-Tek Synergy #1 Microplate Reader. DNA was diluted with MilliQ water (as needed) to obtain 25 µL solutions of 50 ng DNA. To perform PCR, ~655bp of the *COI* gene was amplified in 25 µL reactions containing 50 ng of DNA, 1X buffer, 1.5 mM $MgCl_2$, 0.4 µM of each primer (summarized in Table 1.2), 0.2 mM of each dNTP and 0.05unit/µL Omega Taq polymerase (Bio-Tek). PCR products were visualized in a 2% agarose gel. Products with a single bright band were purified using a PCR clean up kit (MO BIO Laboratories. Inc). The primers and thermocycling conditions used for each gene region are summarized in Table 1.2. Products were sequenced in the forward and reverse directions on an ABI 3730XL at the Hospital for Sick Children, Toronto, ON.

1.2.5 Phylogenetic analyses

Following sequencing, contigs were constructed and consensus sequences were edited by eye using Sequencher 5.4.6. Consensus sequences, along with sequences downloaded from NCBI GenBank, were uploaded to MEGA X and aligned using ClustalW (Kumar et al. 2018). Genetic

distances within and between species were calculated in MEGA using p-distance (Kumar et al. 2018). Sequence networks were constructed using TCS1.21 (Clement et al. 2000) to test if there was more than one network per clade. The best fit model of nucleotide substitution (GTR + I + G) was selected by jModeltest 2.1.7 (Darriba et al. 2012) based on the Akaike information criterion (AIC), which was then used to parametrize maximum likelihood analyses and Bayesian analyses. Maximum likelihood analyses were run in RaxML v.8 using the model GTRGAMMA (Stamatakis 2014) and bootstrap probability (BP) was calculated using 1,000 bootstrap replicates. For the Bayesian analysis, two independent analyses were run in MrBayes v.3.2.6 (Ronquist et al. 2012) with four independent chains for 10,000,000 generations, with samples taken every 1,000 generations. A burn-in of 2,500 was used and the remaining trees were used to calculate Bayesian posterior probabilities (PP). Following Albrecht et al., *Bulinus tropicus* (MN551571, MN551572) was used to root the trees (2007).

1.3 Results

Specimens (n = 257) identified as *Planorbella* or *Helisoma* were collected by hand from 16 field sampling locations in Manitoba, Canada, with additional specimens provided by collaborators from one location each in Minnesota and Oklahoma, USA (Table 1.1). Individuals from two laboratory strains were also included in subsequent analysis: Manitoba (n = 6) and Oklahoma (n = 1). A total of 264 specimens were identified with dichotomous keys based on shell characteristics of Baker (1945) and Burch (1989). These key-based identifications revealed that four species were likely present in the field collections: *H. anceps, P. campanulata, P. pilsbryi,* and *P. trivolvis* (Table 1.1). According to my use of the keys, *Planorbella corpulenta* and *P. subcrenata* were not present within the samples. With the key-based identification, an average of

2.4 putative species were recovered at each site. In addition, the lab strains were identified as *P. duryi* using shell-based morphology (Baker 1945; Burch 1989).

1.3.1 Genetic identification of *Planorbella* and *Helisoma* spp.

In total, 56 individual specimens were successfully sequenced at the mitochondrial *COI* gene (sequence lengths ranging from 586-698 bp). I performed phylogenetic analyses on 586 bp from 48 unique sequences (Figure 1.2), including 29 sequences from my sampling and 19 from GenBank (Table 1.3). A total of five clades were recovered with four (clades A, B, D, and E) having strong support in both ML and BI analyses (>90 and >96, respectively) and one clade (C) with weaker support (60 and 85, respectively). The maximum intraclade p-distances ranged from 0.86-3.4%, while the minimum interclade p-distances ranged from 6.8-24.3% (Table 1.4). The mean intraclade p-distances were <1% for all except for two clades: clade A (1.46%) and clade D (1.94%) (Table 1.5). These two clades also contained multiple putative species: clade A contained GenBank specimens identified as *P. duryi, Planorbella tenue,* and *P. trivolvis,* as well as specimens I identified as *P. trivolvis.* Moreover, clade D contained GenBank specimens identified as *Planorbella oregonensis* and *P. trivolvis*, and specimens I identified as *P. pilsbryi* and *P. trivolvis* (see Tables 1.1 and 1.3). The two clades with the fewest interclade differences were clade C (*P. trivolvis*) and clade D (*P. pilsbryi*), with a mean p-distance of 8.65% (Table 1.5). Network analysis of each clade showed single connected networks for all but one clade (Clade C), which had two disconnected networks (results not shown).

1.3.2 Geometric morphometric analysis of *Planorbella* and *Helisoma* spp.

I performed two separate analyses using geometric morphometric methods. In the first analysis, a total of 37 individuals were included that had both an image voucher and a *COI* sequence (Table 1.6). The landmark configurations were incorporated into a principal component

analysis (PCA) with the genetic clades used for *a priori* group assignment. The first principal component (45.12%) represented aperture shape, while the second principal component (28.85%) represented relative displacement between landmarks #7 and 11 (Figure 1.3A).

In the second analysis, a total of 169 individuals were included with the 37 individuals with an image and sequence voucher as well as 134 individuals with an image voucher, but no associated sequence (Table 1.6). Unsequenced individuals were assigned *a priori* to groups based on morphological and geographic similarity to individuals with sequences. For example, if an individual was collected at the same place and date as a sample with a sequence, and the shell was morphologically similar to the sequenced individual based on the keys (Baker 1945; Burch 1989), then the unsequenced individual was putatively identified as belonging to the same group. In instances where key-based identifications did not match genetic clade assignments, the genetic clade ID was used to assign unsequenced individuals to groups instead of the shell-based identifications. For example, some co-occurring individuals were assigned to more than one species based on the keys (e.g. from the same collection site some individuals were identified as *P. pilsbryi* and others were *P. trivolvis*). However, some of these individuals were subsequently revised based on the genetic clade assignments of an individual that was collected from the same place and time. I used the genetic ID instead of the key-based ID because when I sequenced at least two individuals of the same phenotype from the same site, the sequences were often identical.

In the second PCA, the first PC captured 43% of the overall variation in the samples, while the second PC captured 18% of the overall variation (Figure 1.3B). Both PCAs showed two clusters (Clades A and E) that separated from the other three clusters (Clades B, C, and D). These three clusters greatly overlapped with each other, and also with Clade A to a lesser extent. Only the group containing members from Clade E had no overlap with the other clusters. In both PCAs,

the groups containing members of clades B and D had the most variance across the first axis, which represents the relative shape of the aperture (deformation in landmarks 8-11). The second principal component described the relative displacement between two landmarks at the intersection of the aperture and the body whorl (landmarks 7 and 11) (Figure 1.3, A and B). Measurement error was ~1.5 orders of magnitude smaller than intraspecific shape variation and was considered negligible (Webster and Sheets 2010).

1.4 Discussion

This study is the first to analyze morphological and genetic diversity in *Planorbella* and *Helisoma* at the species level. I hypothesized that species diversity was underestimated based on shell-based identification and that an integrative taxonomic approach would reveal cryptic species within the genus *Planorbella.* Based on phylogenetic analysis, I found preliminary evidence of cryptic species in two of the clades (Clades A and C). Furthermore, the geometric morphometric analysis results indicated the presence of three functionally cryptic species. These three clades (B, C and D) overlapped in shell morphology so that they were indistinguishable based on the landmarks used in the analysis. However, geometric morphometric analysis differentiated the other two clades (A and E) with minimal overlap. The lack of clear and distinct groupings in this PCA findings highlight the importance of using genetic differentiation to confirm species identification among these closely related and morphologically similar freshwater gastropod species.

To estimate the species diversity in my sampling, I integrated information from species descriptions (Baker 1928, 1945; Clarke 1981; Burch 1989), historical and more recent geographic ranges (Clarke 1981; Pip 1987; McKillop 1996; Johnson et al. 2013), and similarity in *COI* sequences. Phylogenetic analysis revealed one of the five clades (clade A) contained sequences

obtained from laboratory strains of unknown geographic origin (specimens: Plsp31 and HeSp_Lab2) (Figure 1.2). The laboratory strain maintained at the University of Manitoba was previously identified as *Planorbella* spp. when only genetic analysis was completed (Eliuk et al. 2020), but after gathering multidisciplinary evidence I now refer to this species as *P. duryi*. Using shell-based identification, I identified the lab strain specimens as *P. duryi* according to criteria by Baker (1945) and Burch (1989). Shell-based identifications were further corroborated by phylogenetic analysis, in which specimens were found in Clade A with specimens from GenBank also identified as *P. duryi*. However, Clade A also included GenBank sequences labeled as another nominal species (*P. trivolvis*) and an invalid species (*P. tenue*) suggesting there is confusion in how to identify *P. duryi*.

The barcode analysis provided preliminary support for the existence of cryptic species in Clade A based on the maximum intraclade distance of 2.7% (Table 1.4) which surpassed the genetic benchmark of 2% that has been used to suggest the presence of cryptic species (Layton et al. 2014). Network analysis revealed that this clade consisted of a single connected network, albeit some individuals were more distantly related, with a maximum of six substitution differences (results not shown). The specimens in GenBank were collected from the US (Maryland, New Mexico, and South Carolina) and Mexico (Veracruz) (Table 1.3). The GenBank locations are consistent with some of the known geographic localities of *P. duryi* (e.g. New Mexico) (Johnson et al. 2013), but also expand the known range to include Maryland, South Carolina, and Mexico. This species has increased its range via the transport of water plants and fish in the aquatic trade (Sitnikova et al. 2010). In fact, the strain in Manitoba originated from such a shipment. Members of this clade formed a cluster in the PCA that had some overlap with *H. anceps* (Figure 1.3B). However, given that four of the five individuals appeared distinct from other species in the PCA

based on the broadness of their apertures, I still recommend the use of landmarks in geometric morphometric analysis to discriminate this species from other *Planorbella* (Figure 1.3).

The four remaining clades in the phylogenetic analysis were comprised of snails collected from Manitoba, Minnesota, and Oklahoma. Shell-based identification and phylogenetic analysis suggested that Clade B is *Helisoma anceps*. Morphologically, the individuals I sampled were smaller (<20 mm diameter) than members of other clades, were carinate, and possessed deeply recessed spires, all of which are characteristic of *H. anceps* (Baker 1945; Clarke 1981; Burch 1989). In addition, these snails were only found in permanent water bodies throughout Manitoba, including lakes and streams, which matched the descriptions of habitat use for *H. anceps* (Baker 1945; Clarke 1981). However, this clade was comprised of GenBank specimens identified as *Helisoma* spp. (MG422674) and *P. trivolvis* (MH087675), not *Helisoma anceps,* though there were *COI* sequences labeled with this species name in other clades of the tree. Network analysis showed that members of this clade formed a single connected network. In the PCA (Figure 1.3B), members of this clade overlapped with members of clade C (*P. trivolvis*) and D (*P. pilsbryi*), suggesting that misidentification between the three species is possible.

Clade C consisted of specimens collected from Minnesota and Oklahoma that I identified as *P. trivolvis* because these snails were larger (<32 mm diameter), with a submersed spire, and lacked lower carina, as is characteristic of *Planorbella trivolvis* (Baker 1945; Clarke 1981). *COI* sequences from GenBank originated from Ontario, Canada (MF544974, KM611923), Oregon, US (MH198421), and New Mexico, US (KP101090), which matches the known distribution of *P. trivolvis*. This clade had the weakest support among the five clades and although I refer to it as one clade, network analysis and p-distances suggest that cryptic species may be present. Network analysis showed that this clade consisted of two disconnected networks that contained individuals

labeled as several nominal taxa: 1) *P. trivolvis* (Plsp43, HeTr2_MN, KP101090) and *H. anceps* (KMF544974); 2) *P. trivolvis* (Plsp29 and Plsp30) and *P. subcrenata* (MH198421). Likewise, the maximum intraclade p-distance of 3.4% (Table 1.4) surpasses the genetic benchmark of >2% that has been used to suggest evidence of cryptic species. Members of this clade formed a cluster in the PCA that overlapped with individuals in clade B and clade D (Figure 1.3), indicating that in terms of the landmark criteria, some individuals from clade C displayed intermediate shell shapes between these two clades. Due to the genetic and phenotypic variation within this clade, more individuals identified as *P. trivolvis* in this study should be sequenced to better assess whether cryptic species are present. In future studies, additional sampling from other geographic locations across the range and the collection of both morphological and genetic vouchers will be essential to characterizing the diversity in this clade.

Snails in clade D were often large (up to 32 mm in diameter), sometimes had lower carina (a trait which distinguished them from *P. trivolvis*), and in many specimens the apical side of the shell was smoothly concave and bowl-like (which distinguished them from *Planorbella corpulenta*) (Baker 1945; Clarke 1981). Using shell-based identification, owing to variation in the above-mentioned traits, I identified individuals within this clade as both *P. pilsbryi* and *P. trivolvis*. The ranges of *P. pilsbryi* and *P. trivolvis* completely overlap in Manitoba, so geographic distribution is not a helpful criteria for species delimitation (Clarke 1981). Based on single gene analysis, I found support for a single clade in phylogenetic and network analyses. However, there was some evidence to suggest that cryptic species are present as the mean intraspecific p-distance of 1.94% approached the genetic benchmark of >2% cryptic species. In addition, based on GenBank sequences, there were multiple nominal species (*P. trivolvis*, *P. pilsbryi*, and *P. oregonensis* within this clade, suggesting that people have trouble distinguishing among these

species using shell-based traits. More detailed morphological analysis of the voucher specimens that represent the GenBank sequences for *P. trivolvis and P. oregonensis* are needed to understand how genotype is linked to phenotype. Another issue is that even when samples were collected from a smaller geographic area, this species showed a high amount of phenotypic variation relative to other species. The majority of specimens in clade D that were also included in the PCA were sampled in Manitoba. Despite originating from only one geographic area (rather than from additional states in the US), this group had the most variation in the PCA (Figure 1.3B). The overlap between specimens in clade A (to a limited extent), C, and D show that my approach to geometric morphometric analysis could not completely discriminate these species. The hypothesis of crypsis could be further tested with the use of additional mitochondrial and nuclear loci (Horsáková et al. 2019).

Finally, clade E was comprised of a single GenBank specimen (KF958031) and a single sequence from Manitoba (Table 1.1). Among the most distinctive of the snails collected, these snails all bear the bell-shaped, or campanulate, aperture that is characteristic of *Planorbella campanulata* (Baker 1945; Clarke 1981; Burch 1989). This distinctive feature is reflected in the PCA groupings shown in Figure 1.3; individuals from clade E cluster together to form a group that is removed from the other four overlapping clusters. The range of *P. campanulata* in Manitoba is restricted to the region east of Lake Winnipeg as well as to the north of this lake (Clarke 1981). Together, this information was sufficient for me to confidently label clade E as *P. campanulata.*

This study showed that using any one approach, shell-based identification, *COI* barcoding, or geometric morphometrics, is insufficient to discriminate *Planorbella* and *Helisoma* species. The most reliable method may be DNA sequencing combined with phylogenetic analysis. I do not recommend using BLAST as a method to assess genetic similarity given that the results suggest

that sequences in GenBank may be misidentified. Four of the five clades contained multiple nominal species according to both GenBank identifications and my own putative identifications based on criteria from Baker (1945) and Burch (1989). The *P. duryi* clade (clade A) contained two other nominal species: *P. tenue* (EF012174) and *P. trivolvis* (AY651208, MH087626, MH087568). The *P. pilsbryi* clade (clade D) included many individuals from Manitoba that I putatively identified as *P. trivolvis* and *P. pilsbryi*. In addition, GenBank sequences identified as *P. trivolvis* appeared in four of the clades. According to Baker (1928), the label *P. trivolvis* has been treated as a "dumping ground", as inexperienced taxonomists place all large planorbids into this taxon. This may be one of the reasons for the frequent misidentifications on GenBank. Finally, Clade B contained one specimen identified as *Helisoma* spp. (MG422674) and another identified as *P. trivolvis* (MH087675), along with several individuals collected from Manitoba and identified as *H. anceps*. To provide some historical taxonomic context, *Helisoma* and *Planorbella* were previously considered synonyms until Burch divided the members into two separate genera based on differences in shell coiling (1989), which may explain why the sequence on GenBank was identified as *Helisoma*. However, there were three GenBank sequences identified as *H. anceps* in the tree and none of them appeared in this particular clade. The abundance of nominal species within clades suggests that *COI* identifications should be treated with skepticism, with additional reference to morphology, phylogenetic relationships, and geographic ranges prior to deciding on a label. To verify GenBank identifications, I recommend that morphological vouchers be included alongside sequences to help make sense of the current taxonomic and phylogenetic state of *Planorbella* and *Helisoma*.

Geometric morphometric analysis differentiated two of the five genetic clades (i.e. A = *P. duryi* and E = *P. campanulata*). However, the landmark configurations were insufficient to

differentiate between clades B, C, and D. Based on my landmark configurations, these three clades can be considered functionally cryptic, as there is no reliable means to identify them with this method. This ambiguity led us to conclude that shell shape alone is insufficient for differentiating species of *Planorbella* and *Helisoma.* It is difficult to determine whether ambiguities in shell morphology have genetic cause or whether they can be attributed to phenotypic plasticity without further experimental evidence. Additional analysis with an ANOVA that includes genetic identifications could help us to understand the extent to which phenotypic plasticity is influencing the morphological variation in shell shape. However, lab-based experimentation that accounts for genetic variation as well as environmentally-induced phenotypic variation would also be key (e.g. Gustafson et al. 2014; Gustafson and Bolek 2015).

1.5 Conclusions

By combining taxonomic descriptions of shell morphology, DNA barcoding, and geometric morphometric analyses of snail shells, I identified four clades that each represent a species of *Planorbella/Helisoma* in North America, and one that represents lab strains maintained at the University of Manitoba and University of Oklahoma for host-parasite interactions research. One significant finding is that two commonly used species in ecological and parasitological studies, *P. trivolvis* and *H. anceps,* occur in several of the clades in the *COI* phylogeny based on GenBank identifications (Figure 1.2). The former, *P. trivolvis,* occurred in four clades (clades A, B, C, and D), while the latter, *H. anceps*, only occurred in one (clade C). I suggest that these species have been misidentified due to the limitations of shell-based identifications. The *COI* phylogeny demonstrated that these two species were lumped together with several other nominal species including *P. duryi*, *P. tenue*, *P. pilsbryi* and *P. oregonensis,* suggesting that solely using shell-

based traits for identification is not a reliable approach to distinguishing species of *Planorbella* and *Helisoma*.

This work illustrates the value of integrative taxonomy in species delimitation of freshwater gastropod species in the genera *Planorbella* and *Helisoma*. I found preliminary support for the presence of cryptic species complexes within some of the focal species, to address morphological variation in the shells of field-collected snails, and to uncover instances of misdiagnoses in public sequence databases. This work has also demonstrated how geometric morphometric analysis can be used in the process of gastropod species identification. In the future, this approach could be performed using machine-learning to automate species identifications, which would greatly save time and resources for researchers that study aspects of gastropod biology. I have demonstrated the feasibility of an integrated approach to taxonomy as a reliable method for species identifications in gastropod species. My sampling efforts have expanded the taxonomic representation on public sequence databases, which will contribute to more complete phylogenetic analyses that include a wider array of species. By focusing on ecologically important genera, this work will benefit researchers in the fields of parasitology and ecology, as expanded understanding of species diversity allows further exploration of the biology of these common North American species.

Literature Cited

Adams, D., Collyer, M., Kaliontzopoulou, A., and Baken, E. 2021. Geomorph: Software for geometric morphometric analyses. R package version 4.0.

Adamson, M.L., and Caira, J.N. 1994. Evolutionary factors influencing the nature of parasite specificity. Parasitology **109**(S1): S85–S95.

Agapow, P.M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., and Purvis, A. 2004. The impact of species concept on biodiversity studies. Q. Rev. Biol. **79**(2): 161–179.

Albrecht, C., Kuhn, K., and Streit, B. 2007. A molecular phylogeny of Planorboidea (Gastropoda, Pulmonata): Insights from enhanced taxon sampling. Zool. Scr. **36**(1): 27–39.

Baker, F.C. 1928. The freshwater Mollusca of Wisconsin. Wisconsin Academy of Sciences, Arts, and Letters.

Baker, F.C. 1945. The molluscan family Planorbidae. University of Illinois Press, Urbana, IL.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., and Das, I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. **22**(3): 148–155.

Boerger, H. 1975. A comparison of the life cycles, reproductive ecologies, and size–weight relationships of *Helisoma anceps, H. campanulatum, and H. trivolvis* (Gastropoda, Planorbidae). Can. J. Zool. **53**(12): 1812–1824.

Bolek, M.G., Detwiler, J.T., and Stigge, H.A. 2019. Selected wildlife trematodes. *In* Digenetic Trematodes, 2nd edition. Springer Nature.

Burch, J.B. 1989. North American freshwater snails. Malacological Publications, Hamburg, MI.

Clarke, A.H. 1981. The freshwater molluscs of Canada. National Museum of Natural Sciences,

National Museums of Canada, Ottowa, CAN.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. 2012. JModelTest 2: More models, new heuristics and parallel computing. Nat. Methods **9**(8): 772. Nature Publishing Group.

Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. **85**(3): 407–415.

deWaard, J.R., Ratnasingham, S., Zakharov, E. V., Borisenko, A. V., Steinke, D., Telfer, A.C., Perez, K.H.J., Sones, J.E., Young, M.R., Levesque-Beaudin, V., Sobel, C.N., Abrahamyan, A., Bessonov, K., Blagoev, G., DeWaard, S.L., Ho, C., Ivanova, N. V., Layton, K.K.S., Lu, L., Manjunath, R., McKeown, J.T.A., Milton, M.A., Miskie, R., Monkhouse, N., Naik, S., Nikolova, N., Pentinsaari, M., Prosser, S.W.J., Radulovici, A.E., Steinke, C., Warne, C.P., and Hebert, P.D.N. 2019. A reference library for Canadian invertebrates with 1.5 million barcodes, voucher specimens, and DNA samples. Sci. data **6**(308): 1–12.

Dillon, R.T.J. 2019. The classification of the Planorbidae. *In* The Freshwater Gastropods of North America, Essays on the Pulmonates. FWGNA Press, Charleston. pp. 127–135.

Eliuk, L.K., Brown, S., Wyeth, R.C., and Detwiler, J.T. 2020. Parasite-modified behaviour in non-trophic transmission: Trematode parasitism increases the attraction between snail intermediate hosts. Can. J. Zool. **98**(7): 417–424.

Gustafson, K.D., and Bolek, M.G. 2015. Effects of trematode parasitism on the shell morphology of snails from flow and nonflow environments. J. Morphol. **277**(3): 316–325.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. **270**: 313–321.

Hebert, P.D.N., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. **54**(5): 852–859.

Hobbs, C.S., Vega, R., Rahman, F., Horsburgh, G.J., Dawson, D.A., and Harvey, C.D. 2021.

Population genetics and geometric morphometrics of the freshwater snail *Segmentina nitida* reveal cryptic sympatric species of conservation value in Europe. Conserv. Genet.: 1–17. Springer Netherlands. doi:10.1007/s10592-021-01369-8.

Horsáková, V., Nekola, J.C., and Horsák, M. 2019. When is a "cryptic" species not a cryptic species: A consideration from the Holarctic micro-landsnail genus *Euconulus* (Gastropoda: Stylommatophora). Mol. Phylogenet. Evol. **132**: 307–320.

Hubendick, B. 1955. Phylogeny in the Planorbidae. Trans. Zool. Soc. London **28**(1): 453–542.

Johnson, P.D., Bogan, A.E., Brown, K.M., Burkhead, N.M., Cordeiro, J.R., Garner, J.T., Hartfield, P.D., Lepitzki, D.A.W., Mackie, G.L., Pip, E., Tarpley, T.A., Tiemann, J.S., Whelan, N. V., and Strong, E.E. 2013. Conservation status of freshwater gastropods of Canada and the United States. Fisheries **38**(6): 247–282.

Jørgensen, A., Kristensen, T.K., and Stothard, J.R. 2004. An investigation of the "Ancyloplanorbidae" (Gastropoda, Pulmonata, Hygrophila): Preliminary evidence from DNA sequence data. Mol. Phylogenet. Evol. **32**(3): 778–787.

Kress, W.J., and Erickson, D.L. 2008. DNA barcodes: Genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. U. S. A. **105**(8): 2761–2762.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547–1549.

Layton, K.K.S., Martel, A.L., and Hebert, P.D.N. 2014. Patterns of DNA barcode variation in Canadian marine molluscs. PLoS One **9**(4): 1–9.

Lockyer, A.E., Jones, C.S., Noble, L.R., and Rollinson, D. 2004. Trematodes and snails: An intimate association. Can. J. Zool. **82**(1): 251–269.

Lotfy, W.M., Dejong, R.J., Abdel-Kader, A., and Loker, E.S. 2005. A molecular survey of *Biomphalaria* in Egypt: Is *B. glabrata* present? Am. J. Trop. Med. Hyg. **73**(1): 131–139.

Macher, J.N., Weiss, M., Beermann, A.J., and Leese, F. 2016. Cryptic diversity and population structure at small scales: The freshwater snail *Ancylus* (Planorbidae, Pulmonata) in the Montseny mountain range. Ann. Limnol. **52**: 387–399.

Madjos, G.G., and Demayo, C.G. 2018. Macro-geographic variations of the invasive giant African snail *Achatina fulica* populations in Mindanao , Philippines. Trans. Sci. Technol. **5**(2): 143–154.

Martin, K.R., Johnson, P., Bowerman, J., and Jingchun, L. 2020. Biogeography of the freshwater gastropod, *Planorbella trivolvis*, in the western United States. PLoS One. **15**(7): 1-18.

McKillop, B. 1996. Geographic and environmental distribution of freshwater gastropods in Manitoba, Canada. Manitoba Museum Man Nat. Occas. Ser. **1**: 1–34.

Mello-Silva, C.C., Grault, C.E., Da Costa, V.A., and Barbosa, F.S. 1998. Possible hybridization of Brazilian planorbid snails and its importance in population dynamics. Mem. Inst. Oswaldo Cruz **93**(SUPPL. I): 227–232.

Morgan, J.A.T., DeJong, R.J., Jung, Y., Khallaayoune, K., Kock, S., Mkoji, G.M., and Loker, E.S. 2002. A phylogeny of planorbid snails, with implications for the evolution of *Schistosoma* parasites. Mol. Phylogenet. Evol. **25**(3): 477–488.

Nantarat, N., Sutcharit, C., Tongkerd, P., Wade, C.M., Naggs, F., and Panha, S. 2019. Phylogenetics and species delimitations of the operculated land snail *Cyclophorus volvulus* (Gastropoda: Cyclophoridae) reveal cryptic diversity and new species in Thailand. Sci. Rep. **9**(7041): 1–12.

Parra, C., and Liria, J. 2017. Shell geometric morphometrics in *Biomphalaria glabrata* (Mollusca:

Planorbidae) uninfected and infected with Schistosoma mansoni. Zool. Syst. **42**(1): 59–64.

Perez, K.E., and Minton, R.L. 2008. Practical applications for systematics and taxonomy in North American freshwater gastropod conservation. J. North Am. Benthol. Soc. **27**(2): 471–483.

Pip, E. 1987. Ecological differentiation within genus *Helisoma* in Central Canada. Nautilus (Philadelphia). **101**(1): 33–44.

Rohlf, F.J. 2015. The tps series of software. Hystrix, Ital. J. Mammology **26**(1): 1–4.

Rohlf, F.J., and Marcus, L.F. 1993. A revolution in morphometrics. Trends Ecol. Evol. **8**(4): 129–132.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61**(3): 539–542.

Saadi, A.J., Davison, A., and Wade, C.M. 2020. Molecular phylogeny of freshwater snails and limpets (Panpulmonata: Hygrophila). Zool. J. Linn. Soc. **190**(2): 518–531.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R.H. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. Annu. Rev. Entomol. **55**: 421–438.

Sheets, H.D. 2014. Integrated Morphometrics Package (IMP) 8.

Sitnikova, T., Soldatenko, E., Kamaltynov, R., and Riedel, F. 2010. The finding of North American freshwater gastropods of the genus *Planorbella* Haldeman, 1842 (Pulmonata: Planorbidae) in East Siberia. Aquat. Invasions **5**(2): 201–205.

Stamatakis, A. 2014. RAxML. Manual/tutorial (1): 1–5.

Tuan, R., and dos Santos, P. 2007. ITS2 variability of *Biomphalaria* (Mollusca, Planorbidae)

species from the Paranapanema Valley (São Paulo State, Brazil): Diversity patterns, population structure, and phylogenetic relationships. Genet. Mol. Biol. **30**(1): 139–144.

Turgeon, D.D., and Quinn, J.F. 1998. Mollusks: Common and scientific names of aquatic invertebrates from the United States and Canada. *In* American Malacological Union, 2nd editio. American Fisheries Society.

Webster, M., and Sheets, H.D. 2010. A practical introduction to landmark-based geometric morphometrics. Paleontol. Soc. Pap. **16**: 163–188.

Yang, L., Tan, Z., Wang, D., Xue, L., Guan, M.X., Huang, T., and Li, R. 2014. Species identification through mitochondrial rRNA genetic analysis. Sci. Rep. **4**: 1–11.

Table 1.1 Summary of *Planorbella* and *Helisoma* field collections from 18 geographic locations. Species identifications were based on descriptive criteria by Burch (1989) and Baker (1945).

Location	GPS coordinates	Putative species (N)
Churchill, MB	58.7684, -94.165	*P. pilsbryi* (1)
Dauphin, MB	51.154449, -99.875924	*P. pilsbryi* (1), *H. anceps* (6)
	51.154821, -100.098574	*P. trivolvis* (1), *P. pilsbryi* (5), *H. anceps* (2)
	51.321161, -100.376896	*P. pilsbryi* (5), *H. anceps* (5)
	51.472697, -100.370085	*P. trivolvis* (1), *P. pilsbryi* (5)
	51.603822, -100. 462917	*P. trivolvis* (1)
Duck Mountain Provincial Park, MB	51.791458, -100.898639	*P. trivolvis* (3)
	51.347079, -100.69808	*H. anceps* (3)
Grand Marais, MB	50.569232, -96.598151	*P. trivolvis* (6), *P. pilsbryi* (6)
	50.561637, -96.586663	*P. trivolvis* (3)
	50.562194, -96.613168	*P. pilsbryi* (2)
Grand Rapids, MB	53.276202, -99.335098	*P. pilsbryi* (1)
Interlakes, MB	52.07284, -98.833126	*P. pilsbryi* (6), *H. anceps* (2)
	51.435157, -98.533016	*P. pilsbryi* (3)
Libau, MB	*	*P. pilsbryi* (10)
Minitonas, MB	51.100522, -101.050312	*H. anceps* (5)
Morden, MB	49.127138, -97.302851	*P. trivolvis* (13), *P. pilsbryi* (4)
Nopiming Provincial Park, MB	50.46879, -95.299175	*P. trivolvis* (2), *P. pilsbryi* (2), *H. anceps* (1)
	*	*P. campanulata (3)*
	*	*P. campanulata (4)*
Pinawa, MB	50.1416, -95.8157	*P. pilsbryi* (2)
Swan River, MB	52. 106389, -101.218616	*P. trivolvis* (1), *P. pilsbryi* (9), *H. anceps* (6)
Thompson, MB	55.65361 -97.9311	*P. trivolvis* (3), *P. pilsbryi* (2)
	558141667 -97.924167	*P. pilsbryi* (1)
	55.5552778 -98.01778	*P. pilsbryi* (1)
Wekusko Falls Provincial Park, MB	54.622381, -100.055222	*P. pilsbryi* (3)
	54.892434, -99.837671	*P. trivolvis* (1), *P. pilsbryi* (1)
	54.612465, -100.171239	*P. trivolvis* (2), *P. pilsbryi* (1)
Whitefish Lake, MB	52.2015, -101.3510	*P. trivolvis* (8), *P. pilsbryi* (6)
Whiteshell Provincial Park, MB	49.8955556 -95.39750	*P. trivolvis* (4), *P. pilsbryi* (1)
	*	*P. trivolvis* (14), *P. pilsbryi* (12), *H. anceps* (6)

	*	*P. trivolvis* (6), *P. pilsbryi* (5)
	*	*P. trivolvis* (6), *P. pilsbryi* (1), *H. anceps* (1)
	50.14084, -95.813769	*P. trivolvis* (4), *P. pilsbryi* (2)
Minnesota Lake, Minnesota	43.8405928 -93.83144	*P. pilsbryi* (12)
Stillwater, Oklahoma	36.2413889 -97.089999	*P. pilsbryi* (1)
	36.074167 -97.4491667	*P. trivolvis* (1)
Lab strain, Manitoba	NA	*P. trivolvis* (6)
Lab strain, Oklahoma	NA	*P. trivolvis* (1)

Table 1.2 Primers and thermocycling conditions for PCR with *Planorbella* and *Helisoma* spp. Conditions used to amplify the partial cytochrome c oxidase I (*COI*) mtDNA gene.

Primers	Sequence (5'-3')	Reference	Thermocycler profile	
LCO1490 F	GGTCAACAAATCATAAAGATATTGG	Folmer *et al.* 1994	95˚C—5 min 95˚C—30 sec	
HCO2198 R	TAAACTTCAGGGTGACCAAAAAAT	Folmer *et al.* 1994	54˚C—45 sec 72˚C—1 min	X 40
			72˚C—5 min	

Table 1.3 Summary of GenBank sequences and accession numbers used in the cytochrome c oxidase I (*COI*) barcode phylogeny.

Clade	Species (as labeled in GenBank)	GenBank accession	Geographic origin of specimen
Clade A	*Planorbella trivolvis*	MH087626	Maryland, US
	Planorbella trivolvis	AY651208	South Carolina, US
	Planorbella trivolvis	MH087568	Maryland, US
	Planorbella duryi	KY514384	New Mexico, US
	Planorbella tenue	EF012174	Veracruz, Mexico
Clade B	*Helisoma* spp.	MG422674	Ontario, Canada
	Planorbella trivolvis	MH087675	Maryland, US
Clade C	*Helisoma anceps*	MF544974	Ontario, Canada
	Planorbella trivolvis	KP101090	New Mexico, US
	Planorbella subcrenata	MH198421	Oregon, US
	Planorbella anceps	KM611923	Ontario, Canada
Clade D	*Planorbella trivolvis*	KT831387	Alberta, Canada
	Planorbella oregonensis	MH509191	Oregon, US
	Planorbella trivolvis	HQ926924	Canada
	Planorbella trivolvis	KM612007	Alberta, Canada
Clade E	*Planorbella campanulata*	KF958031	New York, US
Outgroup	*Bulinus tropicus*	MN551571	Uganda
	Bulinus tropicus	MN551572	Uganda

Table 1.4 Genetic distance (p-distance) of cytochrome c oxidase I (*COI*) within and between *Planorbella*/*Helisoma* spp. Maximum intragroup distances are on the diagonal and values greater than 0.02 are bolded as they indicate cryptic species within the clade. Minimum intergroup distances are off-diagonal values for which values greater than 0.05 are interpreted as support for species-level differences between the clades.

	Clade A	Clade B	Clade C	Clade D	Clade E
Clade A	**0.02679**				
Clade B	0.17584	0.00865			
Clade C	0.06834	0.13554	**0.03377**		
Clade D	0.07535	0.16385	0.07366	0.01567	
Clade E	0.24347	0.17584	0.06834	0.07535	0.008613

Table 1.5 Genetic distance (p-distance) of cytochrome c oxidase I (*COI*) within and between *Planorbella*/*Helisoma* spp. Mean intragroup distances are on the diagonal and values greater than 0.019 are bolded as they could indicate cryptic species within the clade. Mean intergroup distances are off-diagonal values for which values greater than 0.05 are interpreted as support for species-level differences between the clades.

	Clade A	Clade B	Clade C	Clade D	Clade E
Clade A	0.01463				
Clade B	0.22561	0.00486			
Clade C	0.24448	0.14707	**0.01945**		
Clade D	0.27614	0.18069	0.08655	0.00819	
Clade E	0.26076	0.18110	0.07619	0.08820	0.008613

Table 1.6 Summary of species included in geometric morphometric analyses, with species labels assigned based on cytochrome c oxidase I (*COI*) barcode phylogeny.

Clade / Species	Images with COI sequences	Images alone (no sequences)	Totals
A / *P. duryi*	6	0	6
B / *H. anceps*	8	24	32
C / *P. trivolvis*	4	9	15
D / *P. pilsbryi*	12	96	108
E / *P. campanulata*	2	5	8
		Grand Total	**169**

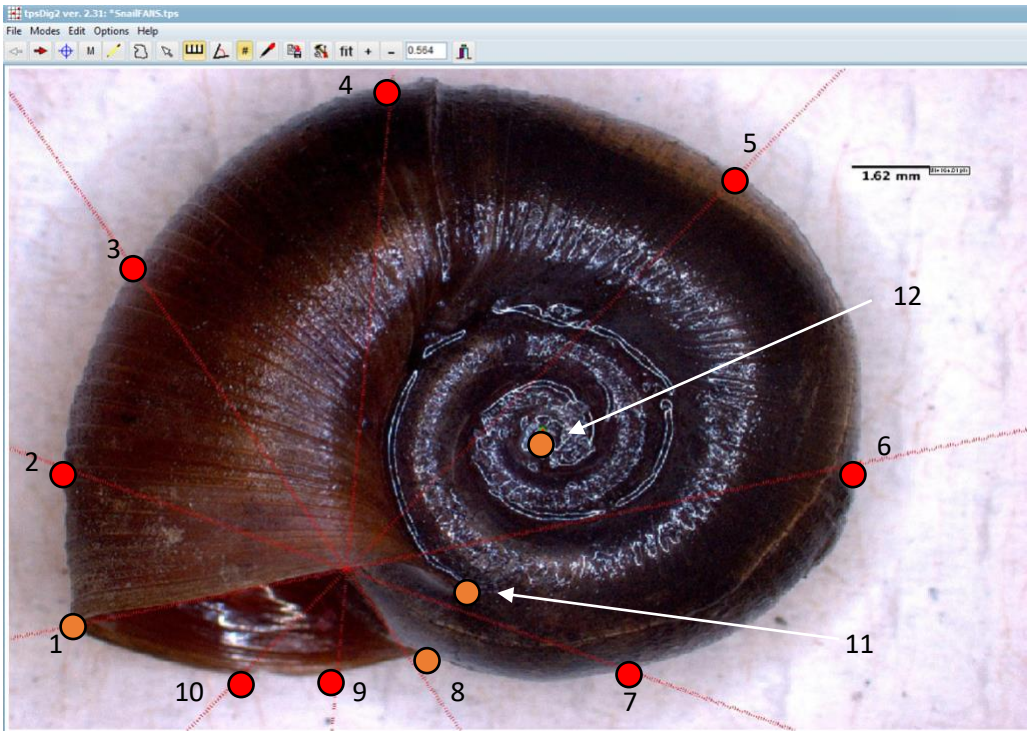


Figure 1.1 Landmark configuration for geometric morphometric analysis of shell shape. Configurations were adapted from Parra and Liria (2017). Configurations comprised of a mix of both traditional type I landmarks (orange: landmarks 1, 8, 11, and 12) and type II sliding semi-landmarks (red: 2-7, 9, and 10) for a total of 12. Any specimens missing one or more of the landmarks were excluded from the analysis.

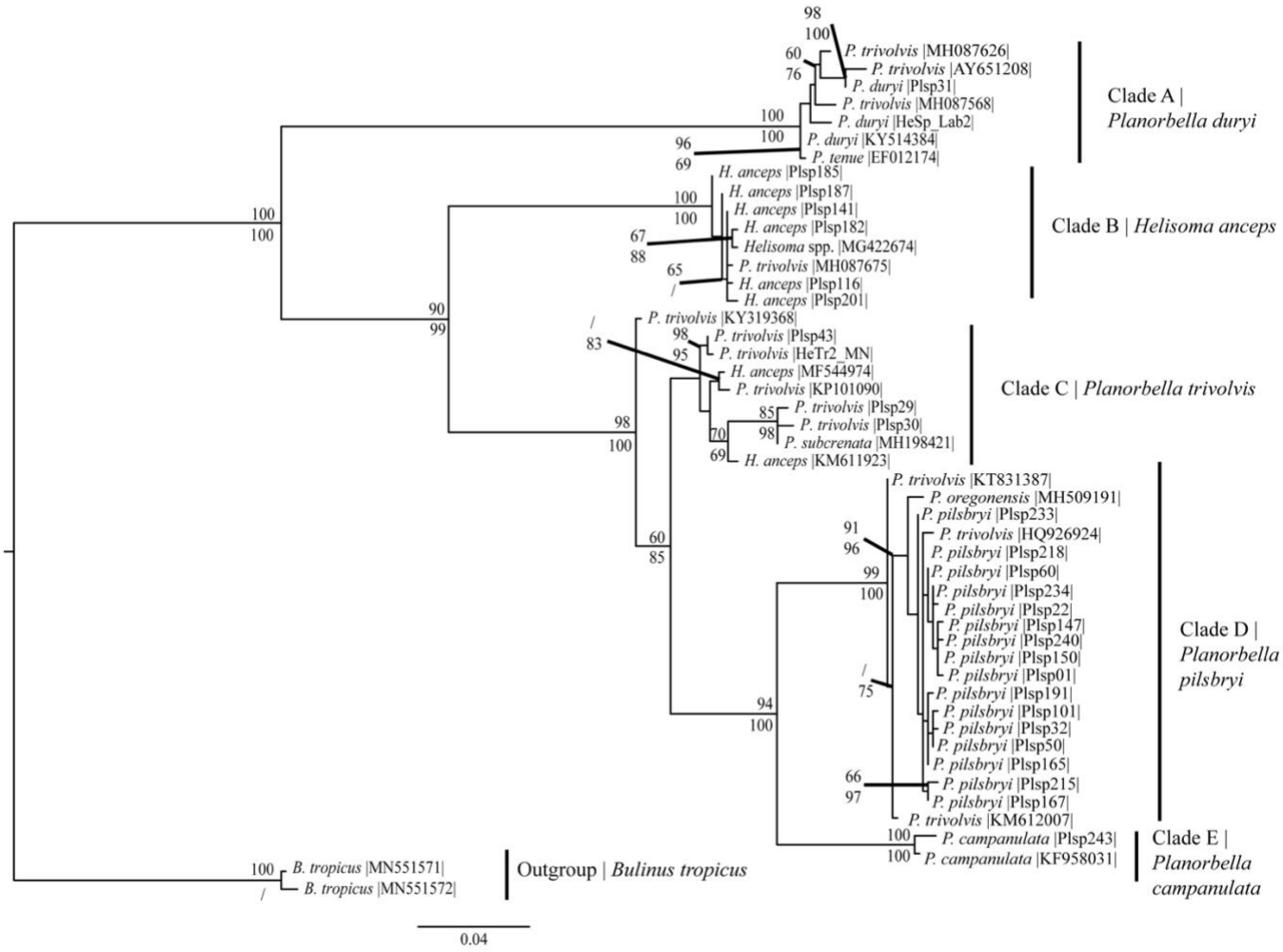
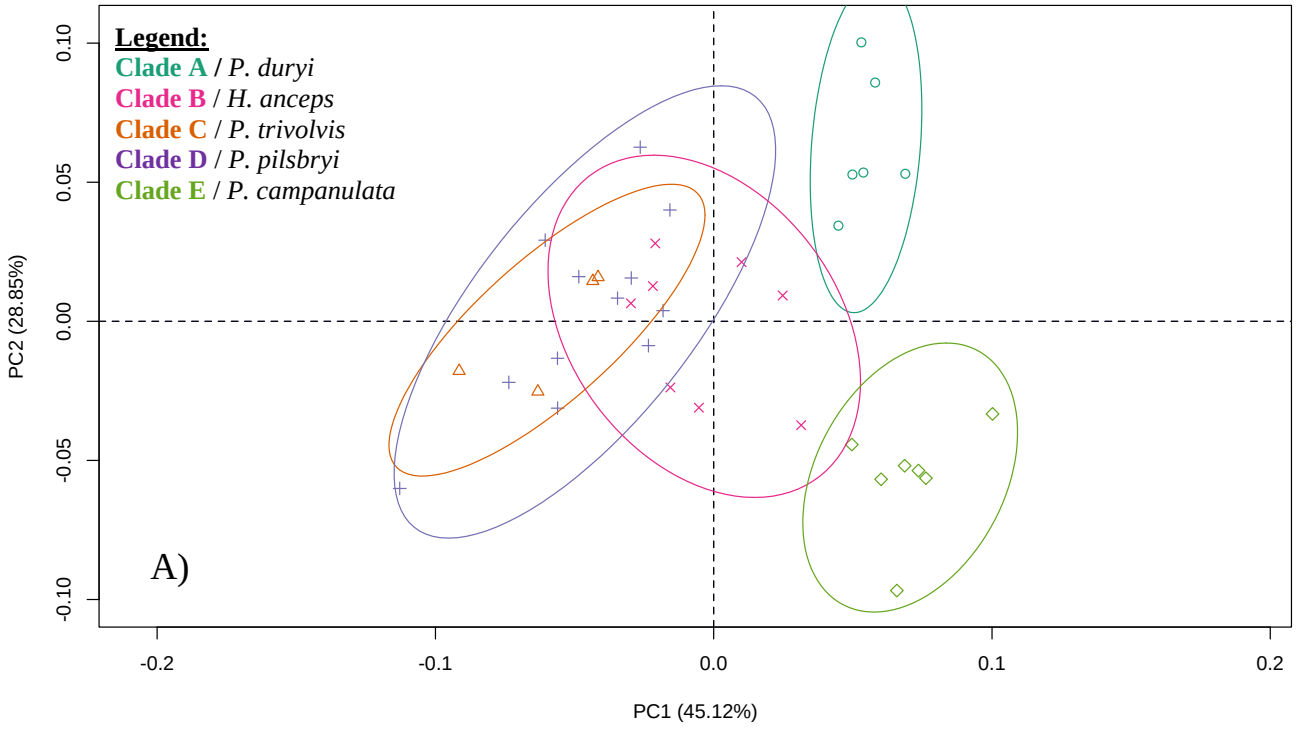


Figure 1.2 Maximum likelihood analysis phylogeny of cytochrome c oxidase I (*COI*) gene region (568bp) of 46 *Planorbella* and *Helisoma* spp. (genera abbreviated in the figure) unique sequences, with one outgroup species from the Tribe Bulinini. Numbers above each node are maximum likelihood bootstrap values, while Bayesian posterior probabilities are displayed under each node (showing values over 60%).


Legend:
Clade A / P. duryi
Clade B / H. anceps
Clade C / P. trivolvis
Clade D / P. pilsbryi
Clade E / P. campanulata
A)
PC2 (28.85%)
PC1 (45.12%)
0.10
0.05
0.00
-0.05
-0.10
-0.2
-0.1
0.0
0.1
0.2

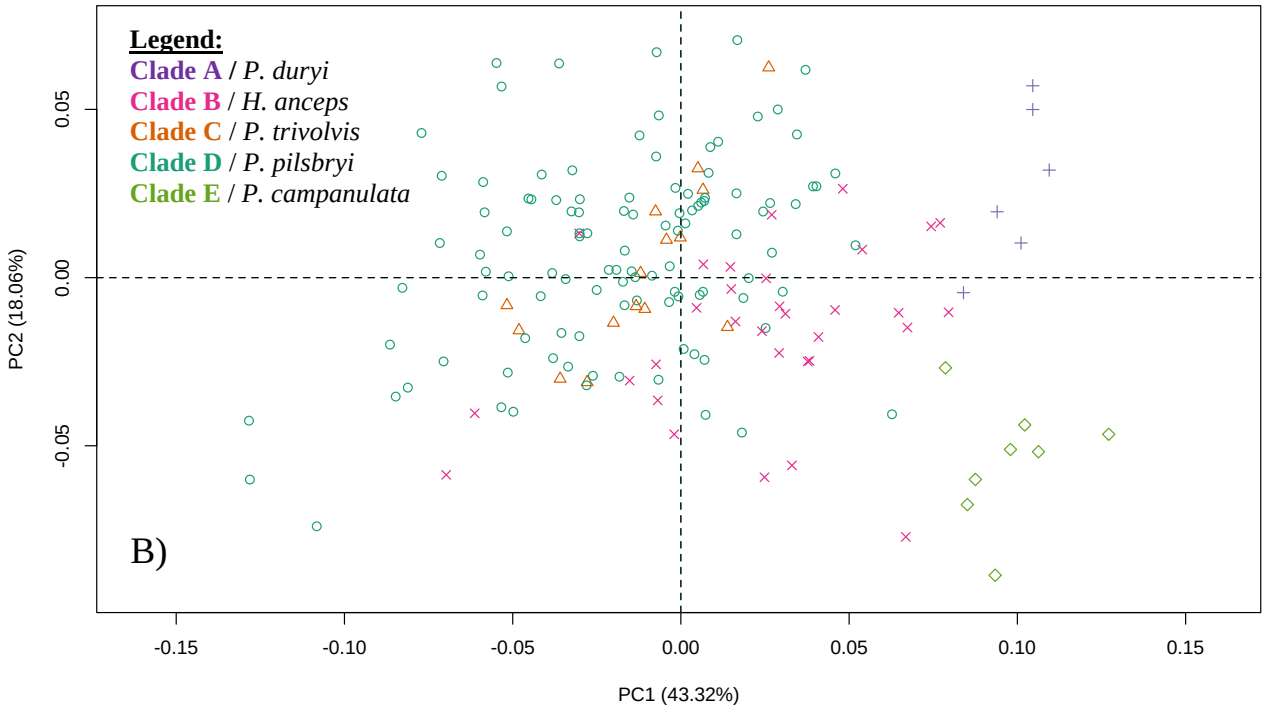

Legend:
Clade A / P. duryi
Clade B / H. anceps
Clade C / P. trivolvis
Clade D / P. pilsbryi
Clade E / P. campanulata
B)
PC2 (18.06%)
PC1 (43.32%)
0.05
0.00
-0.05
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15

Figure 1.3 A) Principal component analysis of 37 *Planorbella* and *Helisoma* spp. that were imaged and sequenced at the cytochrome c oxidase I (*COI*) gene. Molecular clade assignments were used to assign groups *a priori*, as listed in legend. The first principal component (45.12%) represents aperture shape, while the second principal component (28.85%) represents relative displacement between landmarks #7 and 11. Ellipses represent confidence 95% of the standard deviation from the mean of the group. **Figure 1.3 B)** Principal component analysis of 169 *Planorbella* and *Helisoma* spp. Molecular clade assignments were used to assign groups *a priori*, as listed in legend. Principal component axes are 43.32% and 18.06%, with identical interpretations as those in figure A.

Chapter 2: Investigating the phylogenetics of freshwater snails (*Planorbella* and *Helisoma*) with mitogenomes and nuclear loci

2.1 Introduction

Taxonomy and systematics help us to understand the diversity of life on Earth. These scientific disciplines benefit all other biological fields by providing methods to delineate, classify, name species, and reconstruct evolutionary relationships (Wilson 2004; Dayrat 2005). The ability to recognize and classify species has evolved along with the development of phylogenetic analysis methods as more characters have been discovered and found to be phylogenetically informative, including morphological (geometric morphometric analyses) and molecular characters (DNA sequences) (Dayrat 2005; Schlick-Steiner et al. 2010). One of the most fundamental questions of taxonomy is how to delineate species, or where to draw the boundaries between one taxonomic unit and another (Dayrat 2005). Traditionally, taxonomists have delineated species using subjective visual comparisons of morphological traits (Mutanen and Pretorius 2007). Subsequently, the development of DNA barcode methods represented a breakthrough in species delineation, with standardized sequences and universal primers for relatively quick and easy species identification, and the promise of automated species identification and discovery (Hebert et al. 2003; Moritz and Cicero 2004; Hebert and Gregory 2005). However, barcodes are insufficient on their own to rigorously test species hypotheses for a number of reasons (Moritz and Cicero 2004; Wheeler et al. 2004; Marcus 2018). For example, the mitochondrial cytochrome c oxidase (*COI*) gene that is commonly used in metazoan barcoding may have a different evolutionary history than the organism as a whole since it is derived from an uniparentally-inherited organelle (Moritz and Cicero 2004; Marcus 2018). For this reason (among others like

linkage and introgression), the phylogenetic divergence of mtDNA may differ from nuclear gene divergence and lead to false estimates of species boundaries (Moritz and Cicero 2004).

As sequencing technologies have evolved, genomic analysis of non-model species has become more feasible and affordable, making it an attractive alternative to single-locus barcoding (Coissac et al. 2016; Gomes-dos-Santos et al. 2019). There are many advantages to genomic analysis; since PCR is not required for high-throughput sequencing, DNA can be recovered from degraded specimens; and the additional loci will increase the phylogenetic signal of the sequence by increasing the number of variable characters (Coissac et al. 2016). Genome sequencing also greatly expands the number of species that can be included in phylogenetic analysis since numerous loci are uncovered, which allows the researcher more latitude in selecting which loci they use in their analyses (Richards 2015; Coissac et al. 2016). By increasing the species represented in phylogenetic analysis we can approach a more representative sampling of species diversity (Adema 2021).

One taxon where genomic resources are lacking is the Phylum Mollusca, where adult morphological characters have formed the basis for most hypotheses of species delimitation (Kocot 2013). Among metazoan taxa, Phylum Mollusca is the second most species-rich, but it is underrepresented with regards to genomic resources. For instance, genome assemblies have been made available for only ~0.04% the species (Kocot et al. 2011; Richards 2015; Gomes-dos-Santos et al. 2019). Additionally, there is a lag between when molluscan genomes are published and when they are included in robust phylogenetic analyses that help to resolve aspects of their evolutionary ecology (e.g. species diversity) (Gomes-dos-Santos et al. 2019). This lag has critical implications for conservation, as genomic analysis could be used to better define what constitutes a conservation unit (e.g. designatable unit for Species at Risk Act or species for the Endangered Species Act).

Taxonomic clarity is especially essential given the high extinction rates in molluscs and particularly freshwater gastropods (Pip 2000; Lydeard et al. 2004). Freshwater gastropods comprise ~5% of the world's gastropod species but account for ~20% of the recorded molluscan extinctions (Strong et al. 2008). Addressing the extinction rate is difficult due to the shortage of experts studying snail biodiversity, a lack of baseline data for many species' distributions, and a thorough understanding of species life history (Strong et al. 2008). In addition, there may be some methodological barriers that are impeding the characterization of snail genomes. Extracting quality DNA and RNA from molluscs can be difficult due to their unique biochemical composition, namely the polysaccharides and polyphenolic proteins that form complexes with DNA and inhibit DNA-interactive enzymes (Adema 2021). However, recent work has demonstrated that mitochondrial genomes of snails can be obtained using mollusc-specific extraction kits and Illumina sequencing (Adema 2021; Rempel et al., in review).

Compared to nuclear genomes, the mitogenome may be easier to characterize because of the abundance of mitochondria within metazoan cells (Hebert et al. 2003; Ghiselli et al. 2021). Many features of mitochondrial genomes make them phylogenetically informative: different mutation rates or rates of evolution for different mitochondrial genes or regions promote insights into both deep and shallow relationships (Remigio and Blair 1997; Ghiselli et al. 2021). In practical terms, the additional loci recovered with mitogenome sequencing help to overcome the hurdle of selecting a single locus based on availability in public databases, providing more flexibility in phylogenetic analysis (Coissac et al. 2016). The various loci within the mitogenome, including 13 protein-coding regions and two ribosomal subunit genes, evolve at varying rates which allow examination of both recent divergences and assessment of phylogenetic relationships (Remigio and Blair 1997; Coissac et al. 2016). For example, *COI*, the commonly used barcode locus among

metazoa, has a relatively rapid rate of evolution, making it useful for delimiting species-level relationships within a genus; however, this locus does not provide strong bootstrap support for family-level clades (Jørgensen et al. 2004). In contrast, the large ribosomal subunit gene *16S* has a slower rate of evolution relative to *COI* and provides stronger support for more ancient divergences, like those among tribes (Jørgensen et al. 2004). This makes mitochondrial genomes an attractive option for researchers looking to unravel both species-level and genus-level relationships, as well as tribe-level and family-level relationships.

On the other hand, the very features that make mitogenomes useful for examining population structure may limit their utility in revealing phylogenetic relationships (Coissac et al. 2016; Ghiselli et al. 2021). Mitochondria experience uniparental inheritance and are not subject to recombination, which may result in the evolutionary history of the organelle being different from that of the individual as a whole (Coissac et al. 2016; Ghiselli et al. 2021). For these reasons, analysis of mitogenomes alone can fail to reveal introgression between species or lineages (Coissac et al. 2016; Ghiselli et al. 2021). Introgression is defined as the incorporation of alleles from one discrete taxonomic unit, usually a species, into the gene pool of another discrete unit (Harrison and Larson 2014). This generally occurs as a result of hybridization and/or backcrossing between sympatric populations/species (Harrison and Larson 2014). There is evidence that some alleles are more prone to introgress than others, depending on whether they confer adaptive advantages or are under selection (Harrison and Larson 2014). For instance, mitochondrial genes may experience faster rates of introgression than nuclear genes (Vilas et al. 2005). The possibility of introgression, combined with biparental inheritance, can lead to mitochondrial lineages that diverge from those of the individuals as a whole, or from species lineage (Vilas et al. 2005). Introgression is relative, meaning that while alleles at one locus introgress, alleles at another locus must remain constant

(Harrison and Larson 2014). Discordance between mitochondrial and nuclear phylogenies can reveal introgression by indicating that two gene pools have intermixed. An estimate of the occurrence and frequency of introgression can be helpful when delimiting species boundaries (Harrison and Larson 2014; Hibbins and Hahn 2019).

Mitogenomic research within gastropod taxa is still in its early phase, with only a fraction of its results being fully employed (Gomes-dos-Santos et al. 2019). There is an urgency to increase sequence libraries with under-represented and overlooked taxa, since the limited representation of gastropod diversity limits phylogenomic reconstruction (Kocot et al. 2011). One of the issues to be addressed through phylogenomic analysis is the overabundance of synonyms present within many gastropod taxa (Lydeard et al. 2004; Strong et al. 2008). Genomes would provide a more rigorous approach to testing the validity of species delimitation and ensuring that taxonomy reflects phylogeny.

Freshwater gastropods express the greatest levels of diversity and variation in the Holarctic region, a region which includes Canada (Strong et al. 2008). Important taxa in this region include the superorder Hygrophila, with their three major families: Physidae, Lymnaeidae, and Planorbidae (Saadi et al. 2020). Among freshwater limnic pulmonates, the Family Planorbidae has the greatest diversity, while phylogenetic relationships within the family remain largely unresolved, especially at the species-level (Jørgensen et al. 2004; Albrecht et al. 2007; Johnson et al. 2013). There are no precise estimates of global species richness within the Planorbidae, though 40 genera are recognized (Albrecht et al. 2007). Previous estimates by Baker (1945) and Hubendick (1955) placed the number around 350 species, but more recent estimates referred to 200 species (Albrecht et al. 2007). Phylogenomic analyses within the Planorbidae have largely focused on medically and economically important genera, such as *Biomphalaria* species, which

host the human disease-causing parasite *Schistosoma* spp. (Raghavan and Knight 2006; Albrecht et al. 2007; Saadi et al. 2020).

Relationships among planorbid genera have been built on descriptions of their internal anatomy and shell morphology, but the understanding of these relationships is rudimentary (Morgan et al. 2002; Wang et al. 2017). As traditional classification methods have increasingly been supplanted by molecular methods, different hypotheses about the phylogenetic relationships within this family, and more specifically in the genus *Planorbella* were proposed (Morgan et al. 2002; Jørgensen et al. 2004; Albrecht et al. 2007). For example, phylogenetic analysis of members of the Planorbidae using concatenated *COI* and *18S* rDNA sequences revealed that certain relationships proposed on the basis of comparative morphology were not supported with molecular evidence (Albrecht et al. 2007). The genera *Planorbella* and *Planorbarius* were proposed to be closely related based on their shared shell appearance and penile morphology (among other characteristics). However, this relationship was not supported in any of the phylogenetic analyses (Hubendick 1978; Albrecht et al. 2007).

Many species of ecological importance have not been DNA sequenced, such as snails in the genera *Planorbella* and *Helisoma* (Albrecht et al. 2007). These snails act as obligate hosts for many parasites of wildlife, including digenean trematodes (Jørgensen et al. 2004; Lockyer et al. 2004; Gordy et al. 2018). There have been no mitogenomic analyses of the intrageneric relationships among the *Planorbella* and *Helisoma* species, and these species are underrepresented in terms of sequence data on public databases. Incorporating both mitochondrial and nuclear loci in phylogenetic analysis of these species will allow us to explore species-level relationships and understand what processes led to their radiation, including introgression (Albrecht et al. 2007). At present, there is no genetic evidence for introgression between species of *Planorbella* and

Helisoma. However, genetic evidence of hybridization was observed in some *Biomphalaria* (a closely related planorbid genus) species (Mello-Silva et al. 1998; DeJong et al. 2001). If the species in this study have introgressed, I would expect to observe incongruence in the topologies of mitochondrial and nuclear gene trees (Hahn and Hibbins 2019).

In this chapter, I used next-generation-sequencing to characterize the complete mitochondrial genome and select nuclear rRNA repeats for five clades of *Planorbella* and *Helisoma* that were suggested from a *COI* gene phylogeny. Based on the *COI* clades and shell morphology, these clades were identified as *Helisoma anceps, Planorbella campanulata, Planorbella duryi, Planorbella pilsbryi,* and *Planorbella trivolvis*. Preliminary evidence from the barcode phylogeny indicated that misidentification and/or synonyms, as well as potential cryptic species were present within these species. Thus, reconstructing phylogenies from additional loci will help to determine how robustly the *COI* clades may reflect taxonomy. I compared the *COI* gene and mitogenome phylogenies to determine if there were any differences in the topology and support for the monophyly of clades. I predicted that these trees would share congruent topologies because all mitochondrial loci are linked and do not undergo recombination. Further, I compared the *COI* gene and mitogenome trees to nuclear gene phylogenies (*18S* and *28S*) to determine if there was any evidence for historic introgression between species. I hypothesized that closely related and co-occurring species have introgressed, particularly *Planorbella trivolvis* and *Planorbella pilsbryi*.

2.2 Methods

Specimens were hand collected at 16 locations throughout one Canadian province (Manitoba, Canada) and from one population each from two American states (Minnesota and Oklahoma, USA) (see Chapter 1: Table 1.1). I also included snails from two separate laboratory

strains (Oklahoma, Manitoba) as they have been a focus of host-parasite interactions research (Eliuk et al., 2020) (Table 1.1). A total of 269 individuals comprising five putative species were collected or obtained from laboratory colonies and identified with dichotomous keys based on shell characteristics of Baker (1945) and Burch (1989). Key-based identifications revealed that five species were likely present: *H. anceps, P. campanulata, P. duryi, P. pilsbryi,* and *P. trivolvis* (Table 1.1). All snails were preserved in 100% ethanol and stored at -20 ˚C.

2.2.1 DNA extraction of specimens

A 30 mg piece of tissue was removed from the head and mantle region for DNA extractions. In situations where the body mass was less than 30 mg, the entire specimen was used. Preserved tissue samples were rehydrated in MilliQ water, pulverized in liquid nitrogen, and incubated in 350 µL ML1 buffer solution with 25 µL proteinase K at 60 ˚C for a minimum of three hours. DNA extraction protocols followed the E.Z.N.A. Mollusc kit; following incubation, DNA samples were vortexed and eluted in MilliQ water. Eluted samples were stored at -20 ˚C until ready to perform polymerase chain reaction (PCR) or genomic library preparation. A tissue voucher from each of the specimens used in mitogenome sequencing will be deposited at the Manitoba Museum.

2.2.2 Identification of *Planorbella* and *Helisoma* with phylogenetic analysis of *COI* gene

DNA was quantified using the Take3 micro-volume (2 µL) plate absorbance detection application for the Bio-Tek Synergy #1 Microplate Reader and was diluted with MilliQ water (as needed) to obtain 25 µL solutions of 50 ng DNA. Amplifications of ~655bp of the *COI* gene in 25 µL reactions containing 50 ng of DNA, 1X buffer, 1.5 mM $MgCl_2$, 0.4 µM of each primer (summarized in Chapter 1: Table 1.2), 0.2 mM of each dNTP and 0.05unit/µL Omega Taq polymerase (Bio-Tek). PCR products were visualized in a 2% agarose gel; products with a single bright band were selected to be purified using a PCR clean up kit (MO BIO Laboratories. Inc).

Products were sequenced in the forward and reverse directions on an ABI 3730XL at the Hospital for Sick Children, Toronto, ON.

Following sequencing, constructed contigs and consensus sequences were edited by eye using Sequencher 5.4.6. Consensus sequences and sequences that were downloaded from NCBI GenBank were uploaded to MEGA X and aligned using ClustalW (Kumar et al. 2018). Within- and between-species genetic distances were calculated in MEGA using p-distance (Kumar et al. 2018). A total of 56 individual *COI* sequences were obtained for the specimens collected in this study, 29 of which were unique. Along with 19 unique sequences from GenBank (see Chapter 1: Table 1.3), a total of 48 sequences were included in phylogenetic analyses (586 bp). The Akaike information criterion (AIC) was used to select the best fit model of nucleotide substitution (GTR + I + G) in jModeltest 2.1.7 (Darriba et al. 2012). This model was then used to parametrize maximum likelihood (ML) analyses and Bayesian analyses. ML analyses were run with the model GTRGAMMA in RaxML v.8 (Stamatakis 2014) and bootstrap probability (BP) was calculated using 1,000 bootstrap replicates. For the Bayesian analysis, two independent analyses were run in MrBayes v.3.2.6 (Ronquist et al. 2012) with four independent chains for 10,000,000 generations, sampling every 1,000 generations. A burn-in of 2,500 was used and the remaining trees were used to calculate Bayesian posterior probabilities (PP). Following Albrecht et al., *Bulinus tropicus* (Gastropoda: Bulinidae, MN551571, MN551572) was used to root the trees (2007).

2.2.3 Mitochondrial genome sequencing, assembly, and annotation

After conducting *COI* phylogenetic analysis, one individual per clade was selected for additional mitogenomic sequencing, with two individuals (one each from Minnesota and Oklahoma) selected to represent *Planorbella trivolvis* (clade C) (Figure 2.1). To create fragment libraries, DNA samples were sheared by sonication with an S220 Focused-Ultrasonicator (Covaris,

Woburn, MA, USA) and prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, Massachusetts) as described previously (Peters and Marcus 2017). Libraries were sequenced by Illumina NovaSeq6000 (San Diego, California) at Genome Quebec.

Mitogenomes were assembled by mapping sequence libraries to a *Planorbella duryi* reference mitogenome (Gastropoda: Planorbidae, KY514384 (Schultz et al. 2018)) using five iterations of the medium sensitivity settings of Geneious 2020.2 as described by Marcus (2018). Annotations were performed using two reference mitogenomes: *P. duryi* and *Biomphalaria choanomphala* (Hygrophila: Planorbidae, MG431964) (Zhang et al 2018). The locations and structures of tRNAs were determined using ARWEN v.1.2 (Laslett and Canbäck 2008).

2.2.4 Phylogenetic analyses of mitogenome sequences and genetic distance

Phylogenetic analysis included an ingroup of 23 taxa consisting of six taxa from my sequencing and 17 gastropod species within the three families (Planorbidae, Lymnaeidae, and Physidae) in the superorder Hygrophila. The gastropod *Pyramidella dolabrata* (superfamily Pyramidelloidea) served as the outgroup as in Grande et al. (2008). A total of 11 of the 13 protein-coding genes were included in the analysis because the loci ATP6 and ATP8 were missing from four specimens from GenBank (MT628573, MT628577, MT862401, and MT862415); by including these taxa I chose to maximize species representation rather than locus representation. In addition, the analysis included two ribosomal subunit genes (*12S* and *16S*). Mitogenome sequences were aligned in CLUSTAL Omega (Sievers et al. 2011). The best fit model of evolution (GTR + I + G) was determined by jModeltest 2.1.7 (Darriba et al. 2012) using the Akaike information criterion (AIC). I specified these parameters for maximum likelihood (ML) and Bayesian analyses. ML analysis was run in RaxML v.8 with the evolution model GTRGAMMA

(Stamatakis 2014), calculating bootstrap probability (BP) with 10,000 replicates. Bayesian analysis was run in MrBayes v.3.2.6 (Ronquist et al. 2012), with two independent analyses run with four independent chains for 10,000,000 generations. Samples were taken every 1,000 generations, with the first 2,500 trees omitted as burn-in. The remaining trees were used to calculate Bayesian posterior probabilities (PP). Following Schultz et al., *Pyramidella dolabrata* (Gastropoda: Pyramidellidae, NC_012435) (Grande et al. 2008) was used to root the tree (Schultz et al. 2018). In addition to phylogenetic analysis, a table of mean genetic distance (p-distance) subfamilies was generated in MEGA v.X (Kumar et al. 2018).

2.2.5 Nuclear rRNA repeat genome sequencing, assembly, and annotation

Complete nuclear rRNA repeat sequences were also isolated from the sonicated DNA samples, which were assembled and annotated using reference sequences from the *18S* rRNA gene of *Galba cubensis* (Gastropoda: Lymnaeidae, Z83831) (Bargues *et al.* 1997); the partial *18S* rRNA, *ITS1, 5.8S* rRNA, *ITS2*, and partial *28S* rRNA region from *Satsuma polymorpha* (Gastropoda: Camaenidae, AB597368) (Hoso et al. 2010); the *28S* rRNA genes from *Planorbella trivolvis* (KY319366) (unpublished) and *Pomacea bridgesi* (Gastropoda: Ampullariidae, DQ279984) (Giribet et al. 2006); and the complete rRNA repeat from *Macrosoma conifera* (Lepidoptera: Hedylidae, MT878224) (McCullagh et al. 2020).

2.2.6 Phylogenetic analysis of nuclear rRNA repeat sequences and genetic distance

Nuclear rRNA repeat sequences were aligned using Clustal Omega 1.2.2 in Geneious Prime 2021.2.1. A table of p-distances among subfamilies was generated in MEGA v.X (Kumar et al. 2018). Owing to the limited availability of complete rRNA repeat sequences on GenBank, three separate alignments were generated: *18S, 28S,* and concatenated *ITS1, 5.8S,* and *ITS2* sequences. P-distances were calculated for each alignment and subfamilies were used as groups

for the mean within- and between-group calculations. Phylogenies were generated for only the *18S* and *28S* regions, while the concatenated *ITS1-5.8S-ITS2* sequences were analyzed using only pairwise distance calculations. The methods for phylogenetic analysis were similar to the above, with the best model of evolution for both *18S* and *28S* determined to be GTR + I + G. As per Albrecht et al. (2007), *Physella acuta* (Gastropoda: Physidae, KP171533) was used to root the *18S* phylogeny, while *Physa* spp. (Gastropoda: Physidae, AF435654) was used to root the *28S* phylogeny, as per Morgan et al. (2002).

Finally, to directly compare the mitochondrial and nuclear genetic signals, another set of alignments was created containing the sequences from the six individuals for which I obtained the mitogenomes. Alignments were created for the *COI* gene, the complete mitogenome, concatenate 18S-28S genes, and complete rRNA repeat sequences. *Biomphalaria tenagophila* served as an outgroup (Gastropoda: Planorbidae, NC_010220, MT017568, and MT017569). The phylogenetic analyses were identical to those above and the best model of evolution was GTR + I + G.

2.3 Results

Specimens (n = 257) were collected by hand from 16 field sampling locations in Manitoba, Canada (Table 1.1), with additional specimens provided by collaborators from one location each in Minnesota and Oklahoma, USA (Table 1.1). Individuals from two laboratory strains were also included in subsequent analysis: Manitoba (n = 6) and Oklahoma (n = 1).

2.3.1 *COI* identification of *Planorbella* and *Helisoma*

Fifty-six individual specimens were sequenced at the *COI* gene (586-1,541 bp). Phylogenetic analysis was performed on 586 bp of 48 sequences (Chapter 1: Figure 1.1), including 29 unique sequences from my own sampling and 19 from GenBank (Table 1.1). As described in Chapter 1 (sections 1.2 and 1.3), I inferred that there were five clades that each represented the

following species: *H. anceps, P. campanulata, P. duryi, P. pilsbryi,* and *P. trivolvis* (Figure 1.2). The weakest support occurred at the node separating clades C and D.

2.3.2 Mitochondrial genome sequencing, assembly, and annotation

The complete mitochondrial genomes for the target species in this study range in length from 13,026-14,490 base pairs (Table 2.1) and all contain 13 protein-coding genes (*cox1-3, nad1-6, nad4L, atp6, atp8, and cytb*), two rRNA genes (*12S* and *16S*), and 22 transfer RNA genes. The gene order is stable and consistent across all specimens, as is typical of panpulmonate snails (Schultz et al. 2018). Although tRNA gene order is less conserved across panpulmonate taxa and therefore more variable and diverse, the tRNA gene arrangement among the six individuals in this study was stable (Liu et al. 2012). Sequences will be deposited in GenBank (see Table 2.1 for accession numbers). Nucleotide compositions in the six mitogenome sequences are AT-rich as expected (see Table 2.2). GC skew, which is a measure of strand asymmetry, favoured G, while AT skew favoured T, as expected for panpulmonate snails (Liu et al. 2012).

2.3.3 Phylogenetic analysis of mitogenome sequences

Phylogenetic relationships of three families within the superorder Hygrophila were reconstructed for 23 species, with a single non-hygrophilan species (*P. dolabrata*) included as an outgroup. Relationships were reconstructed using sequence reads of 12,907 unambiguous nucleotide positions from 11 protein coding gene regions and two rRNA gene regions; maximum likelihood (ML) and Bayesian inference (BI) analyses were used to reconstruct phylogenetic relationships, following which the topologies of the two trees were assessed for congruence. The trees inferred from ML and BI analysis had identical topologies. The consensus tree from the BI analysis is shown with posterior probabilities and bootstrap values from ML displayed at each node (Figure 2.2).

Within the Family Planorbidae, sequences that had previously fallen into Clade C (*P. trivolvis*: Plsp30, Plsp43) in the *COI* tree instead grouped as sister taxa. The same occurred for members of *P. duryi* (Plsp31, KY514384). Both the mitogenome and the *COI* trees paired *P. pilsbryi* and *P. campanulata* as sister taxa, while *P. trivolvis* more distantly related to the two. Finally, *H. anceps* branched with *P. duryi* in both trees, however it is more derived in the *COI* tree and more ancestral to *P. duryi* in the mitogenome tree.

2.3.4 Nuclear rRNA repeat genome sequencing, assembly, and annotation

The complete nuclear rRNA repeat sequences for the target species in this study vary in length from 8,234-9,572 base pairs (see Table 2.3); all contain two internally transcribed spacer regions (*ITS*1 and *ITS2*) and three ribosomal subunit genes (*5.8S, 18S,* and *28S*). The gene order was stable and consistent across all specimens. Sequences will be deposited in GenBank.

2.3.4 Phylogenetic analysis of nuclear rRNA repeat sequences

For the *18S* locus, phylogenetic analysis was performed on a dataset consisting of 26 sequences (1,685 bp) (Figure 2.2), including 20 sequences from GenBank (Table 2.4). I recovered six clades corresponding to the six subfamilies included in the analysis with moderate support in both BI and ML analyses (>91 and 66, respectively). The weakest support occurred at the node separating the subfamilies Bulininae and Ancylinae. Bayesian and ML trees share similar topologies, but differences occur at a single node, where *Segmentina netidia* (EF012195) and *Bathyomphalus contortus* (EF012184) form a polytomy. For simplicity, only the results of the ML tree are included (Figure 2.2). Genetic distances (p-distance) within and among species were calculated and showed that *H. anceps, P. campanulata, P. pilsbryi,* and *P. trivolvis* were identical with zero nucleotide differences, while *P. duryi* was 0.06% different than the other four species (results not shown).

For the *28S* locus, I ran phylogenetic analysis for 25 sequences (1,479 bp), including 19 sequences from GenBank (Table 2.5). I recovered five clades corresponding to the five subfamilies included in the analysis with strong support and identical topologies in both BI and ML analyses (>99 and 84, respectively) (Figure 2.3). The Bayesian tree is pictured since the branching patterns among planorbids were more visible (longer branch lengths, with congruent topology to the ML tree). Genetic distances within and among planorbid species are shown in Table 2.7 (A). The mean interspecific distances among planorbid species in the genera *Helisoma*, *Planorbella*, and *Biomphalaria* ranged from 0.00-0.46%.

As an alternative to phylogenetic analysis with broad taxon sampling for the region encompassing *5.8S, ITSI* and *ITS2*, pairwise distances were measured for two different alignments: one for partial *5.8S-ITS2* (638 bp, n = 23) containing members of three Hygrophilan families (Planorbidae, Lymnaeidae, and Physidae), and one for the complete *ITS1-5.8S-ITS2* region (1,417 bp, n = 9) containing only members of the genera *Planorbella, Helisoma,* and *Biomphalaria* (Table 2.6). Mean distance values between genera are displayed in Table 2.7 (B-C). The mean interspecific distances for partial *5.8S-ITS2* sequences ranged from 0.56-25.0%. For the complete *ITS1-5.8S-ITS2* region, the mean interspecific distance ranged from 0.64-28%.

Finally, the four phylogenies reconstructed for the six individuals that were sequenced at the mitogenome were compared to determine whether nuclear loci provide sufficient signal for species identification. Five clades were recovered in both mitochondrial analyses, with strong support recovered for some nodes (Figure 2.4A and B). In the *COI* tree, the sister relationship between *H. anceps* and *P. duryi* was weakly supported (55 and 65 in BI and ML analyses). It is notable that the relationship between *H. anceps* and *P. duryi* was different from the trees with broader taxon sampling (Figures 1.2 and 2.1). In the *COI* tree in Fig. 2.4A, *H. anceps* and *P. duryi*

are sister species, while in the mitogenome tree (Fig. 2.4B) they are simply the most closely related. These two relationships are reversed in Figures 1.2 and 2.1: in the *COI* tree, *H. anceps* and *P. duryi* are the most closely related, while in the mitochondrial tree they are sister species. For the nuclear loci (Fig. 2.4C and D), the concatenated *18S/28S* analysis did not recover 5 clades, but grouped *H. anceps, P. campanulata, P. pilsbryi,* and *P. trivolvis* as a shallowly branching polytomy with low support. This was similar to the results in Figures 2.2 and 2.3, where there was insufficient differentiation among these species for branching patterns to be clear. However, with the complete rRNA phylogeny (Fig. 2.4D), five clades were recovered with strong support. The topology is similar to that of the mitogenome, with one major change in the relative positions of *P. pilsbryi* and *P. trivolvis.* In Fig. 2.4B, *P. pilsbryi* is sister species with *P. campanulata*, while *P. trivolvis* is more distantly related; this relationship is reversed in Fig. 2.4D.

2.4 Discussion

The *COI* barcode tree shared similar topology with the mitogenome phylogeny. Both trees indicated that five species were present in my sampling and the relationships between those taxa were similar. *H. anceps* was most closely related to *P. duryi*, while *P. trivolvis*, *P. pilsbryi*, and *P. campanulata* were more similar to each other. In addition, there were no singletons in the mitogenome phylogeny that make inferring monophyly of species more difficult as seen in the *COI* phylogeny (e.g. *P. trivolvis* KY319368). In the mitogenome phylogeny, all nodes possessed uniformly high support (>98), unlike in the barcode phylogeny.

The genera *Helisoma* and *Planorbella* formed a single monophyletic clade with strong support in both the nuclear phylogenies and the mitochondrial phylogenies. Relationships among taxa in the three phylogenies (mitogenome, *18S* and *28S*) were similar, with *H. anceps* and *P. duryi* being closely related while the remaining species were more closely related to each other than the

other two species. One difference occurred in the *18S* tree: two *H. anceps* (Plsp185, HQ659968) paired more closely with *P. campanulata, P. pilsbryi,* and *P. trivolvis* than with *P. duryi*. Additionally, in contrast to the other gene trees, there was a GenBank sequence in the *28S* phylogeny labelled *H. anceps* that was sister to an *H. anceps* sequence from field collections, providing stronger support for my identification of that specimen and the clade itself.

I found little evidence of historic introgression among the five focal species in this study; due to the lack of divergence among the single gene *18S* and *28S* sequences, it is difficult to detect differences in the topologies between the nuclear and mitochondrial gene/mitogenome trees. One difference in the mitogenome tree was that *P. duryi* and *H. anceps* were sister species, while in the nuclear trees *H. anceps* tended to be more closely related to the other three species (*P. campanulata, P. pilsbryi* and *P. trivolvis*). However, relationships among species are less clear in the *18S* tree, considering that two GenBank specimens identified as *H. anceps* are more similar to other species than to each other. In the *28S* tree, the similarity between all the nominal species lead to a lack of a branching pattern, making it difficult to infer species clades from this gene tree. In both *18S* and *28S* phylogenies, four species (*H. anceps, P. campanulata, P. pilsbryi,* and *P. trivolvis*) formed a polytomy. From this I can conclude that *18S* and *28S* are not useful on their own for delimiting *Planorbella* and *Helisoma* at either the genus or species level. However, both trees were consistently separating sequences labelled as *P. duryi* out from other *Planorbella* species, despite forming their own polytomy, so there is some resolution at the genus level.

Bayesian inference and maximum likelihood analyses for mitogenomic loci of the 23 Hygrophilan species and an outgroup recovered strong statistical support for nodes at the family-level, genus-level, and species-level. Within the Family Planorbidae (clade B), there is a high representation of species from the genus *Biomphalaria*, owing to its economic and medical

significance as an obligate host for *Schistosoma* spp. Inclusion of members of this genus was necessary as they are highly represented on the sequence database GenBank. It was one of the goals of this study to increase planorbid representation for those species that have historically been overlooked in phylogenomic analyses, since none of the focal species had mitogenome sequences available on public sequence databases.

While the entire coding region of the nuclear rRNA repeat was collected for the six focal specimens, only two loci (*18S* and *28S*) were used for phylogenetic analysis across a broad sample of taxa. This is a result of the limited samples available on GenBank; in order to maximize taxon sampling, certain less commonly used loci (*ITS1*, *ITS2*, and *5.8S*) had to be excluded from the analysis (Vilas et al. 2005). I will make the complete rRNA repeat sequences collected for this study publicly available, so that future investigations into *Planorbella* and *Helisoma* may make use of the loci when investigating deeper nodes.

The phylogenetic analyses with reduced taxon representation allowed for comparison of the full mitogenome and the full nuclear rRNA repeat sequences (Fig. 2.4). The increased branching of the complete rRNA repeat sequence tree compared to the concatenated *18S/28S* tree demonstrated that the *ITSI, 5.8S*, and *ITSII* loci can increase the resolution in phylogenetic reconstructions involving planorbids. There are differences between the reduced and expanded taxon sampling *COI* and mitogenome trees, namely in the relationships between *H. anceps* and *P. duryi.* This may be a result of the decreased taxon representation, given the low support for the sister relationship between the two species (Fig. 2.4A). Between the mitochondrial and nuclear trees, there is a change in the relative positions of *P. pilsbryi* and *P. trivolvis*; in the mitochondrial tree *P. pilsbryi* is more closely related to *P. campanulata,* while *P. trivolvis* is more distantly related. In contrast, in the rRNA repeat tree, *P. trivolvis* is most closely related to *P. campanulata*,

while *P. pilsbryi* is the most distantly related. This discordance between mitochondrial and nuclear loci might be evidence of past hybridization between *P. pilsbryi* and *P. trivolvis.* However, the disagreement between these trees is difficult to interpret without additional taxa. An alternative experimental crossing approach akin to the methods used by Mello-Silva et al. (1998) could also reveal whether these species have the capacity to successfully hybridize to create viable offspring.

2.5 Conclusions

I have conducted phylogenetic analysis using several loci, both nuclear and mitochondrial, for taxa within the superorder Hygrophila with the intention of more accurate and rigorous species delineation in the genera *Planorbella* and *Helisoma*. I have compared the results of mitochondrial *COI* phylogenetic analysis to that of mitogenome loci and found they have mostly congruent topologies, albeit support for some relationships was stronger with the mitogenome analysis. Using two nuclear loci, I was able to compare mitochondrial and nuclear phylogenies in order to look for introgression. Using the complete mitogenome and rRNA repeat sequences, I found incongruencies in the topologies of the trees, providing some evidence for hybridization or introgression between *P. pilsbryi* and *P. trivolvis*. This work presents the first complete mitochondrial genome sequences for five planorbid species, as well as the first complete nuclear rRNA repeat sequences for these species. The availability of these sequences will increase taxon and locus representation for Planorbidae. These increased genetic resources will allow researchers more freedom when using these species in phylogenetic analyses to choose the loci that best serve to answer the question they are pursuing, whether at shallow or deeper nodes.

Literature Cited

Adema, C.M. 2021. Sticky problems : extraction of nucleic acids from molluscs. Phil. Trans. R. Soc. B **376**.

Albrecht, C., Kuhn, K., and Streit, B. 2007. A molecular phylogeny of Planorboidea (Gastropoda, Pulmonata): Insights from enhanced taxon sampling. Zool. Scr. **36**(1): 27–39.

Baker, F.C. 1945. The molluscan family Planorbidae. University of Illinois Press, Urbana, IL.

Coissac, E., Hollingsworth, P.M., Lavergne, S., and Taberlet, P. 2016. From barcodes to genomes: Extending the concept of DNA barcoding. Mol. Ecol. **25**(7): 1423–1428.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. 2012. JModelTest 2: More models, new heuristics and parallel computing. Nat. Methods **9**(8): 772. Nature Publishing Group.

Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. **85**(3): 407–415.

DeJong, R.J., Morgan, J.A.T., Lobato Paraense, W., Pointier, J.P., Amarista, M., Ayeh-Kumi, P.F.K., Babiker, A., Barbosa, C.S., Brémond, P., Canese, A.P., De Souza, C.P., Dominguez, C., File, S., Gutierrez, A., Nino Incani, R., Kawano, T., Kazibwe, F., Kpikpi, J., Lwambo, N.J.S., Mimpfoundi, R., Njiokou, F., Poda, J.N., Sene, M., Velásquez, L.E., Yong, M., Adema, C.M., Hofkin, B. V., Mkoji, G.M., and Loker, E.S. 2001. Evolutionary relationships and biogeography of Biomphalaria (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. Mol. Biol. Evol. **18**(12): 2225–2239.

Ghiselli, F., Gomes-dos-santos, A., Adema, C.M., Lopes-lima, M., Sharbrough, J., and Boore, J.L. 2021. Molluscan mitochondrial genomes break the rules. Phil. Trans. R. Soc. B 376 **376**(1): 1–18.

Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L.F.C., and Froufe, E. 2019. Molluscan

genomics: the road so far and the way forward. Hydrobiologia **847**(7): 1705–1726.

Gordy, M.A., Cobb, T.P., and Hanington, P.C. 2018. Swimmer's itch in Canada: A look at the past and a survey of the present to plan for the future. Environ. Heal. A Glob. Access Sci. Source **17**(73): 1–18. Environmental Health.

Hahn, M.W., and Hibbins, M.S. 2019. A Three-Sample Test for Introgression. Mol. Biol. Evol. **36**(12): 2878–2882.

Harrison, R.G., and Larson, E.L. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. **105**(S1): 795–809.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and DeWaard, J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. **270**: 313–321.

Hebert, P.D.N., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. **54**(5): 852–859.

Hibbins, M.S., and Hahn, M.W. 2019. The timing and direction of introgression under the multispecies network coalescent. Genetics **211**(3): 1059–1073.

Hubendick, B. 1955. Phylogeny in the Planorbidae. Trans. Zool. Soc. London **28**(1): 453–542.

Hubendick, B. 1978. Systematics and comparative morphology of the Basommatophora. *In* Pulmonates. Academic Press, New York, NY.

Johnson, P.D., Bogan, A.E., Brown, K.M., Burkhead, N.M., Cordeiro, J.R., Garner, J.T., Hartfield, P.D., Lepitzki, D.A.W., Mackie, G.L., Pip, E., Tarpley, T.A., Tiemann, J.S., Whelan, N. V., and Strong, E.E. 2013. Conservation status of freshwater gastropods of Canada and the United States. Fisheries **38**(6): 247–282.

Jørgensen, A., Kristensen, T.K., and Stothard, J.R. 2004. An investigation of the "Ancyloplanorbidae" (Gastropoda, Pulmonata, Hygrophila): Preliminary evidence from

DNA sequence data. Mol. Phylogenet. Evol. **32**(3): 778–787.

Kocot, K.M. 2013. Recent advances and unanswered questions in deep molluscan phylogenetics. Am. Malacol. Bull. **31**(1): 195–208.

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., and Halanych, K.M. 2011. Phylogenomics reveals deep molluscan relationships. Nature **477**(7365): 452–456.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547–1549.

Laslett, D., and Canbäck, B. 2008. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics **24**(2): 172–175.

Lockyer, A.E., Jones, C.S., Noble, L.R., and Rollinson, D. 2004. Trematodes and snails: An intimate association. Can. J. Zool. **82**(1): 251–269.

Lydeard, C., Cowie, R.H., Ponder, W.F., Bogan, A.E., Bouchet, P., Clark, S.A., Cummings, K.S., Frest, T.J., Gargominy, O., Herbert, D.G., Hershler, R., Perez, K.E., Roth, B., Seddon, M., Strong, E.E., and Thompson, F.G. 2004. The global decline of nonmarine mollusks. Bioscience **54**(4): 321–330.

Marcus, J.M. 2018. Our love-hate relationship with DNA barcodes, the Y2K problem, and the search for next generation barcodes. AIMS Genet. **5**(1): 1–23.

Mello-Silva, C.C., Grault, C.E., Da Costa, V.A., and Barbosa, F.S. 1998. Possible hybridization of Brazilian planorbid snails and its importance in population dynamics. Mem. Inst. Oswaldo Cruz **93**(SUPPL. I): 227–232.

Morgan, J.A.T., DeJong, R.J., Jung, Y., Khallaayoune, K., Kock, S., Mkoji, G.M., and Loker, E.S.

2002. A phylogeny of planorbid snails, with implications for the evolution of *Schistosoma* parasites. Mol. Phylogenet. Evol. **25**(3): 477–488.

Moritz, C., and Cicero, C. 2004. DNA barcoding: Promise and pitfalls. PLoS Biol. **2**(10).

Mutanen, M., and Pretorius, E. 2007. Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: A comparison of moth genitalia. Syst. Entomol. **32**(2): 371–386.

Pip, E. 2000. The decline of freshwater molluscs in southern Manitoba. Can. F. Nat. **114**(4): 555–560.

Raghavan, N., and Knight, M. 2006. The snail (*Biomphalaria glabrata*) genome project. Trends Parasitol. **22**(4): 148–151.

Remigio, E.A., and Blair, D. 1997. Molecular systematics of the freshwater snail Family Lymnaeidae (Pulmonata: Basommatophora) utilising mitochondrial ribosomal DNA sequences. J. Molluscan Stud. **63**(2): 173–185.

Richards, S. 2015. It's more than stamp collecting: how genome sequencing can unify biological research. Trends Genet. **31**(7): 411–421.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61**(3): 539–542.

Saadi, A.J., Davison, A., and Wade, C.M. 2020. Molecular phylogeny of freshwater snails and limpets (Panpulmonata: Hygrophila). Zool. J. Linn. Soc. **190**(2): 518–531.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R.H. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. Annu. Rev.

Entomol. **55**: 421–438.

Schultz, J.H., Bansbach, L.M., Bremmer, J.A., Dimmler, K.E., Forde, Q.A., Gagliano, E.M., Glenn, E.M., Greengrass, C.M., Hayes, J.P., Kraus, A.L., Larsen, L.I., Lucero, E., McClendon, M.T., Mercer, H.L., Mims, K.C., Patel, K.N., Patsalis, F.I., Peterson, D.E., Platero, J.M., Rizvi, M.M., Serna, K.I., Steele, T.E., Turner, N.L., Bu, L., Lu, L., and Adema, C.M. 2018. The mitochondrial genome of the planorbid snail *Planorbella duryi*. Mitochondrial DNA Part B Resour. **3**(2): 972–973.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7**(539): 1–6.

Stamatakis, A. 2014. RAxML. Manual/tutorial (1): 1–5.

Strong, E.E., Gargominy, O., Ponder, W.F., and Bouchet, P. 2008. Global diversity of gastropods (Gastropoda; Mollusca) in freshwater. Hydrobiologia **595**(1): 149–166.

Vilas, R., Criscione, C.D., and Blouin, M.S. 2005. A comparison between mitochondrial DNA and the ribosomal internal transcribed regions in prospecting for cryptic species of platyhelminth parasites. Parasitology **131**(6): 839–846.

Wang, J.G., Zhang, D., Jakovlić, I., and Wang, W.M. 2017. Sequencing of the complete mitochondrial genomes of eight freshwater snail species exposes pervasive paraphyly within the Viviparidae family (Caenogastropoda). PLoS One **12**(7): 1–17.

Wheeler, Q.D., Raven, P.H., and Wilson, E.O. 2004. Taxonomy: Impediment or Expedient? Science (80-.). **303**(5656): 285.

Wilson, E.O. 2004. Taxonomy as a fundamental discipline. Philos. Trans. R. Soc. B Biol. Sci.

359(1444): 739.

Table 2.1 Complete mitochondrial genome sequences of Panpulmonata: Hygrophila, including sequences accessed from GenBank and sequences generated in this study (bolded).

Family	Species	Length (bp)	Accession numbers (lab/ GenBank)*	Geographic location
Planorbidae	***Helisoma anceps***	**13,675**	**Plsp185**	**Manitoba, Canada**
	Planorbella campanulata	**13,751**	**Plsp243**	**Manitoba, Canada**
	Planorbella duryi	**13,687**	**Plsp31**	**Lab strain, Oklahoma, USA**
	Planorbella duryi	14,217	KY514384	New Mexico, USA
	Planorbella pilsbryi	**13,720**	**Plsp16/MW8 89961**	**Manitoba, Canada**
	Planorbella trivolvis	**13,826**	**Plsp30**	**Oklahoma, USA**
	Planorbella trivolvis	**13,852**	**Plsp43**	**Minnesota, USA**
	Planorbarius corneus	14,020	MT862415	Denmark
	Biomphalaria choanomphala	13,672	NC_038061	Kenya
	Biomphalaria sudanica	13,671	NC_038060	Kenya
	Biomphalaria pfeifferi	13,624	NC_038059	Kenya
	Biomphalaria glabrata	13,670	NC_005439	Lab strain, UK
	Biomphalaria straminea	13,650	NC_036993	Lab strain, China
	Biomphalaria tenagophila	13,722	NC_010220	Lab strain, Brazil
	Bathyomphalus contortus	13,679	MT628573	Denmark
	Gyraulus laevis	13,685	MT628577	Denmark
	Planorbis planorbis	13,316	MT862401	Denmark
	Planorbis carinatus	13,026	MT483701	Denmark
Lymnaeidae	*Galba pervia*	13,768	NC_018536	China
	Radix auricularia	13,745	NC_026538	USA
	Radix balthica	13,983	KP098541	Germany
	Ampullaceana lagotis	13,751	MN175602	Nanchang, China
Physidae	*Physella acuta*	14,490	NC_023253	New Mexico, USA
Outgroup / Pyramidellidae	*Pyramidella dolbrata*	13,856	NC_012435	Unavailable

*where applicable

Table 2.2 Nucleotide characterization from unambiguous base pairs of complete mitochondrial genome sequences of Hygrophila/Pulmonata.

ID no.	Species	Nucleotide frequency (%)				Whole Genome Sequence		
		A	T	G	C	A+T%	AT skew	GC skew
MW889961, Plsp16	*Planorbella pilsbryi*	33.8	41.5	13.3	11.4	75.4	-0.10212	0.07724
Plsp30	*Planorbella trivolvis*	33.6	41.4	13.6	11.5	74.9	-0.10414	0.08367
Plsp31	*Planorbella duryi*	31.2	41.4	14.3	13.0	72.6	-0.14050	0.04744
Plsp43	*Planorbella trivolvis*	33.5	41.4	13.6	11.4	74.9	-0.10547	0.08765
Plsp185	*Helisoma anceps*	32.2	41.6	14.2	12.0	73.8	-0.12737	0.08397
Plsp243	*Planorbella campanulata*	33.1	41.1	14.1	11.7	74.2	-0.10782	0.09302
KY514384	*Planorbella duryi*	31.0	41.7	14.3	13.0	72.7	-0.14718	0.047619
MT862415	*Planorbarius corneus*	33.2	41.2	13.7	11.2	74.9	-0.10681	0.09960
NC_038061	*Biomphalaria choanomphala*	33.9	42.6	13.0	10.5	76.6	-0.11358	0.10684
NC_038060	*Biomphalaria sudanica*	33.9	42.7	13.0	10.4	76.6	-0.11488	0.11111
NC_038059	*Biomphalaria pfeifferi*	33.8	42.9	13.1	10.3	76.6	-0.11880	0.11966
NC_005439	*Biomphalaria glabrata*	33.1	41.6	14.1	11.3	74.6	-0.11394	0.11024
NC_036993	*Biomphalaria straminea*	33.3	42.0	13.9	10.8	75.3	-0.11554	0.12551
NC_010220	*Biomphalaria tenagophila*	33.7	42.1	13.5	10.7	75.8	-0.11082	0.11570
MT628573	*Bathyomphalus contortus*	33.1	42.0	13.5	11.3	75.1	-0.11851	0.08835
MT628577	*Gyraulus laevis*	32.8	42.0	13.6	11.7	74.8	-0.12299	0.075397
MT862401	*Planorbis planorbis*	32.7	42.9	12.9	11.5	75.6	-0.13492	0.05738
MT483701	*Planorbis carinatus*	32.8	43.4	12.8	11.0	76.2	-0.13911	0.07563
NC_018536	*Galba pervia*	32.2	40.5	14.6	12.7	72.7	-0.11417	0.06960
NC_026538	*Radix auricularia*	30.9	39.7	15.5	13.8	70.7	-0.12447	0.05802
KP098541	*Radix balthica*	31.6	39.7	15.4	13.3	71.3	-0.11360	0.07317
MN175602	*Ampullaceana lagotis*	30.7	39.6	15.7	13.9	70.3	-0.12660	0.06061
NC_023253	*Physella acuta*	32.2	37.0	16.4	14.3	69.2	-0.06936	0.06818
NC_012435	*Pyramidella dolbrata*	27.4	36.0	19.6	17.0	63.4	-0.13565	0.07104

Table 2.3 Nucleotide characterization from unambiguous base pairs of nuclear rRNA repeat genome sequences of Hygrophila/Pulmonata.

ID no.	**Species**	**Nucleotide frequency (%)**				**nuclear rRNA repeat Sequence**		
		A	T	G	C	A+T%	AT skew	GC skew
Plsp16	*Planorbella pilsbryi*	23.8	23.9	27.5	24.7	47.7	-0.00210	0.05354
Plsp30	*Planorbella trivolvis*	23.6	23.8	27.6	24.9	47.5	-0.00421	0.05143
Plsp31	*Planorbella duryi*	23.2	23.3	28.4	25.1	46.5	-0.00215	0.06168
Plsp43	*Planorbella trivolvis*	23.4	24.1	27.9	24.7	47.5	-0.01474	0.06095
Plsp185	*Helisoma anceps*	22.4	22.8	29.4	25.4	45.2	-0.00885	0.07299
Plsp243	*Planorbella campanulata*	23.7	24.0	27.5	24.8	47.7	-0.00629	0.05163

Table 2.4 Summary of GenBank sequences and accession numbers used in *18S* phylogeny.

Family	Species	Accession	Location
Planorbidae	*Ancyclus fluviatilis*	AY282593	Germany
	Anisus soirobis	EF012183	Germany
	Bathyomphalus contortus	EF012184	Lab strain
	Biomphalaria tenagophila	MT017569	Brazil
	Bulinus tropicus	AY282594	Lab strain
	Helisoma anceps	HQ659968	California, USA
	Indoplanorbis exustus	AY282598	Thailand
	Laevapex fuscus	AY282599	USA
	Planorbarius corneus	AY282601	Germany
	Planorbella duryi	HM756307	Lab strain
	Planorbella duryi	KY514382	New Mexico, USA
	Planorbella tenue	EF012191	Mexico
	Planorbis planorbis	EF012192	Germany
	Planorbula armigera	EF012193	Mexico
	Segmentina netidia	EF012195	Germany
Lymnaeidae	*Galba truncatula*	FR797815	Germany
	Lymnaea stagnalis	FR797829	Germany
	Radix auricularia	FR797818	Germany
Physidae	*Physella acuta*	KP171533	Italy

Table 2.5 Summary of GenBank sequences and accession numbers used in *28S* phylogeny.

Family	Species	Accession	Location
Planorbidae	*Biomphalaria glabrata*	AF435694	Brazil
	Biomphalaria sudanica	AF435692	Lab strain
	Biomphalaria tenagophila	AF435690	Paraguay
	Bulinus africanus	AF435658	Kenya
	Ferrissia sp.	AF435664	USA
	Gyraulus sp.	AF435675	USA
	Helisoma anceps	AF435689	USA
	Indoplanorbis exustus	AF435662	Thailand
	Menetus portlandensis	AF435682	USA
	Planorbella campanulata	AY465060	USA
	Planorbella duryi	AF435684	Brazil
	Planorbella duryi	EF152572	USA
	Planorbella duryi	FJ423081	USA
	Planorbella duryi	HM230324	USA
	Planorbella duryi	KY514383	USA
	Planorbella trivolvis	AF435688	USA
	Planorbella trivolvis	KY319366	USA
	Planorbis planorbis	AF435672	Egypt
	Planorbula armigera	AF435683	USA
Lymnaeidae	*Radix auricularia*	AY465067	USA
Physidae	*Physa sp.*	AF435654	Egypt

Table 2.6 Summary of GenBank sequences and accession numbers used in two alignments for complete and partial sequences of *ITS1-5.8S-ITS2*.

Family	Species	Loci	Accession	Location
Planorbidae	*Ancyclus fluviatilis*	*5.8S* partial, *ITS2* partial	MN644839	Scotland
	Biomphalaria glabrata	complete *ITS1, 5.8S, ITS2*	MN644832	Puerto Rico
	Biomphalaria tenagophila	*5.8S* partial, *ITS2* partial	MT017569	Brazil
	Bulinus natalensis	*5.8S* partial, *ITS2* partial	MN644823	South Africa
	Ferrissia californica	*5.8S* partial, *ITS2* partial	MN644840	USA
	Gyraulus parvus	*5.8S* partial, *ITS2* partial	MN644828	Canada
	Menetus callioglyphus	*5.8S* partial, *ITS2* partial	MN644833	USA
	Planorbarius corneus	*5.8S* partial, *ITS2* partial	MN644834	Ukraine
	Planorbella duryi	*5.8S* partial, *ITS2* partial	AY628860	Egypt
	Planorbella duryi	*5.8S* partial, *ITS2* partial	AY628861	Egypt
	Planorbella duryi	complete *ITS1, 5.8S, ITS2*	MT753116	Antigua & Barbuda
	Planorbella duryi	complete *ITS1, 5.8S, ITS2*	MT753117	Antigua & Barbuda
	Planorbella oregonensis	*5.8S* partial, *ITS2* partial	MH492678	USA
	Planorbella subcrenata	*5.8S* partial, *ITS2* partial	MH198420	USA
	Planorbella subcrenata	*5.8S* partial, *ITS2* partial	MN644835	Canada
	Planorbella trivolvis	*5.8S* partial, *ITS2* partial	AY030403	USA
	Planorbis planorbis	*5.8S* partial, *ITS2* partial	MN644830	Ukraine
	Planorbula campestris	*5.8S* partial, *ITS2* partial	MN644836	USA
	Promentus exacuous	*5.8S* partial, *ITS2* partial	MN644837	Canada
	Vorticifex effusus	*5.8S* partial, *ITS2* partial	MN644838	USA
Lymnaeidae	*Lymnaea stagnalis*	*5.8S* partial, *ITS2* partial	MN644814	UK
Physidae	*Haitia mexicana*	*5.8S* partial, *ITS2* partial	MN644808	USA

Table 2.7 Mean pairwise distances for nuclear rRNA repeat genome sequences among species of the family Planorbidae.

A) Mean interspecific p-distance for the *28S* region (1,479 bp) among a subset of five planorbid species (n = 15 ind.).

	Biomphalaria spp.	*Helisoma anceps*	*Planorbella campanulata*	*Planorbella trivolvis*	*Planorbella pilsbryi*
Biomphalaria spp.					
Helisoma anceps	0.0035				
Planorbella campanulata	0.0035	0.0014			
Planorbella trivolvis	0.0046	0.0028	0.0028		
Planorbella pilsbryi	0.0035	0.0007	0.0000	0.0021	

B) Mean interspecific p-distance for the partial *5.8S-ITSII* region (638 bp) among three planorbid genera: *Biomphalaria, Helisoma,* and *Planorbella* (n = 13 ind.).

	B. tenagophila	*B. glabrata*	*H. anceps*	*P. duryi*	*P. subcrenata*	*P. trivolvis*	*P. pilsbryi*	*P. campanulata*
B.tenagophila								
B. glabrata	0.1077							
H. anceps	0.2166	0.2108						
P. duryi	0.2140	0.2503	0.1063					
P. subcrenata	0.1982	0.2157	0.0670	0.1088				
P. trivolvis	0.2035	0.2100	0.0636	0.1237	0.0228			
P. pilsbryi	0.1990	0.2184	0.0680	0.1061	0.0056	0.0291		
P. campanulata	0.2020	0.2088	0.0614	0.1248	0.0191	0.0060	0.0247	

C) Mean interspecific p-distance for the complete *ITSI-5.8S-ITSII* region (1,417 bp) among six planorbid species (n = 9 ind.).

	Biomphalaria tenagophila	*Helisoma anceps*	*Planorbella duryi*	*Planorbella pilsbryi*	*Planorbella trivolvis*	*Planorbella campanulata*
Biomphalaria tenagophila						
Helisoma anceps	0.27757					
Planorbella duryi	0.26196	0.13464				
Planorbella pilsbryi	0.27963	0.11350	0.10479			
Planorbella trivolvis	0.28048	0.11559	0.11291	0.02929		
Planorbella campanulata	0.28093	0.11217	0.11042	0.02813	0.00635	

D) Mean between-group p-distance for the complete *ITSI-5.8S-ITSII* region (1,417 bp) among three planorbid genera (n = 9 ind.).

	Biomphalaria	*Helisoma*	*Planorbella*
Biomphalaria			
Helisoma	0.278		
Planorbella	0.272	0.123	

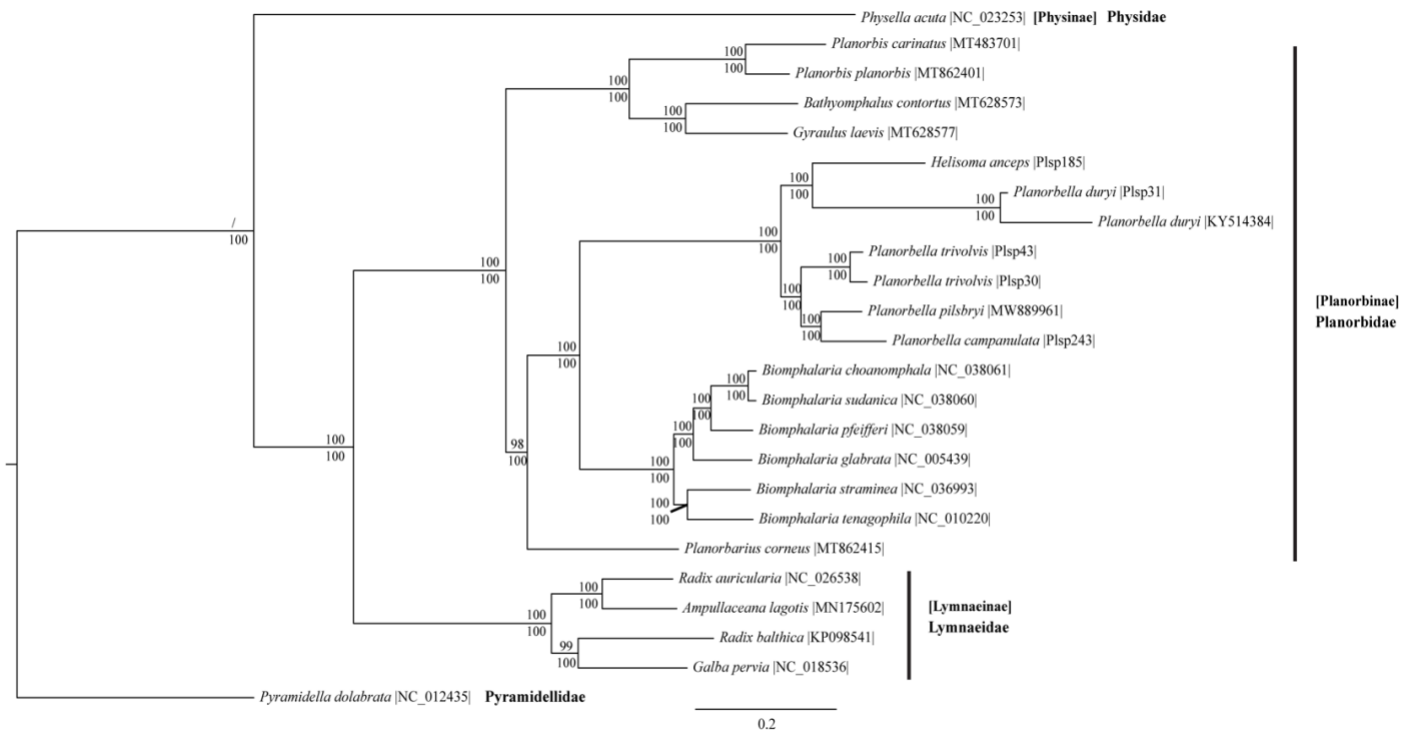


Figure 2.1 Phylogenetic relationship among Hygrophila species in three families based on mitochondrial genomic sequences. Concatenated DNA sequences of 11 protein-coding genes and 2 rRNA genes were analyzed using Bayesian inference (BI) and maximum likelihood (ML) analyses, with *Pyramidella dolabrata* as an outgroup. The Bayesian tree is shown. Numbers above each node are maximum likelihood bootstrap values, while Bayesian posterior probabilities are displayed under each node (showing values over 98%).

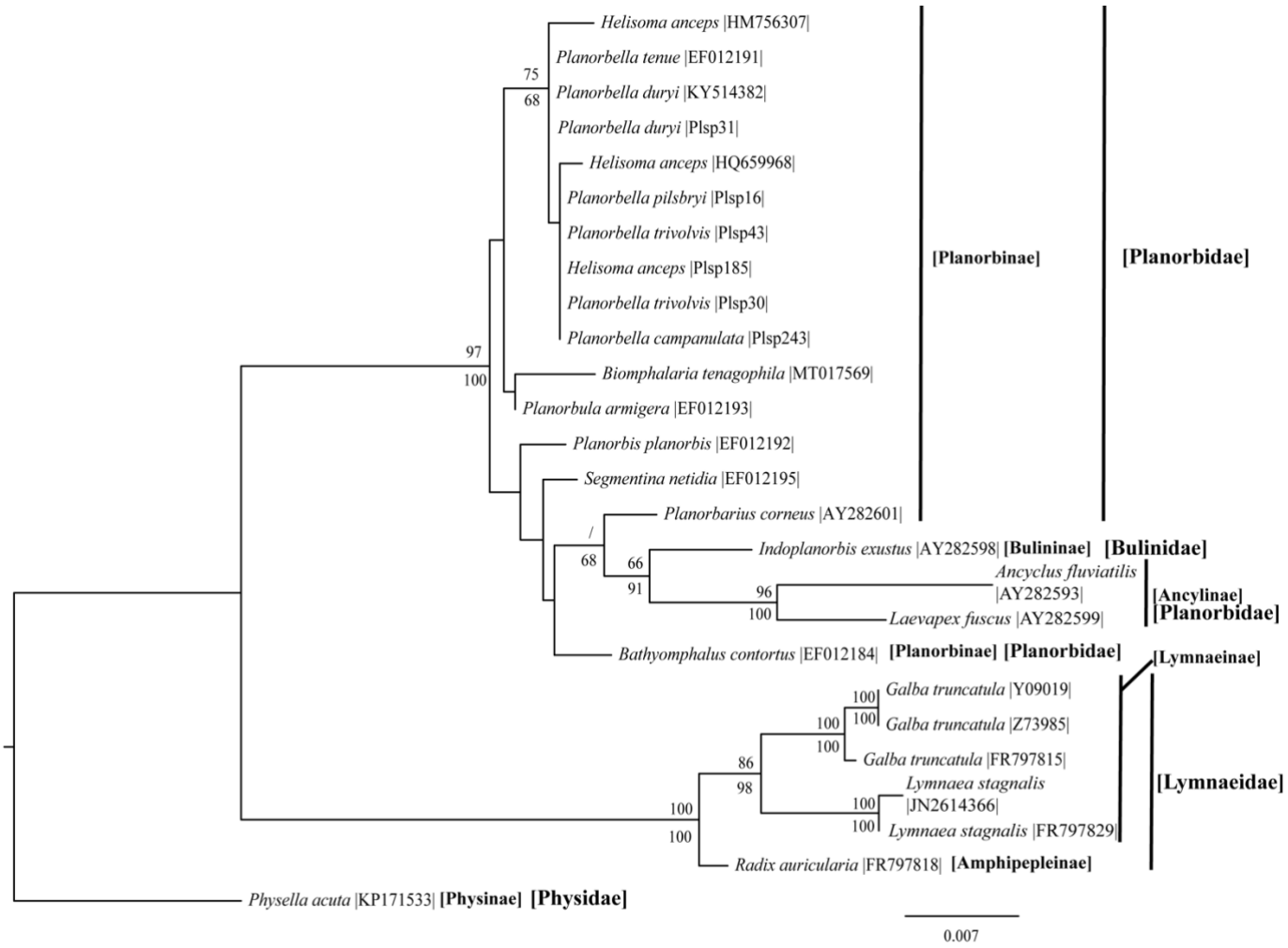


Figure 2.2 Phylogenetic relationship among Hygrophila species in three families based on nuclear rRNA repeat sequence *18S*. Sequences were analyzed using Bayesian inference (BI) and maximum likelihood (ML) analyses, with *Physella acuta* as an outgroup. The Bayesian tree is shown. Numbers above each node are maximum likelihood bootstrap values, while Bayesian posterior probabilities are displayed under each node (showing values over 60%).

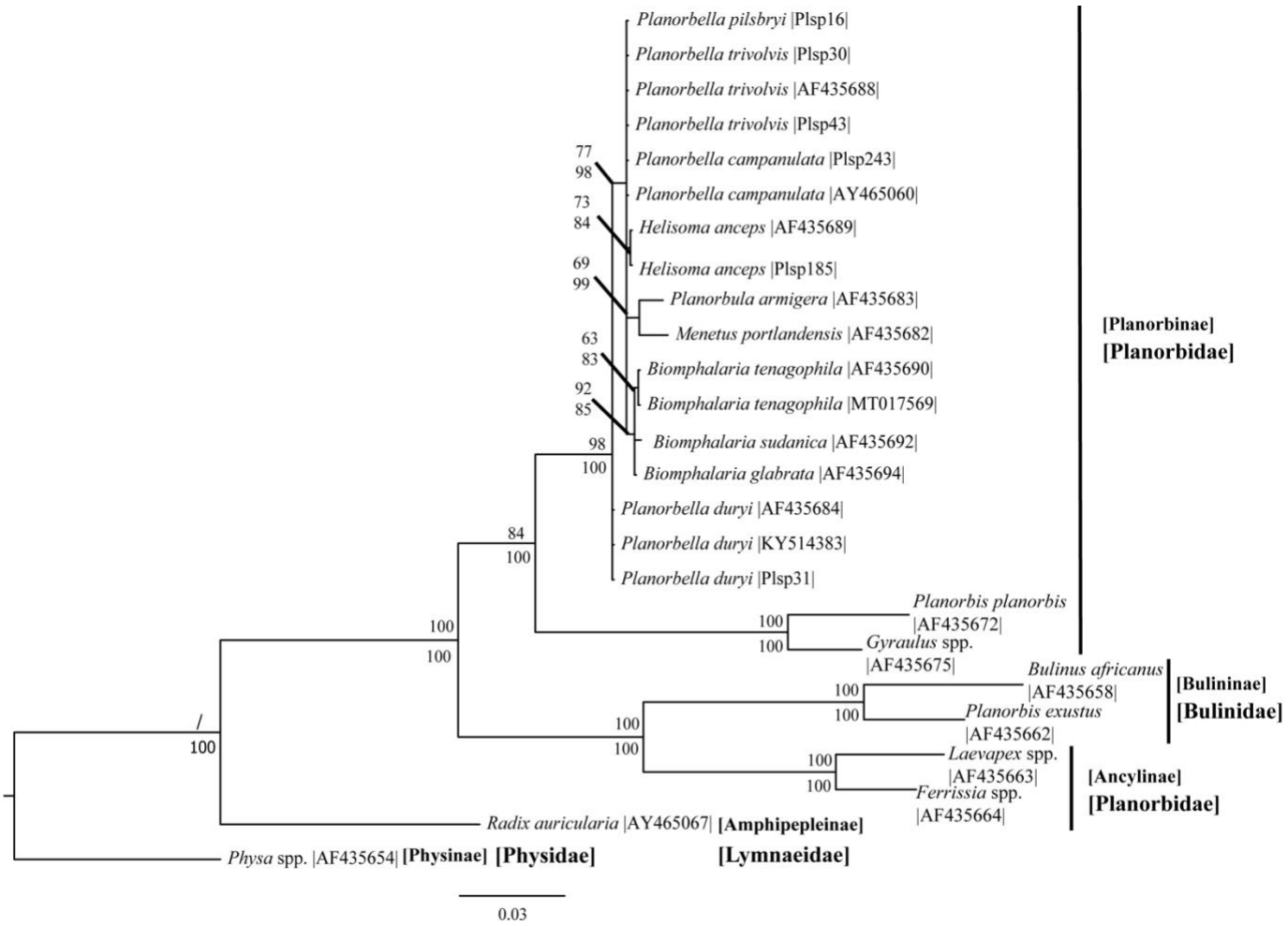


Figure 2.3 Phylogenetic relationship among Hygrophila species in three families based on nuclear rRNA gene *28S*. Sequences were analyzed using Bayesian inference (BI) and maximum likelihood (ML) analyses, with *Physella acuta* as an outgroup. The Bayesian tree is shown for increased branching visibility. Numbers above each node are maximum likelihood bootstrap values, while Bayesian posterior probabilities are displayed under each node (showing values over 60%).

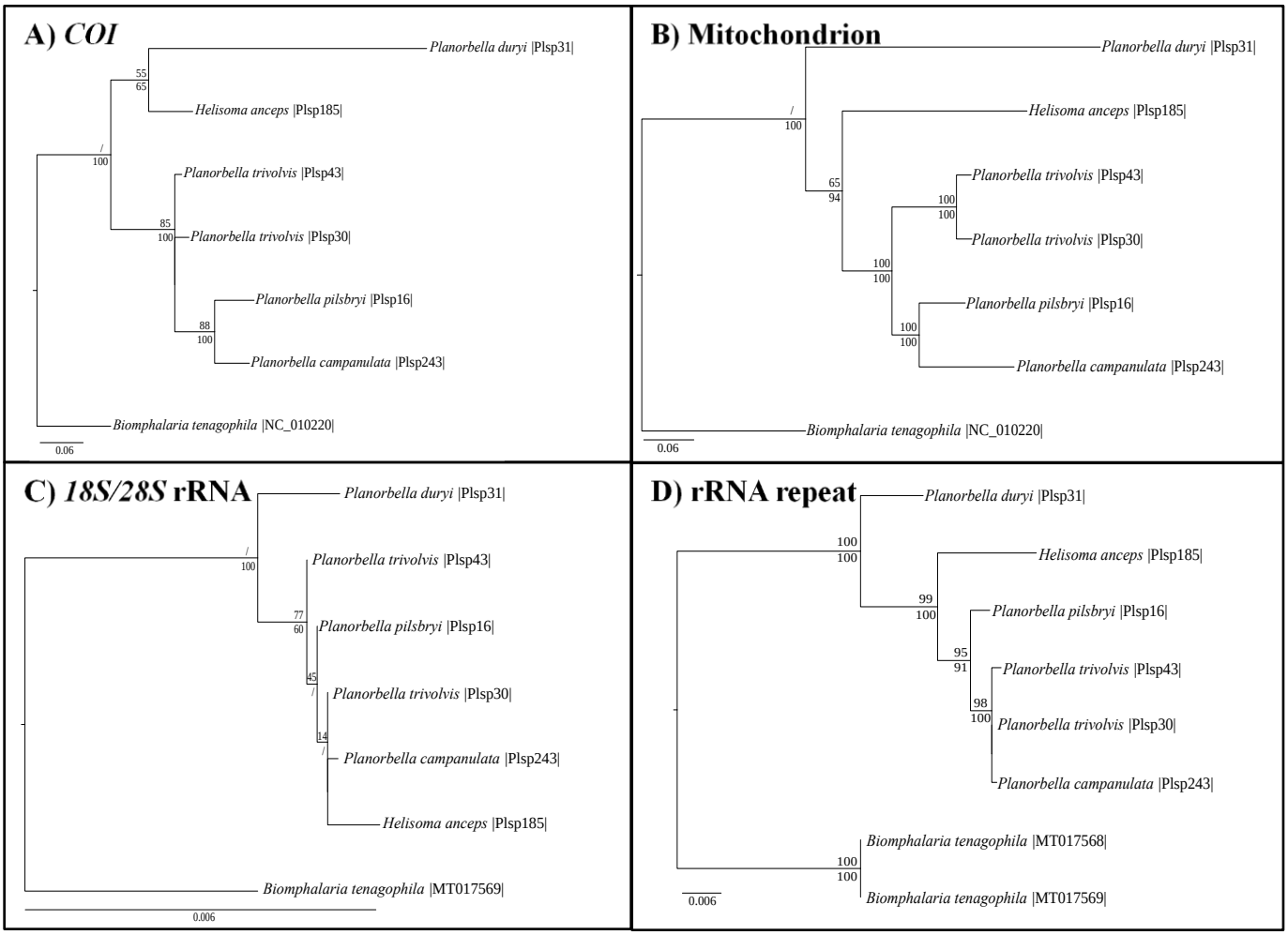


Figure 2.4 Phylogenetic relationships among planorbid species in three genera based on **A)** mitochondrial *COI* sequences*;* **B)** complete mitogenome coding region; **C)** concatenated rRNA *18S-28S* sequences; and **D)** complete nuclear rRNA repeat coding region. Phylogenetic trees were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) analyses with *Biomphalaria tenagophila* as an outgroup. The ML trees are shown. Numbers above each node are maximum likelihood bootstrap values, while Bayesian posterior probabilities are displayed under each node.

Thesis Conclusion

Using integrative taxonomy, I investigated the species diversity of *Planorbella* and *Helisoma* in Manitoba. In Chapter 1, I tested whether cryptic species and phenotypic variation are affecting estimates of species diversity in freshwater snails. I compared identifications of five species based on shell morphology, phylogenetic analysis of the partial *COI* gene, and geometric morphometric analysis of 12 landmarks on the shell. I estimated that my samples included a total of five species when using the shell-based keys and the number of clades present in phylogenetic analysis. Four species of *Planorbella/Helisoma* were recovered while field collecting in Manitoba, but also Minnesota and Oklahoma, and one clade included lab strains maintained at the University of Manitoba and the University of Oklahoma. In contrast, geometric morphometric analysis of 12 landmarks on the shell showed that only two species were morphologically distinguishable, while the other three species had more overlap, indicating how misidentifications could occur among these species due to phenotypic variation and similarity among species at different stages in their development (larval vs adult). The results of the geometric morphometric analysis corroborate previous reports of confusion in identification and taxonomy among these snail species. Further, this analysis also helps to explain why most of the clades in the *COI* tree included several putative species when I incorporated DNA sequences from GenBank. For example, *Planorbella trivolvis* and *Helisoma anceps,* occurred in several of the clades in the *COI* phylogeny based on GenBank identifications, suggesting that identification methods that solely use shell-based traits are not reliable for distinguishing species of *Planorbella* and *Helisoma*. Based on the integrative results, I concluded that there are five valid species represented in this study. This work shows that shell-based identification using descriptive keys can lead to misidentification, as was observed in several cases (multiple putative species from GenBank, my difficulty differentiating *P. trivolvis* vs *P.*

pilsbryi through morphological means). Phylogenetic analysis can facilitate identifications, but this method can be costly and time-consuming, and therefore does not lead to rapid identifications on its own. The results from geometric morphometric analysis with the present landmarks also cannot be used in isolation, and in fact show the importance of including DNA sequences as vouchers. However, further genetic analysis is required to determine if variation at the *COI* locus is reflective of genome-wide relationships within these taxa.

In Chapter 2, I used additional genetic loci to reconstruct the evolutionary relationships of *Planorbella* and *Helisoma* spp within the superorder Hygrophila. I reconstructed two phylogenies for nuclear rRNA genes (*18S* and *28S*) and another phylogeny using the entire mitogenome. I predicted that phylogenies reconstructed from the *COI* barcode region (Chapter 1) and the entire mitogenome would have congruent topologies, given that all mitochondrial DNA loci are linked and there is no recombination. Additionally, I predicted that the topologies of the nuclear rRNA repeat phylogeny would be congruent to the *COI* barcode region phylogeny suggesting no evidence of introgression. I found that the topology of the mitochondrial gene tree was congruent with the mitogenome phylogeny. However, the topologies of the rRNA repeat trees were not completely congruent with the mitochondrial trees suggesting that introgression may occur among some of the taxa in this study.

In all, I generated 56 novel *COI* sequences, six complete mitogenome sequences, and six rRNA repeat sequences. By depositing the mitochondrial and nuclear sequences in a public database, this study increases the number of taxa, geographic sampling locations, and the number of loci available for future studies of planorbid evolutionary ecology. Further, these sequences could help resolve the systematics of this group at both shallow and deeper nodes. These sequences

will help ground ecological investigations in parasite-host relationships and conservation by grounding *Planorbella* and *Helisoma* in a clearer taxonomic framework.

This study not only contributes genetic vouchers, but also morphological vouchers, as all sequenced individuals have a digital image of their shell, and a morphological voucher (preserved tissues). Thus, the vouchers from this study can be used by other researchers to make comparisons and species identifications. For example, future work could compare the specimens collected for this study to collections that include paratypes and holotypes of these species, to better understand extant species diversity in this group. Through the collection and analysis of both morphological and molecular vouchers, I helped to increase species-level resolution for taxa that has been previously overlooked. This knowledge is valuable because accurate identifications of snail species are required for a wide range of further ecological, biological, and evolutionary studies. I hope these results will encourage rigorous analysis in similar questions of species delimitation in freshwater snails, especially in studies investigating whether or not cryptic species are present in a population.