HIDE AND SEEK: INVESTIGATION OF BIODIVERSITY AND DISTRIBUTION OF CRYPTIC DUCKWEED SPECIES IN ALBERTA

KANISHKA MADUSHANKA SENEVIRATHNA
Bachelor of Science, University of Peradeniya Sri Lanka, 2017

A thesis
submitted to the School of Graduate Studies
of the University of Lethbridge
in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE
in
BIOLOGY

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

HIDE AND SEEK: INVESTIGATION OF BIODIVERSITY AND DISTRIBUTION OF CRYPTIC DUCKWEED SPECIES IN ALBERTA

KANISHKA MADUSHANKA SENEVIRATHNA

Date of Defense: November 24, 2020

Dr. R. Laird	Associate Professor	Ph.D.
Dr. T. Burg Thesis Co-Supervisor	Professor	Ph.D.
Dr. J. McCune Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. S. Wiseman Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. T. Russell Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Abstract

This thesis focused on the development of molecular tools to characterize the biogeographical distribution of two cryptic *Lemna* species suspected to be present in Alberta, *Lemna turionifera* and *L. minor*. I developed eight species-specific primers for five morphologically similar *Lemna* species. Our preliminary data show *L. minor* is present in multiple wetlands in southern Alberta. Subsequently, I used genotyping-by-sequencing technology to study the population diversity of these cryptic *Lemna* species. Based on analyses of 103 samples at 16,007 single nucleotide polymorphism loci, I provide strong support for genetic diversity existing between *L. minor* and *L. turionifera* growing in Alberta. I found significant population differentiation in *L. minor* but no genetically distinct populations within *L. turionifera*. In summary, DNA barcoding allowed us to exclusively identify unknown specimens to species level. This study provides an example of using next generation sequencing methods to identify significant fine-scale population genetic variations.

Acknowledgements

First and foremost, I take the immense pleasure in thanking my co-supervisors, Associate Professor Robert Laird and Professor Theresa Burg, Department of Biological Sciences, Faculty of Arts & Science, University of Lethbridge for the tremendous support given throughout my research project with their knowledge, guidance and for providing funding for lab work. Their motivation and patience throughout my research period, in field work and laboratory work, are most appreciated and simply without their supervision and constant guidance this research would not have been a success.

Next, I wish to give away my heartfelt gratitude to all the thesis examination committee members: Assistant Professor Jenny McCune, Associate Professor Steve Wiseman of Department of Biological Sciences, Faculty of Arts & Science, University of Lethbridge for always sharing their knowledge and wisdom. They have always been a great inspiration. They were an integral part of me being able to make it to the end.

Also, I would like to convey my sincere gratitude to Ms. Varina Crisfield, former vascular plant taxonomist at the Alberta Biodiversity Monitoring Institute, Dr. Patrick Barks, Epicentre, Médicins Sans Frontières in Paris, and Ms. Sarah Bogart for providing *Lemna* samples and sharing expertise and valuable guidance throughout the projects.

Moreover, I would like to acknowledge Mr. Ian Gazeley, a master's student in Dr. T. Burg's Lab for always sharing his experience during field work and lab work. His constant helping hand has always been a great support. Without the help that he gave me in developing field work and the emotional support he gave me; I wouldn't have made it.

I express my gratitude to Dr. Matthew Bogard and the members of the Bogard Lab for helping me with the water quality study and allowing me to use his laboratory facilities and laboratory equipment during this research.

Further, I take this opportunity to thank my laboratory partners for their support and companionship. I thank Dr. Brendan Graham, postdoctoral fellow of the Burg Lab, for giving his ideas and comments to improve the analyses and the manuscript. I am much obliged to Ms. Dilini Abeyrama, Ph.D. candidate of the Burg Lab, for her immense support. She supported me in numerous ways by helping me with data analyses and cheering me up when the times were difficult and for been with me to overcome challenges and hardships that I faced throughout my research project. I would like to pay my sincere acknowledgment to Ms. Suzanne Chmilar, master's student of the Laird Lab, for always extending a helping hand during laboratory work. She always worked with enthusiasm to provide me with the required laboratory chemicals, tools, and equipment, and this greatly helped me commence my work on time. Also, I would like to thank Jed Immanuel Lloren, master's student of the McCune Lab, for supporting me with GIS based activities, developing maps and helping me with data analyses. I sincerely appreciate their help with the project.

I would also like to acknowledge with much appreciation, Ms. Amanda Carpenter, Mr. Austin Paiha, Ms. Priyanka Dutt, and Ms. Danika Schramm and all my other colleagues who supported me throughout my research project. They have been a great strength from the beginning to the end of my research project, constantly encouraging its success.

Also, my sincere thanks go to Mr. Barry Rice and Mr. John Crellin who granted me permission to use their photos in my thesis, and Prof. Nele Horemans, Head of

Research unit, Biosphere Impact Studies (BIS) for providing the *Lemna minor* reference genome and annotations file.

I would like to thank Alberta Conservation Association's (ACA) Research Grant program and the Natural Sciences and Engineering Research Council (NSERC) Discovery Grants program for their generous contribution to my project by supporting with funding.

Furthermore, I would also like to acknowledge Dr. Roshan Priyarangana Perera for support with data analysis and helpful comments given to me during the development of this task. I would like to pay my special gratitude to Mr. Sachinthani Karunarathna for cooperating with me in advising and proof reading.

At last but not least, I would like to convey a big thank to both my-families; my family by blood who live in Sri Lanka and my family-like friends who live in Lethbridge, for always being my pillars of support and for standing by my side, giving me all the strength and courage to complete this project successfully with an unfailing interest. Their love and patience always held me strong when the times were rough and full of challenges.

Table of Contents

List of Tables

List of Figures

List of abbreviations, acronyms, and symbols

ABMI	Alberta Biodiversity Monitoring Institute
˚C	Degrees Celsius
1x	One times
BIC	Bayesian Information Criterion
bp	Base pair
BWA tool	Burrows-Wheeler Alignment tool
Ca	Dissolved Calcium
$CaCO_3$	Calcium Carbonate
Ch-a	Chlorophyll-a content
COBL	Consortium for the Barcode of Life
CPCC	Canadian Phycological Culture Centre
DAPC	Discriminant Analysis of Principal Components
DNA	Deoxyribonucleic acid
DOC	Dissolved Organic Carbon
ESU	Evolutionary Significant Unit
F	Dissolved Fluoride
GBS	Genotype by Sequencing
GWAS	Genome-Wide Association Studies
HCO_3	Bicarbonate
K	Genetic cluster
Ma	Million Years
Mbp	Mega base pair
$MgCl_2$	Magnesium chloride
mM	Millimolar
MU	Management Unit
Na	Sodium
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NH_3-N	Ammonia
NO_2-N	Nitrite
NO_3+NO_2-N	Nitrate+Nitrite
PAST	Paleontological Statistics
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RDSC	Rutgers Duckweed Stock Cooperative
SCon	Specific Conductance
SDT	Secchi Disk Transparency
SNP	Single Nucleotide Polymorphisms
TDP	Total Dissolved Phosphorus
TDS	Total Dissolved Solid
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
TN:TP	Total Nitrogen:Total Phosphorus Ratio
TP	Total Phosphorus

VCF file	Variant Call Format file
μL	Microliter
μM	Micromolar

Chapter One
General Introduction

1.1 Thesis organization

This thesis reports the results of a series of studies that investigated the genetic properties of two cryptic duckweed species in Alberta, Canada: *Lemna minor* L. and *L. turionifera* Landolt. Specifically, while the Alberta Biodiversity Monitoring Institute (ABMI) indicates the presence of two *Lemna* species in Alberta, *L. trisulca* and *L. turionifera*, (ABMI, 2020) Preliminary DNA barcoding data indicate a third species, *L. minor*, which bears a very close morphological resemblance to *L. turionifera*, is also present (Barks et al., 2018). In 2014, ABMI stated that *L. minor* was more common than *L. turionifera* in Alberta (i.e., *L. minor* identified at 210 sites versus *L. turionifera* at only four sites). However, current ABMI records (ABMI, 2020) no longer recognize the presence of *L. minor* in the province; their most recent data report (26th October 2020) lists *L. turionifera* at 500 sites and no sites with *L. minor*. It is unknown whether *L. minor* is rare and/or geographically restricted within Alberta or simply 'camouflaged' by its resemblance to *L. turionifera*.

To answer questions related to the biodiversity of *Lemna* in Alberta, it is necessary to use genetic data. Molecular systematic approaches are a common tool to ascertain accurate intra- and inter-species diagnoses,especially between cryptic species (Xiao et al., 2010, Wang et al., 2010, Abdelaziz et al., 2011). Molecular tools, specifically DNA barcoding, have improved our ability to detect cryptic species (Hollingsworth et al., 2011). Moreover, DNA barcoding techniques provide a universal framework to identify

species. They allow us to exclusively identify an unknown specimen to species level using a short section of DNA from a specific gene or genes (Hebert et al., 2003).

There is a need for the development of molecular techniques to identify *Lemna* species. To understand the importance of duckweed in Alberta's ecosystems and their economic and recreational benefits, it is clearly necessary to know what species are present, but even this basic question is in doubt. Thus, in the present study, I developed molecular tools for the rapid identification of cryptic duckweed species to determine the biogeographical distribution of *L. turionifera* and *L. minor* in Alberta. Hence, this thesis focuses on the following three objectives:

(1) Development of molecular tools to identify cryptic *Lemna* species suspected to be present in Alberta, *L. turionifera* and *L. minor*.

(2) Analyses of the biogeographical distribution of the two cryptic *Lemna* species suspected to be present in Alberta, *L. turionifera*, and *L. minor;* and,

(3) Assessment of population level and species level genetic diversity of *L. turionifera* and *L. minor* throughout Alberta.

In this chapter, I provide background information by reviewing cryptic species and DNA barcoding. I then give a brief introduction on gene flow, population genetics, next generation sequencing (NGS) and the natural history of my study species. I used barcoding to answer objectives 1 and 2. I used Population genetics and NGS to answer objective 3. Here I study genetic differences within and among populations and investigate variation in the allele frequencies in populations over space and time.

Chapter 2 focuses on determining the presence of *L. turionifera* and *L. minor* in Alberta, and their biogeographical distribution. Moreover, Chapter 2 elaborates on the development of molecular tools (species-specific PCR primers) for the rapid

identification of *Lemna* species and the use of these PCR primers to distinguish five morphologically similar duckweed species.

Chapter 3 focuses on the genetic diversity and population structure of *L. minor* and *L. turionifera* and their evolutionary and taxonomical relationships. In order to perform population diversity studies, I used Genotyping by Sequencing (GBS), which allows the discovery of single nucleotide polymorphisms (SNPs) in order to identify fine-scale population differences.

The final chapter provides a general discussion and synthesis about the studies, describes and infers the importance of my findings, details limitations, and explains new insights that arise as a result of my study.

1.2 Biodiversity and cryptic species

Climate change, habitat destruction and other anthropogenic disturbances of natural ecosystems are triggering extinctions of species (Thomas et al., 2004)(Van De Wiel et al., 2009). The extraordinary rates of species extinction have focused high attention on biodiversity (Arponen, 2012)(Tokeshi and Arakaki, 2012). Many species and their taxonomic relationships remain understudied. Therefore, it is important to prioritize efforts to identify and describe biodiversity (Arponen, 2012)(Tokeshi and Arakaki, 2012).

Both regional and global biodiversity may be underestimated because of the presence of "cryptic species", species that are morphologically similar but genetically different (Chenuil et al., 2019). The term "cryptic species" can be defined in several ways. Some scientists consider the term cryptic species to be the equivalent of "sibling species" (Sáez and Lozano, 2005, Bickford et al., 2007), but some assert that the term "sibling species" implies more recent common ancestry than "cryptic". Since there is no universal

agreement on these terms, in this thesis I use “cryptic species” if two or more species are or have been catalogued as a single species or they cannot be easily identified using morphological characteristics.

Cryptic species make it difficult to clarify taxonomic relationships, an issue that scientists have struggled with for more than three centuries (Bickford et al., 2007). After the introduction of the Linnaean classification system, most species descriptions adopted the morphological or typological species concept (Winker, 2005). However, since the Modern Synthesis in the mid-1900s, we mainly used the Evolutionary Species Concept or the Biological Species Concept. In the Biological Species Concept, a group of organisms that can interbreed and produce fertile progeny are considered a “species” (Wright, 1940, Dobzhansky, 1950, Mayr, 1999). In the Evolutionary Species Concept, a species is a group of organisms whose members are descended from a common ancestor and who all have a unique evolutionary role and tendencies (De Queiroz, 2007 #290;Simpson, 1951 #377). It is highly important to understand species concepts as it is vital to identify species in an ecosystem separately. This clear identification of species allows us to recognize fundamental units which govern the biodiversity in that ecosystem and to focus on their conservation. However, this is difficult as cryptic species can lead to either conceptual or terminological problems. With the development of new DNA technology and molecular tools, such as DNA barcoding, distinguishing cryptic species is less of an issue than in the past, and our understanding of cryptic species has increased as a result. The discovery of cryptic species has direct effects on biogeography, conservation strategies, and evolutionary theory (Bickford et al., 2007). For example, knowledge of cryptic species helps us to include cryptic species in conservation plans, which in turn helps biodiversity protection and management. Lack of knowledge about cryptic species

could lead to taxonomic uncertainties (Cardoso et al., 2011, Delić et al., 2017). Therefore, correct identification of cryptic species focuses the attention of scientists on accurately evaluating their value for conservation (Pearman, 2001).

1.3 DNA barcoding

DNA barcoding allows the identification of unknown specimens to the species level using short, standardized DNA sequences (400-800 bp) (Hebert et al., 2003). Since its inception in 2003, 8.5 million barcoding sequences have been completed and published online (Ratnasingham and Hebert, 2007, BOLDSYSTEMS, 2020). These data are widely used in biodiversity studies, species identification, population genetics and forensic analyses.

DNA barcoding is a universal framework for identifying species. Barcoding sequences need to have a distinct threshold to differentiate inter- and intra-specific genetic variation (Lahaye et al., 2008). For a DNA fragment to function as a barcode, it must simultaneously contain enough variation across distinct species, be short enough to sequence in a single reaction (minimalism), include conserved regions (standardisation) that can be used to build universal primers and be informative for identification (scalability) (Sass et al., 2007).

Defining a single, ubiquitous barcode for every life form on earth is challenging. The mitochondrial gene *COI*, which codes for subunit 1 of cytochrome oxidase, is the most frequently used barcoding gene for animals (Hebert et al., 2003). The low substitution rates of the mitochondrial DNA in plants (Chase et al., 2005) required us to use different candidate barcoding genes (Kress et al., 2005, Presting, 2006, Chase et al., 2007, Hollingsworth et al., 2011). Many of these markers are contained in plastid

genomes. Some include coding genes such as *matK* (maturase K), *rbcL* (ribulose bisphosphate carboxylase large chain), *rpoB* (beta subunit of the plastid-encoded RNA polymerase), and *rpoC1* (DNA-directed RNA polymerase subunit beta) while others are noncoding spacers like *atpF–atpH* (ATP synthase subunit c), , *trnH–psbA* (Intergenic spacer (photosystem II protein D1), and *psbK–psbI* (photosystem II reaction center protein K). Another commonly sequenced region for land plant phylogenetic studies is nuclear ribosomal *ITS*, the internal transcribed spacers, of the large subunit of ribosomal DNA (Chase et al., 2007, CBOL Plant Working Group et al., 2009)All of these plant barcoding markers have their strengths and weaknesses due to trade-offs among scalability, minimalism, and standardisation. For example, *rbcL* has high universality (the ability to use as a common marker across the land plants), but less discriminating ability (the ability to distinguish two or more species); *trnH–psbA* and *matK* consist of higher resolution power (the ability to differentiate between samples) but result in lower-quality sequences requiring more time editing sequence traces (CBOL Plant Working Group et al., 2009). Even though *ITS* is used for most phylogenetic and population studies, it has some limitations. *ITS* has less variability at the species level (Baldwin et al., 1995, Álvarez and Wendel, 2003). Also, sometimes -*ITS* amplifies the DNA from contaminating fungal species (Kress et al., 2005). Therefore, researchers have focused on using a "two-locus" approach combining *matK* and *rbcL* to design a universal plant barcode. This approach facilitates building a combined plant barcode database to be used for population genetics, taxonomy, conservation, and a large number of other objectives that need identification of plant material (CBOL Plant Working Group et al., 2009).

Since its initial publication in 2009, the Consortium for the Barcode of Life (COBL) Plant Working Group paper has been cited approximately 1400 times (26th

October 2020) showing the impact of this paper on the scientific community. With the new knowledge about plant genomes and genetics, there will be more robust and unique DNA barcodes in the future.

1.4 Gene flow

When studying population genetics, taxonomy, or conservation, it is necessary to understand gene flow. Gene flow is an umbrella term used for all the phenomena that result in the movement of genes from one population to another. Gene flow can result from the movement of individuals, gametes, or even movement of extra-nuclear segments of DNA, such as mitochondria, plasmids, and viruses (Slarkin, 1985, Petit and Excoffier, 2009). Gene flow can also influence evolution by distributing new alleles and gene combinations throughout a population or a species' range (Slatkin, 1987).

Two main methods can be applied to estimate the amount of gene flow: direct and indirect methods. Direct methods are used to screen current gene flow; e.g., direct observation of the movements of individuals between populations (Wright, 1943). Typically, the degree of gene flow is higher among – populations that are near each other geographically. Similarly, gene flow is assumed to be low or negligible among populations that are separated by long distances (Ereshefsky, 1992, Villablanca, 1994). In contrast, indirect methods use allele frequencies to quantify levels and patterns of gene flow with the aid of molecular techniques (Slatkin, 1987). Applying this concept, we can carry out mark-recapture methodology to measure the number of migrating individuals in a study area via indirect methods by using spatial dispersion of gene frequencies to infer gene flow patterns. Here allele frequencies are used to quantify levels and patterns of gene flow with the aid of molecular techniques (Slatkin, 1987).

1.5 Next-generation sequencing

The field of population biology recently has undergone significant advances as a result of the development of next-generation sequencing (NGS). NGS is a high-throughput sequencing technology that generates millions of sequences simultaneously (Elshire et al., 2011). This technology is based on massively parallel sequencing and advanced imaging strategies (e.g., fluorescently labelled nucleotides combined by a polymerase) (Shendure and Ji, 2008). Genotyping by Sequencing (GBS) makes an excellent platform to perform genotyping studies, such as in the analysis of population dynamics (Morris et al., 2013) and the construction of high-density genetic maps (International Cassava Genetic Map Consortium, 2015, Soto et al., 2015). GBS is a robust and powerful tool that allows researchers to screen and genotype thousands of loci for single nucleotide polymorphisms (SNPs). In plant genomics and plant breeding, GBS is used for the assessment of breeding values (expected phenotypic value of an individual's offspring), in genomic selection (Crossa et al., 2013, Glaubitz et al., 2014, He et al., 2014), and in genome-wide association studies (GWAS) (Romay et al., 2013). In this study, I use NGS to study the population genetics of *Lemna* and assess population differentiation within and among *L. minor* and *L. turionifera* (Figure 1.1).

1.6 Study Species

The duckweed subfamily (Araceae: Lemnoideae) comprises a group of perennial aquatic plants (Sree et al., 2016). They are floating or submersed, have extremely reduced morphology and a small number of organs (Landolt, 1986). Moreover, they have a cosmopolitan distribution. These -characteristics lead to difficulties in their taxonomy and

systematics as they are difficult to identify based on morphological features (Landolt, 1986, Les et al., 2002).

Duckweeds originated approximately140 Ma in the early Cretaceous (Nauheimer et al., 2012). Currently the duckweed subfamily contains 36 species (Bog et al., 2020) which belong to five genera (*Landoltia* Les & D. J. Crawford, *Lemna* L., *Spirodela* Schleid, *Wolffia* Horkel ex Schleid, and *Wolffiella* Hegelm) (Azer, 2013, Sree et al., 2016, Bog et al., 2019). The genera *Spirodela, Landoltia*, and *Lemna* are more closely related, while *Wolffia* and *Wolffiella* are sister genera. The newly named genus *Landoltia* has only one species (*L. punctata*) that is morphologically intermediate between *Lemna* and *Spirodela* (Azer, 2013, Sree et al., 2016).

Duckweeds play important ecological roles as a vital food source for numerous species such as wood ducks (*Aix sponsa*), mallard ducks (*Anas platyrhynchos*), Canada geese (*Branta canadensis*), and beavers (*Castor* spp.) (Van der Spiegel et al., 2013). They are economically useful in environmental biotechnology as a feedstock for biofuels (Cheng and Stomp, 2009, Van De Wiel et al., 2009, Xu et al., 2012). Further, duckweeds are used for biomonitoring, phytoremediation, phytohormone biosynthesis, and photosynthesis studies (Bog et al., 2010, Ziegler et al., 2015).

The genus *Lemna* (Greek name of a water plant) includes species with one root per frond. Fronds are free-floating or submersed. The fronds are lanceolate-ovate, flat or gibbous, and 1 to 15 mm in size (Flora of North America, 2020). Their margins are entire or denticulate and the upper surfaces sometimes have small conic papillae along veins. In addition, at the base of the frond are two reproductive pouches on the lateral sides called ‘meristematic pockets’, where daughter fronds and flowers originate (Landolt, 1986). *Lemna* species are widely used as model organisms in biochemistry, evolution, and

ecology (Böttcher and Schroll, 2007, Aliferis et al., 2009, Kielak et al., 2011, Laird and Barks, 2018), and are candidates for quantitative analytical studies in toxicology (Jansen et al., 1996, Cayuela et al., 2007).

The genus *Lemna* contains thirteen species including *L. turionifera* and *L. minor*. *Lemna turionifera* forms small, olive to brown, rootless turions, which are overwintering vegetative plantlets that sink to bottom of the water bodies (Landolt, 1975, Sree et al., 2016). They have root systems < 15 cm and the root tip is mostly rounded (Flora of North America, 2020). *Lemna turionifera* have single to few fronds. Fronds are free-floating and coherent in groups. Fronds are 1-4 mm in width, 1-1.5 times as long as wide, obovate, scarcely gibbous, and have a flat leaf blade. They have papillae, which are distinct on midline of the upper surface (apical papilla scarcely larger than others) (Landolt, 1975, Flora of North America, 2020). In addition, the lower surface is often red in colour (more intensely so than on the upper surface). Flowering in *L. turionifera* is rare as they mainly reproduce through clonal reproduction. This genus inhabits mesotrophic to eutrophic water bodies. They grow mainly in still waters in temperate regions from sea level up to 3700 m (Flora of North America, 2020) (Table 1.1).

Lemna minor has roots with rounded tips which can grow up to 15 cm in length. They contain single to few free-floating fronds. The fronds are coherent in groups and obovate. *Lemna minor* fronds often have a gibbous shape, 1-8 mm wide, and 1.2-3 times as long as wide. *L. minor* is distinguishable from *L. turionifera* due to a lack reddish on lower surface or at least much less so than on upper (Landolt, 1975, Flora of North America, 2020). Flowering is rare and occurs in late spring to early fall. They inhabit mesotrophic to eutrophic, quiet water bodies. Also, they are more frequent in cool to

temperate regions with relatively mild winters. This species is frequently seen at 0 to 2000 m elevation (Flora of North America, 2020).

There are few morphological characters used to distinguish *L. turionifera* and *L. minor*. Though Dudley (1987) found that strains of both *L. minor* and *L. turionifera* (identified as such by Landolt) can produce turions, while others state that *L. turionifera* produces turions and *L. minor* does not (Sinkevičienė, 2011, Halder and Venu, 2012). In addition, *L. turionifera* has more distinct reddish anthocyanin pigmentation, particularly on the lower surface, but according to Chester et al. (2007) reddish anthocyanin pigmentation is not always present in *L. turionifera*, so we must rely on the row of papules along the midline of dorsal surface that are more distinct than those in *L. minor* (Landolt, 1986, Flora of North America, 2020). Still, this is not a very distinct character and can be difficult to see (Chester et al., 2007). Therefore, it appears that these main diagnostic differences between *L. minor* and *L. turionifera* are weak (Dudley, 1987). The distinguishing characteristics between *L. minor* and *L. turionifera* mostly vary continuously rather than discretely, and there are published exceptions to these characteristics (Landolt, 1986, Dudley, 1987, Chester et al., 2007, Flora of North America, 2020). Therefore, it is doubtful that these are truly distinguishing characteristics. In conclusion, genotyping is necessary to settle the conflict. This indicates the importance of molecular tools for accurate identification of *Lemna* species.

Table 1.1 General comparison of habitat diversity and distribution of five morphologically similar duckweed species used in this study (Hilty, 2019, GO BOTANY, 2020, CABI, 2020, Flora of North America, 2020). Photo credits for *L. minor L. turionifera* and *L. gibba*; Kanishka Senevirathna, for *L. minuta*; ©Barry Rice/Sarracenia.com/via CalPhotos - CC BY-NC-SA 3.0, and for *S. polyrhiza*; John Crellin/ Floralimages Distribution maps from Flora of North America (2020).

	L. minor	*L. turionifera*	*L. gibba*	*L. minuta*	*S. polyrhiza*
					
Habitat	Still waters, Mesotrophic to - eutrophic, moist climates, being rare or absent in more arid regions	Similar habitats to *L. minor*, Mesotrophic to - eutrophic, quiet waters, minimum water depth of 1 foot (30 cm) temperate regions	Still or slow-flowing water, mud or damp rocks, Eutrophic, temperate regions with mild winter	Stagnant or slow-moving, calm, eutrophic water, relatively hard-water conditions and high phosphorus levels, temperate to subtropical regions with relatively mild winters	Stagnant or slow-moving water, mildly acidic to alkaline
pH	6.5 - 8	Acceptable 5- 9, optimal 6.5 - 7.5	7.4 - 9.8	7.75	7 - 7.9
Temperatures	6 - 33 °C	20 °C - 28 °C			
Flowering	Rare (late spring - early fall)	Occasional (summer)	Rather frequent (spring - fall)	Very rare (late spring - early fall)	Very rare (early summer- early fall)
Elevation	0 - 2000 m	0 - 3700 m	0 -1900 m	0 - 2600 m	0 - 2500 m
World Distribution	North America, Europe and western Asia	Mainly found in North America, occasionally in Eurasia	Mediterranean climate (dry, mild) and tropical mountains, except Australia	Mediterranean climate, North America and South America	North America, South America, Eurasia, and other parts of the world, nearly worldwide.
					

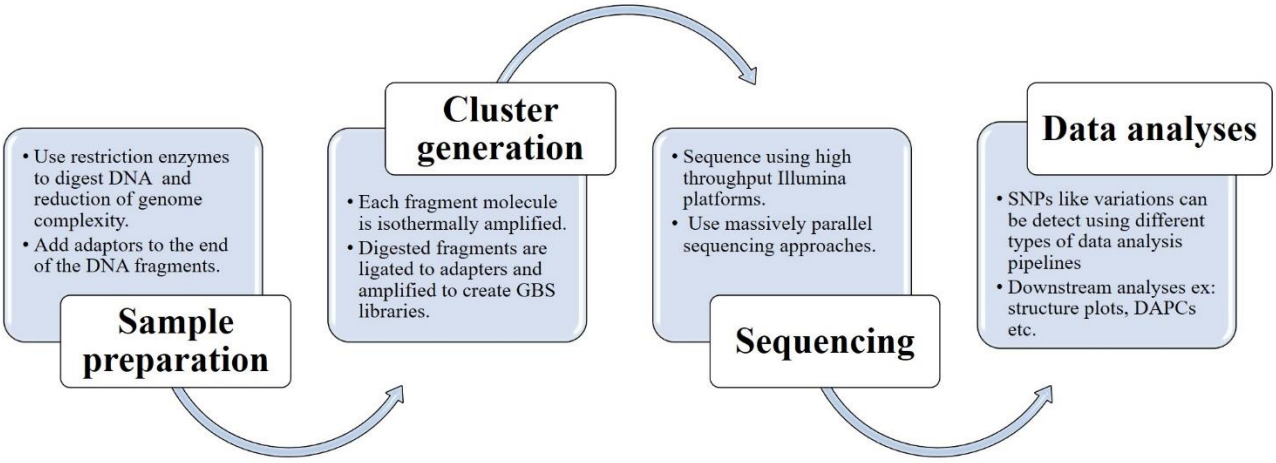


Figure 1.1 General representation of steps of the genotyping-by-sequencing (GBS) protocol.

Chapter Two

HIDE AND SEEK: USING MOLECULAR TOOLS TO INVESTIGATE BIODIVERSITY AND DISTRIBUTION OF CRYPTIC DUCKWEED SPECIES IN ALBERTA

2.1 Abstract

Both regional and global biodiversity may be underestimated because of the presence of cryptic species: species that are morphologically similar, but genetically distinct. Due to their cryptic nature, two duckweed species (*Lemna minor* and *L. turionifera*) have been the focus of discussion in the scientific community. Specifically, while the Alberta Biodiversity Monitoring Institute indicates the presence of two *Lemna* species in Alberta, *L. trisulca* and *L. turionifera*, my DNA barcoding data indicate a third species, *L. minor*, is also present. Molecular tools, specifically DNA barcoding, have improved our ability to detect cryptic species. Thus, in the present study, I developed molecular tools for the rapid identification of monomorphic *Lemna* species allowing us to determine the biogeographical distribution of *L. turionifera* and *L. minor* in Alberta. My data suggest the presence of *L. minor* in the southern part of the province.

2.2 Introduction

Biodiversity studies enrich our understanding for estimating the integrity of different ecosystems, their responses to natural and anthropogenic disturbances, and the set of measures required to conserve or re-establish biodiversity. Both regional and global biodiversity may be underestimated because of the presence of "cryptic species", species that are morphologically identical, but genetically distinct (Chenuil et al., 2019). Molecular tools, specifically DNA barcoding, have improved our ability to identify cryptic species (Hollingsworth et al., 2011). DNA barcoding techniques provide a universal framework to identify species and allow for exclusive identification of unknown specimens to the species level (Lahaye et al., 2008) allowing us to assess biodiversity (Hollingsworth et al., 2011).

Recent climate changes and other anthropogenic events have increased the rate of loss of the world's biodiversity (Midgley et al., 2002). For example, the species richness of some aquatic systems is threatened by global warming and the presence of invasive, exotic species (Van De Wiel et al., 2009). The extraordinary rates of species extinction in aquatic systems have focused attention towards studying aquatic biodiversity (Tokeshi and Arakaki, 2012). Approximately eight million DNA barcodes have been reported from all over the world (Ratnasingham and Hebert, 2007, BOLDSYSTEMS, 2020). Among them, barcodes for aquatic plants have gained a lot of attention due to the significant roles plants play in aquatic ecosystems (Wang et al., 2010). Aquatic plants are vital ecological components of both fresh water and marine ecosystems, maintaining the structure and function of these environments (Lacoul and Freedman, 2006).

Duckweeds (Family Araceae, Subfamily Lemnoideae) are the smallest aquatic monocots. They have extremely simple morphology, being composed of a single frond and zero to several roots, depending on the species (Landolt, 1975, Lemon and Posluszny, 2000). These fast-growing, free-floating plants are widespread throughout freshwater bodies across the world (Keddy, 1976, Landolt, 1986). There are 36 species of Lemnoideae belonging to five genera (*Spirodela* Schleid, *Landoltia*, Les & Crawford, *Lemna* L., *Wolffiella* Hegelm, and *Wolffia* Horkel ex Schleid) (Bog et al., 2020). The genus *Lemna* is found on every continent except Antarctica and is most diverse in North America and Southeast Asia (Landolt, 1975). Recently, species in the genus *Lemna* have attracted attention due to their ecological and economic significance (Xu et al., 2012, Tang et al., 2014, Barks et al., 2018). *Lemna* plays an important ecological role as a fundamental food source for various waterfowl and fish species (Goopy and Murray, 2003) and is economically important for different aspects of environmental biotechnology, such as feedstock and biofuels (Cheng and Stomp, 2009, Xu et al., 2012, Van der Spiegel et al., 2013). Further, these plants are used for biomonitoring, phytoremediation, phytohormone biosynthesis and photosynthesis studies (Bog et al., 2010, Ziegler et al., 2015). Another important aspect of *Lemna* species is their capability of acting as a model organism. *Lemna* species are widely used as a model organism for comparative studies (Böttcher and Schroll, 2007), ecotoxicological studies (Hulsen et al., 2002, Aliferis et al., 2009), and studies of pathogenesis (Zhang et al., 2010). According to Jansen et al. (1996) and Laird and Barks (2018), *Lemna* species are candidate plants for studies based on quantitative analysis because of their simple morphology, rapid and mainly clonal reproduction, and widespread distribution.

Lemna species are widespread across Alberta (VASCAN, 2019, ABMI, 2020, Flora of North America, 2020). They are an important food source for many of Alberta's species that are of interest to hunters, anglers, and trappers (e.g., beavers) (Van der Spiegel et al., 2013). Therefore, it is important to quantify the duckweed species that are present and their biogeographical distribution in the province. To date, the most basic biodiversity question "How many *Lemna* species are there in Alberta?" has not yet been answered conclusively (Barks et al., 2018). According to the Alberta Biodiversity Monitoring Institute (ABMI, 2020) and other databases (VASCAN, 2019, Flora of North America, 2020), Alberta hosts two *Lemna* species: *L. trisulca* and *L. turionifera* (ABMI, 2020). These two species are morphologically distinct and easy to differentiate. However, DNA barcoding data suggest the presence of a third *Lemna* species, *L. minor*, at a single site in southeast Alberta (Crow Indian Lake, Skiff, AB; 49.370 °N, 111.800 °W) (Barks et al., 2018). *Lemna turionifera* and *L. minor* have extremely similar morphologies and are cryptic species (Landolt, 1986, Wang et al., 2010). Morphological studies claim that *L. turionifera* and *L. minor* can be differentiated due to having red pigmentation on the abaxial surface and papules along the midline of the adaxial surface of *L. turionifera* (Chester et al., 2007). In addition, *L. turionifera* produces starchy overwintering buds called "turions" (Chester et al., 2007, Flora of North America, 2020). However, Dudley (1987), Chester et al. (2007), and Wang et al. (2010) reported that these features are also present in *L. minor*; therefore, it is unknown whether *L. minor* is rare and/or geographically restricted within Alberta or simply 'camouflaged' by its resemblance to *L. turionifera*.

In this present study, I developed molecular tools to differentiate between strains of *Lemna* found in Alberta and used these markers to assess the biogeographical distribution of the cryptic species *L. turionifera* and *L. minor*. This will allow us to determine whether *L. minor* is rare in Alberta or simply less known due to its morphological similarity to *L. turionifera*. Development of species-specific primers will allow for rapid screening of large numbers of samples without resorting to expensive and time-consuming DNA sequencing.

2.3 Materials and Methods

2.3.1 Sampling of *Lemna*

To determine the species distribution and assess population genetics of *Lemna* species, I sampled 126 waterbodies throughout Alberta, targeting plants that were either *L. minor* or *L. turionifera*. Sampling was done in collaboration with the Royal Alberta Museum/Alberta Biodiversity Monitoring Institute (ABMI). ABMI provided 57 samples of *Lemna* collected from 2016 to 2018. Another 77 *Lemna* samples were collected at 27 sites in 2013 and initially used in (Barks et al., 2018). In the summer of 2019, I collected another 100 samples from 42 additional sites. Altogether, 234 different *Lemna* samples from 126 sites across the province were used in this study (Supplementary Table 1.1).

In addition, I used seven reference strains originally obtained for the study described in Barks et al. (2018): three from the Rutgers Duckweed Stock Cooperative (RDSC: *L. minor* RDSC 7123, *S. polyrhiza* RDSC 8790, *L. minuta* RDSC 6752) and four

from the Canadian Phycological Culture Centre (CPCC: *L. gibba* CPCC 310, *L. minor* CPCC 490, and CPCC 492, and *L. trisulca* CPCC 399) (Supplementary Table 1.1).

2.3.2 Development of *atpF-atpH* and *psbK-psbI* species-specific primers

I used *atpF-atpH* and *psbK-psbI* sequences from Wang et al. (2010), Tang et al. (2014) and Barks et al. (2018) to create species-specific primers for five duckweed species with varying degrees of morphological similarity, all found in North America (*L. turionifera, L. minor, S. polyrhiza, L. minuta* and *L. gibba*). Although *L. trisulca* is present throughout Alberta, I did not design any primers for this species because it is morphologically distinct, even when examining dried or damaged specimens. I downloaded 33 sequences from NCBI (Wang et al., 2010, Tang et al., 2014, Barks et al., 2018), 17 for *atpF-atpH* (MG000358, MG000413, MG000505, MG000371-373, MG00416, MG000397, MG000405, MG000406, GU454232, GU454233, GU454206-208, KP017659, and KP017648) and 16 for *psbK-psbI* (MG000432, MG000487, MG000479, MG000480, MG000471, MG000490, MG000445-447, GU454338, GU454328-330, and GU454300-302) from the five duckweed species. I aligned sequences using MEGA 6.0 (Tamura et al., 2013) and identified species-specific target sites (Supplementary Figure 1.1, 1.2). I designed eight species-specific primers for *Lemna* species that would allow us to screen large numbers of samples at both loci, without the need to sequence every individual.

2.3.3 Screening *Lemna* species using *atpF-atpH* and *psbK-psbI* species-specific primers

First, I tested the species-specific primers against the reference strains to demonstrate they amplified only the target species (Figure 2.1). For screening and identification, I extracted total DNA from one frond for each sample using a Geneaid Genomic DNA Mini Kit (Plant: GP100; FroggaBio Inc., North York, ON, Canada). I extracted DNA from five individuals for each waterbody and screened them to increase the accuracy of the screening process. I tested each primer on five different duckweed species (*L. turionifera*, *L. minor*, as well as *L. gibba*, *L. minuta*, and *S. polyrhiza*). I performed PCR amplification using an Eppendorf master cycler thermal cycle using the eight species-specific primers. PCRs were performed in 10 µl reactions, with 1x Truin buffer, 0.2 mM dNTP, 0.2 mM of each primer and 1 Unit tru taq (Barks et al., 2018).

All samples were initially screened with an *L. turionifera* specific *psbK-psbI* primer (*LemnaTpsb208*_F) as *L. turionifera* was likely to be the most common species based on previous data (Barks et al., 2018). Any sample that did not amplify with *L. turionifera* primers was screened with *L. minor* specific *atpF-atpH* and *psbK-psbI* primers (*Lemna*Mr*psb*97_F and *Lemna*_Mr*atp*152_R), as *L. minor* is the other morphologically similar species believed to be present in Alberta. If a sample did not amplify with either of those primers, it was reamplified with all eight species-specific primers to determine if the initial PCR failed or if it was another species.

2.3.4 Identifying the relationship between surface water quality variables and distribution of two *Lemna* species

After identifying which species were present in each sample, I carried out a pilot study to identify the relationship between surface water quality variables and distribution of two *Lemna* species. I obtained surface water-quality data for 17 sites from "The Surface Water Quality Data and Online Tool" webpage of the Government of Alberta (https://www.alberta.ca/surface-water-quality-data.aspx). Surface water quality samples in rivers, lakes, and other water bodies across the province of Alberta have been collected by Alberta Environment and Parks, and its partners. Most of the physical and chemical data from the monitoring program are stored in Water Data System, but from the 126 sample sites I used in this study, only 17 sites overlapped with sites in the Surface Water Quality Data and Online Tool (Supplementary Table 1.2). Of the 17 sites, nine contained *L. turionifera*, four contained *L. minor* and four contained both *L. turionifera* and *L. minor*. Twenty-one different surface water-quality variables were used in the data analyses (Supplementary Table 1.2). All the data were subjected to normality testing followed by Principal Component Analysis (PCA) using PAST: Paleontological statistics software package for education and data analysis (Paleontological Statistics) (Hammer et al., 2001). Principal component axes PC1 and PC2 were used to draw an ordination plot. PCA loadings were used to identify the magnitude of the contribution of each surface water-quality variable for the distribution of the sample sites in the ordination.

2.4 Results

2.4.1 Development of *atpF-atpH* and *psbK-psbI* species-specific primers

Nucleotide sequence differences in both *atpF-atpH* and *psbK-psbI* allowed us to develop eight species-specific primers for five morphologically similar North American duckweed

species: *L. turionifera, L. minor, L. gibba, L. minuta,* and *S. polyrhiza* (Supplementary Figure 1.1 and 1.2) each with unique species-specific banding profiles (Figure 2.1).

2.4.2 Screening *Lemna* spp. using *atpF-atpH* and *psbK-psbI* species-specific primers

Of the 126 sites, 91.3% (115) contained *L. turionifera* only- 3.2% (4) contained *L. minor* only and 5.5 % (7) contained both *L. turionifera* and *L. minor*. All but one of the sites containing *L. minor* were in the southern part of the province (the exception being found at 57.243 °N, -115.912 °W) (Figure 2.2).

2.4.3 Identifying the relationship between surface water quality and distribution of *Lemna turionifera* and *L. minor*

The sites that contained *L. minor* formed two distinct clusters in the PCA, one comprising the Keho Lake and Park Lake sites, and the other comprising the other sites (Figure 2.3). Sites that had *L. turionifera* exhibited a more scattered distribution (Figure 2.3). The percentage of variation explained by PC1 and PC2 was 29.72% and 20.37%, respectively. Variables such as Secchi disk transparency, amount of chlorophyll-a, total phosphorus, total dissolved phosphorus, pH, total hardness, bicarbonate content, and carbonate content had the highest PCA loadings among the water quality variables (Figure 2.4). *L. minor* only tended to be on the left side of the ordination in lakes with lower overall hardness, bicarbonate content, and carbonate content.

2.5 Discussion

Due to the morphological similarity and overlapping ranges of *Lemna turionifera* and *Lemna minor*, the identities of *Lemna* species in Alberta has been an open question. I successfully used DNA barcoding to differentiate five morphologically similar *Lemna* species. These primers allowed us to screen large numbers of samples across Alberta and outline the provincial distributions of *L. minor* and *L. turionifera*. My data show *L. minor* is present in multiple wetlands in southern Alberta (see Supplementary Table 1.1) and a single site in the far north of the province (57.243 °N, -115.912 °W). All the southern *L. minor* sites in Alberta are in the South Saskatchewan River Basin, which includes the Red Deer, Bow, and Oldman Rivers.

My finding of the overlapping distribution of *L. minor* and *L. turionifera* in Alberta raises several questions. For example, why do some sites have only one *Lemna* species? In this study I report co-existence of *L. minor* and *L. turionifera* at seven of the 126 water bodies. It is possible that many candidate sites have both *Lemna* species but that one species is more common than the other, such that I only detected one despite screening five individuals for each water body. Alternatively, perhaps differences in water quality are present in ponds and restrict the distribution of one species, such that overlap is rare, or one species is better adapted conferring a fitness advantage. Environmental traits related to water quality (e.g., pH, conductivity, nutrient concentrations), and interspecific competition between species of duckweed can affect their distributions (Landolt, 1975, Armitage and Jones, 2020). According to Landolt (1975) the range of pH for *L. minor* is 5.0-7.5 and for *L. turionifera* it is 3.5-8.2, and the

range for water conductivity for *L. minor* is 70-700 (mS/cm) and 185-1,300 (mS/cm) for *L. turionifera*.

I identified a potential relationship between surface water quality and distribution of *L. minor*. The limited number of *Lemna* collection sites with surface water quality data may adversely affect the PCA analysis. Intensive sampling and collection of a wide range of surface water-quality data will help to determine the underlying factors that limit *Lemna* distribution. However, contrary to this research program, Xu et al. (2015) found no significant relationship between the occurrence of duckweed and the nitrate or phosphate concentration of water bodies, nor their pH. Also, they claim that the presence or absence of a *Lemna* species in a water body is mainly determined by the environmental conditions of the particular water body (e.g., flora, fauna, and soil), but not just the chemical structure of the water body.

At large geographic scales, the distribution of *Lemna* species is limited by climatic boundaries and precipitation (Landolt, 1975). *Lemna turionifera* inhabits the majority of the continental areas of Eastern Asia and North America characterized by lower temperatures and low amounts of precipitation. In contrast, *L. minor* grows abundantly in moister climates, being rare or absent in more arid regions (Landolt, 1975). My results suggest that the geographic distribution of the two *Lemna* species is quite different in Alberta with *L. minor* primarily restricted to the southern part of the province. Climate data collected over past three decades (Daly et al., 2010) show the average annual precipitation in Alberta is generally higher in the northwest and lowest in the southeast. The pattern is reversed for average annual temperatures. Thus, the overall geographic pattern of *Lemna* in Alberta runs counter to that expected based on Landolt (1975).

The DNA barcoding method used in this study allowed for species identification and large-scale screening. The same barcoding techniques in this study can be used to reassess the taxonomic status and herbarium specimens of *Lemna*. In herbaria, most *Lemna* classification has been done by morphological determination alone (Marconi et al., 2019). These assessments will be important to determine whether and how these cryptic species affect regional species richness and range size.

2.6 Conclusions

Biodiversity of *Lemna* in Alberta has been underestimated because of the cryptic nature of *L. minor* and *L. turionifera*. Molecular tools, specifically DNA barcoding, ares a robust and feasible method to detect cryptic species. The molecular tools developed in this study can be used for the rapid identification of five monomorphic *Lemna*. My preliminary data suggest the presence of *L. minor* in the southern part of the province. While *L. turionifera* is widespread in Alberta, *L. minor* is more geographically restricted in the province and probably less known due to its morphological similarity to *L. turionifera. Lemna turionifera* appears to be present in a larger number of lakes in contrast, *L. minor* distribution is much more restricted.

2.7 Figures and Tables

Table 2.1 Species-specific *Lemna* PCR primers and length of PCR product (bp) designed for this study. Universal reverse primers were used for each locus: *psbK-psbI*_R (5' AAACTTTGAGAGTAAGCAT 3') or *atpF-atpH*_R (5'GCTTTTATGGAAGCTTTAACAAT 3') with the exception of LemnaMratp152_R which was amplified using *atpF-atpH*_F (5' ACTCGCACACACTCCCTTTCC 3'). Some mismatches (bold) were introduced to increase PCR specificity for the target species in addition to natural mismatches (underlined). Each primer name contains the species it amplifies, the locus, and location of the primer-binding site.

Locus	Primer Name	Amplifies	Primer Sequence (5'→3')	PCR Product (bp)
psbK-psbI	*Lemna***Mr**psb97_F	*L. minor*	GATTCTAATAAAAATT**C**ATAACG**A**AG	437
	*Lemna***T**psb208_F	*L turionifera*	TCCATTTCGCCATA**TT**CA	326
	*Lemna***Ma**psb178_F	*L minuta*	GTATCTTC**AT**TAAAAA**C**AGCAC	356
	*Lemna***G**psb241_F	*L gibba*	GTGAGGAA**C**TA**G**TTTAA**A**TTT	293
atpF-atpH	*Lemna***Mr**atp152_R	*L. minor*	GGGCGAAGTAATA**G**AA**T**AG	523
	*Lemna***Ma**atp236_F	*L. minuta*	GATTTTTAATGGAATAGAA**G**G**G**AATTTT	439
	*Lemna***P**atp309_F	*S. polyrhiza*	ATTTATTATTTTA**GG**CTAATTAAATTAAAGT	366
	*Lemna***G**atp394_F	*L. gibba*	GTTACAACG**A**ATAC**G**CT	281

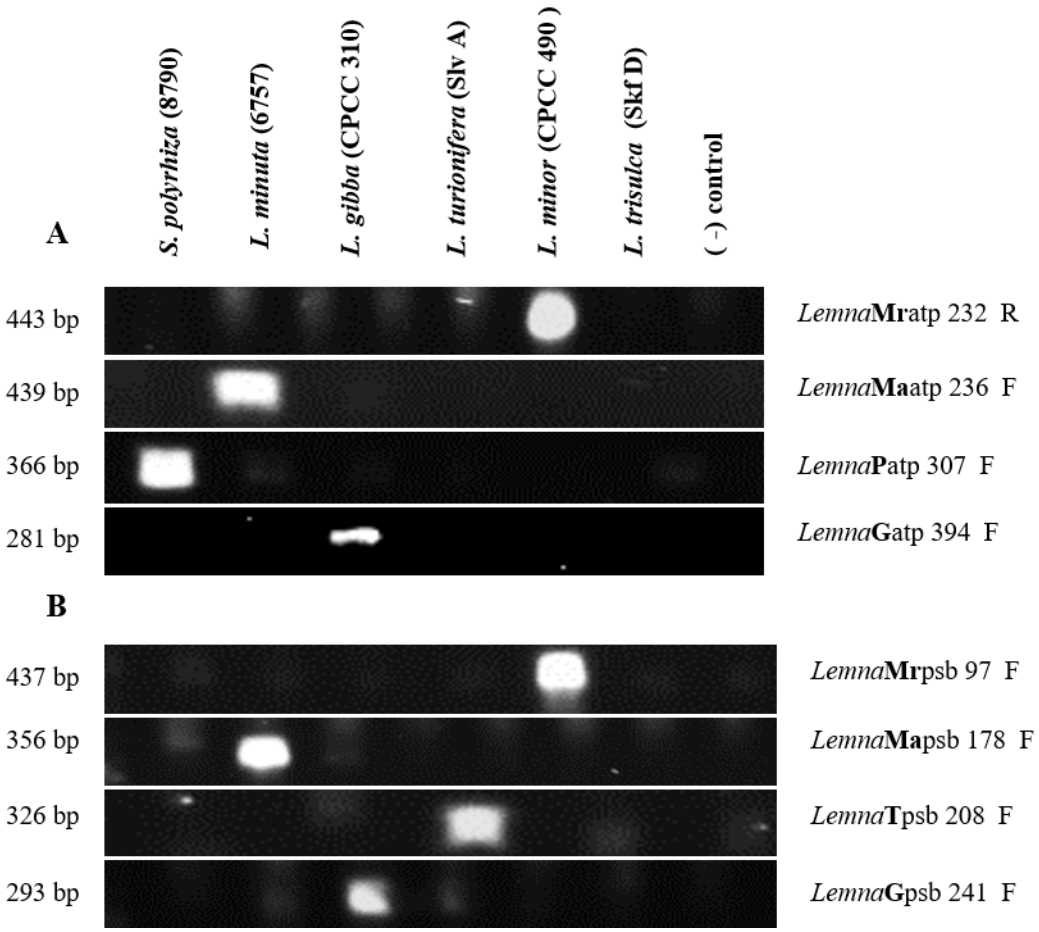


Figure 2.1. Specificity tests. Electrophoresis of DNA extracted from single individuals of *S. polyrhiza, L. minuta. L. gibba, L. turionifera, L. minor,* and *L. trisulca* using the species-specific primer (listed on far right). Each species has a unique banding profile with sizes of PCR products on the left. A) Banding profiles for *atpF-atpH* species-specific primers. B) Banding profiles for *psbK-psbI* species-specific primers.

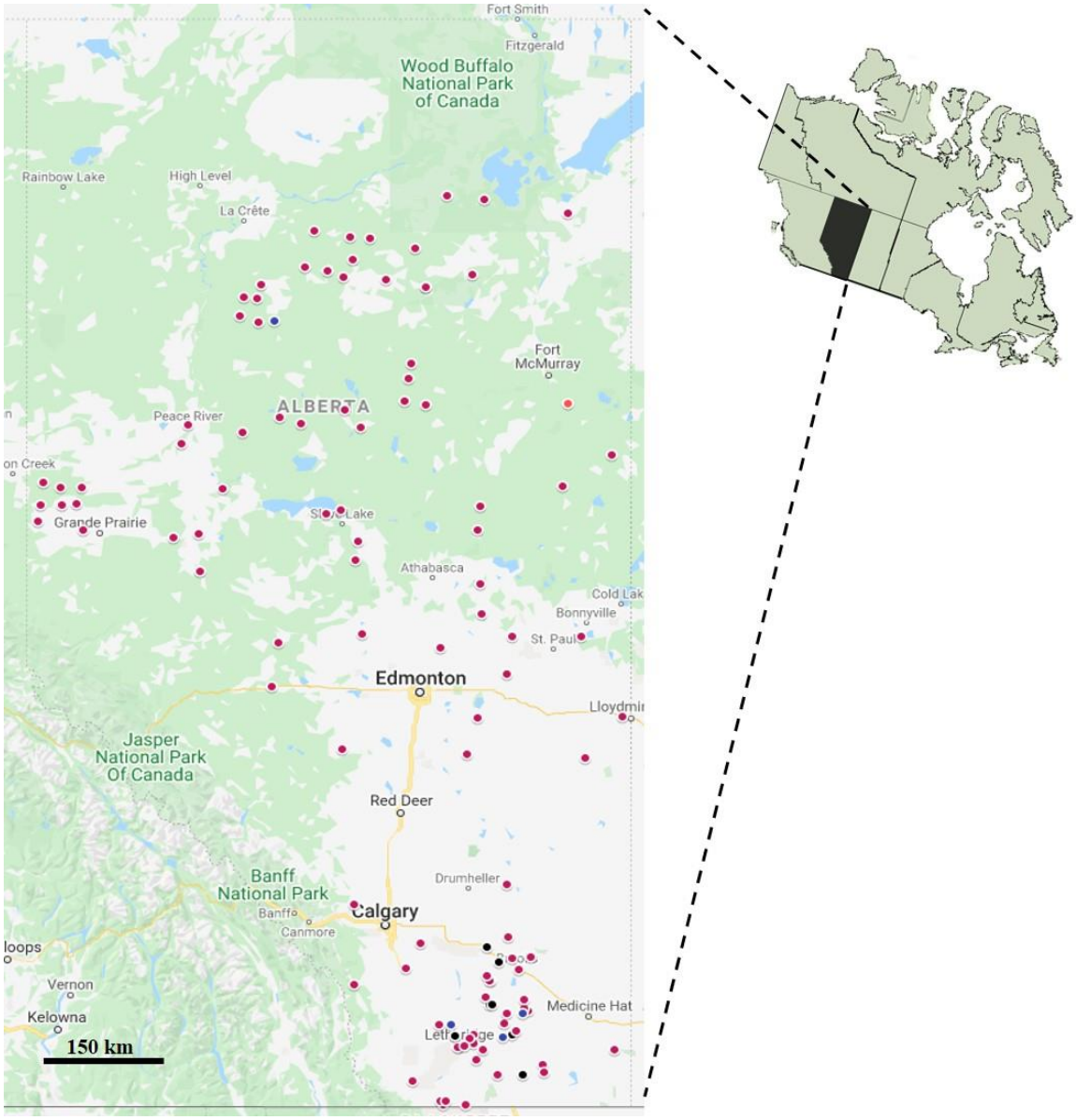


Figure 2.2 Sampling locations for *Lemna* species used in this study. Of the 126 sites, 115 sites contained *L. turionifera* but not *L. minor* (red), four contained *L. minor* but not *L. turionifera* (blue), and seven contained both species (black). Full details of sites are in the Supplementary Table 1.1.

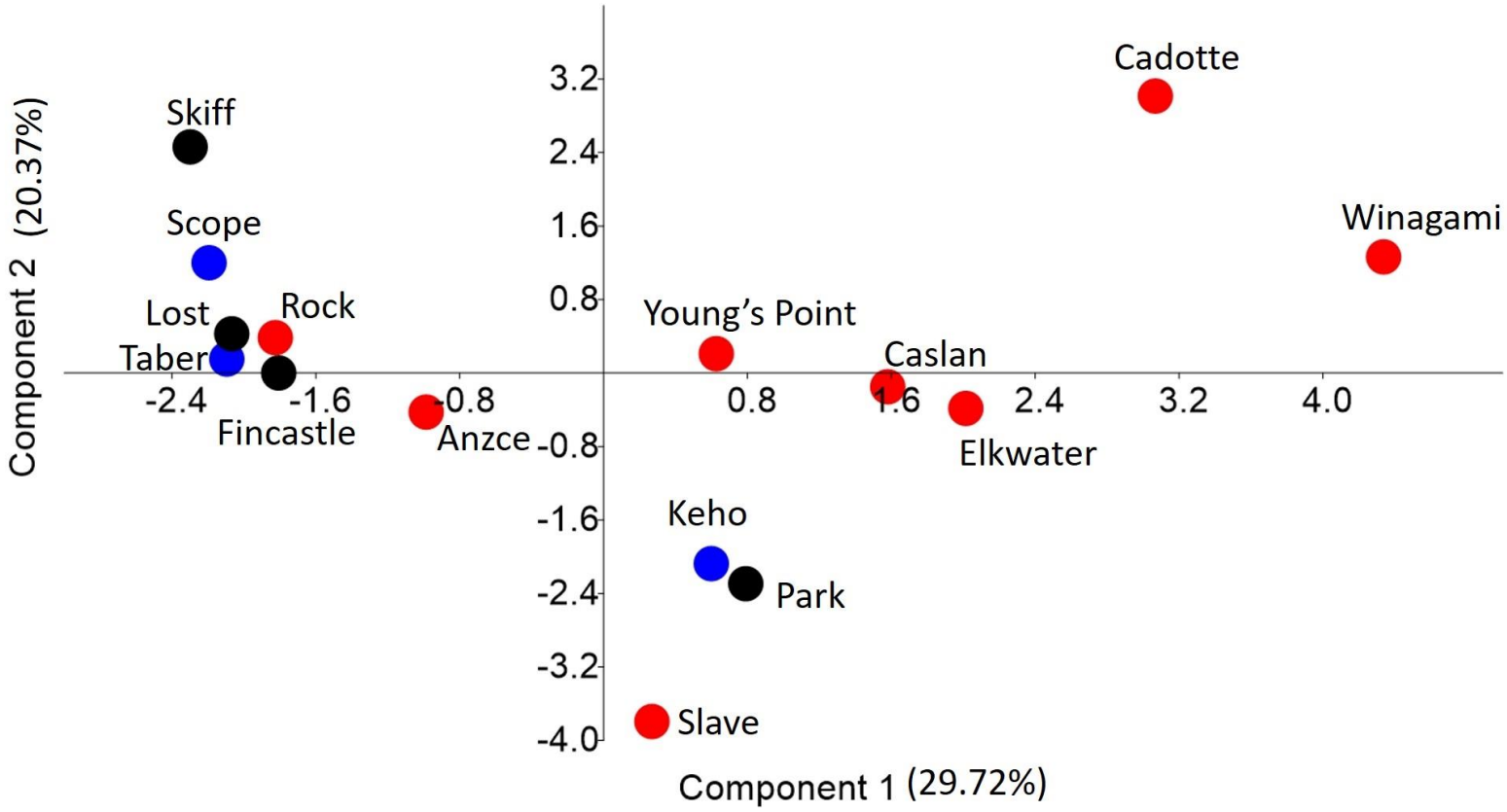


Figure 2.3 Scatter plot of the PCA conducted to identify the relationship between surface water quality and distribution of two *Lemna* species using PAST. Red colour dot represents lakes only have *L. turionifera*, Blue colour dots represent lakes with only *L. minor* and black colour dots represent lakes with both species.

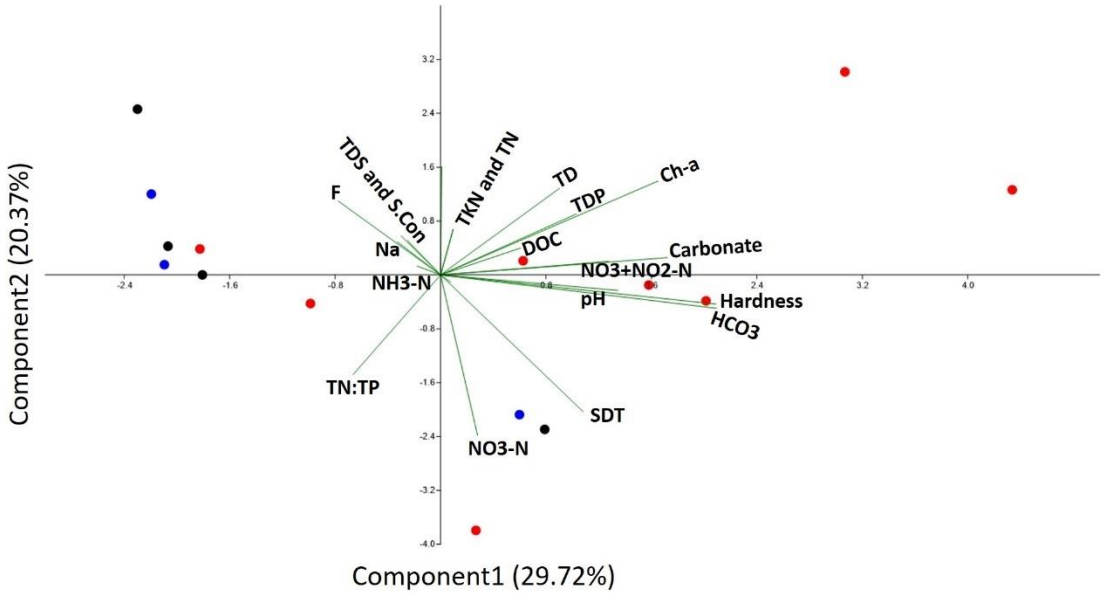


Figure 2.4 Graphical representation of PCA loading for the scatter plot in Figure 2.3. Red colour dots represent lakes that only have *L. turionifera*, blue colour dots represent lakes with only *L. minor*, and black colour dots represent lakes with both species. Green colour lines represent the vectors on the ordination and that show how the different variables are related to the two main ordination axes. The angles between the vectors tell us how characteristics correlate with one another. Abbreviations represent following variables SDT (Secchi Disk Transparency) , Ch-a (Chlorophyll-a content), TP (Total Phosphorus), TDP (Total Dissolved Phosphorus), NH_3-N (Ammonia), NO_3+NO_2-N (Nitrate+Nitrite), NO_2-N (Nitrite), TKN (Total Kjeldahl Nitrogen), TN(Total Nitrogen), TN:TP (TN:TP Ratio), SCon (Specific Conductance), TDS (Total Dissolved Solid), pH, $CaCO_3$ (Phenolphthalein Alkalinity), Hardness (Total Hardness), HCO_3 (Bicarbonate), Carbonate, Ca (Dissolved Calcium), Na (Dissolved Sodium), F (Dissolved Fluoride), and DOC (Dissolved Organic Carbon).

Chapter Three

ANALYZING THE POPULATION GENETIC STRUCTURE OF TWO CRYPTIC DUCKWEED SPECIES (*Lemna minor* & *L. turionifera*) IN ALBERTA USING A GENOTYPING BY SEQUENCING APPROACH

3.1 Abstract

Identifying population genetic structure of different species is important for the development of species-specific management plans. Investigating the population genetics of cryptic species is even more critical with respect to conservation. Results from Chapter 2 suggest that there are two cryptic duckweed species in Alberta: *Lemna minor* and *L. turionifera*. However, their population differences have not been assessed. To address this, I used genotyping by sequencing (GBS), a novel and effective molecular technique to outline and describe phylogenetically distinct groups. Based on analyses of 103 samples at 16,007 single nucleotide polymorphism (SNP) loci, I supply evidence for genetic diversity between *L. minor* and *L. turionifera* growing in Alberta. Taking a within-species perspective, I found no genetically distinct populations within *L. turionifera*. In contrast, significant population differentiation in *L. minor*. This study may provide information for conservation and management strategies for *Lemna* species in Alberta in the future.

3.2 Introduction

Understanding the genetic structure of different populations across their natural geographical ranges is vital to execute species-specific management plans (Allendorf and Luikart, 2009). Population genetic structure can explain how different populations are linked by gene flow (Slatkin, 1987, Manel et al., 2003, Segelbacher et al., 2010), provide insight on the evolutionary relationship between species (Avise et al., 1987), and help identify local adaptations which govern ecological differences (Kawecki and Ebert, 2004). Another vital role of understanding genetic structure is to define boundaries around evolutionarily and demographically distinct populations by describing intraspecific conservation units, such as evolutionarily significant units (Abdelaziz et al., 2011), and therefore identify suitable groups below the species level to which to apply conservation management methods (Funk et al., 2012).

Next-generation sequencing (NGS), a high-throughput, short-read sequencing approach, has improved the capacities of population genomic studies over the last few decades (Kulski, 2016). These recent advances in NGS approaches have led to the development of novel molecular tools to study population genetics, such as genotyping-by-sequencing (GBS) and restriction site-associated DNA sequencing (RAD-seq). GBS is one example of NGS that reduces genome complexity by using restriction enzymes to obtain large number of SNP markers (Elshire et al., 2011). Molecular markers are key in animal and plant genetics and genomics as they allow us to identify genetic differences within and between species. Single nucleotide polymorphisms (SNP) are extremely useful for population genetics, as large SNP datasets can be used increase our ability to resolve fine-scale population structure (Malenfant et al., 2015, Vendrami et al., 2017, Ferchaud et al., 2018).

Recently, GBS approaches have been used as a cost-effective platform for SNP discovery and genotyping (Sonah et al., 2013). These SNP markers can be used for different types of studies, for example genome-wide association studies (GWAS) (Sakiroglu and Brummer, 2017), genetic diversity studies (Lu et al., 2013), genome assembly based studies, and genetic map construction (Ward et al., 2013). GBS provides a platform where marker discovery and genotyping can be performed simultaneously for large numbers of samples, and thereby reduce the cost per sample (He et al., 2014). Since GBS is a cost-effective protocol, it is feasible to use GBS as a tool to support plant breeding programs as well as to study population structure in plants (Küpper et al., 2018, Li et al., 2020, Liu et al., 2020).

Detection of fine-scale population structure of plant species may be challenging due to clonal reproduction (Brzosko et al., 2002, Eckert et al., 2003) as clonally reproducing populations exhibit low genotypic diversity and clonal reproduction produces progeny that are genetically identical to their parent (Jackson et al., 1985). In general, when genetic differentiation is low among populations, greater numbers of samples and genetic markers are required to identify and resolve fine-scale genetic structure (Patterson et al., 2006). Therefore, careful selection of NGS protocols is extremely important as these choices may lead to unfavorably lengthy computational times and higher analytical costs (Shendure and Aiden, 2012).

Duckweeds (Family Araceae, Subfamily Lemnoideae) are the smallest aquatic monocots. Lemnoideae consists of five genera, *Spirodela*, *Landoltia*, *Lemna*, *Wolffia,* and *Wolffiella* (Landolt, 1975, Lemon and Posluszny, 2000). These fast-growing, free-floating aquatic plants are widespread throughout freshwater bodies of North America and worldwide (Keddy, 1976, Landolt, 1986). Duckweeds have been used as a model

organism in numerous fields such as ecology, physiology, and biochemistry (Hillman, 1961, Landolt, 1986, Landolt and Kandeler, 1987).

Duckweeds are predominantly clonally reproducing plants. In theory, duckweeds should have lower genetic diversity than sexually reproducing plants, as clonal reproduction may reduce genetic differentiation between populations (Ashton, 1989, Starfinger and Stöcklin, 1996). However, distinct population genetic variation among different duckweed strains was found from different geographic regions (Bog et al., 2010, Xue et al., 2012, Bog et al., 2013). Studying natural populations is useful for understanding evolutionary processes at the population level (Tang et al., 2014). However, despite the fact that scientists have studied duckweeds for many years, information about their population genetic structure remains limited.

To understand the importance of duckweed in Alberta's ecosystems as well as the economic and recreational values they contain, it is necessary to know their biodiversity within the province. Further, to answer questions related to the biodiversity and conservation of *Lemna* in Alberta, it is necessary to understand their population genomics. A molecular systematic approach is a common tool to ascertain accurate intra- and inter-species differentiations (Xiao et al., 2010, Abdelaziz et al., 2011). Therefore, in this study, I used GBS to determine the biogeographical distribution and population genetic structure of two cryptic *Lemna* species in Alberta; *L. minor* and *L. turionifera*. Specifically, I used SNP data to investigate the population genetic structure of these two cryptic *Lemna* species.

3.3 Materials and Methods

3.3.1 Genetic Materials

I conducted a study to identify gene sequences that differed from each other by one or more sequence polymorphisms within *L. turionifera* and *L. minor* using 12 chloroplast genetic markers; *atpF–atpH, psbK–psbI, matK, rbcL, trnH–psbA, ITS, trnL-trnF* spacer, *trnSGCU-trnGUUC, atpB-rbcL* spacer, *atpB* gene, *matK-trnT* spacer, and *trnL* (*tRNA*-leu gene). I generated chloroplast DNA sequences using polymerase chain reaction (PCR) for samples collected across Alberta. Sequences were aligned and manually edited using MEGA6 (Tamura et al., 2013). I determined that *L. minor* and *L. turionifera* are genetically different from each other, but I failed to detect intraspecific differences. Therefore, I used GBS, a robust and efficient NGS technology to find out intraspecific differences within *L. minor* and *L. turionifera*. I include samples from different points along the watersheds in Alberta (supplementary Figure 2.1) to determine correlations between the population structure of *Lemna* populations and geographical factors. I chose samples along the major watersheds, as I hypothesized, I might see a pattern of population genetic differences along water flow. A total of 192 *Lemna* DNA samples (Supplementary Figure 2.1) from Chapter 2 were used for NGS (48 *L. minor,* 144 *L. turionifera*).

3.3.2 DNA extraction and GBS

I extracted DNA from dry fronds of *L. minor* and *L. turionifera* individuals for GBS using the Geneaid Genomic DNA Mini Kit (Plant: GP100; FroggaBio Inc.). I sent the extracted

DNA to Laval University Genomic Analysis Platform for library preparation (*PstI* and *MspI*) and sequencing on an Ion Torrent sequencer.

3.3.3 SNP discovery using the Fast-GBS pipeline and SNP filtering

SNPs were called using the Fast-GBS pipeline (Torkamaneh et al., 2017). Fast-GBS uses the Sabre programming tool to demultiplex barcoded reads. After demultiplexing, I used the Cutadapt tool to find and remove adapter sequences, primers, and other unwanted sequences from high-throughput sequencing reads (Martin, 2011) prior to sequence alignment using Burrows-Wheeler Aligner tool (BWA) (Li and Durbin, 2010). Three reference genomes (*L. minor* 5500 Genome, *L. minor* 8627 and *L. gibba* 7742a Genome) were available and I ran my data with all three and obtained similar results (An et al., 2018). As such, I did all subsequent analyses using the *L. minor* 8627 (800 Mb) genome sequence from Ernst and Martienssen (2016) as the reference genome. I used SAMtools for file conversion and indexing (Li, 2011), and post-processing of mapped reads, haplotype construction and variant calling were done using Platypus (Rimmer et al., 2014).

I produced filtered SNPs files using VCFtools (Danecek et al., 2011). I constructed the final dataset by removing SNPs with more than 50% missing data (max-missing option of vcftools (–max-missing 0.5)), and individuals with more than 40% missing SNPs.

3.3.4 Data Analyses

3.3.4.1 Discriminant Analysis of Principal Components (DAPC)

For the data analysis, first I investigated a dataset containing both *L. minor* and *L. turionifera* (see supplementary table 2.1). Subsequently, to identify finer-level clusters, I created two additional datasets, containing either *L. minor* or *L. turionifera.* After the SNP detection, I used a Discriminant Analysis of Principal Components (DAPC) (Jombart et al., 2010) to identify and describe the genetic clusters within and between the two target *Lemna* species. To obtain the different genetic clusters from the data set, I used Bayesian information criterion (BIC) model along with the K-means clustering algorithm. Then, as recommended by Jombart et al. (2010), the optimal K-mean was chosen based on the lowest BIC.

3.3.4.2 Inference of individual admixture coefficients

I used individual SNPs to estimate individual admixture coefficients from the genotypic matrix using the LEA package (Frichot and François, 2015) in R 4.0.0 (R Core Team, 2019). I included individual admixture coefficients by the function *snmf()*. I set all parameters to default values recommended by Frichot and François (2015). I used the entropy criterion to identify the number of ancestral populations that best explained the dataset. I selected the results for K = 4 as higher values of K were not biologically meaningful.

3.4 Results

After the final filtering process, the dataset with both species yielded 16007 SNPs for 103 individuals (*L. minor* 29; *L. turionifera* 74) (see supplementary table 2.1). The *L. minor* dataset contained 16294 SNPs from 30 individuals from eight ponds (see supplement table 2.2) while the *L. turionifera* dataset had 11640 SNPs from 67 individuals from 43 ponds (data not shown). I did not observe population genetic differences among watersheds.

3.4.1 Discriminant Analysis of Principal Components

Lemna minor and *L. turionifera* are genetically different from each other (Figure 3.1). All *L. turionifera* individuals formed one cluster, while *L. minor* individuals grouped into at least two distinct clusters (Figure 3.1). Within *L. minor*, the individuals from Keho Lake (*KehB, KehF, KehC, KehD*) separated from the other *L. minor* individuals. Interestingly, the four other individuals from the Circle E wetland complex, Lost Lake, and Enchant are in another cluster (*CirD, CirE5, LosC2, EncB3*). Thus, there are at least two different genetic clusters within the Circle E wetland complex, Lost Lake, and Enchant. I then looked at the SNPs coverage and number of reads of the individuals that clustered in different lineages to check whether they are different because they lack data, or they are indeed genetically different from the rest of the individuals. For each individual the number of sequencing reads are between 400K -700K and contain 12K-15K SNPs. Therefore, this clustering is likely due to genetic variation within the population.

The DAPC generated using only *L. minor* individuals (Figure 3.2) shows a similar pattern as the combined dataset (Figure 3.1). This confirms the presence of at least

two different clusters within the *L. minor* individuals of the Circle E wetland complex, Lost Lake, and Enchant and those from Taber Lake, Skiff, and Bassano. Also, individuals from Keho Lake (*KehB, KehF, KehC, KehD*) separate into another distinct cluster (Figure 3.2). When *L. turionifera* samples were analyzed on their own, only a single cluster was found (results not shown).

3.4.2 Inference of individual admixture coefficients

The ancestry coefficients matrix (Figure 3.3) shows similar patterns to those observed in the DAPC (Figures 3.1 and 3.2). The genetic structure of individuals from Keho Lake (*KehB, KehF, KehC, KehD*) and individuals in cluster 2 from the Circle E wetland complex, Lost Lake, and Enchant (*CirD, CirE5, LosC2, EncB3*) are different from rest of the *L. minor* individuals and from each other.

3.5 Discussion

In this study, I examined intraspecific population differentiation of two cryptic duckweed species in Alberta. To my knowledge this is the first study to identify population genetic differences of *L. minor* and *L. turionifera* in Alberta using NGS methods. All analyses, regardless of dataset and method, confirmed that *L. minor* and *L. turionifera* within Alberta are genetically different from each other. No evidence of genetically different populations within *L. turionifera* were found; however, within *L. minor,* I found evidence of three distinct genetic groups from my samples. In general, duckweed species show extremely low SNP mutation rates per generation leading to

lower genetic variations among species at population level. Compared to other multicellular eukaryotes, *L. minor* and *Spirodela polyrhiza* show the lowest mutation rates per generation (Xu et al., 2019, Sandler et al., 2020).

One distinct cluster formed by *L. minor* included samples from Keho Lake (*KehB, KehF, KehC, KehD*) (49.931°N, 112.993°W). Keho Lake is a large, shallow pond in northwest Lethbridge County. With its extreme wind conditions, Keho Lake is a popular spot for recreational. activities such as canoeing, kayaking, windsurfing, golfing, and swimming. This anthropogenic disturbance could have led to possible contaminations of distinct strains and their subsequent spread. However, the obvious question is where this new strain came from. One possibility is from nearby lakes, but the genetic makeup of Keho Lake individuals (Figure 3.3) is completely different from nearby sampled populations. Therefore, another explanation for this distinct cluster is due to the founder effect. The founder effect is a phenomenon where a new colony is established from a small number of founder individuals (Provine, 2004). The founder effect may be a reason for reduction in genetic variation and cause the new population to be genetically different from others. Human activities or natural events like migratory animals may have brought genetically different *L. minor* individuals to Keho Lake and these individuals may have had characteristics that allowed them to adapt to conditions in the lake leading to them suppressing the growth of other *L. minor* groups in the lake.

Another possible explanation is the occurrence of a population bottleneck. Population bottlenecks occur due to steep reductions in population size due to a natural or anthropogenic event. As with founder effects, bottlenecks may cause allele frequencies of a population to change, and may reduce genetic variation (Catton Jr, 2009). After this reduction in population size, the remaining individuals establish as a new colony with

new genetic characteristics. With the high number of anthropogenic events that occur in Keho Lake, there is a high probability of reduction of the population size of *L. minor* and occurrence of distinct cluster from those who survive from those disturbances.

Some analyses detected finer-scale structure within the Circle E wetland complex, Lost Lake, and Enchant. One of the significant findings of this study is the co-existence of multiple lineages of *L. minor* in the same water body. For example, there were two different genetic clusters present within the Circle E wetland complex, Lost Lake, and Enchant. Sandler et al. (2020) stated that, *L. minor* showed marginally high, but non-significant, mutation rates per generation compared to *S. polyrhiza*. Nonetheless, several other studies show *L. minor* outcrosses more frequently in its natural habitats than *S. polyrhiza* (Vasseur et al., 1993, Ho, 2017, Ho et al., 2019, Xu et al., 2019). This may explain why I observed more intraspecific genetic differences among *L. minor* populations, as well as the existence of multiple lineages of *L. minor* in the same water body.

Initially I used three reference genomes (*L. minor 5500, L. minor 8627,* and *L. gibba 7742a* reference genome) for SNPs calling (An et al., 2018). The results were similar regardless of the reference genome used. Thereafter, I used the *L. minor* 8627 (800 Mb) genome sequence from Ernst and Martienssen (2016) as the reference genome to align the sequences. Gene annotation for this reference genome is still in progress (An et al., 2018). During the SNPs calling, I did not have a reference genome for *L. turionifera*. Therefore, I used the same *L. minor* 8627 genome as the reference sequence for *L. turionifera* because according to the Galla et al. (2019), congeneric and con-

familial references genomes for SNPs calling from closely related species provide high similarity of nucleotide diversity, correlation, and relatedness.

A more thorough and extensive study of *L. minor* in Alberta will be needed to identify different populations or natural strains that are distinct enough to be considered as Evolutionarily Significant Units (Abdelaziz et al., 2011). However, as discussed in Chapter 2, *L. minor* in Alberta is mostly confined to the southern region of the province. As well as in the context of genetics, *L. minor* has greater within-species population diversity than *L. turionifera* in Alberta. As such, there is a timely need to document and carefully monitor *L. minor* within the province, in particular to assess its vulnerability to changes in climate and anthropogenic disturbances in the near future. This study provides evidence of population differences even within clonally reproducing plants. Founder effects or population bottlenecks could be candidate reasons for population differences within species. GBS data are extremely useful for identifying intraspecific genetic differences. Discovering these intraspecific differences will be important to understand their evolution as a species.

3.6 Conclusions

The duckweed species *L. minor* and *L. turionifera* are genetically different from each other in Alberta. From the samples I used in this study I detected no genetically different populations within *L. turionifera*. However, within *L. minor* there are three different genetic groups from the samples collected from the southeastern part of the province. Individuals from the Keho Lake population are genetically different from the other *L. minor* populations. There are two different genetic groups within the individuals from the Circle E wetland complex, Lost Lake, and Enchant populations. The restricted distribution and the high genetic diversity indicate the importance of documentation and careful monitoring of *L. minor* within the province. Finally, GBS is one of the best approaches to determine fine-scale genetic differences within plant species.

3.7 Figures and Tables

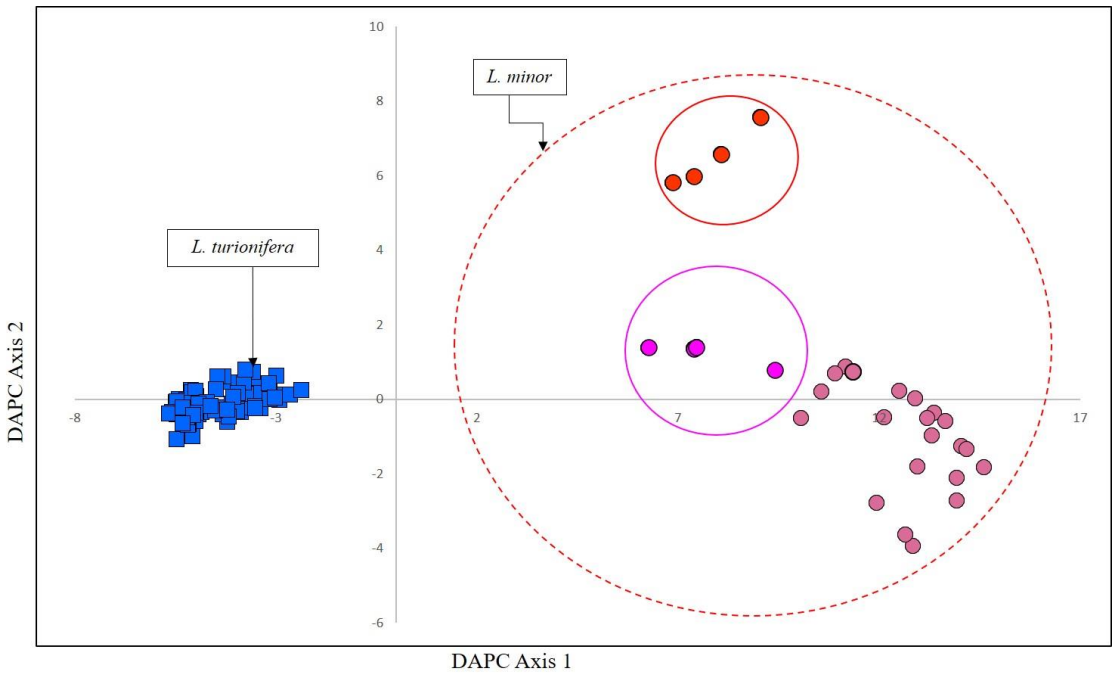


Figure 3.1 Discriminant analysis of principal components (DAPC) based on 16007 SNPs from *L. turionifera* and *L. minor*. Colours represent the different clusters identified by the DAPC. Each point is an individual. The large circle with a dashed line contains the *L. minor* individuals. The solid red circle contains the individuals from Keho Lake (*KehB, KehF, KehC, KehD*). The solid pink circles depict four individuals from the Circle E wetland complex, Lost Lake, and Enchant (*CirD, CirE5, LosC2, Enc B3*). Blue squares represent the *L. turionifera* individuals.

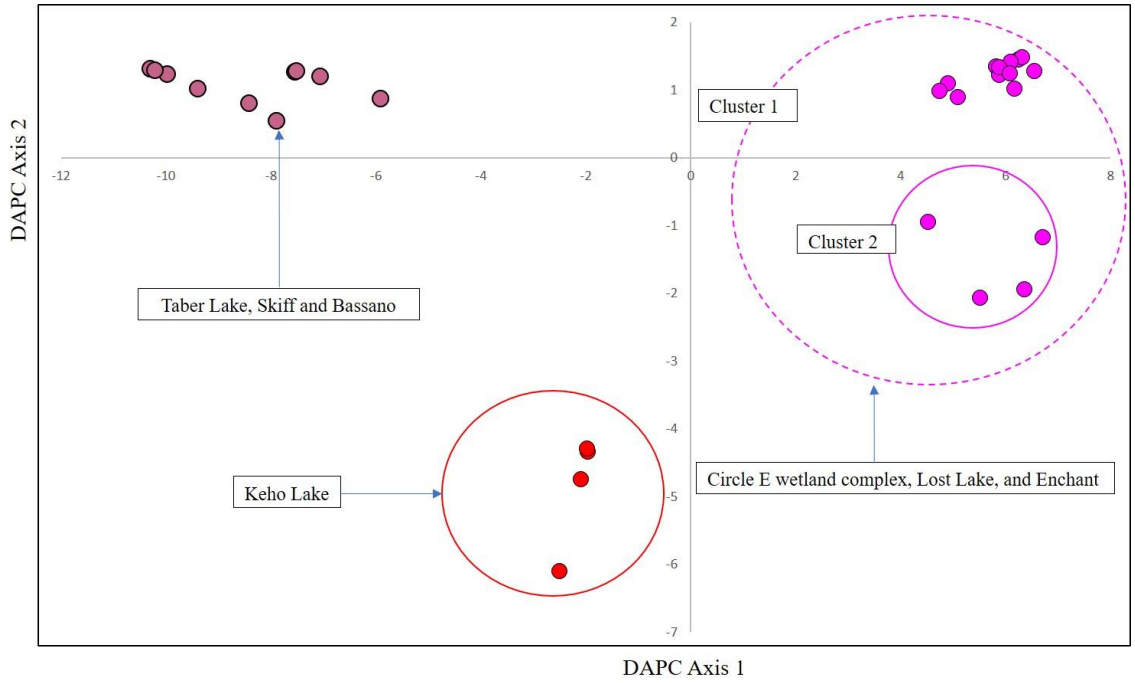


Figure 3.2 DAPC based on 16294 SNPs of *L. minor*. Colours represent the different clusters identified by the DAPC. Each point is an individual. The solid red circle contains the individuals from *Keho* Lake (*KehB, KehC, KehD, KehF*). The large circle with a dashed line (cluster 1) contains all of the individuals belonging to the Circle E wetland complex, Lost Lake, and Enchant. The solid pink circle (cluster 2) depicts four diverged individuals from the Circle E wetland complex, Lost Lake, and Enchant (*CirD, CirE5, LosC2, EncB3*). Maroon color dots (top left) represent the individuals from Taber Lake, Skiff, and Bassano.

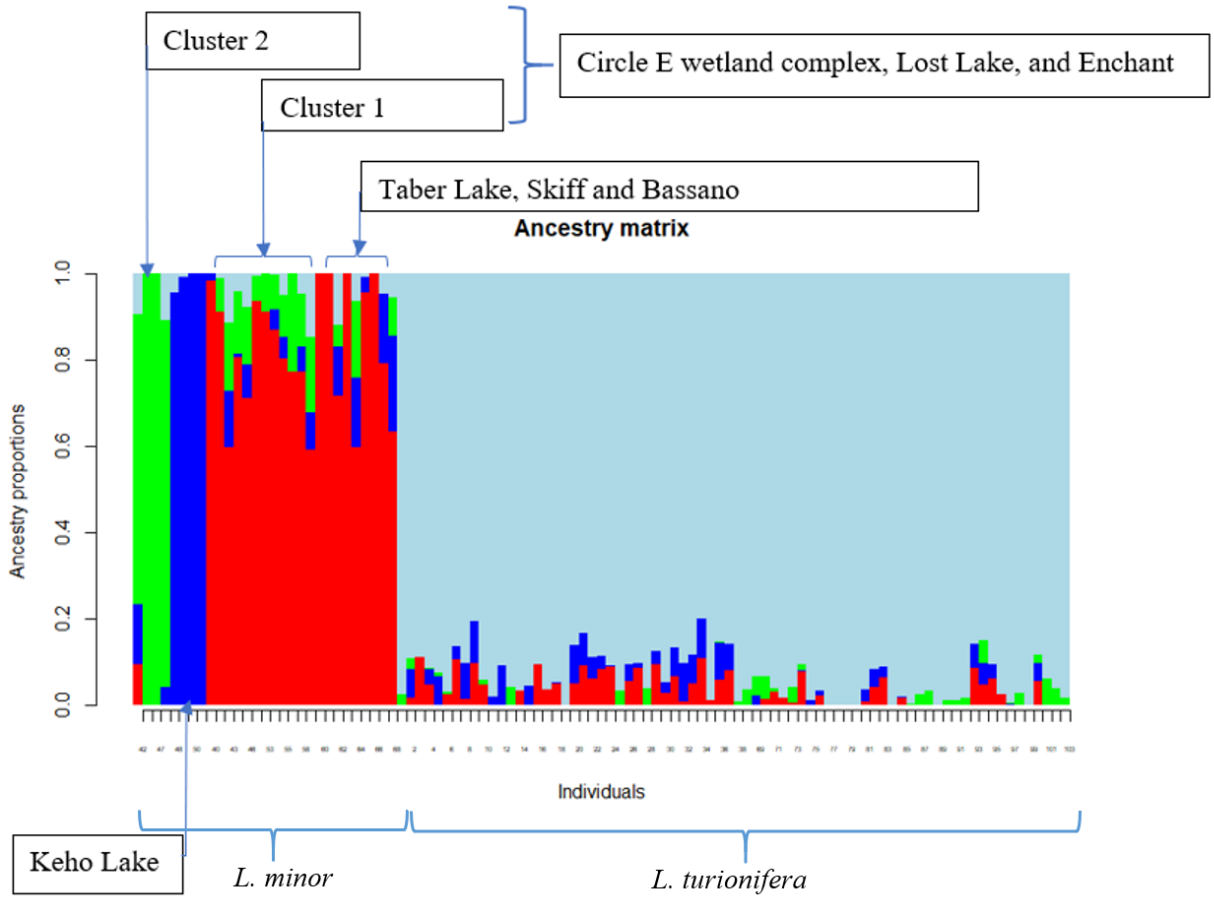


Figure 3.3 Ancestry coefficients based on 16007 SNPs for *L. minor* and *L. turionifera*. The vertical bars in each panel represent individuals and they are coloured to reflect ancestry proportions.

Chapter Four
General Discussion

4.1 Major findings and limitations

Based on DNA barcoding data, three *Lemna* species are found in Alberta, namely *L. minor, L. turionifera,* and *L. trisulca*. Out of the three species *L. minor* is morphologically similar to *L. turionifera* and the former's biogeographical distribution is mostly restricted to the southern part of the province. In addition, based on the NGS data, report three genetically different clusters of *L. minor* within waterbodies of southern Alberta.

Lemna is a small plant with enormous potential. The duckweed family contains some of the oldest model plants. This is in part due to their ease of culturing in laboratory conditions and their simple morphology. A great deal of research has been dedicated to understanding the physiology within the subfamily Lemnoideae (Landolt and Kandeler, 1987). Duckweed also plays a crucial role in food production. In southern Asia, small-scale farmers feed duckweed to farm animals. Duckweed is also used to treat and help convert urban, agricultural, and even industrial sewer water streams into clean water (Fourounjian et al., 2020). Moreover, duckweeds are one of the best candidates for biofuel production (Cheng and Stomp, 2009, Fourounjian et al., 2020).

Traditional morphological diagnostics and DNA technology are key players when it comes to identifying organisms at the species level. With the recent development of DNA barcoding and genomic sequencing, the discovery of cryptic species will likely increase. According to current data, 10-20% of species diagnosed morphologically may be two or more species (Janzen et al., 2017). When considering cryptic species, studying their roles in ecosystems, and their impact on environment and biodiversity is important.

Discovery of cryptic species and their biogeography helps identify their true distributions within ecosystems. This study is an example of how morphological similarity can affect our understanding of biodiversity in an ecosystem. Alberta hosts three *Lemna* species but the morphological resemblance of *L. turionifera* and *L. minor* masked the presence of the latter. Using DNA barcoding, I was able to fill several knowledge gaps about Alberta's *Lemna* diversity.

In this study I saw that *L. turionifera* is distributed throughout the province. However, in the case of *L. minor*, I observed that its distribution is mostly restricted to the southern part of the province. I attempted to understand these distribution patterns with respect to climate and water quality differences in the natural habitats of the two species. However, I do not have exact answers for the observed patterns of biogeographical distribution among *Lemna.* To answer this, I will need more information on the ecological variables that influence *Lemna* species' distributions. These study species are aquatic free-floating plants. Thus, I speculate that their distribution is mostly determined by the surface water quality. I investigated the correlation of surface water quality data with the distribution of *Lemna* species within Alberta and I found that the sites that contained *L. minor* formed two distinct clusters: one comprising the Keho Lake and Park Lake sites, and the other comprising the other sites (Figure 2.3). However, sites containing *L. turionifera* exhibited a more scattered distribution. One of the significant limitations of this study is not having surface water quality data for all of the sampling sites. Out of 126 sites I studied, only 17 sites had surface water quality data. Therefore, collecting water quality data from all those other sites is critical.

According to the GBS data analyses, there are no genetically distinct populations within *L. turionifera* in the Alberta. However, within *L. minor* three distinct genetic

groups were identified. In general, duckweed species show extremely low SNP mutation rate per generation leading to have lower population genetic variation among species. Compared to other multicellular eukaryotes, *L. minor* and *Spirodela polyrhiza* show the lowest mutation rate per generation (Xu et al., 2019, Sandler et al., 2020). Xu et al. (2019) analysed 68 different *S. polyrhiza* whole-genome sequences from all around the world and observed low genetic variation among those individuals, which is associated with low mutation rates. Sandler et al. (2020) compared *L. minor* and *S. polyrhiza* and found low mutation rates per generation for both species. Therefore, I propose that *L. turionifera* might also have low genetic variation due its low mutation rate. Unfortunately, there have been no studies to identify or compare the mutation rate of *L. turionifera* with other duckweed species. In addition, Sandler et al. (2020) stated that, *L. minor* showed marginally high, but non-significant, mutation rate per generation compared to *S. polyrhiza*. Other studies show *L. minor* outcrosses more frequently in their natural habitats than *S. polyrhiza* (Vasseur et al., 1993, Ho, 2017, Ho et al., 2019, Xu et al., 2019). This may explain why I observed more intraspecific genetic differences among *L. minor* populations. Following mutation, any better-adapted individuals can disperse across geographical range.

A possible explanation for the genetically different *L. minor* strains within the province is anthropogenic dispersal events. For example, human activities may have resulted in recent spreading of some duckweed species into non-native habitats such as *L. minor* cultures from outside of the province or from unsampled lakes. The native range of *Lemna minuta* is the Americas, but according to the current records *L. minuta* has expanded to Japan and Europe (Ceschin et al., 2018, Landolt, 2000). Correspondingly, *Landoltia punctata* has invaded North America and Europe from its native habitat range

(Landolt, 1986, Les et al., 1997, Jacono, 2018). According to the records, *L. minor* dispersal in Australasia and *L. gibba* dispersal in Japan were also triggered by anthropogenetic activities (Landolt, 1986).

4.2 Future directions

Building on this study, it would be valuable to investigate the diversity of the entire duckweed family within the province and estimate the distribution of its component species. Furthermore, I will be able to better understand their importance in the ecosystem, recreational interest and genetic variation they contain. In addition, I employed barcoding markers to identify only five morphologically similar species. I can apply the same methodology to employ species-specific markers for all 36 species of duckweed to aid in identification of herbarium samples and reduce the risk of mislabelling plants samples in large culture stocks.

Another essential investigation will be to study transcriptomes, karyotypes, and genome sizes. These analyses are important for identifying how the genome and gene expression changes in different organisms. There is no existing information on the genome size, genome sequences, or annotation of the *L. turionifera* genome. This information will serve as the primary source for gene ontology studies. Gene ontology mainly focuses on functions of genes and gene products and it is important to organize information and develop specific vocabulary of gene and gene products. In the terms of gene ontology, *L. turionifera* is an under-studied species compared to *L. gibba* and *L. minor*. As Alberta has a widespread *L. turionifera* population, it is important to document and report its physiological and genetic characters. Also, studying the phylogenetic relationship and evolutionary history of the entire family would be important. It will be interesting to see the common ancestors, where they originated and to predict the time

and the mode of dispersal to Alberta. This information will help uncover the relationship between Alberta's duckweeds and other duckweed populations around the world.

If we want to build conservation strategies for certain populations, we must have the knowledge of their genetic structure along with their abundance and ecology. The field of conservation genetics incorporates the phylogeographical distribution of species. Furthermore, niche size, location, biogeographical distribution, and evolutionary coherence are also considered during the process. A conservation plan should be established for all *Lemna* species in Alberta by evaluating the above criteria to ensure genetic diversity is maintained for all species and populations. Overall, this work provides essential genomic resources for future researchers working on *Lemna* species or on different strains of certain species with conservation concern. In general, this study provides an example of using next generation sequencing methods to identify potentially important population genetic virions and assist in the development of species-specific management plans.

4.3 Closing Statement

Based on DNA barcoding data, I confirmed the presence of three *Lemna* species (*L. minor*, *L. turionifera* and *L. trisulca*) in Alberta, and *L. minor* is rarer and/or geographically restricted to the southern part of Alberta and 'camouflaged' by *L. turionifera*. NGS studies suggest the presence of significant population differentiation in *L. minor*. In contrast, I found no genetically distinct populations within *L. turionifera*.

References

Abdelaziz, M., Lorite, J., Muñoz- Pajares, A. J., Herrador, M. B., Perfectti, F. & Gómez, J. M. 2011. Using complementary techniques to distinguish cryptic species: a new *Erysimum (*Brassicaceae) species from North Africa. *American Journal of Botany,* 98**,** 1049-1060.

ABMI. 2020. Alberta Biodiversity Monitoring Institute. Vascular plants [Online]. Available: https://abmi.ca/home/data-analytics/biobrowser-home/species-list?groupId=2 [Accessed October 26th 2020].

Aliferis, K. A., Materzok, S., Paziotou, G. N. & Chrysayi-Tokousbalides, M. 2009. *Lemna minor* L. as a model organism for ecotoxicological studies performing 1H NMR fingerprinting. *Chemosphere,* 76**,** 967-973.

Allendorf, F. W. & Luikart, G. 2009. Conservation and Genetics of populations Hoboken, John Wiley & Sons.

Álvarez, I. & Wendel, J. F. 2003. Ribosomal *ITS* sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution,* 29**,** 417-434.

An, D., Li, C., Zhou, Y., Wu, Y. & Wang, W. 2018. Genomes and transcriptomes of duckweeds. *Frontiers in Chemistry,* 6**,** 230.

Armitage, D. W. & Jones, S. E. 2020. Coexistence barriers confine the poleward range of a globally distributed plant. *Ecology Letters,* 23**,** 1838-1848.

Arponen, A. 2012. Prioritizing species for conservation planning. *Biodiversity and Conservation,* 21**,** 875-893.

Ashton, P. J. 1989. Aquatic plants: patterns and modes of invasion, attributes of invading species and assessment of control programmes. *Biological Invasion: A Global Perspective***, no volume?** 111-154.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics,* 18**,** 489-522.

Azer, S. 2013. Taxonomic revision of genus *Lemna* L.(Lemnaceae Gray) in Egypt. *Annals of Agricultural Sciences,* 58**,** 257-263.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. & Donoghue, M. J. 1995. The *ITS* region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden***,** no volume? 247-277.

Barks, P. M., Dempsey, Z. W., Burg, T. M. & Laird, R. A. 2018. Among-strain consistency in the pace and shape of senescence in duckweed. *Journal of Ecology,* 106**,** 2132-2145.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K. K. & Das, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution,* 22**,** 148-155.

Bog, M., Appenroth, K. J. & Sree, K. S. 2019. Duckweed (Lemnaceae): Its Molecular Taxonomy. *Frontiers in Sustainable Food Systems,* 3**,** 117.

Bog, M., Baumbach, H., Schween, U., Hellwig, F., Landolt, E. & Appenroth, K. J. 2010. Genetic structure of the genus *Lemna* L.(Lemnaceae) as revealed by amplified fragment length polymorphism. *Planta,* 232**,** 609-619.

Bog, M., Schneider, P., Hellwig, F., Sachse, S., Kochieva, E. Z., Martyrosian, E., Landolt, E. & Appenroth, K.-J. 2013. Genetic characterization and barcoding of taxa in the genus *Wolffia horkel* ex Schleid.(Lemnaceae) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP). *Planta,* 237**,** 1-13.

Bog, M., Xu, S., Himmelbach, A., Brandt, R., Wagner, F., Appenroth, K.-J. & Sree, K. 2020. Genotyping-by-Sequencing for Species Delimitation in *Lemna* Section Uninerves *Hegelm*.(Lemnaceae). The Duckweed Genomes. Springer.

BOLDSYSTEMS. 2020. Barcode of life data system v4 [Online]. Available: https://www.boldsystems.org/ [Accessed October 26th 2020].

Böttcher, T. & Schroll, R. 2007. The fate of isoproturon in a freshwater microcosm with *Lemna minor* as a model organism. *Chemosphere,* 66**,** 684-689.

Brzosko, E., Wróblewska, A. & Ratkiewicz, M. 2002. Spatial genetic structure and clonal diversity of island populations of lady's slipper (*Cypripedium calceolus*) from the Biebrza National Park (northeast Poland). *Molecular Ecology,* 11**,** 2499-2509.

CABI. 2020. Invasive Species Compendium (*Lemna minuta*) [Online]. Available: https://www.cabi.org/isc/datasheet/108968 [Accessed October 26th 2020].

Cardoso, P., Erwin, T. L., Borges, P. A. & New, T. R. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation,* 144**,** 2647-2655.

Catton Jr, W. R. 2009. Bottleneck: Humanity's Impending Impasse: Humanity's Impending Impasse, USA, Xlibris Corporation.

Cayuela, M. L., Millner, P., Slovin, J. & Roig, A. 2007. Duckweed (*Lemna gibba*) growth inhibition bioassay for evaluating the toxicity of olive mill wastes before and during composting. *Chemosphere,* 68**,** 1985-1991.

CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S. & Erickson, D. L. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences,* 106**,** 12794-12797.

Ceschin, S., Abati, S., Leacche, I. & Zuccarello, V. 2018. Ecological comparison between duckweeds in central Italy: The invasive *Lemna minuta* vs the native *L. minor*. *Plant Biosystems,* 152**,** 674-683.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madriñán, S., Petersen, G., Seberg, O., Jørgsensen, T., Cameron, K. M. & Carine, M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon,* 56**,** 295-299.

Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N. & Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 1889-1895.

Cheng, J. J. & Stomp, A. M. 2009. Growing duckweed to recover nutrients from wastewaters and for production of fuel ethanol and animal feed. *CLEAN–Soil, Air, Water,* 37**,** 17-26.

Chenuil, A., Cahill, A. E., Délémontey, N., du Luc, E. D. S. & Fanton, H. 2019. Problems and questions posed by cryptic species. A framework to guide future studies. *In:* Casetta, E., Marques Da Silva , J. & Vecchi, D. (eds.) From Assessing to

Conserving Biodiversity History, Philosophy and Theory of the Life Sciences. Cham: Springer.
Chester, T., Armstrong, W. & Madore, K. 2007. *Brodiaea santarosae* [Online]. Available: http://tchester.org/plants/analysis/brodiaea/santarosae.html. [Accessed October 26th 2020].
Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., de los Campos, G., Burgueño, J., Windhausen, V. S. & Buckler, E. 2013. Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3: Genes, Genomes, Genetics,* 3**,** 1903-1926.
Daly, C., Conklin, D. R. & Unsworth, M. H. 2010. Local atmospheric decoupling in complex topography alters climate change impacts. *International Journal of Climatology,* 30**,** 1857-1864.
Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T. & Sherry, S. T. 2011. The variant call format and VCFtools. *Bioinformatics,* 27**,** 2156-2158.
De Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology,* 56**,** 879-886.
Delić, T., Trontelj, P., Rendoš, M. & Fišer, C. 2017. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Scientific Reports,* 7**,** 1-12.
Dobzhansky, T. 1950. Mendelian populations and their evolution. *The American Naturalist,* 84**,** 401-418.
Dudley, J. L. 1987. Turion formation in strains of *Lemna minor* (6591) and *Lemna turionifera* (6573, A). *Aquatic Botany,* 27**,** 207-215.
Eckert, C. G., Lui, K., Bronson, K., Corradini, P. & Bruneau, A. 2003. Population genetic consequences of extreme variation in sexual and clonal reproduction in an aquatic plant. *Molecular Ecology,* 12**,** 331-344.
Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One,* 6**,** e19379.
Ereshefsky, M. 1992. Eliminative pluralism. *Philosophy of Science,* 59**,** 671-690.
Ernst, E. & Martienssen, R. Status of the *Lemna gibba* 7742a and *Lemna minor* 8627 genomes. Duckweed Forum, 2016.
Ferchaud, A. L., Laporte, M., Perrier, C. & Bernatchez, L. 2018. Impact of supplementation on deleterious mutation distribution in an exploited salmonid. *Evolutionary Applications,* 11**,** 1053-1065.
Flora of North America. 2020. *Lemna* L. Web documents [Online]. Available: http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=117936#KEY-1-7, http://www.efloras.org/object_page.aspx?object_id=11231&flora_id=1, http://www.efloras.org/object_page.aspx?object_id=11248&flora_id=1, http://www.efloras.org/object_page.aspx?object_id=11244&flora_id=1. [Accessed October 26th 2020].
Fourounjian, P., Fakhoorian, T. & Cao, X. H. 2020. Importance of Duckweeds in Basic Research and Their Industrial Applications. The Duckweed Genomes. Springer.
Frichot, E. & François, O. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution,* 6**,** 925-929.

Funk, W. C., McKay, J. K., Hohenlohe, P. A. & Allendorf, F. W. 2012. Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution,* 27**,** 489-496.

Galla, S. J., Forsdick, N. J., Brown, L., Hoeppner, M. P., Knapp, M., Maloney, R. F., Moraga, R., Santure, A. W. & Steeves, T. E. 2019. Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes,* 10**,** 9.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. & Buckler, E. S. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS One,* 9**,** e90346.

GO BOTANY. 2020. *Lemna minuta* Kunth; least duckweed [Online]. Available: https://gobotany.nativeplanttrust.org/species/lemna/minuta/ [Accessed October 26th 2020].

Goopy, J. P. & Murray, P. 2003. A review on the role of duckweed in nutrient reclamation and as a source of animal feed. *Asian-Australasian Journal of Animal Sciences,* 16**,** 297-305.

Halder, S. & Venu, P. 2012. The taxonomy and report of flowering in Lemna L.(Lemnaceae) in India. *Current Science,* 102**,** 1629-1632.

Hammer, Ø., Harper, D. A. & Ryan, P. D. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica,* 4**,** 9.

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H. & Li, Z. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science,* 5**,** 484.

Hebert, P. D., Cywinska, A., Ball, S. L. & Dewaard, J. R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 270**,** 313-321.

Hillman, W. S. 1961. The Lemnaceae, or duckweeds. *The Botanical Review,* 27**,** 221-287.

Hilty, O. 2019. Available: https://www.illinoiswildflowers.info/wetland/plants/gr_duckweed.html [Accessed October 26th 2020].

Ho, E. K., Bartkowska, M., Wright, S. I. & Agrawal, A. F. 2019. Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. *New Phytologist,* 224**,** 1361-1371.

Ho, K. H. E. 2017. The Effects of Asexuality and Selfing on Genetic Diversity, the Efficacy of Selection and Species Persistence. PhD Thesis, University of Toronto.

Hollingsworth, P. M., Graham, S. W. & Little, D. P. 2011. Choosing and using a plant DNA barcode. *PloS One,* 6**,** e19254.

Hulsen, K., Minne, V., Lootens, P., Vandecasteele, P. & Höfte, M. 2002. A chlorophyll a fluorescence-based *Lemna minor* bioassay to monitor microbial degradation of nanomolar to micromolar concentrations of linuron. *Environmental Microbiology,* 4**,** 327-337.

International Cassava Genetic Map Consortium 2015. High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3: Genes, Genomes, Genetics,* 5**,** 133-144.

Jackson, J. B., Buss, L. W., Cook, R. E. & Ashmun, J. W. 1985. Population biology and evolution of clonal organisms, Yale University Press New Haven.

Jacono, C. 2018. *Landoltia punctata* (G. Mey.) Les & D. J. Crawford: U.S. Geological Survey, Nonindigenous Aquatic Species Database, Gainesville, FL [Online]. Available: https://nas.er.usgs.gov/queries/FactSheet.aspx?SpeciesID=1116 [Accessed October 26th 2020].

Jansen, M. A., Gaba, V., Greenberg, B. M., Mattoo, A. K. & Edelman, M. 1996. Low threshold levels of ultraviolet- B in a background of photosynthetically active radiation trigger rapid degradation of the D2 protein of photosystem-II. *The Plant Journal,* 9**,** 693-699.

Janzen, D. H., Burns, J. M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert, P. D. & Grishin, N. V. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proceedings of the National Academy of Sciences,* 114**,** 8313-8318.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* 11**,** 94.

Kawecki, T. J. & Ebert, D. 2004. Conceptual issues in local adaptation. *Ecology Letters,* 7**,** 1225-1241.

Keddy, P. A. 1976. Lakes as islands: the distributional ecology of two aquatic plants, *Lemna minor* L. and *L. trisulca* L. *Journal of Ecology,* 57**,** 353-359.

Kielak, E., Sempruch, C., Mioduszewska, H., Klocek, J. & Leszczyński, B. 2011. Phytotoxicity of Roundup Ultra 360 SL in aquatic ecosystems: Biochemical evaluation with duckweed (*Lemna minor* L.) as a model plant. *Pesticide Biochemistry,* 99**,** 237-243.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences,* 102**,** 8369-8374.

Kulski, J. K. 2016. Next-generation sequencing-an overview of the history, tools, and "Omic" applications. *Next Generation Sequencing-Advances, Applications and Challenges***,** no volume? 3-60.

Küpper, A., Manmathan, H. K., Giacomini, D., Patterson, E. L., McCloskey, W. B. & Gaines, T. A. 2018. Population genetic structure in glyphosate-resistant and-susceptible Palmer amaranth (*Amaranthus palmeri*) populations using genotyping-by-sequencing (GBS). *Frontiers in Plant Science,* 9**,** 29.

Lacoul, P. & Freedman, B. 2006. Environmental influences on aquatic plants in freshwater ecosystems. *Environmental Reviews,* 14**,** 89-136.

Lahaye, R., Savolainen, V., Duthoit, S., Maurin, O. & Van der Bank, M. 2008. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. *Nature Precedings***,** not right.

Laird, R. A. & Barks, P. M. 2018. Skimming the surface: duckweed as a model system in ecology and evolution. *American Journal of Botany,* 105**,** 1962-1966.

Landolt, E. 1975. Morphological differentiation and geographical distribution of the *Lemna gibba-Lemna minor* group. *Aquatic Botany,* 1**,** 345-363.

Landolt, E. 1986. Biosystematic investigations in the family of duckweeds (*Lemnaceae*), Veröffentlichungen des Geobotanischen Institutes der ETH, Siftung, Rübel, Zürich, Switzerland.

Landolt, E. 2000. Contribution on the Lemnaceae of Ecuador. *Fragmenta Floristica et Geobotanica,* 45**,** 221-237.

Landolt, E. & Kandeler, R. 1987. Biosystematic investigations in the family of duckweeds (Lemnaceae), Vol. 4: the family of Lemnaceae-a monographic study, Vol. 2 (phytochemistry, physiology, application, bibliography). *Veroeffentlichungen des Geobotanischen Instituts der ETH, Stiftung Ruebel (Switzerland),* 2**,** 638.

Lemon, G. D. & Posluszny, U. 2000. Comparative shoot development and evolution in the Lemnaceae. *International Journal of Plant Sciences,* 161**,** 733-748.

Les, D. H., Crawford, D. J., Landolt, E., Gabel, J. D. & Kimball, R. T. 2002. Phylogeny and systematics of Lemnaceae, the duckweed family. *Systematic Botany,* 27**,** 221-240.

Les, D. H., Landolt, E. & Crawford, D. 1997. Systematics of theLemnaceae (duckweeds): Inferences from micromolecular and morphological data. *Plant Systematics and Evolution,* 204**,** 161-177.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-2993.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics,* 26**,** 589-595.

Li, W., Liu, L., Wang, Y., Zhang, Q., Fan, G., Zhang, S., Wang, Y. & Liao, K. 2020. Genetic diversity, population structure, and relationships of apricot (*Prunus*) based on restriction site-associated DNA sequencing. *Horticulture Research,* 7**,** 1-13.

Liu, Y., Yi, F., Yang, G., Wang, Y., Pubu, C., He, R., Xiao, Y., Wang, J., Lu, N. & Wang, J. 2020. Geographic population genetic structure and diversity of *Sophora moorcroftiana* based on genotyping-by-sequencing (GBS). *PeerJ,* 8**,** e9609.

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S. & Costich, D. E. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet,* 9**,** e1003215.

Malenfant, R. M., Coltman, D. W. & Davis, C. S. 2015. Design of a 9K illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Molecular Ecology Resources,* 15**,** 587-600.

Manel, S., Schwartz, M. K., Luikart, G. & Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution,* 18**,** 189-197.

Marconi, G., Landucci, F., Rosellini, D., Venanzoni, R. & Albertini, E. 2019. DNA barcoding as a tool for early warning and monitoring alien duckweeds (*Lemna* sp. pl.): the case of Central Italy. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology,* 153**,** 660-668.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal,* 17**,** 10-12.

Mayr, E. 1999. Systematics and the origin of species, from the viewpoint of a zoologist, Harvard University Press.

Midgley, G., Hannah, L., Millar, D., Rutherford, M. & Powrie, L. 2002. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography,* 11**,** 445-451.

Montana Field Guide. 2020. Lesser Duckweed - *Lemna minor*. Montana Field Guide. Montana Natural Heritage Program. [Online]. Available: http://FieldGuide.mt.gov/speciesDetail.aspx?elcode=PMLEM01020 [Accessed October 26th 2020].

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., Riera-Lizarazu, O., Brown, P. J., Acharya, C. B. & Mitchell, S. E. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences,* 110**,** 453-458.

Nauheimer, L., Metzler, D. & Renner, S. S. 2012. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytologist,* 195**,** 938-950.

Patterson, N., Price, A. L. & Reich, D. 2006. Population structure and eigenanalysis. *PLoS Genet,* 2**,** e190.

Pearman, P. B. 2001. Conservation value of independently evolving units: sacred cow or testable hypothesis? *Conservation Biology,* 15**,** 780-783.

Petit, R. J. & Excoffier, L. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution,* 24**,** 386-393.

Presting, G. G. 2006. Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function. *Botany,* 84**,** 1434-1443.

Provine, W. B. 2004. Ernst Mayr: genetics and speciation. *Genetics,* 167**,** 1041-1046.

R Core Team 2019. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. *URL https://www.R-project.org*.

Ratnasingham, S. & Hebert, P. D. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes,* 7**,** 355-364.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., Wilkie, A. O., McVean, G. & Lunter, G. 2014. Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics,* 46**,** 912-918.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., Elshire, R. J., Acharya, C. B., Mitchell, S. E. & Flint-Garcia, S. A. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology,* 14**,** 55.

Sáez, A. G. & Lozano, E. 2005. Body doubles. *Nature,* 433**,** 111-111.

Sakiroglu, M. & Brummer, E. C. 2017. Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theoretical and Applied Genetics,* 130**,** 261-268.

Sandler, G., Bartkowska, M., Agrawal, A. F. & Wright, S. I. 2020. Estimation of the SNP mutation rate in two vegetatively propagating species of duckweed. *G3: Genes, Genomes, Genetics***,** volume??? 1-23.

Sass, C., Little, D. P., Stevenson, D. W. & Specht, C. D. 2007. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PloS One,* 2**,** e1154.

Segelbacher, G., Cushman, S. A., Epperson, B. K., Fortin, M.-J., Francois, O., Hardy, O. J., Holderegger, R., Taberlet, P., Waits, L. P. & Manel, S. 2010. Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics,* 11**,** 375-385.

Shendure, J. & Aiden, E. L. 2012. The expanding scope of DNA sequencing. *Nature Biotechnology,* 30**,** 1084.

Shendure, J. & Ji, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology,* 26**,** 1135.

Simpson, G. G. 1951. The species concept. *Evolution,* 5**,** 285-298.

Sinkevičienė, Z. 2011. First records of *Lemna turionifera* in Lithuania. *Botanica Lithuanica,* 17**,** 59-61.

Slarkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics,* 16**,** 393-430.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science,* 236**,** 787-792.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., Normandeau, É., Laroche, J., Larose, S. & Jean, M. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PloS One,* 8**,** e54603.

Soto, J. C., Ortiz, J. F., Perlaza-Jiménez, L., Vásquez, A. X., Lopez-Lavalle, L. A. B., Mathew, B., Léon, J., Bernal, A. J., Ballvora, A. & López, C. E. 2015. A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. *BMC Genomics,* 16**,** 190.

Sree, K. S., Bog, M. & Appenroth, K. J. 2016. Taxonomy of duckweeds (Lemnaceae), potential new crop plants. *Emirates Journal of Food and Agriculture***,** 291-302.

Starfinger, U. & Stöcklin, J. 1996. Seed, pollen, and clonal dispersal and their role in structuring plant populations. Progress in Botany. Springer. not the proper citation for a book chapter

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

Tang, J., Zhang, F., Cui, W. & Ma, J. 2014. Genetic structure of duckweed population of *Spirodela, Landoltia* and *Lemna* from Lake Tai, China. *Planta,* 239**,** 1299-1307.

Thomas, C. D., Williams, S. E., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F. & Grainger, A. 2004. Biodiversity conservation: Uncertainty in predictions of extinction risk/Effects of changes in climate and land use/Climate change and extinction risk (reply). *Nature,* 430**,** 1-2.

Tokeshi, M. & Arakaki, S. 2012. Habitat complexity in aquatic systems: fractals and beyond. *Hydrobiologia,* 685**,** 27-47.

Torkamaneh, D., Laroche, J., Bastien, M., Abed, A. & Belzile, F. 2017. Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinformatics,* 18**,** 5.

Van De Wiel, C., Van Der Schoot, J., Van Valkenburg, J., Duistermaat, H. & Smulders, M. J. M. 2009. DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* Lf), from non-invasive relatives. *Molecular Ecology Resources,* 9**,** 1086-1091.

Van der Spiegel, M., Noordam, M. & Van der Fels-Klerx, H. 2013. Safety of novel protein sources (insects, microalgae, seaweed, duckweed, and rapeseed) and legislative aspects for their application in food and feed production. *Comprehensive Reviews in Food Science and Food Safety,* 12**,** 662-678.

VASCAN. 2019. Database of vascular plants of Canada (VASCAN). [Online]. Available: http://data.canadensys.net/vascan/taxon/6489?lang=en, http://data.canadensys.net/vascan/taxon/6491?lang=en, http://data.canadensys.net/vascan/taxon/6492?lang=en [Accessed October 26th 2020].

Vasseur, L., Aarssen, L. & Bennett, T. 1993. Allozymic variation in local apomictic populations of *Lemna minor* (Lemnaceae). *American Journal of Botany,* 80**,** 974-979.

Vendrami, D. L., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., Clark, M. S., Leese, F., McMinn, C. & Moore, H. 2017. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Royal Society Open Science,* 4**,** 160548.

Villablanca, F. X. 1994. Spatial and temporal aspects of populations revealed by mitochondrial DNA. Ancient DNA. Springer. again, not propre citation for a book chapter

Wang, W., Wu, Y., Yan, Y., Ermakova, M., Kerstetter, R. & Messing, J. 2010. DNA barcoding of the Lemnaceae, a family of aquatic monocots. *BMC Plant Biology,* 10**,** 205.

Ward, J. A., Bhangoo, J., Fernández-Fernández, F., Moore, P., Swanson, J., Viola, R., Velasco, R., Bassil, N., Weber, C. A. & Sargent, D. J. 2013. Saturated linkage map construction in *Rubus idaeususing* genotyping by sequencing and genome-independent imputation. *BMC Genomics,* 14**,** 2.

Winker, K. 2005. Sibling species were first recognized by William Derham (1718). *The Auk,* 122**,** 706-707.

Wright, S. 1940. The new systematics, Oxford University Press, London, UK.

Wright, S. 1943. Isolation by distance. *Genetics,* 28**,** 114.

Xiao, J. H., Wang, N. X., Li, Y. W., Murphy, R. W., Wan, D. G., Niu, L. M., Hu, H. Y., Fu, Y. G. & Huang, D. W. 2010. Molecular approaches to identify cryptic species and polymorphic species within a complex community of fig wasps. *PLoS One,* 5**,** e15067.

Xu, J., Zhao, H., Stomp, A.-M. & Cheng, J. J. 2012. The production of duckweed as a source of biofuels. *Biofuels Journal,* 3**,** 589-601.

Xu, S., Stapley, J., Gablenz, S., Boyer, J., Appenroth, K. J., Sree, K. S., Gershenzon, J., Widmer, A. & Huber, M. 2019. Low genetic variation is associated with low mutation rate in the giant duckweed. *Nature Communications,* 10**,** 1-6.

Xu, Y., Ma, S., Huang, M., Peng, M., Bog, M., Sree, K. S., Appenroth, K.-J. & Zhang, J. 2015. Species distribution, genetic diversity and barcoding in the duckweed family (Lemnaceae). *Hydrobiologia,* 743**,** 75-87.

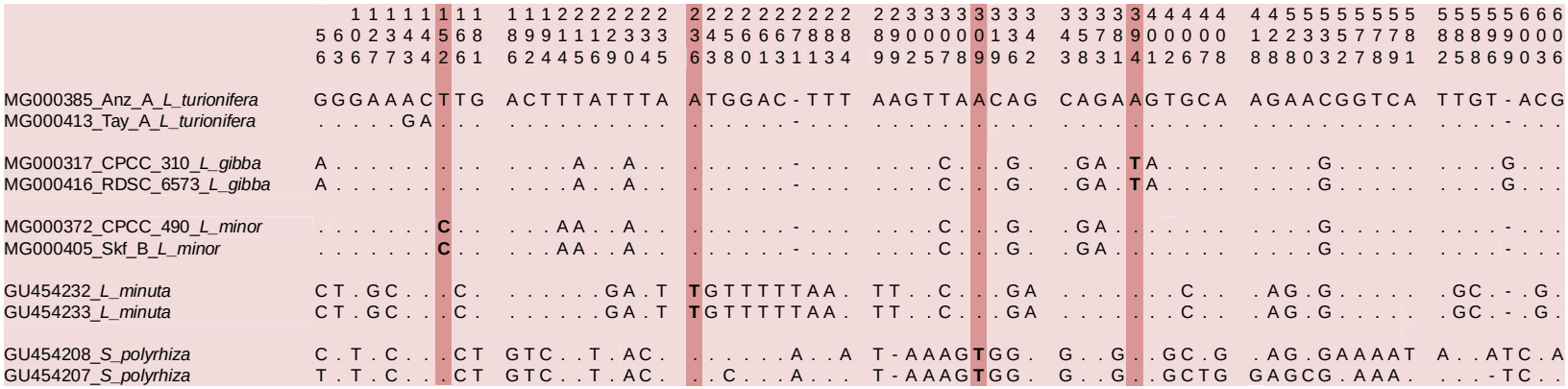
Xue, H., Xiao, Y., Jin, Y., Li, X., Fang, Y., Zhao, H., Zhao, Y. & Guan, J. 2012. Genetic diversity and geographic differentiation analysis of duckweed using inter-simple sequence repeat markers. *Molecular Biology Reports,* 39**,** 547-554.

Zhang, Y., Hu, Y., Yang, B., Ma, F., Lu, P., Li, L., Wan, C., Rayner, S. & Chen, S. 2010. Duckweed (*Lemna minor*) as a model plant system for the study of human microbial pathogenesis. *PLoS One,* 5**,** e13527.

Ziegler, P., Adelmann, K., Zimmer, S., Schmidt, C. & Appenroth, K. J. 2015. Relative in vitro growth rates of duckweeds (Lemnaceae)-the most rapidly growing higher plants. *Plant Biology,* 17**,** 33-41.

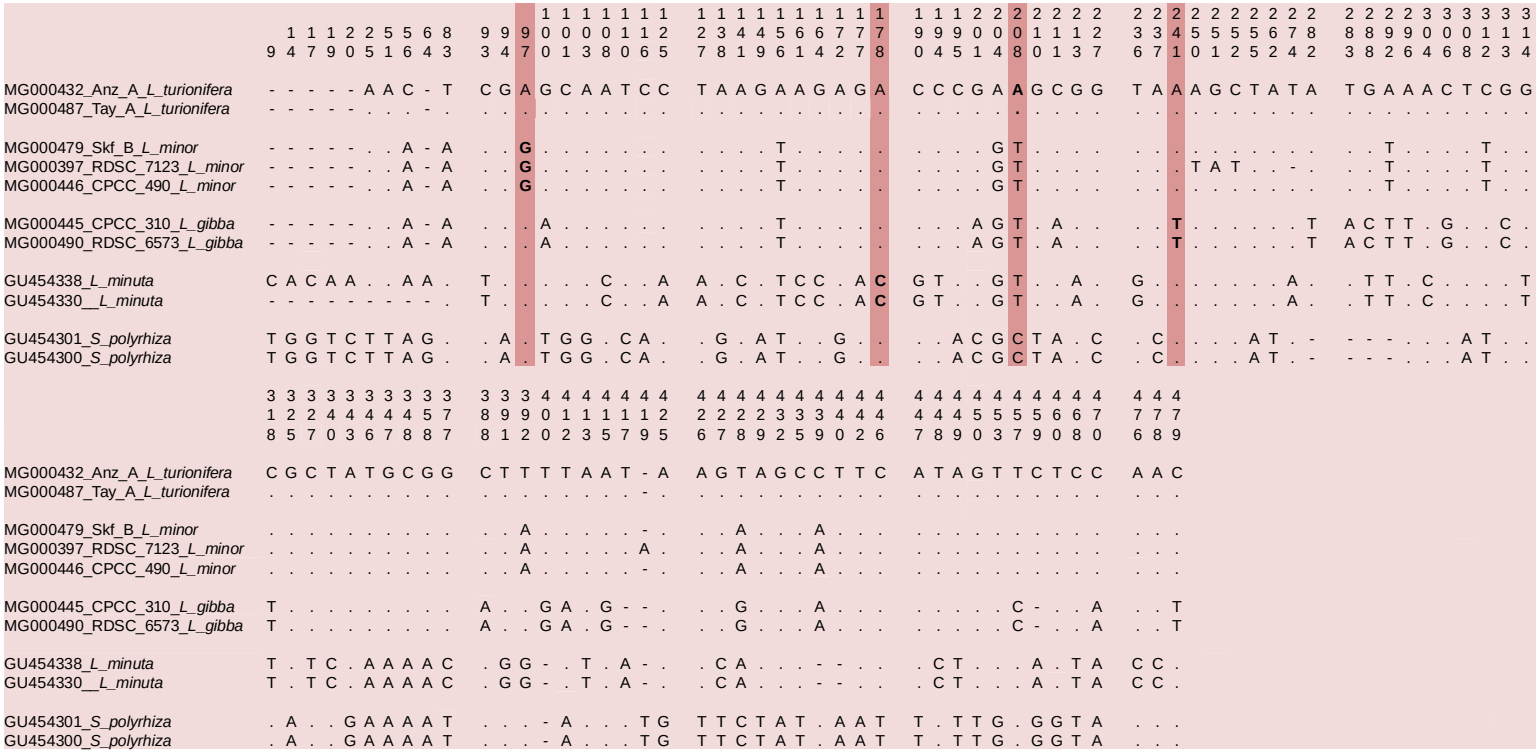
Appendix 1: Supplementary documents for Chapter 2

atpF-atpH



Supplementary Figure 1.1 Variable sites targeted to design primers for the five target *Lemna* species used in this study. The primers designed on the *atpF-atpH* sequences are shown in Figure 2.1. The positions of the *Lemna*-specific primers are shown above the sequence. NCBI accession numbers are included in the name of the *Lemna* species. Non-variable sites are indicated using dots and insertion-deletions are indicated using dashes.

psbK-psbI



Supplementary Figure 1.2 Variable sites targeted to design primers for the five target *Lemna* species used in this study. The primers designed on the *psbK-psbI* sequences are shown in Figure 2.1. The positions of the *Lemna*-specific primers are shown above the sequence. NCBI accession numbers are included in the name of the *Lemna* species. Non variable sites are indicated using dots and insertion-deletions are indicated using dashes.

Supplementary Table 1.1 Location, Sample ID, coordinates, collection dates and source for each individual used in this study. (A) strains collected by Barks et al., 2018, (B) additional strains supplied by ABMI; each sample was treated as a separate site as they are at least 50 km away from each other, (C) strains collected in southern Alberta as part of this study, and (D) strains provided by reference collections.

(A)

Location Name	Sample ID	Species ID	Museum Reference	Latitudes	Longitudes
Anzac	Anz A	LT	N/A	56.45	-111.04
Anzac	Anz B	LT	N/A	56.45	-111.04
Anzac	Anz C	LT	N/A	56.45	-111.04
Barrhead	Bar A	LT	N/A	54.15	-114.46
Barrhead	Bar B	LT	N/A	54.15	-114.46
Barrhead	Bar C	LT	N/A	54.15	-114.46
Camrose	Cam A	LT	N/A	52.89	-112.71
Camrose	Cam B	LT	N/A	52.89	-112.71
Camrose	Cam C	LT	N/A	52.89	-112.71
Caslan	Cas A	LT	N/A	54.66	-112.51
Caslan	Cas B	LT	N/A	54.66	-112.51
Caslan	Cas C	LT	N/A	54.66	-112.51
Coaldale	Cld A	LT	N/A	49.73	-112.62
Coaldale	Cld B	LT	N/A	49.73	-112.62
Coaldale	Cld C	LT	N/A	49.73	-112.62
Del Bonita	Dbn A	LT	N/A	49.03	-112.75
Del Bonita	Dbn B	LT	N/A	49.03	-112.75
Del Bonita	Dbn C	LT	N/A	49.03	-112.75
Denwood	Dwd A	LT	N/A	52.86	-110.76
Denwood	Dwd B	LT	N/A	52.86	-110.76
Denwood	Dwd C	LT	N/A	52.86	-110.76
Edson	Eds A	LT	N/A	53.61	-115.95
Edson	Eds B	LT	N/A	53.61	-115.95

Edson	Eds C	LT	N/A	53.61	-115.95
Elkwater	Elk A	LT	N/A	49.66	-110.27
Elkwater	Elk B	LT	N/A	49.66	-110.27
Elkwater	Elk C	LT	N/A	49.66	-110.27
Hairy Hill	Hhl A	LT	N/A	53.74	-112.07
Hairy Hill	Hhl B	LT	N/A	53.74	-112.07
Hairy Hill	Hhl C	LT	N/A	53.74	-112.07
Hanna	Han A	LT	N/A	51.50	-112.06
Hanna	Han B	LT	N/A	51.50	-112.06
Hanna	Han C	LT	N/A	51.50	-112.06
Hwy 28	Hwy A	LT	N/A	54.01	-113.15
Hwy 28	Hwy B	LT	N/A	54.01	-113.15
Hwy 28	Hwy C	LT	N/A	54.01	-113.15
Kehiwin	Khw A	LT	N/A	54.12	-110.82
Kehiwin	Khw B	LT	N/A	54.12	-110.82
Kehiwin	Khw C	LT	N/A	54.12	-110.82
Michelsen Marsh	Mch A	LT	N/A	49.54	-112.56
Michelsen Marsh	Mch B	LT	N/A	49.54	-112.56
Michelsen Marsh	Mch C	LT	N/A	49.54	-112.56
Park Lake	Prk A	LT	N/A	49.81	-112.92
Park Lake	Prk B	LT	N/A	49.81	-112.92
Park Lake	Prk C	LT	N/A	49.81	-112.92
Patricia	Pat A	LT	N/A	50.69	-111.67
Patricia	Pat B	LT	N/A	50.69	-111.67
Patricia	Pat C	LT	N/A	50.69	-111.67
Saskatoon Island	Stn A	LT	N/A	55.20	-119.06
Saskatoon Island	Stn B	LT	N/A	55.20	-119.06
Saskatoon Island	Stn C	LT	N/A	55.20	-119.06
Skiff	Skf A	LT	N/A	49.37	-111.80
Skiff	Skf B	LM	N/A	49.37	-111.80
Skiff	Skf C	LM	N/A	49.37	-111.80
Slave Lake	Slv A	LT	N/A	55.41	-114.80

	Slave Lake	Slv B	LT	N/A	55.41	-114.80
	Slave Lake	Slv C	LT	N/A	55.41	-114.80
	Taylorville	Tay A	LT	N/A	49.03	-113.12
	Taylorville	Tay B	LT	N/A	49.03	-113.12
	Taylorville	Tay C	LT	N/A	49.03	-113.12
	Valleyview	Val A	LT	N/A	55.17	-117.16
	Valleyview	Val B	LT	N/A	55.17	-117.16
	Valleyview	Val C	LT	N/A	55.17	-117.16
	Wandering River	Wan A	LT	N/A	55.20	-112.54
	Wandering River	Wan B	LT	N/A	55.20	-112.54
	Wandering River	Wan C	LT	N/A	55.20	-112.54
	Whitecourt	Wht A	LT	N/A	54.06	-115.83
	Whitecourt	Wht B	LT	N/A	54.06	-115.83
	Whitecourt	Wht C	LT	N/A	54.06	-115.83
	Winagami	Win A	LT	N/A	55.61	-116.76
	Winagami	Win B	LT	N/A	55.61	-116.76
	Winagami	Win C	LT	N/A	55.61	-116.76
	Young's point	Yng A	LT	N/A	55.13	-117.57
	Young's point	Yng B	LT	N/A	55.13	-117.57
	Young's point	Yng C	LT	N/A	55.13	-117.57
	Parkway Services, University of Lethbridge	Pws A	LT	N/A	49.67	-112.86
	Water Building, University of Lethbridge	Wat A	LT	N/A	49.68	-112.87
(B)	Chipewyan Lake-Wood Buffalo	Chi A	LT	VS-382-LEMNMIN	57.57	-113.39
	Chipewyan Lake-Wood Buffalo	Chi B	LT	VS-T382-LEMNMIN	57.57	-113.39
	Fort Chipewyan- Wood Buffalo NP	Fot A	LT	VS-236-LEMNMIN	58.37	-112.44
	Fort Chipewyan- Wood Buffalo	Fot B	LT	VS-240-LEMNMIN	58.25	-111.05
	Fort Mackay-Wood Buffalo	Fom A	LT	VS-354-LEMNMIN	57.69	-112.63
	Garden Creek-Wood Buffalo NP	Gar A	LT	VS-234-LEMNMIN	58.42	-113.05
	Grimshaw- Northern Lights	Gri A	LT	VS-649-LEMNMIN	56.25	-117.34
	Loon Lake- Northern Sunrise	Loo A	LT	VS-593-LEMNMIN	56.39	-114.75
	Ryley-Beaver	Ryl A	LT	UIS-1120-28	53.28	-112.54

St. Isidore- Northern Sunrise	Sti A	LT	VS-652-LEMNMIN	56.17	-116.44
Trout Lake-Opportunity	Tro A	LT	VS-626-LEMNMIN	56.22	-114.48
Blackfoot	Blf A	LT	UIS-1062-15	53.29	-110.14
Cadotte Lake	Cad A	LT	VS-W464-LEMNATUR	57.23	-116.16
Calling Lake	Cal A	LT	UIS-727-14	55.44	-112.50
Chipewyan Lake	Chi C	LT	VS-W534-LEMNATUR	56.69	-113.69
Chipewyan Lake	Chi D	LT	VS-W380-LEMNATUR	57.64	-114.06
Chipewyan Lake	Chi E	LT	VS-W503-LEMNATUR	56.83	-113.63
Chipewyan Lake	Chi F	LT	VS-W566-LEMNATUR	56.44	-113.39
Cochrane	Coc A	LT	VS-W1449LEMNATUR	51.28	-114.58
Conklin	Con A	LT	UIS-607-16	55.95	-110.32
Conklin	Con B	LT	UIS-669-7	55.64	-111.13
Fox Lake	Fox A	LT	UIS-W319-8	58.03	-114.65
Fox Lake	Fox B	LT	UIS-W322-LEMNTUR	57.92	-113.58
Fox Lake	Fox C	LT	UIS-W348-5	57.82	-114.62
Fox Lake	Fox D	LT	VS-W320-LEMNATUR	58.02	-114.33
Fox Lake	Fox E	LT	VS-W378-LEMNATUR	57.66	-114.75
Gordondale	Gor A	LT	VS-W767-LEMNATUR	55.68	-119.74
Gordondale	Gor B	LT	VS-W768-LEMNATUR	55.63	-119.44
Grimshaw	Gri B	LT	UIS-680-5	56.06	-117.45
Hythe	Hyt A	LT	VS-W799-LEMNATUR	55.45	-119.77
Hythe	Hyt B	LT	VS-W800-LEMNATUR	55.46	-119.42
Hythe	Hyt C	LT	VS-W831-LEMNATUR	55.29	-119.81
John D'Or Prairie	Joh A	LT	UIS-317-25	58.09	-115.24
John D'Or Prairie	Joh B	LT	VS-W377-LEMNATUR	57.72	-115.02
Kikino	Kik A	LT	UIS-920-17	54.36	-112.47
La Crête	Lac A	LT	VS-W433-LEMNATUR	57.47	-116.42
La Crête	Lac B	LT	VS-404-LEMNATUR	57.59	-116.12
La Crête	Lac C	LT	VS-W376-LEMNATUR	57.75	-115.41
La Crête	Lac D	LT	VS-W434-LEMNATUR	57.45	-116.18
La Crête	Lac E	LT	VS-433-LEMNATUR	57.47	-116.42

	La Glace	Lag A	LT	VS-W801-LEMNATUR	55.47	-119.18
	Little Smoky	Lit A	LT	UIS-W905-7	54.79	-117.14
	Longview	Lon A	LT	UIS-W1547-20	50.39	-114.59
	Loon Lake	Loo B	LT	UIS-622-LEMNTUR	56.32	-115.82
	Loon Lake	Loo C	LT	VS-622-SPIRPOL	56.32	-115.82
	Loon Lake	Loo D	LT	VS-623-LEMNATUR	56.25	-115.46
	Loon Lake	Loo E	LM	VS-W465-LEMNATUR	57.24	-115.91
	Loon Lake	Loo F	LT	VS-W622-LEMNATUR	56.32	-115.82
	Loon Lake	Loo G	LT	VS-W623-LEMNATUR	56.25	-115.46
	Manning	Man A	LT	VS-W463-LEMNATUR	57.30	-116.48
	Slave Lake	Slv D	LT	UIS-848-20	54.90	-114.57
	Slave Lake	Slv E	LT	VS-W816-LEMNATUR	55.09	-114.52
	Slave Lake	Slv F	LT	VS-W848-LEMNATUR	54.90	-114.57
	Trout Lake	Tro B	LT	VS-W565-LEMNATUR	56.47	-113.76
	Vilna	Vil A	LT	UIS-T955-15	54.13	-111.97
	Widewater	Wid A	LT	VS-W782-DUCKWEED	55.36	-115.05
	Woking	Wok A	LT	VS-W769-LEMNATUR	55.63	-119.09
(C)	Bassano	Bas A	LT	N/A	50.81	-112.38
	Bassano	Bas B	LT	N/A	50.81	-112.38
	Bassano	Bas C	LM	N/A	50.80	-112.40
	Bassano	Bas D	LT	N/A	50.78	-112.41
	Bow City	Bow A	LT	N/A	50.43	-112.32
	Brooks	Bro A	LT	N/A	50.55	-111.87
	Buck Lake	Buc A	LT	N/A	52.95	-114.78
	Carseland	Car A	LT	N/A	50.85	-113.49
	Circle E Wetland Complex	Cir A	LM	N/A	50.63	-112.20
	Circle E Wetland Complex	Cir B	LT	N/A	50.63	-112.20
	Circle E Wetland Complex	Cir C	LT	N/A	50.64	-112.20
	Circle E Wetland Complex	Cir D	LM	N/A	50.37	-112.21
	Circle E Wetland Complex	Cir E	LM	N/A	50.37	-112.21
	Circle E Wetland Complex	Cir F	LM	N/A	50.36	-112.23

Deer Creek Ranch	Dee A	LT	N/A	49.09	-111.52
Eight Mile Lake	Eig A	LT	N/A	49.78	-112.69
Enchant	Enc A	LT	N/A	50.16	-112.34
Enchant	Enc B	LM	N/A	50.17	-112.34
Enchant	Enc C	LM	N/A	50.17	-112.34
Fincastle Lake	Fin A	LM	N/A	49.82	-111.97
Fincastle Lake	Fin B	LT	N/A	49.82	-111.97
Fincastle Lake	Fin C	LM	N/A	49.82	-111.98
Foremost	For A	LT	N/A	49.48	-111.46
Forestvill	Fov A	LT	N/A	50.43	-112.34
Forestvill	Fov B	LT	N/A	50.44	-112.41
Frank Lake	Fnk A	LT	N/A	50.57	-113.73
Frank Lake	Fnk B	LT	N/A	50.57	-113.73
Frank Lake	Fnk C	LT	N/A	50.57	-113.73
Frank Lake	Fnk D	LT	N/A	50.57	-113.73
Frank Lake	Fnk E	LT	N/A	50.57	-113.73
Hays	Hay A	LT	N/A	50.09	-111.72
Hays Transfer Station	Hts A	LT	N/A	50.12	-111.79
Hill Spring	Hil A	LT	N/A	49.29	-113.62
Jefferson	Jef A	LT	N/A	49.06	-113.08
Keho Lake	Keh A	LM	N/A	49.93	-112.99
Keho Lake	Keh B	LM	N/A	49.93	-112.99
Keho Lake	Keh C	LM	N/A	49.93	-113.00
Keho Lake	Keh D	LM	N/A	49.92	-113.01
Keho Lake	Keh E	LM	N/A	49.93	-113.03
Keho Lake	Keh F	LM	N/A	49.95	-113.05
Kimball	Kim A	LT	N/A	49.07	-113.16
Lethbridge	Lth A	LT	N/A	49.69	-112.89
Lost Lake	Los A	LT	N/A	50.16	-112.30
Lost Lake	Los B	LM	N/A	50.16	-112.30
Lost Lake	Los C	LM	N/A	50.15	-112.30
Marr Lake	Mrr A	LT	N/A	49.32	-113.85

Marsh Lake South	Msh A	LT	N/A	49.38	-112.21
Mary Lake	Mar A	LT	N/A	49.03	-113.21
Mary Lake	Mar B	LT	N/A	49.03	-113.21
Mary Lake	Mar C	LT	N/A	49.03	-113.21
Medicine Wheel	Med A	LT	N/A	50.48	-112.41
Medicine Wheel	Med B	LT	N/A	50.48	-112.41
Medicine Wheel	Med C	LT	N/A	50.48	-112.40
Medicine Wheel	Med D	LT	N/A	50.50	-112.41
Park Lake	Prk D	LT	N/A	49.82	-112.92
Park Lake	Prk E	LT	N/A	49.81	-112.93
Park Lake	Prk F	LT	N/A	49.81	-112.93
Park Lake	Prk G	LM	N/A	49.82	-112.93
Prouty Lake	Pro A	LT	N/A	50.25	-112.42
Prouty Lake	Pro B	LT	N/A	50.25	-112.43
Prouty Lake	Pro C	LT	N/A	50.25	-112.43
Purple Springs Wetland Complex	Pur A	LT	N/A	49.87	-111.90
Purple Springs Wetland Complex	Pur B	LT	N/A	49.87	-111.90
Purple Springs Wetland Complex	Pur C	LT	N/A	49.87	-111.90
Purple Springs Wetland Complex	Pur D	LT	N/A	49.87	-111.90
Purple Springs Wetland Complex	Pur E	LT	N/A	49.82	-111.83
Purple Springs Wetland Complex	Pur F	LT	N/A	49.84	-111.86
Purple Springs Wetland Complex	Pur G	LT	N/A	49.84	-111.86
Purple Springs Wetland Complex	Pur H	LT	N/A	49.84	-111.86
Purple Springs Wetland Complex	Pur I	LT	N/A	49.83	-111.86
Range Road 243-Taylorville	RGE 243 A	LT	N/A	49.06	-113.14
Range Road 243-Taylorville	RGE 243 B	LT	N/A	49.06	-113.14
Range Road 243-Taylorville	RGE 243 C	LT	N/A	49.06	-113.14
Raymond Reservoir West	Ray A	LT	N/A	49.41	-112.65
Raymond Reservoir East	Ray B	LT	N/A	49.41	-112.69
Rock Lake	Roc A	LT	N/A	50.68	-111.97
Rolling Hills	Rol A	LT	N/A	50.22	-111.78
Scope Lake	Sco A	LM	N/A	50.06	-111.80

	Skiff	Skf D	LT	N/A	49.37	-111.79
	Skiff	Skf E	LT	N/A	49.37	-111.80
	Skiff	Skf F	LT	N/A	49.37	-111.79
	Skiff	Skf G	LM	N/A	49.37	-111.79
	Spring Coulee	Spc A	LT	N/A	49.34	-113.02
	Spring Coulee	Spc B	LT	N/A	49.33	-112.97
	Sugar Factory Pond	Sug A	LT	N/A	49.49	-112.65
	Taber Lake	Tab A	LM	N/A	49.80	-112.12
	Taber Lake	Tab B	LM	N/A	49.80	-112.12
	Taber Lake	Tab C	LM	N/A	49.80	-112.12
	Taylorville	Tay D	LT	N/A	49.04	-113.01
	Taylorville	Tay E	LT	N/A	49.04	-113.01
	Taylorville	Tay F	LT	N/A	49.03	-113.12
	Township Rd 114	Twp 114 A	LT	N/A	49.93	-113.18
	Township Road 12	TWP 12 A	LT	N/A	49.03	-113.18
	Tyrrell Lake	Tyr A	LT	N/A	49.37	-112.23
	Tyrrell Lake	Tyr B	LT	N/A	49.37	-112.23
	Tyrrell Lake	Tyr C	LT	N/A	49.37	-112.23
	Vauxhall	Vau A	LT	N/A	49.95	-112.10
	Verger	Vrg A	LT	N/A	50.92	-112.03
	Vernon Pond	Ver A	LT	N/A	49.39	-111.44
	Waterton River	Wtr A	LT	N/A	49.12	-113.85
(D)	L. minor RDSC 7123	RDSC 7123	*L. minor*	N/A	N/A	N/A
	L. turionifera RDSC 6573	RDSC 6573	****see note*	N/A	N/A	N/A
	L. gibba CPCC 310	CPCC 310	*L. gibba*	N/A	N/A	N/A
	L. minor CPCC 490	CPCC 490	*L. minor*	N/A	N/A	N/A
	L. minor CPCC 492	CPCC 492	*L. minor*	N/A	N/A	N/A
	L. trisulca CPCC 399	CPCC 399	*L. trisulca*	N/A	N/A	N/A
	S. polyrhiza RDSC 8790	RDSC 8790	*S. polyrhiza*	N/A	N/A	N/A
	L. minuta RDSC 6752	RDSC 6752	*L. minuta*	N/A	N/A	N/A

Sample ID gives the abbreviated name of each study site

LT = *L. turionifera* and LM = *L. minor*

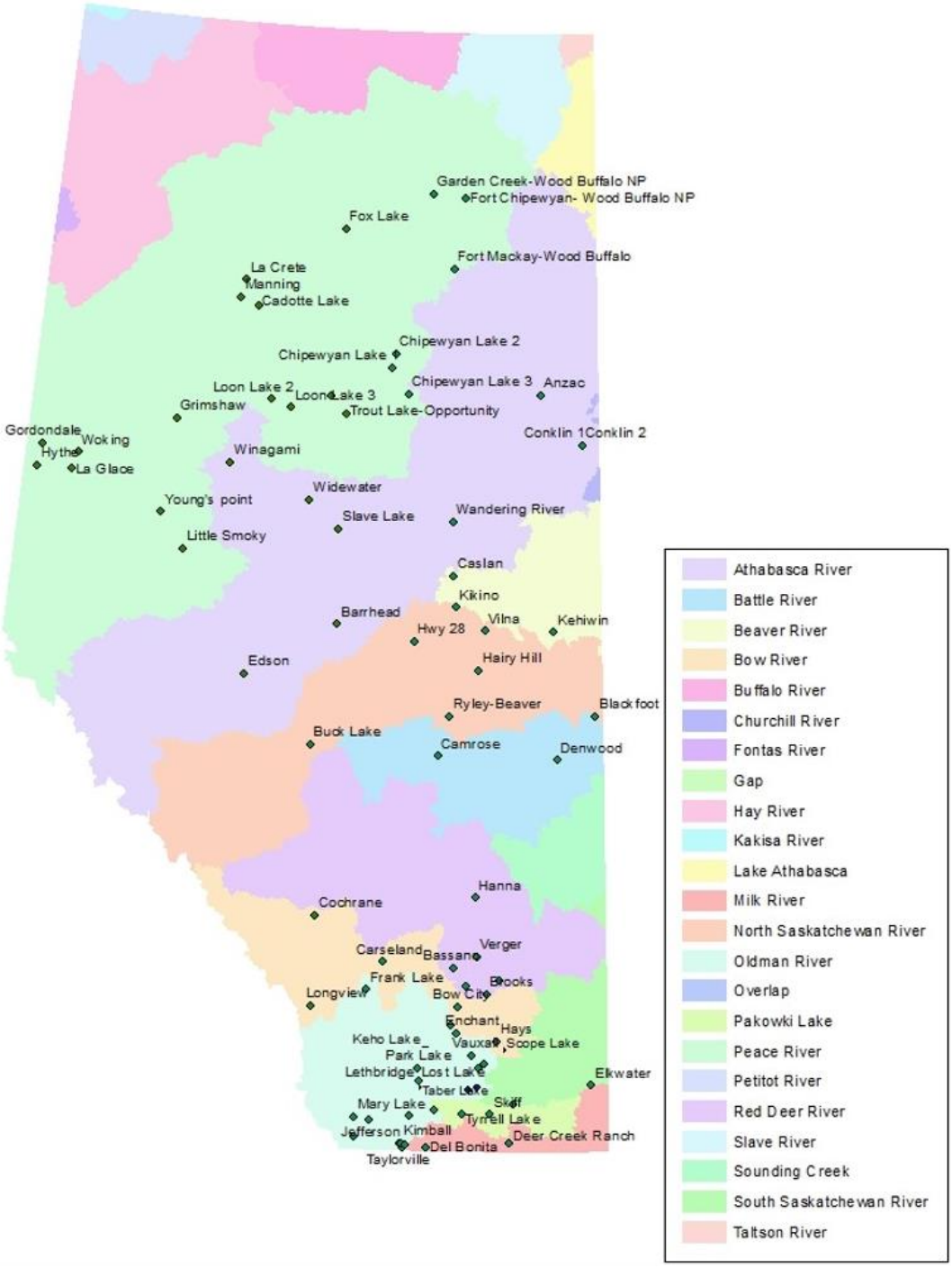
*** *L. turionifera* RDSC 6573; provided by the Rutgers Duckweed Stock Cooperative was identified as *L. gibba* after the molecular analyses done a) by Barks et al., 2018 and b) by the screening process of *Lemna* species using *atpF-atpH* and *psbK-psbI* species-specific primers developed in this study.

Supplementary Table 1.2 Surface water quality samples obtained from The Surface Water Quality Data and Online Tool webpage of The Alberta Government (https://www.alberta.ca/surface-water-quality-data.aspx). Sites with missing data are noted (-) and unless otherwise stated, all units are mg/L.

Location Name	Species Present	Latitude (Decimal Degrees)	Longitude (Decimal Degrees)	Secchi Disk Transparency (m)	Chlorophyll-a (µg/L)	Phosphorus Total (TP)	Phosphorus Total Dissolved (TDP)	Ammonia (NH_3-N)	Nitrate+Nitrite (NO_3+NO_2-N)	Nitrite (NO_2-N)	Total Kjeldahl Nitrogen	Total Nitrogen (TN)	TN:TP Ratio	Specific Conductance (Lab) (µS/cm)	Total Dissolved Solids (TDS)	pH (Lab)	Phenolphthalein Alkalinity ($CaCO_3$)	Total Hardness ($CaCO_3$)	Bicarbonate (HCO_3)	Carbonate (CO_3)	Calcium Dissolved (Ca)	Sodium Dissolved (Na)	Fluoride Dissolved (F)	Dissolved Organic Carbon (DOC)
Anzac	LT	56.44	-111.10	1.8	3.8	0.03		-	-	-	-	-	-	113	60	6.8	0.1	54.5	58	0.5	15.2	2.6	0.15	20.7
Caslan	LT	54.68	-112.57	1.3	9.8	0.03		-	-	-	-	-	-	386	222	8.38	7.9	190	226	9.5	36.1	15	0.28	-
Elkwater	LT	49.67	-110.29	1.4	8.4	0.05	-	-	0.02	0.005	-	-	-	462	254	8.6	-	229	225	1.5	26	17	0.27	-
Slave Lake	LT	55.44	-115.02	2.4	6.4	0.02	0.01	0.01	0.005	0.002	0.59	0.595	27	200	112	8.19	1	81.7	110	1	-	7.6	0.1	9.6
Winagami	LT	55.62	-116.73	1	107.5	0.42	-	-	0.02	0.001	-	-	-	450	273	8.82	-	219	229	17	43	13	0.19	-
Young's point	LT	55.10	-117.54	1.2	88.5	0.11	-	-	0.02	0.002	-	-	-	172	92	7.64	-	63	93	-	17	9	0.07	-
Cadotte Lake	LT	57.20	-116.17	0.3	185.4	0.41	0.08	0.03	0.006	0.003	4.13	4.136	10	273	158	8.38	-	132	140	2	38	9	0.12	29.2
Frank Lake	LT	50.57	-113.71	-	-	0.69	-	0.28	0.016	0.016	9	9.016	13	5510	3539	8.85	60	-	-	-	44	860	0.13	-
Rock Lake	LT	50.68	-111.97	-	-	0.07	-	0.02	0.003	0.003	0.76	0.763	10	223	123	8.21	0.1	-	-	-	21.4	7.1	0.31	-
Taber Lake	LM	49.80	-112.08	-	-	0.07	-	0.01	0.003	0.003	1	1.003	14	356	211	7.58	0.1	-	-	-	21.1	27	0.16	-
Keho Lake	LM	49.95	-112.98	1.2	7.3	0.01	0.01	0.05	0.005	0.003	0.24	0.245	22	280	180	8.45	2.2	140	140	2.6	35	20	-	2.3
Scope Lake	LM	50.07	-112.08	-	-	0.07	-	0.01	0.003	0.003	0.54	0.543	8	330	206	8.29	0.1	-	-	-	31	17	0.8	-
Park Lake	LM+LT	49.81	-112.93	2.2	5.7	0.02	0.01	0.05	0.005	0.003	0.25	0.255	13	270	150	8.45	2.1	120	120	2.5	30	7.2	-	1.9
Skiff	LM+LT	49.37	-111.80	-	-	0.18	-	0.09	0.003	0.003	1.28	1.283	7	2570	1917	7.56	0.1	-	-	-	88	400	0.24	-
Fincastle Lake	LM+LT	49.83	-111.98	-	-	0.05	-	0.01	0.008	0.003	0.74	0.748	15	297	144	7.76	0.1	-	-	-	22.2	13	0.11	-
Lost Lake	LM+LT	50.13	-112.30	-	-	0.06	-	0.01	0.003	0.003	0.68	0.683	12	434	263	7.84	0.1	-	-	-	27.4	30	0.29	-

LT = *L. turionifera* and LM = *L. minor*

Appendix 2: Supplementary documents for Chapter 3



Supplementary Figure 2.1 Sampling localities for *Lemna* species used in chapter 3. A total of 192 *Lemna* DNA samples were used in this study (48 belong to *L. minor* and 144 belong to *L. turionifera*). This design included sampling points along the watersheds in Alberta.

Supplementary Table 2.1 Samples included in the analyses from both *L. minor* and *L. turionifera* with major watersheds, river basins, location, **s**ample ID and coordinates (latitude and longitude).

Major Watersheds	**River Basins**	**Location Name**	**Sample ID**	**Latitude**	**Longitude**
L. minor					
South Saskatchewan	Pakowki Lake	Skiff	SkfB	49.37	-111.80
South Saskatchewan	Pakowki Lake	Skiff	SkfC	49.37	-111.80
South Saskatchewan	Pakowki Lake	Skiff	SkfG1	49.37	-111.80
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirA	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirD	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirE1	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirE5	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirF	50.63	-112.20
South Saskatchewan	Red Deer River	Bassano	BasC2	50.81	-112.38
South Saskatchewan	Oldman River	Enchant	EncB1	50.17	-112.34
South Saskatchewan	Oldman River	Enchant	EncB3	50.17	-112.34
South Saskatchewan	Oldman River	Lost Lake	LosB1	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosB4	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosB5	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosC1	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosC2	50.16	-112.30
South Saskatchewan	Oldman River	Taber Lake	TabA	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabB1	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabB2	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabC	49.80	-112.12
South Saskatchewan	Oldman River	Keho Lake	KehB	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehC	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehD	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehF	49.93	-112.99
South Saskatchewan	Oldman River	Taber Lake	TabC2	49.82	-112.92
South Saskatchewan	Oldman River	Taber Lake	TabC3	49.82	-112.92

South Saskatchewan	Bow River	Scope Lake	ScoA1	50.06	-111.80
South Saskatchewan	Bow River	Scope Lake	ScoA2	50.06	-111.80
South Saskatchewan	Bow River	Scope Lake	ScoA3	50.06	-111.80
Milk River	Milk River	Del Bonita	DbnC	49.03	-112.75
Milk River	Milk River	Del Bonita	DbnA	49.03	-112.75
Milk River	Milk River	Deer Creek Ranch	DeeA1	49.09	-111.52
Milk River	Milk River	Deer Creek Ranch	DeeA2	49.09	-111.52
Milk River	Milk River	Deer Creek Ranch	DeeA3	49.09	-111.52
Beaver river	Beaver River	Caslan	CasA	54.66	-112.51
Beaver river	Beaver River	Caslan	CasB	54.66	-112.51
Beaver river	Beaver River	Kehiwin	KhwA	54.12	-110.82
Beaver river	Beaver River	Kehiwin	KhwC	54.12	-110.82
North Saskatchewan	Battle River	Denwood	DwdA	52.86	-110.76
North Saskatchewan	Battle River	Denwood	DwdB	52.86	-110.76
North Saskatchewan	Battle River	Denwood	DwdC	52.86	-110.76
North Saskatchewan	Battle River	Camrose	CamA	52.89	-112.71
North Saskatchewan	Battle River	Camrose	CamB	52.89	-112.71
North Saskatchewan	North Saskatchewan River	Blackfoot	BlfA1	53.29	-110.14
North Saskatchewan	North Saskatchewan River	Vilna	VilA2	54.13	-111.97
North Saskatchewan	North Saskatchewan River	Hairy Hill	HhlA	53.74	-112.07
North Saskatchewan	North Saskatchewan River	Hairy Hill	HhlB	53.74	-112.07
North Saskatchewan	North Saskatchewan River	Kikino	KikA1	54.36	-112.47
North Saskatchewan	North Saskatchewan River	Kikino	KikA3	54.36	-112.47
North Saskatchewan	North Saskatchewan River	Hwy 28	HwyC	54.01	-113.15
Athabasca	Athabasca River	Edson	EdsA	53.61	-115.95
Athabasca	Athabasca River	Edson	EdsB	53.61	-115.95
Athabasca	Athabasca River	Barrhead	BarA	54.15	-114.46
Athabasca	Athabasca River	Wandering River	WanB	55.20	-112.54
Athabasca	Athabasca River	Wandering River	WanC	55.20	-112.54
Athabasca	Athabasca River	Anzac	AnzC	56.45	-111.04
Athabasca	Lesser Slave	Slave Lake	SlvE	55.09	-114.52
Athabasca	Lesser Slave	Widewater	WidA1	55.36	-115.05

Athabasca	Lesser Slave	Widewater	WidA2	55.36	-115.05
Athabasca	Lesser Slave	Widewater	WidA3	55.36	-115.05
Athabasca	Lesser Slave	Winagami	WinA	55.61	-116.76
Peace River	Peace River	Fort Mackay-Wood Buffalo	FomA3	57.69	-112.63
Peace River	Peace River	Loon Lake	LooB	56.32	-115.82
Peace River	Peace River	Loon Lake	LooG	56.25	-115.46
Peace River	Peace River	Chipewyan Lake	ChiE	56.83	-113.64
Peace River	Peace River	Chipewyan Lake	ChiF	56.44	-113.39
Peace River	Peace River	Grimshaw	GriB2	56.06	-117.45
Peace River	Peace River	Grimshaw	GriB	56.06	-117.45
Peace River	Peace River	Young's point	YngA	55.13	-117.57
Peace River	Peace River	Gordondale	GorB	55.68	-119.74
Peace River	Peace River	Hythe	HytA	55.46	-119.77
Peace River	Peace River	Hythe	HytB	55.46	-119.77
Peace River	Peace River	Hythe	HytC	55.46	-119.77
Peace River	Peace River	Cadotte Lake	CadA2	57.23	-116.16
Peace River	Peace River	Cadotte Lake	CadA3	57.23	-116.16
Peace River	Peace River	Manning	ManA1	57.30	-116.48
Peace River	Peace River	Manning	ManA2	57.30	-116.48
Peace River	Peace River	Manning	ManA3	57.30	-116.48
Peace River	Peace River	Manning	ManA4	57.30	-116.48
Peace River	Peace River	Manning	ManA5	57.30	-116.48
Peace River	Peace River	Conklin	ConA	55.95	-110.32
Peace River	Peace River	Conklin	ConB	55.64	-111.13
Peace River	Peace River	Little Smoky	Lit A	54.79	-117.14
South Saskatchewan	Bow River	Bow City	BowA	50.43	-112.32
South Saskatchewan	Bow River	Carseland	CarA	50.85	-113.49
South Saskatchewan	Bow River	Cochrane	CocA2	51.28	-114.58
South Saskatchewan	Red Deer River	Brooks	BroA	50.55	-111.87
South Saskatchewan	Red Deer River	Patricia	PatA	50.69	-111.67
South Saskatchewan	Red Deer River	Hanna	HanA	51.50	-112.06
South Saskatchewan	Red Deer River	Hanna	HanB	51.50	-112.06

South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirE4	50.63	-112.20
South Saskatchewan	Oldman River	Foremost	ForA	49.48	-111.46
South Saskatchewan	Oldman River	Tyrrell Lake	TyrB	49.37	-112.23
South Saskatchewan	Oldman River	Waterton River	WtrA	49.12	-113.85
South Saskatchewan	Oldman River	Hill Spring	HilA	49.29	-113.62
South Saskatchewan	Oldman River	Range Road 243-Taylorville	RGE243B	49.06	-113.14
South Saskatchewan	Oldman River	Enchant	EncA1	50.17	-112.34
South Saskatchewan	Oldman River	Park Lake	PrkD	49.82	-112.92
South Saskatchewan	Oldman River	Park Lake	PrkE	49.82	-112.92
South Saskatchewan	Oldman River	Prouty Lake	ProA	50.25	-112.42
South Saskatchewan	Pakowki Lake	Skiff	SkfE	49.37	-111.80
South Saskatchewan	South Saskatchewan	Elkwater	ElkA	49.66	-110.27
South Saskatchewan	South Saskatchewan	Elkwater	ElkB	49.66	-110.27

Supplementary Table 2.2 1 Samples included in the analyses from *L. minor* with Major watersheds, River Basins, Location, Sample ID and Coordinates.

Major Watersheds	River Basins	Location Name	Sample ID	Latitude	Longitude
South Saskatchewan	Pakowki Lake	Skiff	SkfB	49.37	-111.80
South Saskatchewan	Pakowki Lake	Skiff	SkfC	49.37	-111.80
South Saskatchewan	Pakowki Lake	Skiff	SkfG1	49.37	-111.80
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirA	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirD	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirE1	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirE5	50.63	-112.20
South Saskatchewan	Red Deer River	Circle E Wetland Complex	CirF	50.63	-112.20
South Saskatchewan	Red Deer River	Bassano	BasC2	50.81	-112.38
South Saskatchewan	Red Deer River	Bassano	BasC3	50.81	-112.38
South Saskatchewan	Oldman River	Enchant	EncB1	50.17	-112.34
South Saskatchewan	Oldman River	Enchant	EncC	50.17	-112.34
South Saskatchewan	Oldman River	Lost Lake	LosB1	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosB4	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosB5	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosC1	50.16	-112.30
South Saskatchewan	Oldman River	Lost Lake	LosC2	50.16	-112.30
South Saskatchewan	Oldman River	Taber Lake	TabA	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabB1	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabB2	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabC	49.80	-112.12
South Saskatchewan	Oldman River	Taber Lake	TabC2	49.82	-112.92
South Saskatchewan	Oldman River	Taber Lake	TabC3	49.82	-112.92
South Saskatchewan	Oldman River	Keho Lake	KehB	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehC	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehD	49.93	-112.99
South Saskatchewan	Oldman River	Keho Lake	KehF	49.93	-112.99
South Saskatchewan	Bow River	Scope Lake	ScoA1	50.06	-111.80
South Saskatchewan	Bow River	Scope Lake	ScoA2	50.06	-111.80
South Saskatchewan	Bow River	Scope Lake	ScoA3	50.06	-111.80