IMPROVING GENOMIC RESOURCES FOR THE STUDY OF INVASIVENESS IN EURASIAN WATERMILOFIL (*MYRIOPHYLLUM SPICATUM*) AND THEIR HYBRIDS

by

Jeffrey John Pashnick

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Plant Science – Plant Genetics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2021

ACKNOWLEDGEMENTS

There are far too many people who have supported me over the long, long journey that was my Ph.D. and I cannot hope to thank everyone in the room I have here. First, I'd like to thank my parents who have supported me from the very first day I stepped into academia through the very last step. Alex McMenamin and Devin Jones, I cannot thank you enough for the amount of support both of you provided in taking those steps (both literally and figuratively), and for being the best non-blood family anyone could hope for.

Academic lineage couldn't be a more accurate term for the amount of time I've had with my advisor, Dr. Ryan Thum. From meeting during a guest lecture in 2009 at GVSU, all the way through the final hours of my Ph.D. here at MSU you've been molding and mentoring at every step of my higher education. So much so that I'll write the same thing I wrote in my master's thesis "I cannot even begin to thank Dr. Ryan Thum…".

I would also like to thank my Ph.D. committee both past and present, Dr. Jennifer Lachowiec, Dr. Norman Weeden, Dr. Matthew Lavin, Dr. Bruce Maxwell, and Dr. Zach Gompert for all the study design, data collection and analysis, and writing help you've all provided. During my Ph.D. journey I developed a passion for teaching, and I would like to thank Dr. Tracy Dougher, Dr. Tracy Sterling, and Dr. Michael Giroux for all their support in giving me the opportunity to teach courses and develop as an instructor. The experience you all helped provide is second to none for my career beyond this Ph.D.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Genomics has revolutionized the way biologists ask fundamental questions about evolution. The thousands to tens of thousands of molecular markers generated through modern genomics increase the likelihood of detecting traits associated with a phenotype of interest. While genomics provides ever increasing evidence detecting these traits, they must be developed in each new system. *Myriophyllum spicatum* L. (Eurasian watermilfoil, EWM) and their hybrids with native *Myriophyllum sibiricum* Komorov (northern watermilfoil, NWM) are heavily managed aquatic plants in the United States. Genotypes both within and across these taxa and their hybrids can differ in their growth and herbicide response, prompting interest in determining which specific genotypes and genes will respond best to specific control tactics. However, because genotypes are unable to be distinguished by morphology, distinguishing genotypes requires molecular markers. EWM, NWM, and their hybrid are hexaploid (2n=6x=42) and developing these molecular markers requires accurately genotyping in a hexaploid with unknown chromosomal inheritance. The first manuscript of this dissertation empirically tested the genotyping information obtained from three commonly used molecular marker types, AFLPs, microsatellites, and GBS data. We found that while GBS markers have the lowest error rate, all molecular marker types provide the same genotype information. In the second chapter we used a mapping population, GBS data, and likelihood models to determine if watermilfoil was an allohexaploid, autohexaploid, or a mix between them. We found overwhelming evidence that watermilfoil is an allohexaploid across the genome. Finally, using the characteristics of each molecular marker type, the third chapter developed a cost-effective and information dense panel of microhaplotypes to genotype in watermilfoil. Microhaplotyping data can be shared across laboratories and promotes collaboration with weed managers by informing management with genetic information. Together, the work in this dissertation provides diploidized molecular markers and polyploid mode of inheritance to begin to connect genotype to herbicide response traits in watermilfoil.

GENERAL INTRODUCTION

Genomics has revolutionized the way biologists ask fundamental questions about evolution. Speciation as a result of whole genome duplication, or divergence in genetic variation, can be detected across whole genomes, whereas prior genetic information was limited to a few molecular markers. Hybridization, especially when cryptic diversity is present, can be more accurately distinguished due to the increase of markers. The thousands to tens of thousands of molecular markers generated through modern genomics increase the likelihood of detecting traits associated with a phenotype of interest. While genomics provides ever increasing evidence for these questions, they must be developed in each new system. Features in genomes, such as size, ploidy, and repetitive regions, complicate the development of these molecular markers and the use of these genomic techniques.

Polyploidy is an extremely common genome complexity in plants, with an estimated two-thirds of extant angiosperms being polyploid (Wood et al. 2009, Van de Peer et al. 2017). Further, it's been estimated that around 15% of all speciation in angiosperms is due to polyploidization (Wood et al. 2009). Speciation in polyploids can occur due to reproductive isolation, and genetic and morphological variation (Wood et al. 2009). There are two main types of polyploidy, allopolyploidy and autopolyploidy, with a third intermediate type, mixosomy (Soltis et al. 2016). Autopolyploids are defined by whole genome duplications within the same taxon, whereas allopolyploids result from whole genome duplication either before or after the hybridization of two taxa (Doyle and Sherman-Broyles 2017, Bourke 2018). The resolution of molecular markers, and in-turn, the genomic resources created from them, requires taking polyploid formation into consideration in each new system.

One of the first roadblocks in the development of genomic resources in polyploid plants is correctly analyzing molecular markers (Bourke 2018). The use of previous technologies for generating molecular markers, such as fragment-based sequencing, must treat polyploid data as dominant. While genotyping can be done with dominant molecular markers, dosage of those alleles cannot be calculated accurately in a polyploid and could lead to inaccurate genotyping (Bourke 2018). In fact, the use of dominant molecular markers in a polyploid severely limits studies on population genetics (allele frequencies cannot be calculated), bottom-up genetics approaches (linking genotype to phenotype through QTL mapping), and top-down genetics approaches (population level scans for selection) (Van de Peer et al. 2017, Bourke et al. 2018, Ravet et al. 2018). Advancements in sequencing technologies has drastically changed the way we can collect data for allele frequencies and further genetics approaches. Genotyping-by-sequencing via single or double-digest restriction associated DNA sequencing produces thousands of markers in comparison to tens to hundreds (Davey et al. 2011, Peterson et al. 2012). Markers produced using these methodologies are scattered across the genome and SNPs can be analyzed as haplotypes (dosage can be calculated), leading to higher accuracy in genotyping of polyploids. However, while all of these properties are beneficial, polyploidy complicates previously mentioned analyses depending on how that polyploid was formed.

Chromosomal inheritance directly impacts how quickly allele frequencies change in a population, experimental or natural, by the way chromosomes pair during meiosis. In an experimental population, the rate at which alleles segregate are different under an allopolyploid model than they are with an autopolyploid mode of inheritance. In a natural population, the rate at which novel alleles are combined in different genetic backgrounds is much faster in an

autopolyploid than they are in an allopolyploid due to pairing with homeologs versus homologs. On the other hand, heterosis is often able to be fixed in allopolyploids with a buffered breakdown effect in comparison to the random pairing of homeologs during meiosis in autopolyploids (Washburn and Birchler, 2014). Understanding the chromosomal inheritance pattern each novel system has is information critical to the models used to experimentally test if a phenotypic trait is linked to a specific genotype. For example, applying a model of allopolyploid inheritance to an autopolyploid system may lead to false positives indicating an area of the genome under low levels of selection or areas of the genome incorrectly associated with a trait of interest due fitting a true model of polysomic inheritance with a diploid model.

Herbicide resistance is increasingly problematic in the management of invasive plants, with 502 unique species by herbicide mode of action resistance cases confirmed (weedscience.org 2019). Understanding the rapid evolution of herbicide resistance in weed species and the genetic architecture of herbicide resistance requires genetic mapping (Ravet et al. 2018). However, the genomic resources available in weed species is limited with only four reference genomes available in *Conyza canadensis*, *Thlaspi arvense*, *Echinochloa crus-galli*, and *Raphnus raphanistrum* (Ravet et al. 2018). For many other weed species of interest without reference genomes (non-model weed species), polyploidy is a challenge that must be considered when building genomic resources (Ravet et al. 2018). In addition to genomic resources, gathering phenotypic data in many other weed species provides its own set of challenges due to accessibility of weed accessions (e.g. aquatic or rangeland weeds), laboratory growth conditions, and repeatability. Therefore, in order to understand the genetic basis of herbicide resistance in non-model weedy species, these resources must be built.

For example, understanding the genetic basis (the specific genes and alleles) of traits related to herbicide response has promise to develop genetic assays that could be used inform the management of invasive species. The development of genetic assays requires linking genotype to phenotype. The most commonly used and statistically powerful method to do this is quantitative trait locus (QTL) mapping. However, implementing a QTL mapping strategy requires two main components, segregating phenotypes for a trait or traits of interest, and accurate high-density genotype data (Basu et al. 2004, Delye 2013, Bourke et al. 2017, Ravet et al. 2018).

Myriophyllum spicatum L. (Eurasian watermilfoil, EWM) and their hybrids with native *Myriophyllum sibiricum* Komorov (northern watermilfoil, NWM) are heavily managed aquatic plants in the United States. Genotypes both within and across these taxa and their hybrids can differ in their growth and herbicide response (Berger et al. 2012, 2015, Thum et al. 2012, LaRue et al. 2013, Netherland and Willey 2017, Taylor et al. 2017), prompting interest in determining which specific genotypes will respond best to specific control tactics. However, because genotypes are unable to be distinguished by morphology, distinguishing genotypes requires molecular markers. EWM, NWM, and their hybrid are hexaploid (2n=6x=42) and developing these molecular markers requires accurately genotyping in a hexaploid with unknown chromosomal inheritance. Historically molecular markers (amplified fragment length polymorphisms and microsatellites) had to be treated as dominant data in watermilfoil due to hexaploidy. Therefore, accurately genotyping in watermilfoil therefore requires the use of sequence-based data as opposed to fragment-based data.

The aim of this dissertation is to develop molecular markers to accurately genotype in polyploid watermilfoil. Ultimately, those genomic resources developed here will enable studies

which require calculation of allele frequencies and linking genotype to phenotypes of herbicide resistance and invasive characteristics.

COMPARISON OF MOLECULAR MARKERS FOR DISTINGUISHING GENOTYPES OF EURASIAN WATERMILFOIL, NORTHERN WATERMILFOIL, AND THEIR HYBRIDS

Contribution of Authors and Co-Authors

Manuscript in Chapter: 2

Author: Jeff Pashnick

Contributions: Project inception and design, molecular marker data collection (AFLPs, microsatellites, and GBS), data analysis (some microsatellites, all AFLP and GBS data analysis), manuscript writing.

Co-Author: Ryan A. Thum

Contributions: Project inception and design, data analysis of some microsatellite data, and manuscript writing.

Manuscript Information

Jeff Pashnick and Ryan A. Thum

Journal of Aquatic Plant Management

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__x__ Published in a peer-reviewed journal

Aquatic Plant Management Society, Inc.
Submitted 5/31/2019
Published 01/2020

INTRODUCTION

Managed aquatic plant taxa can exhibit genetic variation that is relevant to their growth, impacts, and control efficacy. For example, genotypes of hydrilla (*Hydrilla verticillata* L.f. Royle) differ in their degree of sensitivity to fluridone (Michel et al. 2004), and genetic screening of hydrilla populations can therefore be used to predict whether fluridone treatment will be efficacious on the specific genotypes present in a water body (Benoit and Les 2013). For most species, where, when, and how genetic variation will be important for management decisions and outcomes is still unknown. However, molecular markers can be used to quantify and monitor genetic variation across space and time, which holds potential to identify specific genotypes of interest (e.g., putatively herbicide resistant genotypes) or changes in genetic variation that could signal important changes in population-level responses to management.

Invasive Eurasian watermilfoil (*Myriophyllum spicatum* L; EWM) is a highly managed aquatic plant in the United States. EWM hybridizes with native northern watermilfoil (*Myriophyllum sibiricum* Komarov; NWM), and EWM and hybrid genotypes can differ in their growth and herbicide response (Berger et al. 2012, 2015, Thum et al. 2012, LaRue et al. 2013, Netherland and Willey 2017, Taylor et al. 2017), prompting interest in determining which specific genotypes will respond best to specific control tactics. However, because genotypes are unable to be distinguished by morphology, distinguishing genotypes requires molecular markers.

Previous studies of genetic variation in EWM, NWM, and hybrid watermilfoil have used amplified fragment length polymorphisms (AFLPs) and microsatellite markers. AFLPs have been used to identify hybridization and demonstrate genetic diversity in EWM, NWM, and hybrids (Zuellig and Thum 2012, LaRue et al. 2013). AFLPs are relatively cheap, but they have

limited precision to distinguish closely-related genotypes from sequencing or scoring errors. This means that individuals that are the same genotype (i.e., ramets of the same genet) can be mistakenly considered as different genotypes due to these errors. Further, slight differences among laboratories in methods or scoring makes it difficult to compare AFLP data collected from different laboratories. Microsatellite markers have also been used to study genetic variation in EWM, NWM, and hybrids (Wu et al. 2013, 2015a,b, Taylor et al. 2017, Guastello and Thum 2018). Microsatellite scoring is generally more precise and repeatable than AFLPs, but inter-laboratory collaboration is still challenging due to slight differences in bench methods, scoring methods, and fragment analysis parameters. A potentially more serious limitation of microsatellites is the limited number of loci available to distinguish genotypes. This means that there is the potential for two unrelated individuals to share the same multi-locus microsatellite genotype (MMG) by chance through sexual reproduction, as opposed to by common ancestry through asexual reproduction.

Next-generation sequencing methods for genotyping by sequencing (GBS), such as ddRAD-Seq (double digest restriction associated DNA sequencing, Peterson et al. 2012), offer promise to improve molecular genotyping over AFLPs and microsatellites. These methods can produce hundreds to thousands of markers, which could be better at distinguishing genotypes than the limited number of microsatellite markers. At the same time, since ddRAD-Seq is sequence-based, it has the potential to be more precise and repeatable than AFLPs. To date, we are unaware of any studies of watermilfoil that have utilized ddRAD-Seq.

In this study, we compare genotype assignments using microsatellites, AFLPs, and GBS-SNPs. Specifically, because there is the potential for two unrelated individuals to share the same

MMG by chance instead of via common ancestry through asexual propagation, we tested whether individuals with the same MMG were estimated to be the same or different genotypes for AFLP and GBS-SNP markers.

MATERIALS AND METHODS

Sampling Strategy and DNA Extraction

Samples used for the genetic analysis in this study were collected for separate projects examining spatial and temporal patterns of genetic diversity in watermilfoil in Michigan and Minnesota (Thum, unpublished data). Briefly, plant samples for those projects were collected by randomly sampling one plant from a rake toss at each location on a point-intercept grid of the littoral zone of the lake (e.g., see Parks et al. 2016). This sampling typically resulted in 50-100 plants collected from each lake. Then, approximately 20 plants were randomly subsampled for genetic analysis using microsatellite markers (see description of methods below). All of these samples were extracted using the Qiagen DNeasy Plant Mini Kit[1] following the standard plant protocol.

In this study, we specifically tested the hypothesis of whether individuals with the same MMGs represented ramets of the same genets (i.e., clones, or individuals that share common ancestry through asexual reproduction) versus having the same MMG by chance through sexual reproduction. Therefore, we chose individuals for this study by identifying individuals that shared the same MMG within and/or among lakes in the Michigan and Minnesota datasets described above. To test whether genotypes were the same due to chance versus ramets of the same genet, we collected AFLP and SNP data for eight to 14 individuals sharing each unique

MMG in our study (see Table 1). For each MMG, we also gathered molecular marker data on two to three duplicate DNA extractions in order to quantify scoring and sequencing error.

Microsatellite Data Collection and Analysis

Eight microsatellite loci from Wu et al. (2013) were collected for the 192 samples in this project (Myrsp 1, Myrsp 5, Myrsp 9, Myrsp 12, Myrsp 13, Myrsp 14, Myrsp 15, and Myrsp 16). Each microsatellite locus was amplified using the protocols detailed in Wu et al. (2013). Fluorescently labeled microsatellite PCR products were sent to University of Illinois – Urbana-Champaign's Core Sequencing Facility for fragment analysis on an ABI 3730xl sequencer. Microsatellites were scored using GeneMapper (version 5.0). We identified unique MMGs using the R-package "PolySat" (Clark and Jasieniuk 2011). EWM, NWM, and hybrids are hexaploid, and therefore their exact genotypes cannot be determined because of potential differences in allele dosage in heterozygotes. Therefore, we treated microsatellite data as dominant (presence or absence of all possible alleles at each locus) (see also Wu et al. 2015a,b). We delineated distinct MMGs using in PolySat using Lynch distances and a threshold of 0.

AFLP Data Collection and Analysis

AFLP data were collected on 192 individuals (Table 1) using methods described in Zuellig and Thum (2012). We used three primer pairs during the selective amplification step of AFLPs, EcoRI-AGG/MseI-CAT, EcoRI-AGG/MseI-CAG, and EcoRI-AGG/MseI-CAC. AFLP data were run on an ABI3730xl DNA Sequencer at the University of Illinois – Urbana-Champaign Core Sequencing Facility. Fragment data were scored in GeneMapper version 5.0. We assessed the repeatability of AFLP loci using the methods of Ley and Hardy (2013)

comparing duplicates using the SpAgeDi software (Hardy and Vekemans 2002). Loci with low repeatability ($F_{ST} \leq 0.80$) were removed from the dataset, ultimately leaving us with 108 AFLP molecular markers.

We compared the estimated number of distinct AFLP genotypes to the number of distinct MMG genotypes and GBS genotypes (see below). We used AFLPop (Duchense and Bernatchez 2002) to distinguish genotypes while accounting for scoring error. We calculated the genetic distance between all individuals, and individuals that differed by less than the estimated scoring error rate (six differences) were considered to be the same genotype. To visualize relatedness among genotypes, and confirm that individuals belonging to the same MMG were most closely related to each other as opposed to other MMGs, we ran a principal coordinates analysis (PCoA) using GenAlEx version 6.503 (Peakall and Smouse 2006).

ddRAD Data Collection and Analysis

We also used a next-generation genotyping by sequencing approach, ddRAD sequencing (Peterson et al. 2012), to genotype our 192 individuals. This method is particularly useful for species that do not have a reference genome, like watermilfoil. Sequencing reads produced through ddRAD-Seq can be clustered together to create contigs leading to a consensus sequence of each locus in the data set, this *de novo* assembly is able to be done in the absence of a true reference genome. DNA from each individual was quantified using a Qubit v3 fluorometer, and 250ng of total DNA was used for the DNA library preparation and sequencing, which was conducted at the University of Texas-Austin Genomic Sequencing and Analysis Facility. The libraries were prepared using EcoRI and SphI restriction enzymes, and 350bp fragments were

selected. Data were sequenced using the Illumina HiSeq4000 for this project using a 2x125 paired-end run type.

We processed the ddRAD-Seq reads using a bioinformatics pipeline to produce a panel of single nucleotide polymorphisms (SNPs) for each of our 192 individuals. ddRAD data were analyzed using the GBS-SNP-CROP (Melo et al. 2016) pipeline, with the addition of SWEEP (Clevenger and Ozias-Akins 2015), and TASSEL v5.0 (Bradbury et al. 2007) software on Montana State Universities' Hyalite Computing Cluster. First, raw sequencing reads were trimmed of Illumina adapters using Trimmomatic v0.33 (Bolger et al. 2014) part of the GBS-SNP-CROP pipeline. After trimming, reads were filtered for quality and demultiplexed based on a unique barcode sequence for each individual. Next, reads were clustered and assembled into a mock reference using VSearch version 2.9.1 (Rognes et al. 2016) through GBS-SNP-CROP. In the clustering and mock reference assembly step, we used the following parameters: sequence identity of 0.90, read lengths of 300bp, and a p-value threshold of 0.01 from PEAR version 0.9.8 (paired-end read merging (Zhang et al. 2014)). After a mock reference was created, we aligned the reads from each individual to the mock reference using BWA version 0.7.12 (Li and Durbin 2010). SNPs were called with SAMtools version 1.7 (Li 2011), and reads were sorted and indexed with BWA. After SNPs were called they were then filtered using SWEEP (Clevenger and Ozias-Akins 2015), a tool designed to filter diploidized SNPs from polyploid SNPs.

The read processing and filtering performed above resulted in a final dataset of 542 diploid SNPs. We used the R-Package "poppr" (Kamvar et al. 2014) to calculate genetic distances among all individuals. We used the genetic distance among duplicate DNA extractions to estimate a genotyping error rate, and used this as the threshold (four differences) to distinguish

genets and ramets. We constructed a PCoA using TASSEL version 5.0 (Bradbury et al. 2007) to visualize relatedness among genotypes and confirm that individuals belonging to the same MMG were most closely related to each other as opposed to other MMGs. Finally, we compared estimated genets and ramets after correcting for sequencing error using SNPs to those estimated using MMGs and AFLPs.

RESULTS AND DISCUSSION

Microsatellites and AFLPs agreed on their estimates of genets and ramets. For AFLPs, individuals with the same MMG were most similar to one another (Fig. 1 A, C, and E). Moreover, the range of differences found between duplicates of the same individuals (n=32) ranged from zero to six differences. All individuals with the same MMG, regardless of whether they were from the same or different lakes, differed by fewer than this, suggesting that individuals with the same MMG are ramets of the same genet. In contrast, individuals with different MMGs differed by greater than six AFLP markers. Similarly, the estimated genets and ramets for GBS agreed with those estimated by MMGs. Individuals with the same MMG were most similar to each other with GBS markers (Fig. 1 B, D, and F). Our GBS error rate among individuals with duplicate extractions was just four SNPs. As with AFLP molecular markers, individuals with the same MMG were never more than four SNPs different, whereas individuals with different MMGs were always greater than four SNPs different.

This study therefore provides evidence that microsatellite, AFLP, and SNP based molecular markers are all effective at distinguishing ramets and genets in watermilfoil. Distinguishing genotypes of watermilfoil is important because different genotypes can respond differently to herbicides (Berger et al. 2012, 2015, Thum et al. 2012, Netherland and Willey

2017, Taylor et al. 2017). This recognition has sparked interest in characterizing the growth and herbicide response properties of different genotypes. For example, two lakes in our dataset (Bald Eagle and Coon, Table 1) contained distinct Eurasian and hybrid watermilfoil genotypes that would not likely be distinguished without molecular genotyping data (Moody and Les 2003, Parks et al. 2016). Since it is possible that different genotypes occurring in the same lake could exhibit different herbicide responses, herbicide studies of specific lakes should incorporate genotyping to ensure that different genotypes are tested separately.

As we have shown here, the same genotype can also occur in multiple lakes (see Table 1). For example, we included a genotype in this study that was isolated from Townline Lake, Michigan, and is known to exhibit fluridone resistance (see Berger et al. 2012, 2015, Thum et al. 2012), and diquat resistance (Netherland and Willey 2017). In an ongoing survey of genetic variation in Michigan, we identified this same MMG in several lakes (Thum, unpublished data), which raises concern that these lakes may also exhibit resistance to these herbicides. We therefore included individuals from two other lakes found to have this MMG in order to test with additional markers (AFLPs and SNPs) whether these same lakes harbored the same genotype, as opposed to having the same MMG by chance. Indeed, AFLPs and SNPs identified individuals in different lakes with this same MMG as ramets of the same genet, indicating that plants in lakes with the same MMG share common ancestry through asexual (clonal) reproduction.

We recognize that individuals that share common ancestry through asexual reproduction may differ by somatic mutations that determine their response to treatment (e.g., Michel et al. 2004). Herbicide experiments are therefore the only way to be certain that individuals with the same molecular genotype, such as the Townline genotype, in fact exhibit the same level of

resistance. Nevertheless, in the absence of herbicide information, it seems prudent for managers to assume that individuals that share ancestry via asexual reproduction will exhibit similar characteristics. Thus, our study illustrates the potential for molecular genotyping to identify genotypes that have been characterized previously, potentially eliminating the need for herbicide studies on each lake where a genotype is found, or at least providing important, interim information to managers unless and until an herbicide study proves otherwise.

While this study provides evidence that microsatellite, AFLP, and SNPs can all distinguish ramets and genets, we believe that SNPs provide several advantages. In our study, the genotyping error rate for GBS was 0.7% (four out of 542 SNPs differed among duplicate samples), while AFLP error rates were 5.6% (six out of 108 markers). Although microsatellite error rates were negligible for our study because we chose individuals known to have the same MMG, we have estimated microsatellite error rates from auto-scoring to be ~15% (Thum, unpublished data). This difference in error rate can be attributed to SNPs being generated via DNA sequencing, whereas AFLPs and microsatellites are based on assays of DNA sequence via fragment analysis. AFLPs and microsatellites can therefore be more heavily impacted by laboratory methods, instrumentation, and human judgement in comparison to DNA sequences. SNP data will therefore be more easily shared and integrated across laboratories.

A current challenge for employing GBS methods in watermilfoil genotyping is the prohibitive cost and turnaround time. The development of a cheaper and faster SNP assay should therefore be a priority for genetic surveys and monitoring of watermilfoil. For example, SNP assays such as microhaplotyping (Kidd et al. 2013) offer promise to significantly reduce the cost

per sample, but retain the resolution and accuracy of SNP-based analyses (Kidd et al. 2013, Campbell et al. 2015).

SOURCES OF MATERIALS

[1]DNeasy Plant Mini Kit. Qiagen Corp. 27220 Turnberry Lane, Suite 200, Valencia, CA 91355.

ACKNOWLEDGEMENTS

We would like to thank R. Newman and J. Eltawely for their help with the collection of the Minnesota samples used in this study. Additionally we would like to thank Jake Britton, Rick Buteyn, Jamiee Conroy, Breanne Garbill, Paul Hausler, Matt Novotny, and Syndell Parks for collection of the Michigan samples. This project was supported by the Minnesota Aquatic Invasive Species Center, with funding provided by the Minnesota Environmental and Natural Resources Trust Fund, as recommended by the Legislative-Citizen Commission on Minnesota Resources, and the Michigan Invasive Species Grant Program (www.michigan.gov/invasives).

LITERATURE CITED

Benoit LK, Les DH. 2013. Rapid identification and molecular characterization of phytoene desaturase mutations in fluridone-resistant Hydrilla *(Hydrilla verticillata*). Weed Sci. 61(01):32–40.

Berger ST, Netherland MD, MacDonald GE. 2012. Evaluating fluridone sensitivity of multiple hybrid and Eurasian watermilfoil accessions under mesocosm conditions. J. Aquat. Plant Manage. 50: 135-141.

Berger ST, Netherland MD, MacDonald GE. 2015. Laboratory documentation of multiple-herbicide tolerance to fluridone, norflurazon, and topramazone in a hybrid watermilfoil *(Myriophyllum spicatum* × *M. sibiricum*) population. Weed Sci. 63(01):235–241.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 23(19):2633–2635.

Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15:855-867.

Clark L V., Jasieniuk M. 2011. Polysat: An R package for polyploid microsatellite analysis. Mol. Ecol. Resour. 11(3):562–566.

Clevenger JP, Ozias-Akins P. 2015. SWEEP: A tool for filtering high-quality SNPs in polyploid crops. G3; Genes|Genomes|Genetics. 5(9):1797–1803.

Duchesne P, Bernatchez L. 2002. AFLPop: A computer program for simulated and rea population allocation, based on AFLP data. Mol. Eco. Notes. 2: 380-383.

Guastello PR, Thum RA. 2018. Mesocosm and field evaluation of Eurasian and hybrid watermilfoil response to endothall in Jefferson Slough, Montana. J. Aquat. Plant Manage. 56: 63-68.

Hardy OJ, Vekemans X. 2002. Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Eco. Notes. 2(4): 618-620.

Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, Ihuegbu N. 2013. Microhaplotype loci are a powerful new type of forensic marker. For. Sci. Int.: Genetics Supp. Series 4. e123-e124.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. *Poppr* : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281.

LaRue EA, Zuellig MP, Netherland MD, Heilman MA, Thum RA. 2013. Hybrid watermilfoil lineages are more invasive and less sensitive to a commonly used herbicide than their exotic parent (Eurasian watermilfoil). Evol. Appl. 6(3):462–471.

Ley AC, Hardy OJ. 2013. Improving AFLP analysis of large-scale patterns of genetic variation - A case study with the Central African lianas *Haumania spp*. (Marantaceae) showing interspecific gene flow. Mol. Ecol. 22(7):1984–1997.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21):2987–2993.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26(5):589–595.

Melo ATO, Bartaula R, Hale I. 2016. GBS-SNP-CROP: A reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics. 17(1):1–15.

Michel A, Arias RS, Scheffler BE, Duke SO, Netherland M, Dayan FE. 2004. Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). Mol. Ecol. 13(10):3229–3237.

Netherland MD, Willey L. 2017. Mesocosm evaluations of three herbicides on Eurasian watermilfoil (*Myriophyllum spicatum*) and hybrid watermilfoil (*Myriophyllum spicatum* x *Myriophyllum sibiricum*): Developing a predictive assay. J. Aquat. Plant Manage. 55: 39-42.

Peakall R, Smouse PE. 2006. GENALEX: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Eco. Notes. 6: 288-295.

Parks SR, McNair JN, Hausler P, Tyning P, Thum RA. 2016. Divergent responses of invasive watermilfoil to treatment with auxinic herbicides in a large Michigan lake. Lake and Res. Manage. 32: 366-372.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7(5).

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Taylor L-AL, Mcnair JN, Guastello P, Pashnick J, Thum RA. 2017. Heritable variation for vegetative growth rate in ten distinct genotypes of hybrid watermilfoil. J. Aquat. Plant Manage. 55(July).

Thum RA, Heilman, MA, Hausler PJ, Huberty LE, Tyning P, Wcisel DJ, Zuellig MP, Berger ST, Glomski LM, Netherland MD. 2012. Field and laboratory documentation of reduced fluridone sensitivity of a hybrid watermilfoil biotype (*Myriophyllum spicatum x Myriophyllum sibiricum*). J. Aquat. Plant Manage. 50: 141-146.

Wu Z-G, Yu D, Xu X-W. 2013. Development of microsatellite markers in the hexaploid aquatic macrophyte, *Myriophyllum spicatum* (Haloragaceae). Appl. Plant Sci. 1(2):1200230.

Wu Z, Ding Z, Yu D, Xu X. 2015a. Influence of niche similarity on hybridization between *Myriophyllum sibiricum* and *M. spicatum*. J. Evol. Biol. 28(8):1465–1475.

Wu Z, Yu D, Wang Z, Li X, Xu X. 2015b. Great influence of geographic isolation on the genetic differentiation of *Myriophyllum spicatum* under a steep environmental gradient. Sci. Rep. 5(October):1–9.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 30(5):614–620.

Zuellig MP, Thum RA. 2012. Multiple introductions of invasive Eurasian watermilfoil and recurrent hybridization with northern watermilfoil in North America. J. Aquat. Plant Manage. 50: 1-19.

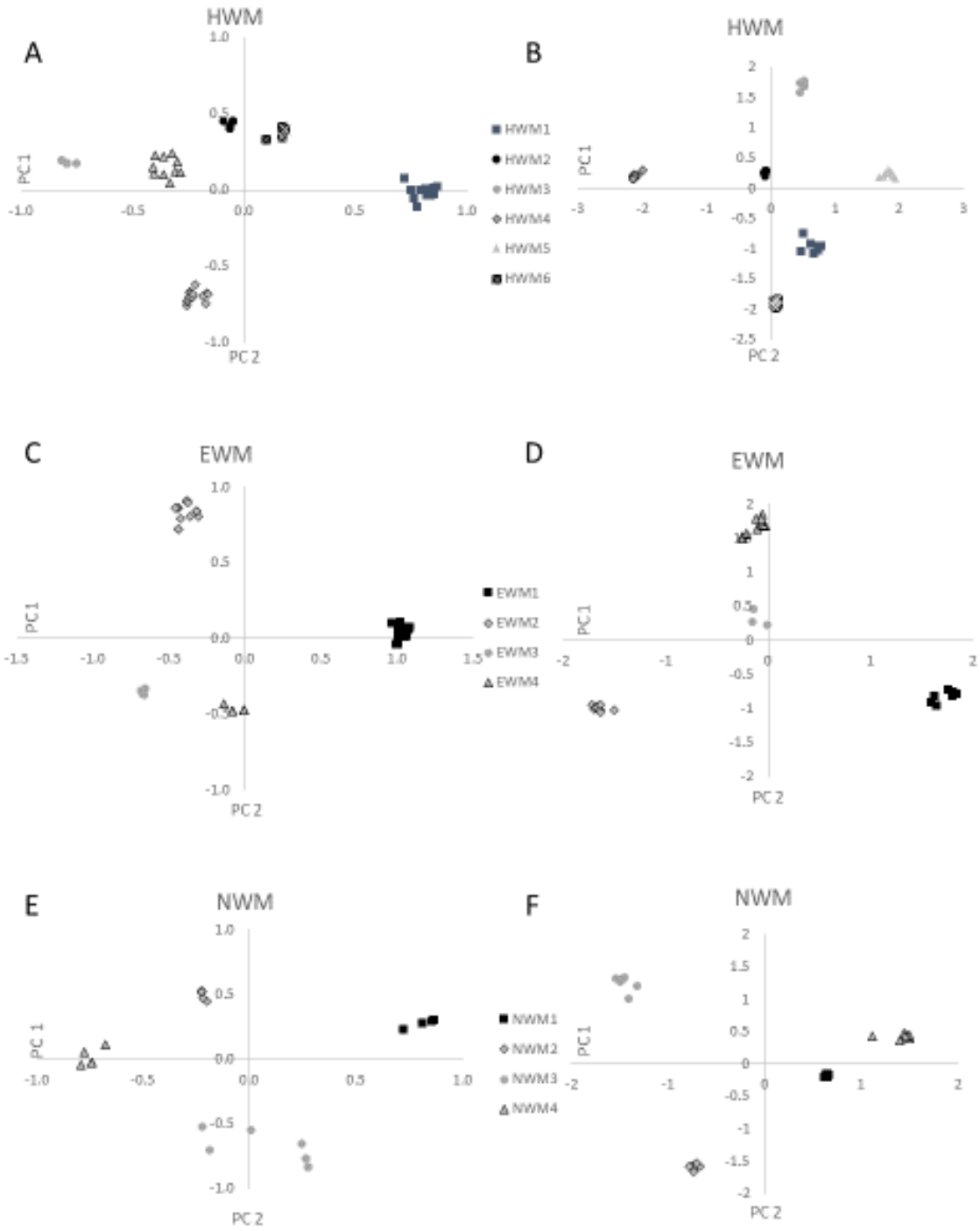
Table 1: The microsatellite multi-locus genotype (MMG), lake of sample origin, the county of origin, and the number of individuals sampled for all watermilfoil used in this study. Samples are grouped by MMG, and lakes sharing those genotypes are listed in order.

Genotype (MMG)	Lake	County of Origin	Number of Individuals
HWM1	Coon Lake, MN	Anoka	9
HWM1	Elmo Lake, MN	Anoka	9
HWM2	Ham Lake, MN	Anoka	9
HWM3	Bald Eagle Lake, MN	Ramsey	2
HWM3	Lake Josephine, MN	Ramsey	2
HWM3	Otter Lake, MN	Anoka	2
HWM3	Fish Lake, MN	Dakota	2
HWM3	Bone Lake, MN	Washington	2
HWM3	South Lindstrom Lake, MN	Chisago	2
HWM4	Lake Minnetonka, North Arm Bay, MN	Hennepin	11
HWM4	Lake Minnetonka, Gray's Bay, MN	Hennepin	2
HWM4	Lake Minnetonka, St. Alban's Bay, MN	Hennepin	2
HWM5	Sage Lake, MI	Ogemaw	5
HWM5	Budd Lake, MI	Clare	4
HWM6	Townline Lake, MI	Montcalm	3
HWM6	Lake Templene, MI	Sherman	3
HWM6	Muskellunge Lake, MI	Montcalm	3

EWM1	Rebecca Lake, MN	Hennepin	11
EWM1	Bald Eagle Lake, MN	Ramsey	2
EWM1	Mitchell Lake, MN	Hennepin	2
EWM1	Coon Lake, MN	Anoka	2
EWM1	Big Marine Lake, MN	Ramsey	2
EWM1	Riley Lake, MN	Carver	2
EWM2	Fish Lake, MI	Sherman	14
EWM3	Jordan Lake, MI	Ionia	14
EWM4	Lansing Lake, MI	Ingham	14
NWM1	Bald Eagle Lake, MN	Ramsey	9
NWM2	Lake Minne-Belle, MN	Meeker	9
NWM3	Mitchell Lake, MN	Sherburne	9
NWM4	Spectacle Lake, MN	Isanti	9

Figure 1: Principal Coordinates Analysis (PCoA) of AFLP (A, C, and E) and GBS (B, D, and F) molecular markers labeled by microsatellite multi-locus genotype. PC 1 and PC 2 are the first two axes of the PCA. In each panel, the first two PC axes combined explained the following percentages of the data A) 53.59%, B) 47.80%, C) 64.29%, D) 70.74%, E) 64.95%, and F) 55.67%. Each symbol used corresponds to a microsatellite multi-locus genotype (see Table 1), and each point on the graph represents an individual.

PCoA - AFLP Molecular Markers **PCoA - GBS Molecular Markers**



GENOTYPING-BY-SEQUENCING PROGENY FROM A CONTROLLED CROSS IDENTIFIES GENOME-WIDE ALLOPOLYPLOIDY AND DIPLOIDIZED MOLECULAR MARKERS FOR HEXAPLOID EURASIAN WATERMILFOIL

Contribution of Authors and Co-Authors

Manuscript in Chapter: 3

Author: Jeff Pashnick

Contributions: Project inception and study design, mapping population breeding and molecular marker data collection, data analysis, writing.

Co-Author: Jennifer Lachowiec

Contributions: Data analysis in R for allele counts and figure construction. Manuscript writing.

Co-Author: Alex Harkess

Contributions: Draft reference genome assembly, and writing.

Co-Author: Ryan A. Thum

Contributions: Project inception and study design, writing.

Manuscript Information

Jeff Pashnick, Jennifer Lachowiec, Alex Harkess, and Ryan A. Thum

The American Journal of Botany

Status of Manuscript:
_____ Prepared for submission to a peer-reviewed journal
__x__ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Botanical Society of America
Submitted 4/26/2021

Genotyping-by-Sequencing progeny from a controlled cross identifies genome-wide allopolyploidy and diploidized molecular markers for hexaploid Eurasian watermilfoil

Jeff Pashnick[1*], Jennifer Lachowiec[1], Alex Harkess[2], and Ryan A. Thum[1]

[1]Montana State University – Department of Plant Science and Plant Pathology. Plant Bioscience Building PO Box 173150 Bozeman, MT 59717

[2]Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, 36849 HudsonAlpha Institute for Biotechnology, Huntsville, AL, 35806

***Corresponding Author:** jeffpashnick@montana.edu

ABSTRACT

Allopolyploidy and autopolyploidy are common genome complexities in plants and can be distinguished on the basis of how their duplicated sets of chromosomes align during meiosis. Differences in chromosomal inheritance pattern changes the way molecular markers are generated and studied under each model. Eurasian watermilfoil (*Myriophyllum spicatum* L.; EWM) and interspecific hybrids with northern watermilfoil (*Myriophyllum sibiricum* Komarov; NWM) are invasive aquatic weeds in the United States. However, EWM, NWM, and their hybrids are hexaploid (2n = 6x = 42), which complicates genotyping and genetic analysis. We created a mapping population by self-pollinating a single hybrid EWM × NWM watermilfoil and determined genotypes genome-wide using genotyping-by-sequencing. These sequencing data were analyzed using both a de novo and reference-based approach to determine the most likely inheritance type in these polyploid hybrids. We found overwhelming evidence for allopolyploid inheritance of chromosomes in Eurasian, northern, and hybrid watermilfoil. Our confirmation of disomic inheritance and diploidized markers will improve the ability to conduct basic population genetic analyses of watermilfoil, including distinguishing sexual versus asexual (clonal) reproduction, population structure, hybridization, and introgression.

INTRODUCTION

Polyploidy is a common phenomenon in plants (Bourke et al. 2017, Van de Peer 2017, Gerard 2018, Clark et al. 2018). Two types of polyploidy – allopolyploidy and autopolyploidy – can be distinguished by how duplicated sets of chromosomes align during meiosis. In autopolyploidy, the duplicated chromosomes are homologous, and during meiosis they either form polyvalents with all homologous chromosomes, or bivalents that are random pairs of all homologous chromosomes (polysomic inheritance) (Doyle and Sherman-Broyles 2017). For allopolyploids, the duplicated chromosomes are homeologous, and homologous chromosomes from the different subgenomes form bivalents separately from each other such that each subgenome exhibits disomic inheritance (Doyle and Sherman-Broyles 2017). In addition, polyploids may exhibit a mixture of the two types of inheritance (mixosomy or segmental allopolyploidy) (Soltis et al. 2016, Bourke et al. 2017). Chromosomal inheritance in polyploid crops such as wheat (*Triticum aestivum*) and rapeseed (*Brassica napus*) have been determined, but chromosomal inheritance in many non-model polyploid plants remains uncharacterized (Bourke et al. 2017, 2018).

Historically, chromosomal inheritance was detected by visual observation of the karyotype of chromosomes during meiosis, where allopolyploids are likely to form bivalents, whereas autopolyploids are likely to form polyvalents (Bourke et al. 2017). However, this method is not always reliable indicator of chromosomal inheritance because autopolyploids may form bivalents via random pairing of homologous chromosomes instead of forming polyvalents. For example, autotetraploid potato forms bivalents instead of quadravalents (Bourke et al., 2017). In addition, one to several genes may be used to detect chromosomal inheritance through phylogenetic

relationships among those genes, but these studies are often limited to a subset of the total chromosomes in an organism and in-turn may not detect mixosomy (Lu et al. 2017). More reliably, auto- versus allopolyploid inheritance can be inferred from segregation ratios in genotype data (Clark et al. 2018, Gerard 2018). With next-generation genotyping-by-sequencing (GBS) approaches, segregation-based inference of auto- versus allopolyploidy can be used to detect inheritance patterns for thousands of molecular markers spread across chromosomes, providing a more complete picture of inheritance than could be inferred from one to several genes.

Eurasian watermilfoil (*Myriophyllum spicatum* L.; EWM), and interspecific hybrids with its sister species, northern watermilfoil (*Myriophyllum sibiricum* Komarov; NWM), are invasive aquatic weeds in the United States. Genetic variation in these taxa has become of interest to aquatic plant managers because genotypes vary in their responses to several herbicides that are commonly used to control them (Berger et al. 2012, 2015, Thum et al. 2012, Netherland and Willey 2017, Taylor et al. 2017, Chorak and Thum 2020). Distinguishing different genotypes enables characterization of herbicide response (e.g., Chorak and Thum 2020) and identifying genes or chromosomal regions that are associated with herbicide response. However, EWM, NWM, and their hybrids are hexaploid ($2n = 6x = 42$), which complicates genotyping and genetic analysis. Phylogenetic analysis of two nuclear genes (G3dph, adh, and internal transcribed spacer), and chloroplast markers by Lu et al. (2017) suggested that EWM is allopolyploid without currently extant diploid progenitor species, but that study covered only a very small portion of the overall genome. Further, because inheritance pattern of specific markers is unknown, previous genotyping studies with microsatellites (Thum et al. 2020, Taylor

et al. 2017, Wu et al. 2013, 2015AB) and AFLPs (Zuellig and Thum 2012, LaRue et al. 2013) were limited to treating marker data as dominant, which obscures the estimation of actual genotypes (e.g., homologous vs homeologous loci, allele dosage, etc.), allele frequencies, and any inferences based on these measures.

In this study we collected GBS data on progeny from a controlled cross to determine the most likely chromosomal inheritance type (auto- versus allopolyploid) for SNP-based molecular markers across the genome of hexaploid watermilfoil. These data were used to test whether EWM is allopolyploid across the genome, is autopolyploid across the genome, or whether it exhibits mixosomy. We find overwhelming evidence for allopolyploidy, and in doing so, we identify numerous diploidized molecular markers that can be used for future genotyping studies.

METHODS

Construction and Sequencing of the Mapping Population -

Mapping populations are effective for determining inheritance type from molecular markers, because segregation of alleles can be measured directly in the progeny of a known cross (Clark et al. 2018, Bourke et al. 2017). We created a mapping population by self-pollinating a single genotype of hybrid EWM × NWM watermilfoil from Hayden Lake, Idaho. Because this genotype is an interspecific hybrid, we expected it to be highly heterozygous, and therefore that its progeny would segregate at a large number of molecular markers. This parental genotype of hybrid watermilfoil has a genome-wide heterozygosity of 0.30 and are combinations of alleles from the northern and Eurasian watermilfoil parents specific to that population.

Sequencing data were collected for the mapping population using Restriction Associated Digest sequencing (Davey et al. 2010, RADseq). DNA from each individual was quantified using a Qubit v3 fluorometer, and 250 ng of total DNA was used for DNA library preparation and sequencing, which was conducted at the University Minnesota Genomics Center. The libraries were prepared using the ApeKI restriction enzyme and 159 individuals (139 unique and 20 randomized duplicate DNA extractions) were barcoded for each well and then plate-indexed for sequencing. Libraries were sequenced using Illumina NextSeq using 2x150 paired-end reads. Raw sequencing data were cleaned and de-multiplexed via unique barcode sequences using custom perl scripts (Dr. John Garbe, University of Minnesota Genomics Core).

***Genotyping and Chromosomal Inheritance* – de novo**

Allelic variants were called from the RADseq data using STACKs software (v2.53; Catchen et al. 2013), specifically designed for RADseq data analysis without a reference genome, and using the default parameters. A catalog of loci for the mapping population parent was created, allowing for up to six nucleotide mismatches for a read to be added to a locus. Progeny reads were aligned to this catalog to define "matches" for each progeny individual using the default alignment parameters.

***Reference-based Approach* –**

Without a reference sequence, STACKs filters out many potentially useful reads. To increase the number of molecular markers and compare chromosomal inheritance to the de novo analysis, we also performed variant calling using a draft reference genome for EWM (Harkess et al., unpublished). The details of the sequencing and assembling of the reference genome are beyond the scope of this paper and will be presented elsewhere. Briefly, an individual of EWM was

sequenced using PacBio long-read sequencing and Hi-C short read sequencing, and the draft reference genome was assembled with Flye v2.8.2 (Kolmogorov et al. 2019), scaffolded into draft pseudomolecules with Juicer (Durand et al. 2016), and manually polished with Juicebox (Durand et al. 2016). We aligned the sequencing data to the draft reference genome using bowtie2 v.2.3.5.1 using default parameters (Langmead and Salzberg 2012). Alignment files were converted to bamfiles, sorted, and variant called using samtools v.1.9 (Li et al, 2008, Li 2011). Data were then exported as a VCF file.

Chromosomal Inheritance and Mapping of Molecular Markers –

The genotype data from both the de novo and reference-based variant calling approaches were analyzed using R version 3.6.1 with "PolyRAD" v. 2.1 (Clark et al. 2018). The genotype data were filtered, requiring a minimum of 130 individuals with a locus to retain it, and a minimum number of individuals with a minor allele to be 5. PolyRAD calls genotypes in polyploids by using genotype likelihood. Genotype likelihoods (based on probability) at each locus are calculated for each individual under both the allo- and autopolyploid models of chromosomal inheritance. Finally, the genotype likelihood for all individuals at each locus is compared against the parents' genotypes with a chi-square analysis to determine if that locus is more likely to fit a model of allopolyploidy or autopolyploidy.

We used two different stringencies of filtering in our variant calling: a 'low stringency' filter at read depths of 3x for each allele to retain as many markers as possible for genotyping, and a 'high stringency' filter at read depths of 10x for each allele to ensure higher quality genotyping data.

RESULTS

Genotyping was completed in the progeny resulting from self-pollinating a hybrid of EWM and NWM to determine the pattern of chromosomal inheritance in hybrids genome-wide. RAD-seq produced on average 753,350 reads per individual at an average read depth of 20x. Principal coordinates analysis as part of *PolyRAD* was used to determine if an individual's sequencing error was too great to keep in the analysis, and 15 individuals were removed (Supp. Fig. 1).

With de novo variant calling and filtering, we obtained a total of 385 markers with our 3x coverage criterion, and a total of 152 markers with our 10x coverage criterion. Of the 385 markers in the 3x coverage dataset, 379 (98.4%) had segregation patterns that were more consistent with allopolyploid inheritance (Fig. 1A). Of the remaining six markers (1.6%) that had a higher likelihood autohexaploid inheritance, four were identified as chloroplast markers, and the other two did not map to any of the 21 chromosomes in the draft genome. For the higher stringency read depth (10x), all 152 molecular markers had a higher likelihood of allopolyploid inheritance (Fig. 1B).

With reference-based variant calling and filtering, we obtained a much larger number of markers: 26,219 using the 3x coverage criterion, and 12,739 using the 10x coverage criterion. Of the markers in the 3x analysis, 25,359 (96.7%) had a higher likelihood of allopolyploidy (Fig. 3A). The remaining molecular markers (860) showed a higher likelihood to be inherited as autohexaploid. These 860 molecular markers with higher likelihoods of autohexaploidy tended to be alleles with higher read depths, but that rule does not hold throughout the dataset (Supp. Figure 2). For example, the marker with the highest average read depth (Chromosome6_15506826 at 33x), had a p-value of 0.08 for the allopolyploid model and a p-

value of $5.72x10^{-207}$ for autopolyploid. With the 10x coverage filtering criterion, the proportion of markers with higher likelihoods of allohexaploid inheritance increased to 99.7% (12,703 markers; Fig. 3B). Of those 36 molecular markers that did not support allohexaploid inheritance, 20 of them map to the 21 largest chromosomes of the draft reference genome. The 16 remaining molecular markers with higher evidence for autohexaploidy map to contigs and debris unable to be incorporated into the largest chromosomes of the reference genome, and when run through nBLAST return hits with various other organisms or no significant results (Supp. Table 1).

For the de novo approach, none of the molecular markers mapped to the 21 largest chromosomes of the reference genome. We suspect that while this is real watermilfoil DNA sequence variation, that de novo analysis would be more likely to call variants in highly variable regions which are difficult to uniquely assign to the largest 21 chromosomes of the draft reference. From the reference-based analyses, the molecular markers that were most likely to be inherited as allopolyploid were found across all 21 chromosomes (Table 1). Only the reference-based analysis identified markers as autopolyploid, and these mapped to eight of the 21 chromosomes in the draft genome assembly (Table1).

DISCUSSION

In this study, we found overwhelming evidence for allopolyploid inheritance of chromosomes in Eurasian, northern, and hybrid watermilfoil. The vast majority of alleles (91% in the de novo and 99% in the reference-based Figs.2 and 4) found only one or two alleles per individual, which in and of itself suggests diploidized molecular markers. Nevertheless, across our de novo and reference-based analyses at two different stringencies of filtering based on sequence coverage, 97-100% of all of the molecular markers that we analyzed had higher

likelihoods of allopolyploid inheritance compared to autopolyploid inheritance. In addition, these allopolyploid molecular markers were distributed across all 21 chromosomes in the Eurasian watermilfoil genome assembly, strongly suggesting that each of the 21 pairs of chromosomes is inherited disomically. Our study therefore corroborates Lu et al.'s (2017) inference that Eurasian watermilfoil is allopolyploid but deepens our understanding of inheritance as allopolyploid across the genome, whereas their inference was based on only three nuclear genes that could have missed any autopolyploid inheritance in the case of mixosomy or segmental allopolyploidy.

We did find a small proportion of molecular markers across our analyses that had higher likelihoods of autopolyploid inheritance. In many cases, these molecular markers could be dismissed as errors. For example, in our de novo analyses, all six of the markers estimated to be autopolyploid did not map to any of the 21 chromosomes in the draft genome assembly; four of these were determined to be chloroplast markers, and the other two are suspected contaminants. Similarly, many of the molecular markers identified as autopolyploid in our reference-based analyses did not map to any of the 21 chromosomes. Furthermore, increasing the sequencing coverage stringency increased the proportion of markers that were estimated to be allopolyploid over autopolyploid, and therefore the autopolyploid markers may be artifacts of low sequencing coverage for some molecular markers. To check for biases in read depth versus chi-square ratio for determining the more likely model, we graphed those values across the genome and no pattern was found between the two (Supp. Fig. 2). Nevertheless, we did identify 20 molecular markers in our reference-based analysis that had higher likelihoods of autopolyploidy (Table 1, Supp. Fig. 2). It is therefore possible that a small proportion of the several chromosomes may be

inherited as segmental allopolyploids or loci have not, through evolution, fully resolved their inheritance between the two models (Ward 2000).

The confirmation of disomic inheritance across the genome, along with genotyping by sequencing approaches, stands to drastically improve genotyping studies for Eurasian, northern, and hybrid watermilfoil. Previous genotyping studies using microsatellites have treated them as dominant data because several alleles were present at each marker, and thus it was impossible to distinguish whether and which alleles were homologous versus homeologous. As such, inferences based on allele frequencies, or exact genotypes, were impossible. However, in this study we show that genotyping by sequencing using RADseq yields many diploidized molecular markers that can provide accurate genotype and allele frequency information. Even our de novo variant calling yielded a relatively large number of diploidized molecular markers (~150-350) compared to the 20 currently available microsatellite markers developed by Wu et al. (2013) that are relegated to analysis as dominant markers. This approach will improve the ability to conduct basic population genetic analyses of watermilfoil, including distinguishing sexual versus asexual (clonal) reproduction (e.g., calculating probabilities of identity with molecular marker data), population structure, hybridization, and introgression. Variant calling using a draft genome assembly vastly increased the density of diploidized markers across the genome, which should facilitate studies looking to associate genotypes and phenotypes using quantitative trait locus mapping or genome wide association approaches.

Acknowledgments

The authors thank Dr. Norman Weeden, Gregory Chorak, Zach Kuzniar, Hannah Hoff, Katie Gannon, and Leah Simantel for thoughtful feedback in the writing of this manuscript, and for the many hours of greenhouse assistance for setup and take-down. Funding for this project was provided by the Robert L. Johnson memorial research grant awarded to Jeff Pashnick from the Midwest Aquatic Plant Management Society. Additional support was provided by the Montana Agricultural Experiment Station (Project MONB00249).

Author Contributions

Author: Jeff Pashnick

Contributions: Project inception and design, mapping population breeding and molecular marker data collection, data analysis, writing.

Co-Author: Jennifer Lachowiec

Contributions: Data analysis in R for allele counts and figure construction. Manuscript writing.

Co-Author: Alex Harkess

Contributions: Draft reference genome assembly, and writing.

Co-Author: Ryan A. Thum

Contributions: Project inception and design, manuscript writing.

Data Availability Statement

Raw sequencing data for this project has been uploaded to the NCBI Sequence Read Archive at: https://submit.ncbi.nlm.nih.gov/subs/sra/SUB9507880/overview

Custom scripts used in raw sequence data processing may be found at: https://github.com/pashnickj/Scripts-for-Pashnick-et-al.-2021

LITERATURE CITED

Berger S.T., Netherland M.D., MacDonald G.E. 2012. Evaluating fluridone sensitivity of multiple hybrid and Eurasian watermilfoil accessions under mesocosm conditions. *J. Aquat. Plant Manage.* 50: 135-141.

Berger, S. T., M. D. Netherland, and G. E. MacDonald. 2015. Laboratory Documentation of Multiple-Herbicide Tolerance to Fluridone, Norflurazon, and Topramazone in a Hybrid Watermilfoil (*Myriophyllum spicatum* × *M. sibiricum*) Population. *Weed Science* 63: 235–241.

Bourke, P. M., P. Arens, R. E. Voorrips, G. D. Esselink, C. F. S. Koning-Boucoiran, W. P. C. van't Westende, T. Santos Leonardo, et al. 2017. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant Journal* 90: 330–343.

Bourke, P. M., R. E. Voorrips, R. G. F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. *Frontiers in Plant Science* 9: 1–17.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Chorak G and Thum RA. 2020. Identification of resistant clones of Eurasian (Myriophyllum spicatum) and hybrid (*M. spicatum X M. sibricum*) watermilfoil to an operational rate of fluridone. *Invasive Plant Science and Management*.

Clark, L. V., A. E. Lipka, and E. J. Sacks. 2018. polyRAD: Genotype calling with uncertainty from sequencing data in polyploids and diploids. *bioRxiv* 9252: 380899.

Davey, J. W., and M. L. Blaxter. 2011. RADSeq: Next-generation population genetics. *Briefings in Functional Genomics* 10: 108.

Doyle, J. J., and S. Sherman-Broyles. 2017. Double trouble: taxonomy and definitions of polyploidy. *New Phytologist* 213: 487–493.

Durand, N. C., M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* 3: 95–98.

Gerard, D., L. F. V. Ferrão, A. A. F. Garcia, and M. Stephens. 2018. Genotyping Polyploids from Messy Sequencing Data. *Genetics* 210: 789–807.

Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 37: 540–546.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

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., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lü, T. F., X. Li, S. N. Wang, L. Han, F. Hao, C. L. Fu, P. Zhang, and X. Liu. 2017. Allohexaploid speciation of the two closely related species *Myriophyllum spicatum* and *M. sibiricum* (Haloragaceae). *Aquatic Botany* 142: 105–111.

Netherland, M. D., and L. Willey. 2017. Mesocosm evaluation of three herbicides on Eurasian watermilfoil (Myriophyllum spicatum) and hybrid watermilfoil (Myriophyllum spicatum ×myriophyllum sibiricum): Developing a predictive assay. *Journal of Aquatic Plant Management* 55: 39–41.

Soltis, D. E., C. J. Visger, D. Blaine Marchant, and P. S. Soltis. 2016. Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany* 103: 1146–1166.

Taylor, L.-A. L., J. N. Mcnair, P. Guastello, J. Pashnick, and R. A. Thum. 2017. Heritable variation for vegetative growth rate in ten distinct genotypes of hybrid watermilfoil. *Journal of Aquatic Plant Management* 55.

Thum R.A., Heilman, M.A., Hausler P.J., Huberty L.E., Tyning P., Wcisel D.J., Zuellig M.P., Berger S.T., Glomski L.M., Netherland M.D.. 2012. Field and laboratory documentation of reduced fluridone sensitivity of a hybrid watermilfoil biotype (*Myriophyllum spicatum x Myriophyllum sibiricum*). *J. Aquat. Plant Manage*. 50: 141-146.

Thum, R. A., G. M. Chorak, R. M. Newman, J. A. Eltawely, J. Latimore, E. Elgin, and S. Parks. 2020. Genetic diversity and differentiation in populations of invasive Eurasian (Myriophyllum spicatum) and hybrid (Myriophyllum spicatum × Myriophyllum sibiricum) watermilfoil. *Invasive Plant Science and Management* 13: 59–67.

Van de Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411–424.

Ward, S. M. 2000. Allotetraploid segregation for single-gene morphological characters in quinoa (Chenopodium quinoa willd.). *Euphytica* 116: 11–16.

Wu, Z., Z. Ding, D. Yu, and X. Xu. 2015A. Influence of niche similarity on hybridization between Myriophyllum sibiricum and M. spicatum. *Journal of Evolutionary Biology* 28: 1465–1475.

Wu, Z., D. Yu, Z. Wang, X. Li, and X. Xu. 2015B. Great influence of geographic isolation on the genetic differentiation of Myriophyllum spicatum under a steep environmental gradient. *Scientific Reports* 5: 1–9.

Wu, Z.-G., D. Yu, and X.-W. Xu. 2013. Development of Microsatellite Markers in the Hexaploid Aquatic Macrophyte, *Myriophyllum spicatum* (Haloragaceae). *Applications in Plant Sciences* 1: 1200230.

Zuellig M.P., and Thum R.A.. 2012. Multiple introductions of invasive Eurasian watermilfoil and recurrent hybridization with northern watermilfoil in North America. *J. Aquat. Plant Manage*. 50: 1-19.

Table 1: The distribution of the molecular markers in the 3x read depth analysis across the largest 21 chromosomes of the draft reference genome. The number of markers that map to each chromosome are listed. Chromosomes are arbitrarily labeled and currently do not have any structure.

	Reference - Autohexaploid	**Reference - Allohexaploid**
Chromosome 1	0	178
Chromosome 2	2	635
Chromosome 3	2	666
Chromosome 4	0	532
Chromosome 5	0	718
Chromosome 6	2	534
Chromosome 7	1	642
Chromosome 8	0	792
Chromosome 9	4	748
Chromosome 10	0	842
Chromosome 11	4	656
Chromosome 12	0	744
Chromosome 13	0	696
Chromosome 14	0	510
Chromosome 15	0	576
Chromosome 16	0	424

Chromosome 17	0	416
Chromosome 18	0	562
Chromosome 19	1	538
Chromosome 20	2	566
Chromosome 21	0	544

Figure 1- Scatter plots showing the chi-square value each marker being inherited under an allo- or autohexaploid marker in the de novo analysis at both the 3x (A) and 10x (B) stringencies. The red line is an indication of where the models are equally likely. Points that fall above the line have a higher likelihood of being inherited as diploidized markers, or allohexaploid. Markers that fall below the line are more likely to be inherited polysomically, or autohexaploid. All molecular markers are mapped to these graphs, however, many points lie directly on top of each other.

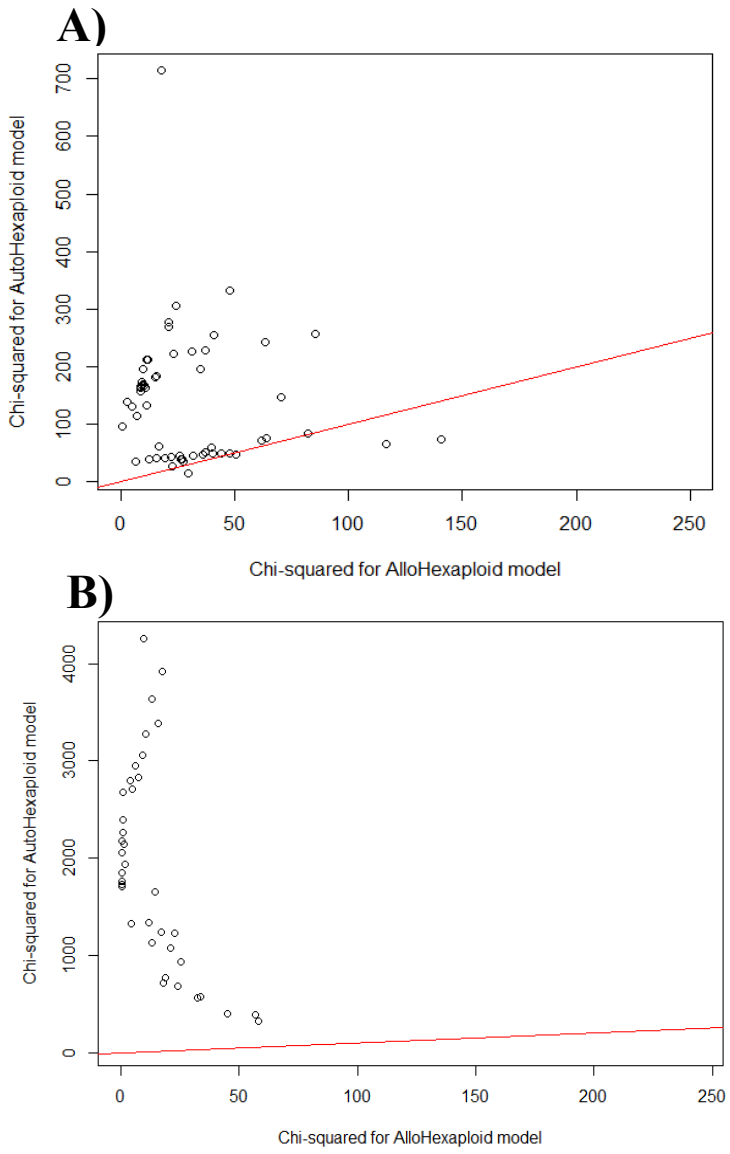


Figure 2 – A histogram of the number of alleles observed in the data under the less stringent cutoffs for the de novo analysis. A significant majority of our markers were diploidized.

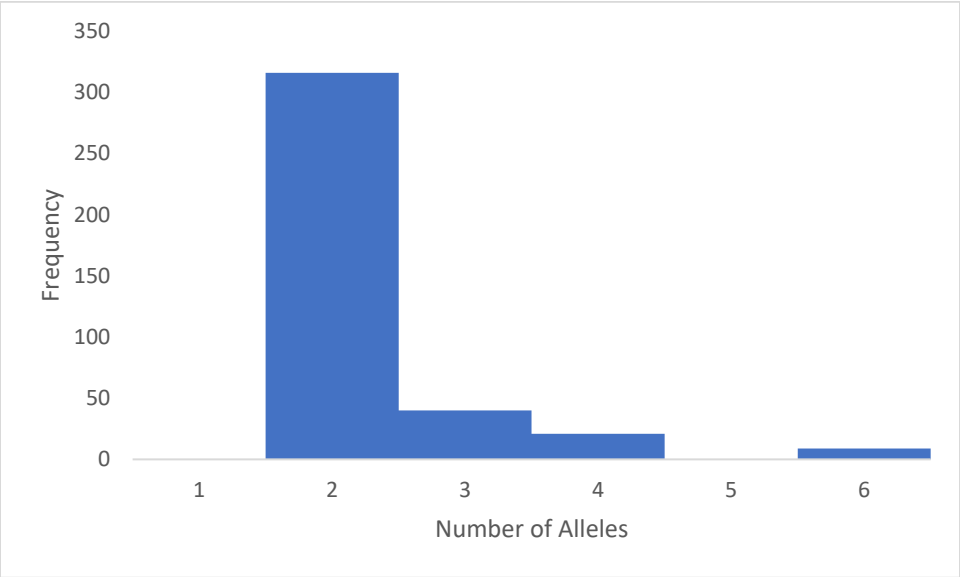


Figure 3- Scatter plots of each marker's chi-square values under an allo- or autohexaploid model for the reference-based analysis at both the 3x (A) and 10x (B) read depth stringencies. The red line is an indication of where the models are equally likely. Points that fall above the line have a higher likelihood of being inherited as diploidized markers, or allohexaploid. Markers that fall below the line are more likely to be inherited polysomically, or autohexaploid.

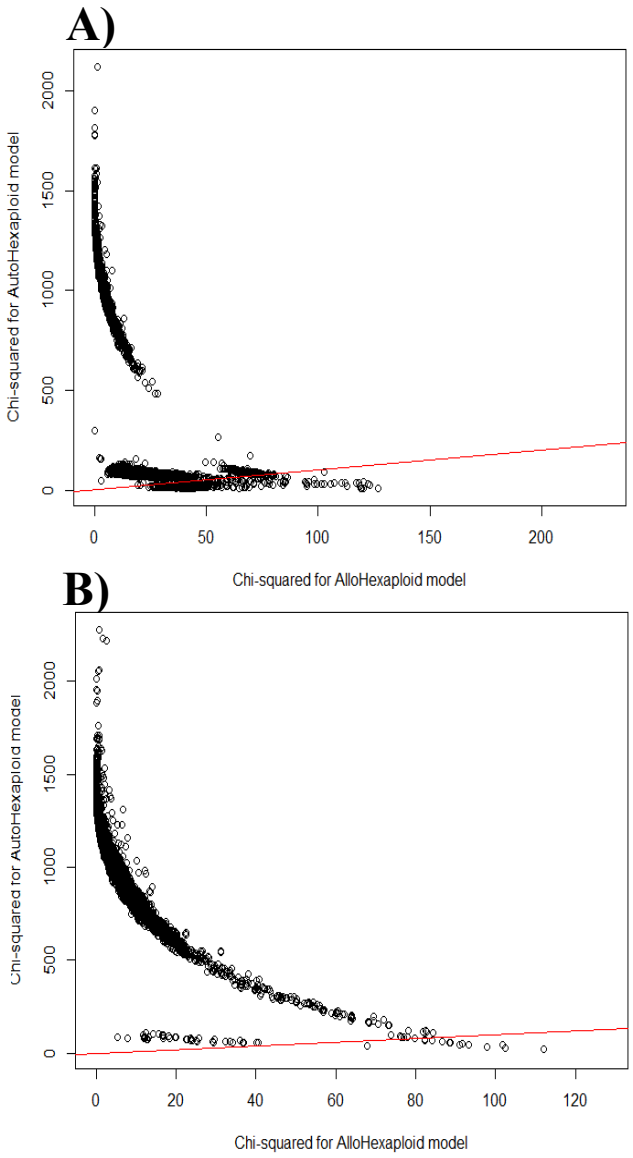
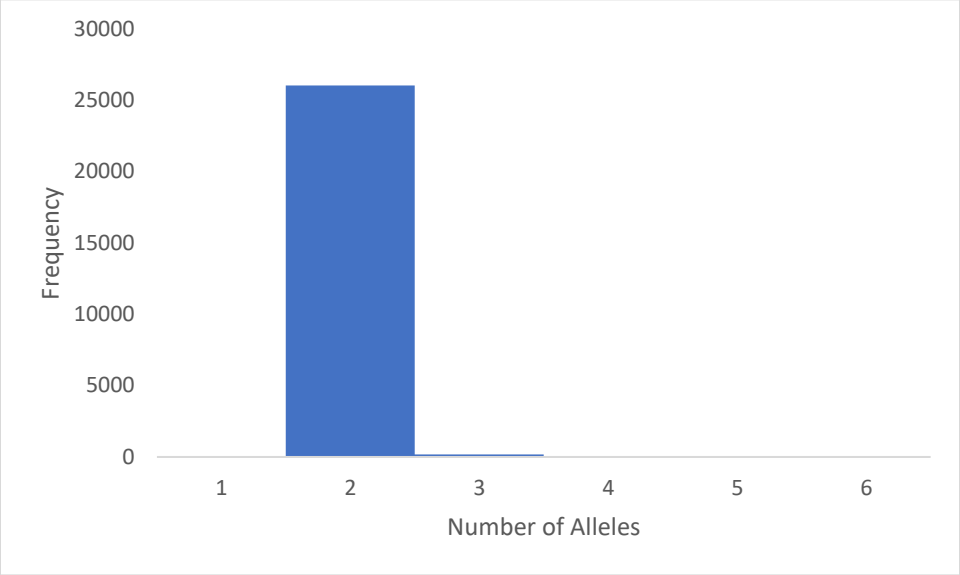


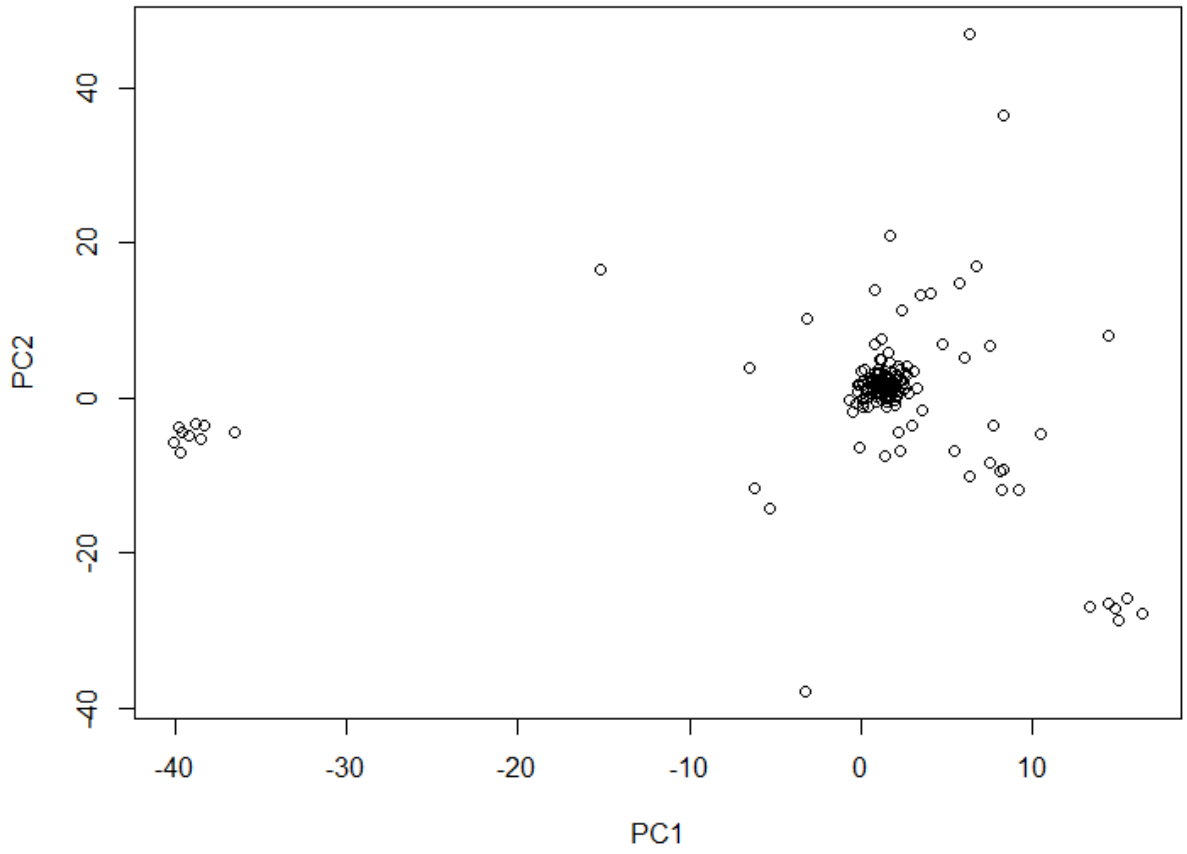
Figure 4 – A histogram of the number of alleles present in the less stringent reference-based analysis of our data. Similar to the de novo analysis, a large number of our markers are diploidized.



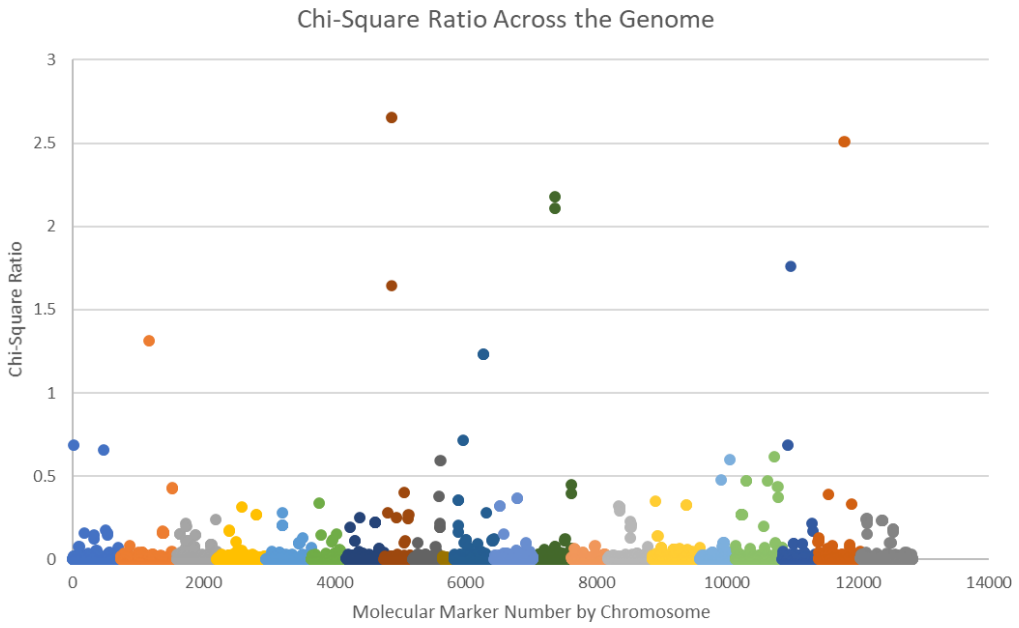
Supplemental Table 1- List of the 16 alleles with evidence for autohexaploidy that do not map to the 21 largest chromosomes of the draft reference genome.

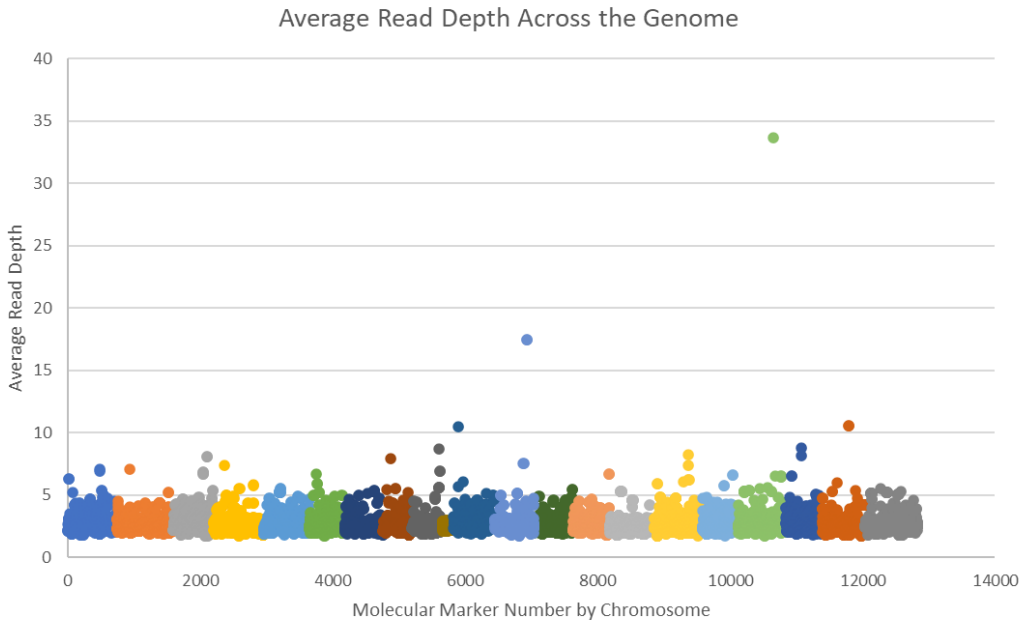
Locus	**Top BLAST**	**Number of Markers**
scaffold_2383___fragment	Horse orchid, brassicas, rice	4
Chromosome170	no significant blast	2
Chromosome178	*Solanum tuberosum*	4
scaffold_3189___fragment_4___debris	*Myriophyllum spicatum* ITS	2
contig_1704___fragment_2___debris	*Apteryx mantelli mantelli*	2
contig_639	no significant blast	2

Supplemental Figure 1- Principal coordinates analysis of the individuals in the analysis prior to genotyping. Individuals which were closely related were chosen (between -10 and 10 on PC1) to remain in the analysis, excluding individuals with likely high sequencing error rates.



Supplemental Figure 2- Top) Chi-square ratio for each molecular marker across the genome. A ratio between zero and one is a marker more likely to be inherited as a diploid, and a ratio above one an autopolyploid. Bottom) Read depth for each locus across the genome. Each color block is one of the 21 chromosomes of the draft reference genome for Eurasian watermilfoil.





DEVELOPMENT OF A MICROHAPLOTYPING BASED PANEL FOR STANDARDIZED GENOTYPING IN WATERMILFOIL

Contribution of Authors and Co-Authors

Manuscript in Chapter:4

Author: Jeff Pashnick

Contributions: Project inception and study design, molecular marker data collection for panel development. Analysis of molecular markers for panel design. Primer design for 128 markers to distinguish species and genotype. Pilot sequencing library preparation. Writing on the development of the panel for the manuscript.

Co-Author: Neil Thompson

Contributions: Custom R scripts for determining power and inclusion of molecular markers. Protocol assistance.

Co-Author: Gregory Chorak

Contributions: Pilot sequencing data analysis.

Co-Author: Ryan A. Thum

Contributions: Project inception and study design, writing.

Manuscript Information

Jeff Pashnick, Neil Thompson, Gregory Chorak, and Ryan A. Thum

TBD

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Introduction

Accurately genotyping is a critical component of distinguishing cryptic species or linking genotype to phenotype (Hendry et al. 2011). Genotyping is done through the identification and analysis of molecular markers; however, there are many types of molecular markers (e.g. amplified fragment length polymorphisms (AFLPs), microsatellites, genotyping by sequencing (GBS)), each with their own advantages and disadvantages. For example, GBS provides thousands of SNP markers which can accurately genotype but is one of the most expensive methods for genotyping (Campbell et al. 2015). Therefore, the necessity for cheaper SNP-based genotyping assay has led to the development of methods such as microhaplotyping and genotyping by thousands (GT-seq) which significantly reduce the genotyping cost per individual (Kidd et al. 2013, Campbell et al. 2015, Baetscher et al. 2017). Higher numbers of molecular markers are more accurately able to distinguish genotypes within the same taxon and are more powerful for linking genotype to phenotype through forward and reverse genetics (Byers et al. 2016).

Eurasian watermilfoil (*Myriophyllum spicatum* L.; EWM) is a heavily managed aquatic invasive species in the United States which can hybridize with native northern watermilfoil (*Myriophyllum sibiricum* Komarov; NWM). Molecular markers have been used to identify taxa as they are morphologically difficult to distinguish, and more recently molecular markers have been used to genotype in watermilfoil (Thum et al. 2012, Zuellig and Thum 2012, Pashnick and Thum 2020). In both the laboratory and the field, genetic variation in watermilfoil has been shown to impact their response to herbicide application, the most commonly used management tool (Parks et al. 2016, Taylor et al. 2017, Chorak and Thum 2020). Therefore, building an

accurate, rapid, and cost-effective assay can inform managers on the genotypes of watermilfoil present in their waterbodies. Waterbodies have also been shown to share genotypes of watermilfoil, and in some cases, herbicide data is linked to those genotypes (Berger et al. 2015, Netherland and Willey 2017, Chorak and Thum 2020). When herbicide response data has shown a genotype to be less susceptible and that genotype is shared in a waterbody to be managed, that information can be used to inform management in that new system.

AFLP-, microsatellite-, and SNP (single nucleotide polymorphism)-based molecular markers have all been used to genotype watermilfoil; however, each marker has several disadvantages (see Table 1, Pashnick and Thum 2020). AFLP and microsatellite molecular markers are both fragment-based and thus difficult to interpret because human judgement when scoring chromatograms, and differing methodologies in protocol and instrumentation. In polyploids, such as watermilfoil, an allohexaploid with up to six alleles possible at each locus (Pashnick et al. *In Prep.*), fragment-based genotyping has severe limitations. AFLP and microsatellites must be treated as dominant molecular markers in polyploids, possibly missing important genetic information while genotyping in watermilfoil. GBS and other SNP-based methodologies solve some of these issues due to being sequence-based assays of genetic variation; however, GBS data is currently the most expensive molecular marker data to generate. Therefore, a molecular assay retaining the advantages of SNP-based methodologies while reducing the time and money required for genotyping should be a priority for the management of EWM.

Table 1: A table of the advantages and disadvantages of each of the molecular markers used to genotype in watermilfoil.

Molecular Marker	Advantages	Disadvantages
AFLPs	• Cheapest to generate • Genome-wide assay (at specific restriction enzyme cutsites). • Produces ~100 molecular markers.	• Fragment-based. • Time consuming (scoring chromatograms). • Error rates approximately 7% in clean data, and ~50% when auto-scored. • Cannot be analyzed for polyploid systems.
Microsatellites	• Cheap to generate. • Fewer peaks on chromatograms make them easier than AFLPs to score.	• Fragment based. • Limited to specific loci • Time consuming (scoring chromatograms). • Error rates approximately ~15% when auto-scored. • ~10-20 markers produced. • Cannot be analyzed for allele dosage in polyploid systems.
GBS	• Sequence based data. • Can be analyzed for polyploid systems. • Produces thousands of molecular markers. • Assays the whole genome.	• The most expensive molecular markers to generate. • Data variable due to restriction enzyme usage and sequencing protocol.

	• Lowest error rates (0.7%).	• Turnaround time is typically months as opposed to weeks.
Microhaplotyping	• Sequence based data. • Can be analyzed accurately in a polyploid. • A specifically chosen, information dense, selection of microhaplotypes. • Assays the whole genome. • Low error rate due to high read depth. • Lowest per sample cost. • Easy to add additional markers to a pre-existing panel. • Data easily shared across laboratories.	• Significant development time.

We developed and piloted a panel of SNPs for genotyping in watermilfoil, using microhaplotyping (Kidd et al. 2013). Microhaplotyping, or using closely linked SNPs (i.e. within a few hundred base pairs) can be an effective way to genotype while significantly reducing the cost per sample, but retaining the resolution and accuracy of SNP-based analyses (Kidd et al 2013, Campbell et al 2015).

Methods

Samples and ddRAD Genotyping

We took previously analyzed GBS data from Pashnick and Thum, 2020 and combined that data with 96 northern, Eurasian, and hybrid watermilfoil sequenced to detect introgression, therefore increasing the intraspecific genetic variation for the construction of the microhaplotyping panel(Table 3, Pashnick and Thum 2020). To obtain genotype data and generate sequence data for the design of microhaplotyping markers, all 288 individuals were sequenced using double-digest restriction site-associated DNA sequencing and the methods described in Pashnick and Thum, 2020 (ddRAD-seq, Peterson et al. 2012). Bioinformatics was performed using Stacks version 2.41 to trim the raw sequencing reads for quality (phred score >33), remove Illumina adapters, and demultiplex individuals based on their unique DNA barcode (Catchen et al. 2013). The denovo_map.pl pipeline of the Stacks software was used to generate "stacks" or contigs of sequencing reads using default parameters. Finally, we ran the "populations" module of the STACKs software to keep stacks that 1) were present in all three taxa, and 2) had data for at least 70% of individuals within a taxon to ensure a locus would be able to be amplified from all watermilfoil in the final microhaplotyping panel leaving us with 20,158 stacks.

Microhaplotyping Loci Selection

We first selected loci with ≥ 2 SNPs to generate diploidized microhaplotypes (closely linked SNPs, i.e. within the same sequencing read) to keep the resolution of GBS-based markers while overcoming the limitations of determining allele dosage in fragment based data (Baetshcer et al. 2017). To do so, we used a custom R script to filter the stacks to contain ≥ 2 SNPs. After filtering for microhaplotype loci, we filtered further for microhaplotypes able to distinguish interspecific variation and intraspecific variation using the R package *Rubias* (Moran and

Anderson, 2019). Interspecific genetic variation is critical for distinguishing between northern, Eurasian, and hybrid watermilfoil since morphological variation is often difficult to distinguish taxa. Intraspecific genetic variation is important for distinguishing genotypes within a given taxon (Pashnick and Thum 2020, Thum et al. 2020). After filtering with custom R scripts and *Rubias,* we were left with 342 microhaplotype loci to design primers for the microhaplotyping panel.

Primer Design

We designed primers for all 342 remaining filtered loci for the microhaplotyping panel using the Primer3 software (Untergasser et al. 2012, Baetscher et al 2018). Since primers needed to be multiplexed into a single PCR reaction, we designed primers for each marker within 2°C of our chosen 60°C annealing temperature. To retain as much of a haplotype as possible, we also designed primers as close to the start and end of each locus as possible. We were able to design 128 primer pairs to use in the pilot run of the panel (Supp. Table 1). Primers created were then modified with the Nextera sequencing adapters for use in library construction; however, it should be noted that TruSeq or other sequencing chemistries may be used as adapter sequence here (Supp. Table 1).

Microhaplotyping- Piloting the Panel

We selected 12 individuals (three EWM, three NWM, and six HWM) from the population of 288 individuals that generated this panel for a pilot run to test the 128 markers (Table 3). Following a modified protocol for high-throughput amplicon sequencing, a sequencing library was constructed for the 12 samples selected for this pilot run (Gohl et al. 2016). Briefly, primer pairs for each of the 128 molecular markers were combined equally into a single multiplex

containing the forward and reverse primers. We performed an initial PCR for each individual with the 128 marker primer multiplex. PCR products were diluted 1:10 and then used as template for a second PCR adding the barcode and index sequences for each sample. Finally, samples were standardized using a SequelPrep kit (Invitrogen), cleaned using Ampure XP beads (Beckman Coulter) and a magnetic plate, following the manufacturer instructions, and combined into a single library for sequencing. For the full, detailed protocol please see the supplemental protocol (SupplementalProtocol_Microhaplotyping). Sequencing was completed on the MiSeq platform using 2x300 paired-end sequencing at the University of Minnesota Genomics Center.

Raw data produced was cleaned by using bbduk to remove Nextera adapter sequence (Bushnell, B. https://sourceforge.net/projects/bbmap). After adapter removal, data were processed with MICROHAPLOT software to variant call at each locus for each individual (Baetsher et al. 2017). Final markers for the panel were chosen based on the following criteria:

1. Markers amplified in all 12 individuals.
2. Markers had a minimum read depth of 100x in each individual, and no greater than 10,000x.

Leaving a total of 59 markers chosen to comprise the final genotyping set.

To test the ability for the remaining 59 markers to distinguish both taxon and genotype, microhaplotyping data for the 12 pilot individuals were imported into GenAlEx v6.5 to test genetic distance and construct principal coordinates analyses (Peakall and Smouse, 2006).

Results and Discussion

Principal coordinates analysis of the 12 pilot individuals show that the 59 molecular markers were able to distinguish between EWM, NWM, and HWM (Figure1A). The panel was also

accurate at assigning different genotypes within all three taxa for the individuals included in the pilot (Figure 1B, C, and D, Table 2). However, while genotypes were distinguishable within taxon for the pilot data, validation of the markers used in the panel need to be tested on a sample with larger genetic variation and duplicate DNA extractions need to be included. If a clone is always a clone, genetic distance between duplicate DNA extractions should never exceed the error rate of sequencing (Pashnick and Thum, 2020). Finally, the 59 markers aligned to 20 of the 21 chromosomes of the draft Eurasian watermilfoil reference genome, ensuring our data was not biased to any chromosome or subgenome (Table 4).

Here we present the development of a cost-effective and information dense panel of markers for genotyping in watermilfoil. This panel provides the base for a standardized way to genotype watermilfoil at a fraction of the cost of previous methods (AFLPs, microsatellites, and GBS). We show evidence that the panel can accurately determine the taxon, genotype within taxon, and data is not biased to any chromosome or subgenome. However, we recognize that this panel needs to be tested on a population with more genetic diversity within taxon, and duplicate DNA extractions to resolve sequencing error. Further, this panel may be used to directly share information across laboratories previously unavailable with other methods of genotyping due to differences in lab protocols or human analysis.

Distinguishing taxon between EWM, NWM, and HWM has often been the management unit for watermilfoil, and studies for herbicide response often stop at distinguishing taxon (Moody and Les 2000, Berger et al. 2012, 2015). Modern studies into the herbicide response in watermilfoil has begun to use the genotype as the unit of study as opposed to the much broader taxon (Netherland and Wiley 2017, Taylor et al. 2017). In our final panel of 59 markers, 10 of

them are diagnostic to the taxon level and the remaining 49 markers are able to distinguish genotype. Genetic data can be collected for many more samples for the same cost as previous studies, enabling managers to have more accurate estimates for the frequency of each taxon and genotype of watermilfoil in their managed waterbodies. Collecting genetic data on more samples increases the likelihood of detecting genotypes which may currently be rare or at the start of their invasion. If management response data has already been collected on those genotypes, management can be informed based on the frequency and presence of those genotypes within new waterbodies. In addition to clone sharing across waterbodies, a standardized genotyping panel allows herbicide screens in multiple laboratories to genotype. A major advantage of using microhaplotyping over standard SNP-based analyses such as any GBS methodology is the diploidized nature of all markers in the panel. Due to comparisons of haplotypes, or closely linked SNPs within ~250bp, each variable SNP is linked within the same sequencing read. Diploidized markers allow accurate calculations of allele frequencies, enabling studies in population genetics previously impossible in watermilfoil. For example, allele frequencies can be used to determine the amount of introgression or the hybrid class of hybrid watermilfoil out in natural populations. Differences in introgression can provide evidence for possible shifts in reproduction in populations in response to management. As an allohexaploid, watermilfoil may also have buffered breakdown of heterosis in more introgressed generations, possibly allowing them to exhibit nuisance characteristics associated with heterosis much longer.

Far fewer loci are amplified and sequenced using a microhaplotyping approach compared to GBS, which allows many more individuals to be sequenced in a single lane. For the cost of genotyping each individual, this is a drastic improvement over other sequencing methods.

Additionally, since fewer markers are targeted than GBS, error rates are kept low through increased depth of coverage (Campbell et al. 2015). The adaptability of the sequencing chemistry also allows each lab and each sequencing core to tailor the library construction to their specific sequencer. While we used a Nextera adapter system here, this may be adapted to any other sequencing chemistry such as TruSeq. Sequencing platform can also be modified and can be matched to the cost tradeoff between the number of individuals to be haplotyped and the depth of coverage necessary. For example, one could choose the MiSeq for sample numbers under ~200, and NovaSeq for greater than 200.

Previously, genotyping in watermilfoil had to be done in a single laboratory for each project due to subjectivity in calling genotypes, meaning that sharing data across laboratories was challenging or impossible. This microhaplotyping method of genotyping watermilfoil can be easily standardized across laboratories. Genetic data can now be generated in any population and compared to genotype information from previous studies to determine if a population of interest shares genotypes with other previously managed populations. Changes to workflow described above will not affect the haplotypes identified, allowing microhaplotyping information to be generated and directly shared across laboratories unlike with microsatellites or AFLPs.

Figure 1: A principal coordinates analysis of the 12 individuals in the pilot run with the panel of 59 molecular markers. Panel A shows that the 59 molecular markers are able to distinguish taxon across PC1. Panel B, C, and D show principal coordinates analysis of individuals within taxon, showing distinguishable genotypes within taxon.

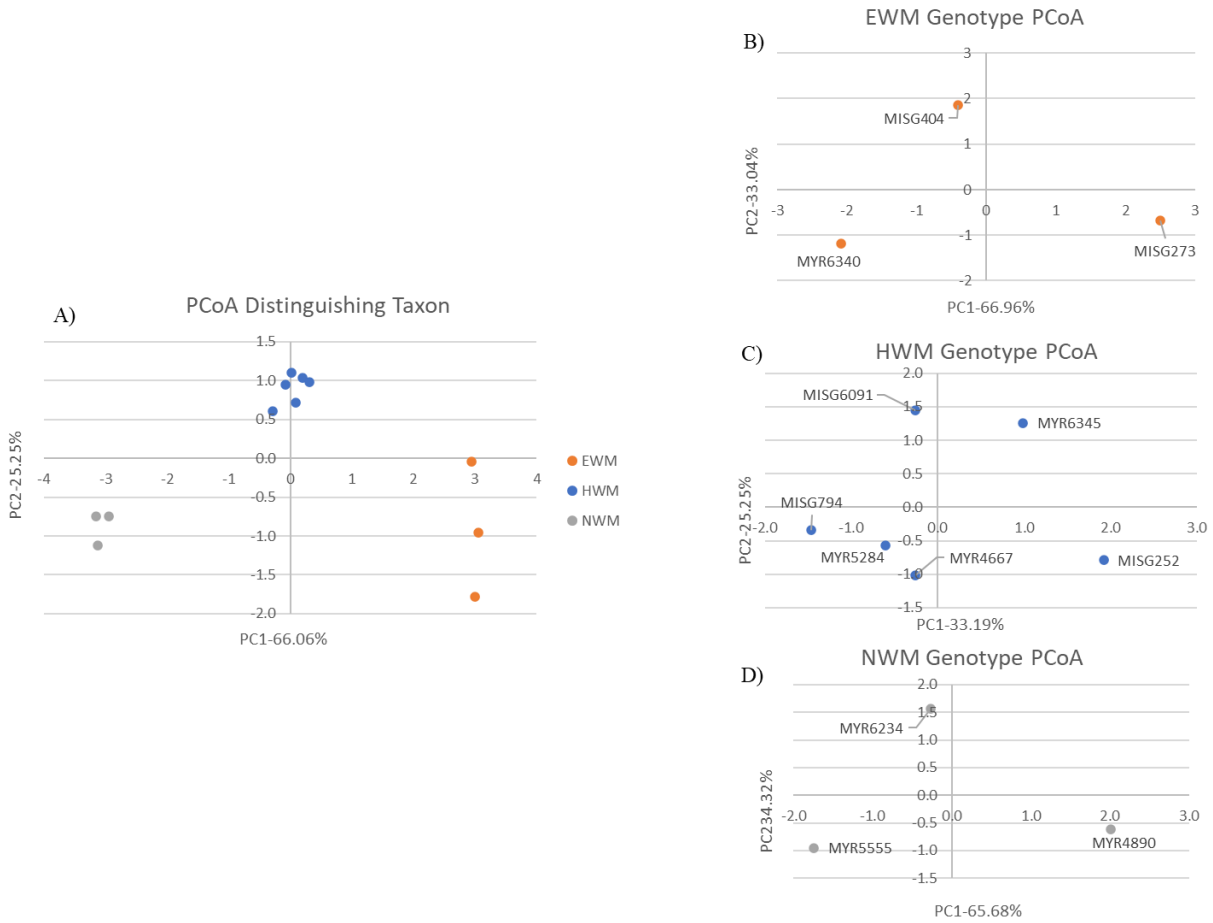


Table 2: Pairwise genetic distance of the 12 individuals in the pilot run of the 59 molecular marker panel. The taxon (Eurasian watermilfoil (EWM), northern watermilfoil (NWM), and hybrid watermilfoil (HWM)) and genotype for each individual are listed on the right of the table. Loci are diploidized and each individual may have up to two haplotypes per maker.

MYR6340	MISG273	MISG404	MISG252	MYR6345	MISG794	MYR5284	MYR4667	MISG6091	MYR4890	MYR6234	MYR5555	Clone
0												**EWM1**
30	0											**EWM7**
17	21	0										**EWM4**
46	64	54	0									**HWM1**
44	66	55	19	0								**HWM6**
51	63	55	30	25	0							**HWM15**
49	67	55	23	20	17	0						**HWM3**
47	67	56	19	22	18	11	0					**HWM5**
56	66	62	28	18	21	18	19	0				**HWM16**
130	133	131	62	58	54	50	53	46	0			**NWM1**
134	137	135	68	62	61	55	59	50	14	0		**NWM4**
136	142	138	64	63	60	53	56	51	20	12	0	**NWM2**

Table 3*:* The taxon (Eurasian watermilfoil (EWM), northern watermilfoil (NWM), and hybrid watermilfoil (HWM)) and lake of origin information for the 288 individuals used to build the panel of molecular markers. Bolded and italicized populations denote that an individual from that population was chosen for the pilot test of the 128 marker panel.

Sample	**Taxon**	**Genotype**	**State**	**Population (Lake)**
6368_A1	EWM	EWM1	MN	All over MN
6374_B1	EWM	EWM1	MN	All over MN
5484_C1	EWM	EWM1	MN	All over MN
5491_D1	EWM	EWM1	MN	All over MN
6340_E1	***EWM***	***EWM1***	***MN***	***All over MN***
6342_F1	EWM	EWM1	MN	All over MN
4880_G1	EWM	EWM1	MN	All over MN
6369_H1	EWM	EWM1	MN	All over MN
6309_A2	EWM	EWM1	MN	All over MN
6312_B2	EWM	EWM1	MN	All over MN
6364_C2	EWM	EWM1	MN	All over MN
6139_D2	EWM	EWM1	MN	All over MN
5379_E2	EWM	EWM1	MN	All over MN
5400_F2	EWM	EWM1	MN	All over MN
MISG258_G2	EWM	EWM1	MI	All over
MISG265_H2	EWM	EWM2	MI	Fish Lake
MISG273_A3	***EWM***	***EWM2***	***MI***	***Fish Lake***

MISG275_B3	EWM	EWM2	MI	Fish Lake
MISG276_C3	EWM	EWM2	MI	Fish Lake
MISG303_D3	EWM	EWM2	MI	Fish Lake
MISG282_E3	EWM	EWM2	MI	Fish Lake
MISG283_F3	EWM	EWM2	MI	Fish Lake
MISG284_G3	EWM	EWM2	MI	Fish Lake
MISG285_H3	EWM	EWM2	MI	Fish Lake
MISG288_A4	EWM	EWM2	MI	Fish Lake
MISG291_B4	EWM	EWM2	MI	Fish Lake
MISG292_C4	EWM	EWM2	MI	Fish Lake
MISG299_D4	EWM	EWM2	MI	Fish Lake
MISG003_E4	EWM	EWM3	MI	Fish Lake
MISG017_F4	EWM	EWM3	MI	Lake Jordan
MISG025_G4	EWM	EWM3	MI	Lake Jordan
MISG039_H4	EWM	EWM3	MI	Lake Jordan
MISG045_A5	EWM	EWM3	MI	Lake Jordan
MISG046_B5	EWM	EWM3	MI	Lake Jordan
MISG048_C5	EWM	EWM3	MI	Lake Jordan
MISG065_D5	EWM	EWM3	MI	Lake Jordan
MISG070_E5	EWM	EWM3	MI	Lake Jordan
MISG077_F5	EWM	EWM3	MI	Lake Jordan
MISG079_G5	EWM	EWM3	MI	Lake Jordan

MISG080_H5	EWM	EWM3	MI	Lake Jordan
MISG010_A6	EWM	EWM3	MI	Lake Jordan
MISG073_B6	EWM	EWM3	MI	Lake Jordan
MISG380_C6	EWM	EWM4	MI	Lake Jordan
MISG382_D6	EWM	EWM4	MI	Lansing Lake
MISG383_E6	EWM	EWM4	MI	Lansing Lake
MISG387_F6	EWM	EWM4	MI	Lansing Lake
MISG391_G6	EWM	EWM4	MI	Lansing Lake
MISG392_H6	EWM	EWM4	MI	Lansing Lake
MISG393_A7	EWM	EWM4	MI	Lansing Lake
MISG417_B7	EWM	EWM4	MI	Lansing Lake
MISG397_C7	EWM	EWM4	MI	Lansing Lake
MISG400_D7	EWM	EWM4	MI	Lansing Lake
MISG404_E7	***EWM***	***EWM4***	***MI***	***Lansing Lake***
MISG409_F7	EWM	EWM4	MI	Lansing Lake
MISG410_G7	EWM	EWM4	MI	Lansing Lake
MISG413_H7	EWM	EWM4	MI	Lansing Lake
5491D_A8	EWM	EWM1	MN	Many MN Lakes
6340D_B8	EWM	EWM1	MN	Many MN Lakes
MISG292D_C8	EWM	EWM2	MI	Fish Lake
MISG283D_D8	EWM	EWM2	MI	Fish Lake
MISG003D_E8	EWM	EWM3	MI	Lake Jordan

MISG017D_F8	EWM	EWM3	MI	Lake Jordan
MISG380D_G8	EWM	EWM4	MI	Lansing Lake
MISG383D_H8	EWM	EWM4	MI	Lansing Lake
4864_A9	NWM	NWM1	MN	Bald Eagle Lake
4877_B9	NWM	NWM1	MN	Bald Eagle Lake
4890_C9	***NWM***	***NWM1***	***MN***	***Bald Eagle Lake***
4907_D9	NWM	NWM1	MN	Bald Eagle Lake
4930_E9	NWM	NWM1	MN	Bald Eagle Lake
4934_F9	NWM	NWM1	MN	Bald Eagle Lake
4954_G9	NWM	NWM1	MN	Bald Eagle Lake
4960_H9	NWM	NWM1	MN	Bald Eagle Lake
5537_A10	NWM	NWM2	MN	Minne-Belle Lake
5543_B10	NWM	NWM2	MN	Minne-Belle Lake
5552_C10	NWM	NWM2	MN	Minne-Belle Lake
5555_D10	***NWM***	***NWM2***	***MN***	***Minne-Belle Lake***
5559_E10	NWM	NWM2	MN	Minne-Belle Lake
5568_F10	NWM	NWM2	MN	Minne-Belle Lake
5576_G10	NWM	NWM2	MN	Minne-Belle Lake
5577_H10	NWM	NWM2	MN	Minne-Belle Lake
5581_A11	NWM	NWM3	MN	Mitchell (ShCo)
5587_B11	NWM	NWM3	MN	Mitchell (ShCo)
5591_C11	NWM	NWM3	MN	Mitchell (ShCo)

5601_D11	NWM	NWM3	MN	Mitchell (ShCo)
5643_E11	NWM	NWM3	MN	Mitchell (ShCo)
5614_F11	NWM	NWM3	MN	Mitchell (ShCo)
5626_G11	NWM	NWM3	MN	Mitchell (ShCo)
5635_H11	NWM	NWM3	MN	Mitchell (ShCo)
6228_A12	NWM	NWM4	MN	Spectacle
6229_B12	NWM	NWM4	MN	Spectacle
6232_C12	NWM	NWM4	MN	Spectacle
6234_D12	***NWM***	***NWM4***	***MN***	***Spectacle***
6239_E12	NWM	NWM4	MN	Spectacle
6242_F12	NWM	NWM4	MN	Spectacle
6244_G12	NWM	NWM4	MN	Spectacle
6245_H12	NWM	NWM4	MN	Spectacle
4934D_A1	NWM	NWM1	MN	Bald Eagle Lake
4954D_B1	NWM	NWM1	MN	Bald Eagle Lake
5559D_C1	NWM	NWM2	MN	Minne-Belle Lake
5542_D1	NWM	NWM2	MN	Minne-Belle Lake
5601D_E1	NWM	NWM3	MN	Mitchell (ShCo)
5343D_F1	NWM	NWM3	MN	Mitchell (ShCo)
6232D_G1	NWM	NWM4	MN	Spectacle
6234D_H1	NWM	NWM4	MN	Spectacle
MISG314_A2	HWM	HWM1	MI	Townline/Templene/Muskellunge

MISG349_B2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG360_C2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG241_D2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG245_E2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG252_F2	***HWM***	***HWM1***	***MI***	***Townline/Templene/Muskellunge***
MISG445_G2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG452_H2	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG454_A3	HWM	HWM1	MI	Townline/Templene/Muskellunge
MISG087_B3	HWM	HWM2	MI	Sage/Budd
MISG090_C3	HWM	HWM2	MI	Sage/Budd
MISG091_D3	HWM	HWM2	MI	Sage/Budd
MISG162_E3	HWM	HWM2	MI	Sage/Budd
MISG530_F3	HWM	HWM2	MI	Sage/Budd
MISG535_G3	HWM	HWM2	MI	Sage/Budd
MISG542_H3	HWM	HWM2	MI	Sage/Budd
MISG551_A4	HWM	HWM2	MI	Sage/Budd
MISG562_B4	HWM	HWM2	MI	Sage/Budd
5261_C4	HWM	HWM3	MN	Ham Lake MN
5263_D4	HWM	HWM3	MN	Ham Lake MN
5270_E4	HWM	HWM3	MN	Ham Lake MN
5273_F4	HWM	HWM3	MN	Ham Lake MN
5274_G4	HWM	HWM3	MN	Ham Lake MN

5277_H4	HWM	HWM3	MN	Ham Lake MN
5278_A5	HWM	HWM3	MN	Ham Lake MN
5284_B5	***HWM***	***HWM3***	***MN***	***Ham Lake MN***
5290_C5	HWM	HWM3	MN	Ham Lake MN
2_D5	HWM	HWM4	MN	Lake Minnetonka, Various Bays
13_E5	HWM	HWM4	MN	Lake Minnetonka, Various Bays
23_F5	HWM	HWM4	MN	Lake Minnetonka, Various Bays
5018_G5	HWM	HWM4	MN	Lake Minnetonka, Various Bays
5022_H5	HWM	HWM4	MN	Lake Minnetonka, Various Bays
5052_A6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
2725_B6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
2806_C6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
2967_D6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
161_E6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
165_F6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
181_G6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
1765_H6	HWM	HWM4	MN	Lake Minnetonka, Various Bays
1785_A7	HWM	HWM4	MN	Lake Minnetonka, Various Bays
1793_B7	HWM	HWM4	MN	Lake Minnetonka, Various Bays
4885_C7	HWM	HWM5	MN	MN Many Lakes
4920_D7	HWM	HWM5	MN	MN Many Lakes
5366_E7	HWM	HWM5	MN	MN Many Lakes

5360_F7	HWM	HWM5	MN	MN Many Lakes
6217_G7	HWM	HWM5	MN	MN Many Lakes
6223_H7	HWM	HWM5	MN	MN Many Lakes
4992_A8	HWM	HWM5	MN	MN Many Lakes
5000_B8	HWM	HWM5	MN	MN Many Lakes
4667_C8	***HWM***	***HWM5***	***MN***	***MN Many Lakes***
4772_D8	HWM	HWM5	MN	MN Many Lakes
5445_E8	HWM	HWM5	MN	MN Many Lakes
5462_F8	HWM	HWM5	MN	MN Many Lakes
6155_G8	HWM	HWM6	MN	Coon/Elmo
6159_H8	HWM	HWM6	MN	Coon/Elmo
6161_A9	HWM	HWM6	MN	Coon/Elmo
6165_B9	HWM	HWM6	MN	Coon/Elmo
6338_C9	HWM	HWM6	MN	Coon/Elmo
6343_D9	HWM	HWM6	MN	Coon/Elmo
6345_E9	***HWM***	***HWM6***	***MN***	***Coon/Elmo***
6347_F9	HWM	HWM6	MN	Coon/Elmo
6355_G9	HWM	HWM6	MN	Coon/Elmo
6390_H9	HWM	HWM6	MN	Coon/Elmo
6380_A10	HWM	HWM6	MN	Coon/Elmo
6384_B10	HWM	HWM6	MN	Coon/Elmo
6385_C10	HWM	HWM6	MN	Coon/Elmo

6389_D10	HWM	HWM6	MN	Coon/Elmo
6188_E10	HWM	HWM6	MN	Coon/Elmo
6189_F10	HWM	HWM6	MN	Coon/Elmo
6193_G10	HWM	HWM6	MN	Coon/Elmo
6204_H10	HWM	HWM6	MN	Coon/Elmo
MISG314D_A11	HWM	HWM1	MI	Duplicate DNA extraction
MISG241D_B11	HWM	HWM1	MI	Duplicate DNA extraction
MISG087D_C11	HWM	HWM2	MI	Duplicate DNA extraction
MISG542D_D11	HWM	HWM2	MI	Duplicate DNA extraction
5277D_E11	HWM	HWM3	MN	Duplicate DNA extraction
5290D_F11	HWM	HWM3	MN	Duplicate DNA extraction
161D_G11	HWM	HWM4	MN	Duplicate DNA extraction
1793D_H11	HWM	HWM4	MN	Duplicate DNA extraction
5022D_A12	HWM	HWM4	MN	Duplicate DNA extraction
4885D_B12	HWM	HWM5	MN	Duplicate DNA extraction
5000D_C12	HWM	HWM5	MN	Duplicate DNA extraction
4772D_D12	HWM	HWM5	MN	Duplicate DNA extraction
6347D_E12	HWM	HWM6	MN	Duplicate DNA extraction
6355D_F12	HWM	HWM6	MN	Duplicate DNA extraction
6380D_G12	HWM	HWM6	MN	Duplicate DNA extraction
6193D_H12	HWM	HWM6	MN	Duplicate DNA extraction
MYR1079_A1	EWM	EWM5	ID	Pend Oreille

MYR1133_B1	EWM	EWM5	ID	Pend Oreille
MYR1914_C1	EWM	EWM6	ID	CDA
MYR1969_D1	EWM	EWM6	ID	CDA
MYR1969D_E1	EWM	EWM6	ID	CDA
MYR1978_F1	EWM	EWM6	ID	CDA
GN19A_G1	EWM	EWM7	MI	Gun
GB55B1_H1	EWM	EWM7	MI	Gun
GN28B_A2	EWM	EWM7	MI	Gun
MSQ1_B2	EWM	EWM8	WI	Moshawquit
MOS45A_C2	EWM	EWM8	WI	Moshawquit
MSQ56_D2	EWM	EWM8	WI	Moshawquit
MYR158D_E2	EWM	EWM1	MN	Minne-St.Alb
MYR158_F2	EWM	EWM1	MN	Minne-St.Alb
MYR2810D_G2	EWM	EWM1	MN	Minne-Smiths
MYR2810_H2	EWM	EWM1	MN	Minne-Smiths
MYR2754_A3	EWM	EWM1	MN	Minne-Smiths
HOU687_B3	EWM	EWM9	MI	HoughtonPre
HOU931B_C3	EWM	EWM9	MI	HoughtonPre
HOU965_D3	EWM	EWM9	MI	HoughtonPre
HOU982A_E3	EWM	EWM9	MI	HoughtonPre
MYR3178_F3	EWM	EWM10	MT	Noxon
MYR3258_G3	EWM	EWM10	MT	Noxon

MYR3266_H3	EWM	EWM10	MT	Noxon
MYR81_A4	EWM	EWM1	MN	Minnetonka - St. Albans
MYR81D_B4	EWM	EWM1	MN	Minnetonka - St. Albans
MYR2089_C4	EWM	EWM1	MN	Minnetonka - St. Albans
Long66B1_D4	EWM	EWM11	MI	Long Lake, MI
Long197A1_E4	EWM	EWM11	MI	Long Lake, MI
MYR2925D_F4	EWM	EWM1	MN	Minnetonka - Vets
MYR2825_G4	EWM	EWM1	MN	Minnetonka - Vets
MYR1725_H4	EWM	EWM1	MN	Minnetonka - Grays
MYR2968_A5	NWM	NWM5	MN	Minnetonka - Vets
MYR2828_B5	NWM	NWM5	MN	Minnetonka - Vets
HAN105_C5	NWM	NWM6	MI	Hanbury Lake
HAN41_D5	NWM	NWM6	MI	Hanbury Lake
HAN48_E5	NWM	NWM6	MI	Hanbury Lake
MYR1156_F5	NWM	NWM7	ID	Pend Oreille
MYR1211_G5	NWM	NWM7	ID	Pend Oreille
MYR3452_H5	NWM	NWM8	MT	Noxon
MYR3432_A6	NWM	NWM9	MT	Noxon
MYR3534_B6	NWM	NWM10	MT	Noxon
MYR2041_C6	NWM	NWM11	MN	Minnetonka - Christmas
MYR2059_D6	NWM	NWM11	MN	Minnetonka - Christmas
MYR2062_E6	NWM	NWM11	MN	Minnetonka - Christmas

MYR2038_F6	NWM	NWM11	MN	Minnetonka - Christmas
MYR2032_G6	NWM	NWM11	MN	Minnetonka - Christmas
MOS181_H6	NWM	NWM12	WI	Moshawquit Lake
MOS181D_A7	NWM	NWM12	WI	Moshawquit Lake
MOS109_B7	NWM	NWM12	WI	Moshawquit Lake
MOS1_C7	NWM	NWM12	WI	Moshawquit Lake
MSQ12_D7	NWM	NWM12	WI	Moshawquit Lake
MSQ23_E7	NWM	NWM12	WI	Moshawquit Lake
MYR3000_F7	NWM	NWM13	MN	Lake Minnetonka - Smith's Bay
MYR2984_G7	NWM	NWM13	MN	Lake Minnetonka - Smith's Bay
MYR4011_H7	NWM	NWM14	MT	Bever Lake
MYR4017_A8	NWM	NWM14	MT	Bever Lake
MYR4019_B8	NWM	NWM14	MT	Bever Lake
MYR4022_C8	NWM	NWM14	MT	Bever Lake
MYR4022D_D8	NWM	NWM14	MT	Bever Lake
MYR4023_E8	NWM	NWM14	MT	Bever Lake
MYR4023D_F8	NWM	NWM14	MT	Bever Lake
MYR2169_G8	NWM	NWM15	MT	Noxon
MYR2142_H8	NWM	NWM15	MT	Noxon
MI102-07_A9	HWM	HWM1	MI	Townline Lake
MI102-08_B9	HWM	HWM1	MI	Townline Lake
MI102-09_C9	HWM	HWM1	MI	Townline Lake

MI102-13_D9	HWM	HWM1	MI	Townline Lake
MI102-15_E9	HWM	HWM1	MI	Townline Lake
MI102-17_F9	HWM	HWM1	MI	Townline Lake
SAW5TT7_G9	EWM	EWM12	MI	Sawyer Lake
SAW5TT7D_H9	EWM	EWM12	MI	Sawyer Lake
HOU288b_A10	HWM	HWM7	MI	Houghton Lake
HOU474_B10	HWM	HWM7	MI	Houghton Lake
HOU608_C10	HWM	HWM7	MI	Houghton Lake
HOU611A_D10	HWM	HWM7	MI	Houghton Lake
HOU623_E10	HWM	HWM7	MI	Houghton Lake
HOU697_F10	HWM	HWM7	MI	Houghton Lake
HOU427_G10	HWM	HWM8	MI	Houghton Lake
HOU257_H10	HWM	HWM8	MI	Houghton Lake
HTN190_A11	HWM	HWM7	MI	Houghton Lake
HTN83_B11	HWM	HWM7	MI	Houghton Lake
HTN257_C11	HWM	HWM8	MI	Houghton Lake
HTN296B_D11	HWM	HWM9	MI	Houghton Lake
HTN490_E11	HWM	HWM9	MI	Houghton Lake
HTN611_F11	HWM	HWM8	MI	Houghton Lake
HTN777_G11	HWM	HWM8	MI	Houghton Lake
HTN920_H11	HWM	HWM7	MI	Houghton Lake
936AT8_A12	EWM	EWM13	MI	Houghton Lake

936AT8D_B12	EWM	EWM13	MI	Houghton Lake
984AT7D_C12	EWM	EWM13	MI	Houghton Lake
984AT7_D12	EWM	EWM13	MI	Houghton Lake
963BT7_E12	EWM	EWM13	MI	Houghton Lake
963BT7D_F12	EWM	EWM13	MI	Houghton Lake
25BT7_G12	HWM	HWM7	MI	Houghton Lake
25BT7D_H12	HWM	HWM8	MI	Houghton Lake
MISG794	***HWM***	***HWM15***	***MI***	***Thunder Lake***
MYR6091	***HWM***	***HWM16***	***MN***	***Budd Lake***

Table 4: Locations of the 59 marker final panel when aligned to the watermilfoil draft reference genome. Chromosomes are arbitrarily labeled for the 21 chromosomes of watermilfoil.

Chromosome Number	**Markers**
1	1
2	2
3	1
4	2
5	0
6	2
7	2
8	1
9	4
10	4
11	2
12	5
13	5
14	5
15	2
16	4
17	5
18	2

19	3
20	3
21	4
Total	**59**

Literature Cited

Baetscher, D. S., A. J. Clemento, T. C. Ng, E. C. Anderson, and J. C. Garza. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources* 18: 296–305.

Berger ST, Netherland MD, MacDonald GE. 2012. Evaluating fluridone sensitivity of multiple hybrid and Eurasian watermilfoil accessions under mesocosm conditions. J. Aquat. Plant Manage. 50: 135-141.

Berger ST, Netherland MD, MacDonald GE. 2015. Laboratory documentation of multiple-herbicide tolerance to fluridone, norflurazon, and topramazone in a hybrid watermilfoil *(Myriophyllum spicatum* × *M. sibiricum*) population. Weed Sci. 63(01):235–241.

Byers, K. J. R. P., S. Xu, and P. M. Schlüter. 2017. Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology* 26: 277–290.

Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15:855-867.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

Gohl, D., D. M. Gohl, A. MacLean, A. Hauge, A. Becker, D. Walek, and K. B. Beckman. 2016. An optimized protocol for high-throughput amplicon-based microbiome profiling. *Protocol Exchange*: 1–26.

Guastello PR, Thum RA. 2018. Mesocosm and field evaluation of Eurasian and hybrid watermilfoil response to endothall in Jefferson Slough, Montana. J. Aquat. Plant Manage. 56: 63-68.

Hendry, A. P., M. T. Kinnison, M. Heino, T. Day, T. B. Smith, G. Fitt, C. T. Bergstrom, et al. 2011. Evolutionary principles and their practical application. *Evolutionary Applications* 4: 159–183.

Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, Ihuegbu N. 2013. Microhaplotype loci are a powerful new type of forensic marker. For. Sci. Int.: Genetics Supp. Series 4. e123-e124.

LaRue EA, Zuellig MP, Netherland MD, Heilman MA, Thum RA. 2013. Hybrid watermilfoil lineages are more invasive and less sensitive to a commonly used herbicide than their exotic parent (Eurasian watermilfoil). Evol. Appl. 6(3):462–471.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21):2987–2993.

Moody, M. L., and D. H. Les. 2002. Evidence of hybridity in invasive watermilfoil (Myriophyllum) populations. *Proceedings of the National Academy of Sciences of the United States of America* 99: 14867–14871.

Moran, B. M., and E. C. Anderson. 2019. Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences* 76: 551–560.

Netherland MD, Willey L. 2017. Mesocosm evaluations of three herbicides on Eurasian watermilfoil (*Myriophyllum spicatum*) and hybrid watermilfoil (*Myriophyllum spicatum* x *Myriophyllum sibiricum*): Developing a predictive assay. J. Aquat. Plant Manage. 55: 39-42.

Parks SR, McNair JN, Hausler P, Tyning P, Thum RA. 2016. Divergent responses of invasive watermilfoil to treatment with auxinic herbicides in a large Michigan lake. Lake and Res. Manage. 32: 366-372.

Pashnick, J., and R. A. Thum. 2020. Note Comparison of molecular markers to distinguish genotypes of Eurasian watermilfoil , northern watermilfoil , and their hybrids. 61–66.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7(5).

Taylor L-AL, Mcnair JN, Guastello P, Pashnick J, Thum RA. 2017. Heritable variation for vegetative growth rate in ten distinct genotypes of hybrid watermilfoil. J. Aquat. Plant Manage. 55(July).

Thum RA, Heilman, MA, Hausler PJ, Huberty LE, Tyning P, Wcisel DJ, Zuellig MP, Berger ST, Glomski LM, Netherland MD. 2012. Field and laboratory documentation of reduced fluridone sensitivity of a hybrid watermilfoil biotype (*Myriophyllum spicatum x Myriophyllum sibiricum*). J. Aquat. Plant Manage. 50: 141-146.

Wu Z-G, Yu D, Xu X-W. 2013. Development of microsatellite markers in the hexaploid aquatic macrophyte, *Myriophyllum spicatum* (Haloragaceae). Appl. Plant Sci. 1(2):1200230.

Wu Z, Ding Z, Yu D, Xu X. 2015a. Influence of niche similarity on hybridization between *Myriophyllum sibiricum* and *M. spicatum*. J. Evol. Biol. 28(8):1465–1475.

Wu Z, Yu D, Wang Z, Li X, Xu X. 2015b. Great influence of geographic isolation on the genetic differentiation of *Myriophyllum spicatum* under a steep environmental gradient. Sci. Rep. 5(October):1–9.

Zuellig MP, Thum RA. 2012. Multiple introductions of invasive Eurasian watermilfoil and recurrent hybridization with northern watermilfoil in North America. J. Aquat. Plant Manage. 50: 1-19.

CONCLUSIONS AND FUTURE WORK

In this dissertation, I developed genomic resources in *M. sibiricum, M. spicatum* and the hybrid between them. Eurasian watermilfoil and their hybrids with northern watermilfoil are heavily managed in aquatics and therefore each of the studies conducted here attempt to inform the management of these plants through the use of genetics. The work shown here began investigating the resolution of the molecular markers used, and ultimately ended with the development of a tool to increase the amount of genetic information available for these weedy plants.

At the onset of my work in *Myriophyllum*, the molecular markers that were commonly used were Sanger sequencing data and AFLPs. Sanger sequencing of internal transcribed spacer and chloroplast DNA markers was limited to distinguishing Eurasian watermilfoil, northern watermilfoil, and hybrids. If the unit of management was at the taxon level (i.e. the presence of hybrid watermilfoil), Sanger sequence data would be sufficient. However, this technique did not provide any information on the genotypes and intraspecific genetic variation in watermilfoil. Fragment based data in the form of AFLPs does provide information on the different genotypes of watermilfoil, but data is error prone (see chapter 2). AFLP markers did provide a data to test for introgression in watermilfoil, which could have implications for the way watermilfoil reproduce (LaRue et al. 2013). The major downside with both of these molecular markers is the inability to calculate allele frequencies for use in population genetics.

Because watermilfoil is hexaploid, fragment-based markers that are able to be used to calculate allele frequencies in diploids, such as microsatellite markers, must be treated as dominant molecular markers. Dosage of a marker is unable to be determined, and it is unknown

if each locus is truly amplified across all subgenomes, or just a fraction of them. Microsatellites, like AFLPs, do provide information on the genetic variation present in watermilfoil but are also error prone, difficult to share data across laboratories due to differences in protocols, and must be treated as dominant data in watermilfoil (see chapter 2).

To solve these issues, I took a sequence-based approach to generate molecular markers in watermilfoil. With DNA sequence attached to each molecular marker, and read depth resolving error through replicate sequencing of each marker, allele frequencies can be calculated with these data providing information not before available with other marker types. The use of sequence based molecular markers opens the door for studies that were previously impossible in watermilfoil, dosage can be determined through haplotypes and read depths, allele frequencies can be calculated, and numbers of molecular markers provide higher density genotyping across the genome. As such, I used these molecular markers to develop a microhaplotype panel to cheaply genotype hundreds to thousands of individuals, providing a standardized method to provide genetic information on watermilfoil across laboratories. Synthesis of all of the work laid out in this dissertation provides an important piece of that panel allowing markers to 1) distinguish taxon, 2) distinguish genotypes within taxon, 3) be inherited in a diploidized fashion, and 4) be distributed across the entire genome.

Future Work in Myriophyllum

The work shown here provides diploidized molecular markers for genotyping in watermilfoil, and on the mode of chromosomal inheritance. While beyond the scope of this dissertation, the diploidized molecular markers generated here can be used to ask questions about introgression at many more areas of the genome than previous studies, and be used to link genotype to phenotype. Introgression can provide insights into the source of genetic variation present in populations of watermilfoil. Coupled with studies on the fitness of individual genotypes within a population, information on areas of the genome important for that fitness can be determined using a genomic clines analysis (Gompert and Burkele 2010). The density of molecular markers provided with a SNP-based method of genotyping can also be used as high-density marker genotype information to begin to link important traits (such as herbicide response or fitness) to specific genes in a quantitative trait locus mapping approach.

CUMULATIVE REFERENCES CITED

Baetscher, D. S., A. J. Clemento, T. C. Ng, E. C. Anderson, and J. C. Garza. 2018. Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Molecular Ecology Resources 18: 296–305.

Basu, C., M. D. Halfhill, T. C. Mueller, and C. N. Stewart. 2004. Weed genomics: New tools to understand weed biology. Trends in Plant Science 9: 391–398.

Benoit L.K., Les D.H. 2013. Rapid identification and molecular characterization of phytoene desaturase mutations in fluridone-resistant Hydrilla *(Hydrilla verticillata*). Weed Sci. 61(01):32–40.

Berger S.T., Netherland M.D., MacDonald G.E. 2012. Evaluating fluridone sensitivity of multiple hybrid and Eurasian watermilfoil accessions under mesocosm conditions. J. Aquat. Plant Manage. 50: 135-141.

Berger S.T., Netherland M.D., MacDonald G.E. 2015. Laboratory documentation of multiple-herbicide tolerance to fluridone, norflurazon, and topramazone in a hybrid watermilfoil *(Myriophyllum spicatum* × *M. sibiricum*) population. Weed Sci. 63(01):235–241.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.

Bourke, P. M. 2018. Genetic mapping in polyploids.

Bourke, P. M., P. Arens, R. E. Voorrips, G. D. Esselink, C. F. S. Koning-Boucoiran, W. P. C. van’t Westende, T. Santos Leonardo, et al. 2017. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. Plant Journal 90: 330–343.

Bourke, P. M., G. Van Geest, R. E. Voorrips, J. Jansen, T. Kranenburg, A. Shahin, R. G. F. Visser, et al. 2018. PolymapR - Linkage analysis and genetic map construction from F 1 populations of outcrossing polyploids. Bioinformatics 34: 3496–3502.

Bourke, P. M., R. E. Voorrips, R. G. F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. Frontiers in Plant Science 9: 1–17.

Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 23(19):2633–2635.

Byers, K. J. R. P., S. Xu, and P. M. Schlüter. 2017. Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology* 26: 277–290.

Campbell N.R., Harmon S.A., Narum S.R. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15:855-867.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. Molecular Ecology 22: 3124–3140.

Chorak G and Thum RA. 2020. Identification of resistant clones of Eurasian (Myriophyllum spicatum) and hybrid (*M. spicatum X M. sibricum*) watermilfoil to an operational rate of fluridone. Invasive Plant Science and Management.

Clark L. V., Jasieniuk M. 2011. Polysat: An R package for polyploid microsatellite analysis. Mol. Ecol. Resour. 11(3):562–566.

Clark, L. V., A. E. Lipka, and E. J. Sacks. 2018. polyRAD: Genotype calling with uncertainty from sequencing data in polyploids and diploids. bioRxiv 9252: 380899.

Clevenger J.P., Ozias-Akins P. 2015. SWEEP: A tool for filtering high-quality SNPs in polyploid crops. G3; Genes|Genomes|Genetics. 5(9):1797–1803.

Davey, J. W., and M. L. Blaxter. 2011. RADSeq: Next-generation population genetics. Briefings in Functional Genomics 10: 108.

Délye, C. 2013. Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: A major challenge for weed science in the forthcoming decade. Pest Management Science 69: 176–187.

Doyle, J. J., and S. Sherman-Broyles. 2017. Double trouble: taxonomy and definitions of polyploidy. New Phytologist 213: 487–493.

Duchesne P., Bernatchez L. 2002. AFLPop: A computer program for simulated and rea population allocation, based on AFLP data. Mol. Eco. Notes. 2: 380-383.

Durand, N. C., M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems 3: 95–98.

Gerard, D., L. F. V. Ferrão, A. A. F. Garcia, and M. Stephens. 2018. Genotyping Polyploids from Messy Sequencing Data. *Genetics* 210: 789–807.Gohl, D., D. M. Gohl, A. MacLean, A. Hauge, A. Becker, D. Walek, and K. B. Beckman. 2016. An optimized

protocol for high-throughput amplicon-based microbiome profiling. Protocol Exchange: 1–26.

Gompert, Z., and C. Alex Buerkle. 2010. Introgress: A software package for mapping components of isolation in hybrids. Molecular Ecology Resources 10: 378–384.

Guastello P.R., Thum R.A. 2018. Mesocosm and field evaluation of Eurasian and hybrid watermilfoil response to endothall in Jefferson Slough, Montana. J. Aquat. Plant Manage. 56: 63-68.

Hardy O.J., Vekemans X. 2002. Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Eco. Notes. 2(4): 618-620.

Hendry, A. P., M. T. Kinnison, M. Heino, T. Day, T. B. Smith, G. Fitt, C. T. Bergstrom, et al. 2011. Evolutionary principles and their practical application. *Evolutionary Applications* 4: 159–183.

Kidd K.K., Pakstis A.J., Speed W.C., Lagace R., Chang J., Wootton S., Ihuegbu N. 2013. Microhaplotype loci are a powerful new type of forensic marker. For. Sci. Int.: Genetics Supp. Series 4. e123-e124.

Kamvar Z.N., Tabima J.F., Grünwald N.J. 2014. *Poppr* : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

LaRue E.A., Zuellig M.P., Netherland M.D., Heilman M.A., Thum R.A. 2013. Hybrid watermilfoil lineages are more invasive and less sensitive to a commonly used herbicide than their exotic parent (Eurasian watermilfoil). Evol. Appl. 6(3):462–471.

Ley A.C., Hardy O.J. 2013. Improving AFLP analysis of large-scale patterns of genetic variation - A case study with the Central African lianas *Haumania spp*. (Marantaceae) showing interspecific gene flow. Mol. Ecol. 22(7):1984–1997.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21):2987–2993.

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26(5):589–595.

Lü, T. F., X. Li, S. N. Wang, L. Han, F. Hao, C. L. Fu, P. Zhang, and X. Liu. 2017. Allohexaploid speciation of the two closely related species Myriophyllum spicatum and M. sibiricum (Haloragaceae). Aquatic Botany 142: 105–111.

Melo A.T.O., Bartaula R., Hale I. 2016. GBS-SNP-CROP: A reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC Bioinformatics. 17(1):1–15.

Michel A., Arias R.S., Scheffler B.E., Duke S.O., Netherland M., Dayan F.E. 2004. Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). Mol. Ecol. 13(10):3229–3237.

Moody, M. L., and D. H. Les. 2002. Evidence of hybridity in invasive watermilfoil (Myriophyllum) populations. Proceedings of the National Academy of Sciences of the United States of America 99: 14867–14871.

Moran, B. M., and E. C. Anderson. 2019. Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences* 76: 551–560.

Netherland M.D., Willey L. 2017. Mesocosm evaluations of three herbicides on Eurasian watermilfoil (*Myriophyllum spicatum*) and hybrid watermilfoil (*Myriophyllum spicatum* x *Myriophyllum sibiricum*): Developing a predictive assay. J. Aquat. Plant Manage. 55: 39-42.

Parks S.R., McNair J.N., Hausler P., Tyning P., Thum R.A.. 2016. Divergent responses of invasive watermilfoil to treatment with auxinic herbicides in a large Michigan lake. Lake and Res. Manage. 32: 366-372.

Pashnick, J., and R. A. Thum. 2020. Note Comparison of molecular markers to distinguish genotypes of Eurasian watermilfoil , northern watermilfoil , and their hybrids. 61–66.

Peakall R., Smouse P.E. 2006. GENALEX: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Eco. Notes. 6: 288-295.

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., Hoekstra H.E. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7(5).

Ravet, K., E. L. Patterson, H. Krähmer, K. Hamouzová, L. Fan, M. Jasieniuk, A. Lawton-Rauh, et al. 2018. The power and potential of genomics in weed biology and management. *Pest Management Science* 74: 2216–2225.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Soltis, D. E., C. J. Visger, D. Blaine Marchant, and P. S. Soltis. 2016. Polyploidy: Pitfalls and paths to a paradigm. American Journal of Botany 103: 1146–1166.

Taylor L-AL, Mcnair J.N., Guastello P., Pashnick J., Thum R.A. 2017. Heritable variation for vegetative growth rate in ten distinct genotypes of hybrid watermilfoil. J. Aquat. Plant Manage. 55(July).

Thum R.A., Heilman, M.A., Hausler P.J., Huberty L.E., Tyning P., Wcisel D.J., Zuellig M.P., Berger S.T., Glomski L.M., Netherland M.D.. 2012. Field and laboratory documentation of reduced fluridone sensitivity of a hybrid watermilfoil biotype (*Myriophyllum spicatum x Myriophyllum sibiricum*). J. Aquat. Plant Manage. 50: 141-146.

Thum, R. A., G. M. Chorak, R. M. Newman, J. A. Eltawely, J. Latimore, E. Elgin, and S. Parks. 2020. Genetic diversity and differentiation in populations of invasive Eurasian (Myriophyllum spicatum) and hybrid (Myriophyllum spicatum × Myriophyllum sibiricum) watermilfoil. *Invasive Plant Science and Management* 13: 59–67.

Van De Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. Nature Reviews Genetics 18: 411–424.

Ward, S. M. 2000. Allotetraploid segregation for single-gene morphological characters in quinoa (Chenopodium quinoa willd.). Euphytica 116: 11–16.

Washburn, J. D., and J. A. Birchler. 2014. Polyploids as a 'model system' for the study of heterosis. Plant Reproduction 27: 1–5.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences of the United States of America 106: 13875–13879.

Wu Z-G., Yu D., Xu X-W. 2013. Development of microsatellite markers in the hexaploid aquatic macrophyte, *Myriophyllum spicatum* (Haloragaceae). Appl. Plant Sci. 1(2):1200230.

Wu Z., Ding Z., Yu D., Xu X. 2015a. Influence of niche similarity on hybridization between *Myriophyllum sibiricum* and *M. spicatum*. J. Evol. Biol. 28(8):1465–1475.

Wu Z., Yu D., Wang Z, Li X., Xu X. 2015b. Great influence of geographic isolation on the genetic differentiation of *Myriophyllum spicatum* under a steep environmental gradient. Sci. Rep. 5(October):1–9.

Zhang J., Kobert K., Flouri T., Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 30(5):614–620.

Zuellig M.P., Thum R.A. 2012. Multiple introductions of invasive Eurasian watermilfoil and recurrent hybridization with northern watermilfoil in North America. J. Aquat. Plant Manage. 50: 1-19.