2011

Analysis of the Genetic Structure of Bithynia tentaculata Snail Populations in Wisconsin and Minnesota

Sarah J. Whalen
Minnesota State University, Mankato

Analysis of the Genetic Structure of *Bithynia tentaculata* Snail Populations in Wisconsin and Minnesota

By

Sarah J. Whalen

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

In

Biology

Department of Biological Sciences
Minnesota State University
Mankato, Minnesota 56001, USA

July 2011

Analysis of the Genetic Structure of *Bithynia tentaculata* Snail Populations in Wisconsin and Minnesota

Sarah J. Whalen

This thesis has been examined and approved by the following members of the thesis committee.

Dr. Robert E. Sorensen, Advisor

Dr. Bradley J. Cook

Dr. John D. Krenz

ABSTRACT

In recent years, there have been tens of thousands of waterfowl mortalities in Wisconsin and Minnesota. An invasive species of snail, *Bithynia tentaculata*, is a host for the trematode parasites (*Cyathocotyle bushiensis* and *Sphaeridiotrema globulus*) that have caused these deaths. A microsatellite-enriched genomic library was detected using DNA from a *B. tentaculata* specimen from Lake Onalaska (Pool 7 of the Upper Mississippi River). Seven polymorphic microsatellite loci were used to genotype snails collected from Lake Butte des Morts, Shawano Lake, and Lake Onalaska in Wisconsin, as well as Lake Winnibigoshish in Minnesota. The genetic diversity of each population was measured as the number of alleles detected at each locus (N_A), observed and expected heterozygosity (H_O and H_E), and allelic richness (A_R). Populations were then differentiated by pairwise F_{ST} values, and the number of genetically distinct populations (K) was estimated. A consensus tree showing the relationship between geographical populations was created using matrices of Nei's distance after repeatedly subsampling (bootstrapping) the data. Cluster analysis showed the genetic data from these snails was best explained by two groups, one containing the eastern Wisconsin populations and the other containing snails from Lake Onalaska and Lake Winnibigoshish. Furthermore, genetic distance and F_{ST} data suggests that the population of *B. tentaculata* in Shawano Lake likely founded the population in Lake Butte des Morts, which then contributed individuals to both Lake Onalaska and Lake Winnibigoshish.

ACKNOWLEDGEMENTS

I would like to take the time to thank my advisor, Dr. Robert E. Sorensen, for his enthusiasm, patience, and willingness to assist me in every aspect of this study. I would also like to thank my committee members, Dr. Bradley J. Cook and Dr. John D. Krenz, for their input during the review process. Many people were involved in helping me collect snails, including Dr. Robert E. Sorensen, Dr. Bradley J. Cook, Emily Koppel, Jeff Lawrence (Minnesota Department of Natural Resources), and Brett Chirempes (UW-Oshkosh). Zac Ruzycki provided assistance with the extraction of DNA from snail tissue and David Hedlund assisted me with the preparation of competent cells for microsatellite cloning. Christina Mangan helped with the amplification of snail DNA using PCR. Zack Bateson provided me with a tremendous amount of assistance during this study, particularly with analyzing genotyping gels and performing statistical analyses with multiple software programs that I was not familiar with. In addition, I would like to thank Kathy Perez (UW-Lacrosse) for agreeing to share the sequences for several *Bithynia tentaculata* microsatellite primers before they were published so that I could use them in my research. Finally, I would like to express thanks to my family, friends, peers, and co-workers for your encouragement and support. Thank you!

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Recently, invasive species such as the emerald ash borer (*Agrilus planipennis*), silver carp (*Hypophthalmichthys molitrix*) and zebra mussel (*Dreissena polymorpha)* have received public attention. An invasive species is defined as an exotic species that occurs outside of its native range due to human activity and increases in abundance at the expense of native species (Primack, 2006). Invasive species are capable of causing immense environmental damage, estimated to cost the United States over $138 billion per year (Pimental *et al.*, 1999). Monitoring and preventing the spread of invasive species is essential in the conservation of biodiversity. However, invasive species themselves are not the only threat to biodiversity. Many invasive species also host and transport parasites that may cause harm to native species. For example, the small hive beetle (*Aethina tumida*) poses little or no harm as a parasite of African honeybees but can cause great harm to European honeybee subspecies by destroying their nests (Neumann, Elzen, 2004).

Fortunately, with advances in molecular biology, the spread of invasive species can be monitored and possibly mitigated. A well-accepted means of inferring population structure and determining ancestral populations is by using microsatellite DNA sequences, otherwise referred to as simple sequence repeats (SSRs), short tandem repeats (STRs) or variable number tandem repeats (VNTRs; King *et al.*, 1997; Schlötterer, 2000). Microsatellite loci are small segments of DNA with 1-6 base pair sequences repeated sequentially 10-30 times (for example, CATACATACATACATA…), which can be used as genetic markers (King *et al.*, 1997; Schlötterer, 2000).

Microsatellite markers can be amplified using the Polymerase Chain Reaction (PCR; Mullis *et al.,* 1986). A single pair of PCR primers is generally sufficient to detect the alleles at a given microsatellite locus in most individuals of the species being studied (Fisher *et al.*, 1996). Most microsatellite regions of DNA are not known to code for specific protein products and provide several advantages for studies of microevolution because they tend to evolve more rapidly than transcribed genetic sequences (King *et al.*, 1997). Mutations occur more frequently during replication of microsatellite DNA than in coding regions due to the repetitive nature of the sequence. These mutations occur as a result of slipped-strand mispairing during the replication process, which changes the number of repeats that a microsatellite allele possesses (Tautz, Schlötterer, 1994). This leads to variation in allele length (in this case, an allele is the length of a repetitive sequence measured in base pairs). Microsatellite markers typically detect greater genetic diversity than other molecules such as mitochondrial DNA and allozymes (Davies *et al.*, 1999).

Another advantage of microsatellite use in studies of population genetics, as compared to mitochondrial DNA, is that they are inherited in a Mendelian fashion. In sexually reproducing populations, each individual inherits one of its alleles at a given locus from each parent. Microsatellites are relatively abundant, and allele frequencies differ among populations. The use of microsatellites as genetic markers has proven an effective tool in the studies of a variety of species, from mapping the diversity of the pathogenic rice fungus *Magnaporthe grisea* to population genetics studies of polar bears,

bottlenose whales, apostle birds, and many others (Crompton *et al.*, 2008; Woxvold *et al.*, 2006; Zheng *et al.*, 2008).

In Lake Onalaska (Pool 7 of the Upper Mississippi River) in western Wisconsin, two trematode parasite species have been responsible for the deaths of over 30,000 waterfowl since mortality was first detected in 2002 (Sauer *et al.*, 2007). These parasites, *Sphaeridiotrema globulus* and *Cyathocotyle bushiensis*, both use the same species of snail, *Bithynia tentaculata*, as an intermediate host in their life cycles. *Bithynia tentaculata*, otherwise referred to as the faucet snail, is a native of Europe (Kipp, Benson, 2010). As an invasive species, *B. tentaculata* has spread parasitic infections in central North America. The faucet snail was first reported in 1871 in Lake Michigan and was later reported in Shawano Lake in 1996, Lake Butte des Morts during surveys for the snails performed between 1998 and 2000, and Lake Onalaska in 2002 (Cole, 2001; Mills *et al.*, 1993; Sauer *et al.*, 2007). According to the Minnesota Department of Natural Resources (MNDNR), *B. tentaculata* has also been found in Lake Winnibigoshish in northern Minnesota, where lesser scaup (*Aythya affinis)* mortalities occurred in both 2007 and 2008 (Lawrence *et al.*, 2008; MNDNR, 2009). It is not known how long these snails have inhabited these lakes, as their presence is usually not discovered until there has been noteworthy waterfowl mortality. According to the National Wildlife Health Center (NWHC; 2011), the first report of massive waterfowl mortality in Wisconsin due to either *C. bushiensis* or *S. globulus* was in Shawano Lake in 1996. The next places waterfowl deaths due to these parasites were reported in Minnesota and Wisconsin were Lake Onalaska in 2002 and Lake Winnibigoshish in 2007 (NWHC, 2011).

Several explanations have been proposed to explain how *B. tentaculata* came to inhabit North America, however the most widely accepted explanation is that *B. tentaculata* first arrived in Lake Michigan in the ballast of timber ships from Europe (Jokinen, 1992; Mills *et al.*, 1993). The presence of *B. tentaculata* was first recorded in Lake Michigan in 1871 (Mills *et al.*, 1993). It spread to Lake Ontario by 1879, the Hudson River by 1892, and to other bodies of water in the Finger Lakes region in New York soon after (Jokinen, 1992; Mills *et al.*, 1993). *B. tentaculata* was introduced to Lake Oneida in New York State between 1910 and 1918 and Lake Erie by 1930 (Carr, Hiltunen, 1965; Harman, 2000; Krieger, 1985). After being introduced into the Erie Canal, *B. tentaculata* began to replace other species of snails (Jokinen, 1992). In addition to negatively impacting the diversity of snail species in the Erie Canal, *B. tentaculata* has also impacted the species richness of mollusks in Lake Oneida (Harman, 2000). In Wisconsin, the presence of *B. tentaculata* was recorded only in the Lake Michigan area and the Wolf River drainage, including Lake Butte des Morts and Shawano Lake, until its presence was noted in Lake Onalaska in 2002 (Sauer *et al.*, 2007).

Historical accounts of the pattern and timing of how *B. tentaculata* has spread throughout the United States are valuable, yet they cannot account for the current geographic distribution of this snail and do little to explain how its range continues to expand. Since *B. tentaculata* is a sexually reproducing species, microsatellite loci should be useful in the study of the genetic structure of its various populations. From the point of introduction, the diversity of alleles within a population will gradually increase due to the mutational changes in microsatellite allele size and the immigration of individuals with

unique alleles from additional populations. Although the magnitude of this increase will be limited by the diversity present at the time of the introduction, it is expected that the older populations of *B. tentaculata* will be more genetically diverse than more recently established and dispersed populations unless gene flow masks the ability to detect such differences. As an aquatic species that appears to be found only in permanent, freshwater lake systems in North America, dispersal of this snail is presumably limited. Avenues of dispersal for this snail might include water currents, boat traffic, or even the birds themselves. A study of the microsatellite loci in *B. tentaculata* should provide insight as to the snail's method of migration.

Mortality of water birds in Pool 7 due to *C. bushiensis* and *S. globulus* was unknown prior to 2002, which raises questions about the origin of these parasites and their host snail, *B. tentaculata,* at this site. Understanding how *B. tentaculata* has spread throughout Wisconsin and Minnesota would prove to be very useful in understanding connections between these populations of snails and could help explain which populations are most likely the source of the *B. tentaculata* population(s) in Pool 7 and Lake Winnibigoshish. Once this information is obtained, various studies could be performed to determine how these snails invade. An awareness of likely modes of transport may prove useful in limiting the spread of these snails and mitigating potentially fatal parasitic infections in other populations of waterfowl in the future.

In the present study, microsatellite loci will be used to assess the genetic diversity of *B. tentaculata* populations at geographic locations in Wisconsin and Minnesota with the goal of understanding the relationship between genetic diversity and a population's

source or age. Ultimately, it will be determined if the population of *B. tentaculata* in Pool 7 descended relatively recently from older populations of these snails adjacent to Lake Michigan in eastern Wisconsin. DNA extracted from *B. tentaculata* will be used to isolate microsatellites, and those microsatellite markers will be used to determine the genetic structure of the *B. tentaculata* populations in Lake Onalaska in western Wisconsin and in various lakes in eastern Wisconsin (Shawano Lake and Lake Butte des Morts), as well as Lake Winnibigoshish in northern Minnesota. The genetic structures of the *B. tentaculata* populations will be determined by the diversity of alleles in isolated microsatellite loci, as well as genotype frequencies in accordance with Hardy-Weinberg equilibrium. Wright's Fixation Index (F_{ST}; Weir, Cockerham, 1984) will then be used to compare the genetic structures of the populations to each other and assignment tests will be performed to distinguish genetically distinct populations (K). In addition, Nei's genetic distance (Nei, 1972) will be used to determine the phylogenetic relationship between the Wisconsin and Minnesota *B. tentaculata* populations.

Insights obtained through the proposed research may be used to understand the mechanisms that enable the spread of *B. tentaculata*. The goal of this study is to perform an analysis of the genetic structure of *B. tentaculata* populations within Wisconsin and Minnesota, which would help determine how these populations are related to each other, and could possibly present clues as to their origins. The working hypothesis of this study is that descendant populations of *B. tentaculata* populations will contain only a subset of the allelic diversity present in their founding populations. Therefore, based on the historical records for the presence of these snails at various sites in Wisconsin and

Minnesota, it is predicted that populations in eastern Wisconsin lakes possess more allelic diversity than populations in Lake Onalaska and Lake Winnibigoshish in northern Minnesota. A corollary of this hypothesis is that the western populations are descendants of the populations in eastern Wisconsin and that the alleles in the more recent populations were originally derived from their ancestor population(s). If this hypothesis is supported it would provide evidence that the Pool 7 and Lake Winnibigoshish snail populations were established more recently. Likewise, it is predicted that the population of *B. tentaculata* in eastern Wisconsin lakes will be more diverse than those in Lake Onalaska and Lake Winnibigoshish. It is also predicted that Lake Onalaska and Lake Winnibigoshish will give the largest F_{ST} values when compared to other populations. If the aforementioned hypothesis is supported by the results of the proposed study, the idea that *B. tentaculata* first reached the United States in or around Lake Michigan and spread west throughout Wisconsin and Minnesota would be supported.

Research Questions

The following research questions will be answered by the proposed study:

1. What are the genetic structures of the *B. tentaculata* populations in eastern Wisconsin lakes (Lake Butte des Morts and Shawano Lake), Lake Onalaska (Pool 7), and Lake Winnibigoshish?
2. How do the genetic structures of the *B. tentaculata* populations in eastern Wisconsin lakes, Lake Onalaska, and Lake Winnibigoshish compare to each other? More specifically, which populations (when compared to each other) have

the largest F_{ST} values and which geographic populations are grouped into the same cluster according to individual genotypes?

MATERIALS AND METHODS

I. Sampling of Bithynia tentaculata Populations.

In May of 2009, 9 water bodies were sampled for *Bithynia tentaculata* specimens: Pool 13 of the Upper Mississippi River, Lake Geneva, Lake Koshkonong, Lake Michigan, Fox Lake, Green Lake, Green Bay, Shawano Lake and Petenwell Lake (Figure 1). Pool 13 was sampled from Thomson, IL and the remainder of the sites are found in Wisconsin. These water bodies were chosen to sample for *B. tentaculata* either because these snails were reported there in the past (such as Lake Shawano), they were near other bodies of water where these snails had been reported and of similar size, or they were directly connected by waterways to other bodies of water where these snails had been reported. Additional snails were collected from Lake Onalaska, Lake Winnibigoshish and the Lake Butte des Morts area, including Lake Poygan and Lake Winneconne.

Snails were categorized as coming from 4 geographic areas (Lake Butte des Morts, Shawano Lake, Lake Onalaska, and Lake Winnibigoshish; Figure 1). Snails were collected using a variety of methods including: deep water sampling of bottom-dwelling snails using a ponar from a boat, picking snails off rocks or vegetation along the shoreline, and using the ponar in shallow water just off the shoreline to collect bottom-dwelling snails. Snails were stored on ice in plastic bags filled with approximately 500 mL of water from their original habitat while they were transported to Minnesota State University Mankato (MSUM). Snails were dissected and examined under a dissecting microscope (6.5x to 45x magnification), to determine the presence or absence of

parasites. The anterior and posterior portions of individual snails were placed in separate 1.5-mL microfuge tubes and stored at -80°C until DNA was extracted.

II. Detection of a Microsatellite-Enriched Genomic Library

The DNA from an individual *B. tentaculata* snail from Lake Onalaska (Pool 7) was used as a source of high molecular weight, genomic DNA. The DNA from this snail (and all other snails used in this study) was extracted according to Hamburger *et al.* (1987). Snail tissue was removed from storage at -80°C and ground in the microcentrifuge tube using a chilled pestle. Once homogenized, 250 μL of lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM EDTA, 5% SDS, and 250 μg/mL proteinase K) were added to the tube and the mixture incubated for 2 hr at 37°C. A phenol-chloroform extraction and ethanol precipitation were performed according to standard procedures (Ausubel *et al.*, 1999), after which the pellet was resuspended in 100 μg/mL RNAse in TE (10 mM Tris-Cl, pH 7.5, 1mM EDTA) and incubated for 3 hr at 37°C. The phenol-chloroform extraction and ethanol precipitation were repeated and the DNA pellet was resuspended in 40 μL of distilled water and quantified using a BioPhotometer (Eppendorf). When electrophoresed on a 1% agarose gel containing ethidium bromide (0.57 g/mL) the sample appeared to contain a large amount of high-molecular-weight DNA.

The extracted DNA from this snail was treated to produce a microsatellite-enriched genomic library using DeWoody's Microsatellite Cloning Protocol (DeWoody, 2002). 1 μg of DNA was digested with 0.5 μL (5 units, 5U) of each of the following

restriction enzymes: *Nhe*I, *Rsa*I, and *Alu*I (Promega). The reaction included 1X Buffer B (Promega) and 20 μg of BSA (Roche) in a total reaction volume of 20 μL. Ten separate microfuge tubes prepared with this reaction mixture were incubated at 37°C for 4 hr. After incubation, 4 μL from each tube was electrophoresed on a 1% agarose gel containing ethidium bromide (0.57 g/mL) to ensure satisfactory digestion of the DNA. The remaining 16 μL of each reaction were pooled.

The ends produced by *Nhe*I in the digested DNA were blunted with mung bean nuclease. To do so, 160 μL of digested DNA, 1X Reaction Buffer (Promega), and 1 μL (10U) of mung bean nuclease (Promega) were incubated for 40 min at 37°C. This DNA was purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and eluted in 50 μL nuclease-free water. The 5'ends were dephosphorylated using shrimp alkaline phosphatase (SAP). This reaction contained 40 μL of purified DNA, 5 μL (5U) of SAP (Promega) and 1X SAP Reaction Buffer (Promega) and was incubated at 37°C, followed by 15 min at 65°C. The digested, dephosphorylated DNA was once again purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and eluted in 37 μL of nuclease-free water. These dephosphorylated restriction fragments were ligated to SNX linkers (SNX forward: 5'CTAAGGCCTTGCTAGCAGAAGC-3'; SNX reverse: 3'-AAAAGATTCCGGAACGATCGTCTTCGp-5'; Hamilton *et al.,* 1999) using T4 DNA ligase. This reaction involved mixing 10 μL of digested and dephosphorylated DNA, double-stranded SNX linkers (forward and reverse, 1.95 μM each), 1X Buffer B (Promega), 10 mM rATP (Promega), 1 μL (20U) high concentration T4 DNA ligase (Promega), 1 μL (20U) *Xmn*I (Promega), and 1X T4 ligase buffer (Promega) in a total

volume of 30 μL which was incubated in a Mastercycler® gradient thermocycler (Eppendorf), where it underwent 30 cycles of alternation between 30 min at 16°C and 10 min at 37°C, followed by incubation at 65°C for 20 min. This product was used as a template for PCR reactions to verify the success of the ligation reaction. Each 15 μL PCR mixture contained: 1X magnesium-free thermophilic DNA polymerase buffer (Promega), dNTPs (320 μM each, Promega), 1.5 mM $MgCl_2$ (Promega), 0.3 μL (1.5U) *Taq* DNA polymerase (Promega), 0.4 μM each SNX double-stranded F & R (forward and reverse) primer, and the remaining volume was made up of a mixture of DNA template (50 ng) and nuclease-free water. The reaction mixtures underwent 35 cycles of 45 s at 95°C, 60 s at 60°C and 60 s at 72°C, preceded by 5 min at 95°C and followed by 2 min at 72°C. These PCR amplifications were electrophoresed in a 1% agarose gel to detect success of the ligation reaction.

The ligated DNA was hybridized to biotinylated microsatellite oligonucleotides ($(GATA)_7$, $(CATA)_7$, $(GATC)_7$, and $(AC)_{10}$,; Integrated DNA Technologies, Inc. (IDT®)). SNX-ligated DNA (500 ng) was added to 150 pmol of biotinylated oligonucleotides and 173.4 μL of 6X SSC (0.1% SDS). The final volume was adjusted to 250 μL with nuclease-free water (giving a final concentration of 4.2X SSC, 0.07% SDS) and the reaction mixture underwent a 5-min denaturation period at 95°C followed by a 2-hr incubation at 55°C in a Mastercycler® gradient thermocycler (Eppendorf). Streptavidin-coated beads (Promega) were used to isolate microsatellite-bearing fragments from the hybridized genomic DNA. In this process, 300 μL of beads were prepared by washing them 3 times in 150 μL of TEN100 (10 mM Tris-HCl, 1 mM

EDTA, and 100 mM NaCl). They were incubated with the hybridized DNA at room temperature for 30 min and magnetized. The mixture was washed 3 times in 500 μL of TEN100 followed by 3 washes in 500 μL 0.2X SSC, 0.1% SDS. The beads were incubated for 5 min at 95°C in 200 μL TE and the supernatant was ethanol-precipitated as in DeWoody's Microsatellite Cloning Protocol (Spring 2002). Next, 22 μL of 3M sodium acetate were added followed by the addition of 500 μL ice-cold 100% ethanol. The mixture was frozen for > 15 min and spun at 13,200 rpm in a Centrifuge (5415 D, Eppendorf) for 10 min. The supernatant was removed and 400 μL of room temperature 70% ethanol were added, followed by a 2-min spin at 13,200 rpm. The supernatant was again removed and the pellet was air-dried, resuspended in 40 μL of TE and stored at -4°C.

Success of the hybridization was tested in a 15-μL PCR amplification containing 1.5, 3, or 6 μL of template (hybridization product). Each reaction mixture contained 1X magnesium-free Thermophilic DNA Polymerase buffer (Promega), dNTPs (320 μM each, Promega), 1.5 mM $MgCl_2$ (Promega), 0.3 μL (1.5U) *Taq* DNA Polymerase (Promega), 0.6 μM SNX F (forward) primer and the final volume was adjusted to 15 μL with nuclease-free water. The reactions underwent 45 cycles of 45 s at 95°C, 45 s at 58°C and 60 s at 72°C, preceded by 3 min at 95°C and followed by 10 min at 72°C in a Mastercycler® gradient thermocycler (Eppendorf). The entire 15 μL reactions were electrophoresed in a 1% agarose gel and visible bands > 200 base pairs (bp) were excised from the gel and refrigerated overnight at 20°C in microcentrifuge tubes. They were

cleaned using a Zymoclean™ Gel DNA Recovery kit (Zymo Research) and eluted in 12 μL of nuclease-free water.

50 μL PCR amplifications were set up according to DeWoody (2002). Each reaction mixture contained: 10 μL of microsatellite-enriched (hybridized) DNA, 1X magnesium-free Thermophilic DNA Polymerase buffer (Promega), dNTPs (320 μM each, Promega), 1.5 mM $MgCl_2$ (Promega), 1.5U *Taq* DNA Polymerase (Promega), and 0.6 μM SNX F primer. Reaction mixtures underwent 32 cycles of 45 s at 95°C, 45 s at 58°C and 60 s at 72°C, preceded by 3 min at 95°C and followed by 10 min at 72°C in a Mastercycler® gradient thermocycler (Eppendorf). 10 μL of each reaction were electrophoresed in a 1% agarose gel and the remaining PCR products of hybridized DNA were cleaned with the Wizard SV Gel and PCR Clean-Up System (Promega) and eluted in 50 μL TLE (0.005 M EDTA, 0.1 M Lithium Chloride, 0.2 M Tris).

The cleaned microsatellite-enriched DNA was digested with *Nhe*I to prepare it for subsequent *Xba*I ligation into the pBluescript® (pBS) II SK (+) plasmid (Stratagene): 10U *Nhe*I (Fermentas), 10 ng/μL BSA (Promega), and 1X Buffer Tango (Fermentas) were added to 43.5 μL eluted DNA and the mixture incubated for > 2 hr at 37°C. Prior to ligation, the pBS II SK (+) plasmid was digested with *Xba*I: 10 μg of pBS II SK(+),10 ng/μL BSA (Promega), 1X *Xba*I buffer (Promega), 0.5 μL (50U) *Xba*I (Promega) and 34 μL nuclease-free water were combined and incubated for > 1 hour at 37°C. Digestion products were cleaned with the Wizard® SV Gel and PCR Clean-Up System (Promega) and eluted in 40-45 μL nuclease-free water. The cleaned, digested pBS II SK(+) was digested with SAP. 40 μL of eluted pBS II SK (+), 5U SAP (Promega) and 1X SAP

buffer were combined and incubated for 30 min at 37°C, followed by 15 min at 60°C. The resulting digested pBS II SK (+) was cleaned using ZYMO DNA Clean & Concentrator™-5 (Zymo Research), quantified using a BioPhotometer (Eppendorf), and a 3:1 (insert:vector) ligation was performed: 300 ng of insert (microsatellite-enriched DNA), 100 ng of pBS II SK (+), 1 μL (20U) of High Concentration T4 DNA Ligase (Promega), 1 μL (5U) of *Nhe*I (Promega), 10 mM rATP (Promega), 1X T4 DNA Ligase (Promega) and 1X Buffer B (Promega) were combined and underwent 30 cycles of 30 min at 16°C and 10 min at 37°C in a Mastercycler® gradient thermocycler (Eppendorf). The pBS II SK (+) plasmid, now containing microsatellite-enriched DNA inserts, was transformed into XL1-Blue competent cells via electroporation. 40 μL of XL1-Blue competent cells (previously frozen at -80°C and thawed for 5 min on ice) and 1.5 μL of ligated pBS II SK (+) plasmid were added to a cuvette, electroporated at 2.5 kV in a GenePulser Xcell (Bio-Rad) and incubated for 30 min at 37°C in 1 mL SOC broth. 100 μL of the transformed bacterial cultures were grown overnight at 37°C on plates of 20 mL of LB agar with 100 μg/mL of ampicillin, 80 μg/mL of X-gal and 20 mM IPTG.

After incubation, colonies containing plasmids with inserts appear white and colonies containing plasmids without inserts appear blue. Therefore, white colonies (i.e. colonies containing microsatellite-enriched DNA fragments) were selected and used in a 15-μL PCR amplification with T3 and T7 primers (specific to the pBS II SK(+) plasmid) to determine the approximate size of the insert. The colony of choice was touched gently with an autoclaved toothpick, which was dipped into a tube containing the 15 μL PCR reaction mixture (dNTPs (320 μM each, Promega), 1X DyNAzyme™ II DNA

Polymerase buffer (Finnzymes), 167 nM T3 primer, 167 nM T7 primer, and 0.24U DyNAzyme™ II DNA Polymerase (Finnzymes)). The tube was incubated for 3 min at 95°C, followed by 32 cycles of 30 s at 95°C, 30 s at 55°C, and 72°C for 30 s, followed by 10 min at 72°C in a Mastercycler® gradient thermocycler (Eppendorf). 7 μL of each PCR reaction was electrophoresed in a 1% agarose gel and PCR products producing bands > 370 bp were used for sequencing.

Approximate sizes of PCR products were determined using LabWorks 4.5 software (Ultra-Violet Products, Cambridge, UK) and an AutoChemi imaging system (Ultra-Violet Products, Cambridge, UK). The remaining 8 μL of the selected PCR products were cleaned with the ZYMO DNA Clean & Concentrator™-5 kit (Zymo Research), resuspended in 15 μL of nuclease-free water and quantified using a BioPhotometer (Eppendorf). For each 100 base pair (bp) of product, 2 ng of PCR product were used in the sequencing reaction (*e.g.* a 300-bp product would require 6 ng of PCR product for sequencing). The required volume of PCR product was added to a 0.2 mL tube and the volume was brought up to 3 μL with nuclease-free water. The remaining 4 μL of the 7μL sequencing reaction consisted of: 2.86 μM of either T3 or T7 primer, 0.5 μL of ABI PRISM™ BigDye® Cycle Sequencing Mix (Applied Biosystems, Foster City, California, USA), and 1.5 μL of 5X ABI buffer (Applied Biosystems, Foster City, California, USA). Sequencing reactions cycled in a Mastercycler® gradient thermocycler (Eppendorf) as following B.A. Roe (unpublished data): 5 min at 95 °C, followed by 99 cycles of 30 s at 95 °C, 20 s at 48 °C, and 4 min at 60 °C. The reaction contents were held at 4 °C until further use.

Sequenced PCR products were cleaned using ZR DNA Sequencing Clean Up Kit™ (Zymo Research) and eluted in 6 μL of nuclease-free water. Entire cleaned samples were mixed with 2 μL of a 5:1 (deionized formamide:ABI loading dye (Applied Biosystems, Foster City, California, USA)) mix in the well of a 96-well plate. Once all of the samples were loaded into the 96-well plate, the mixtures were boiled for 5 min on a hot plate to evaporate excess water. Samples were loaded into a polyacrylamide gel between two glass plates and electrophoresed on an ABI PRISM™ 377 DNA Sequencer (Applied Biosystems, Foster City, California, USA). Sequences were analyzed using 4Peaks version 1.7.2 (Griekspoor, Groothuis, 2005) BaseFinder version 6.2.3 (Giddings *et al.*, 2006), and primers were designed using MSATCOMMANDER version 0.8.1 (Faircloth, 2008) and Primer3 (Rozen, Skaletsky, 2000).

Once primers had been designed, forward and reverse primers were ordered (without the presence of a fluorescent tag) so their polymorphism could be tested. This was done using Rhodamine Green in a 15-μL PCR amplification containing 267 nM Rhodamine Green, 12 ng/mL BSA (Roche), 1X DyNAzyme™ II DNA Polymerase buffer (Finnzymes), dNTPs (320 μM each, Promega), 167 nM forward (F) primer, 167 nM R (reverse) primer, 0.24U DyNAzyme™ II DNA Polymerase (Finnzymes), and 50ng of template DNA. Reactions underwent an initial incubation of 5 min at 95°C followed by followed by 32 cycles of 45 s at 95°C, 45 s at 55°C, and 72°C for 60 s, followed by 10 min at 72°C in a Mastercycler® gradient thermocycler (Eppendorf). Temperature was held constant at 15°C until further use. Samples were electrophoresed on an ABI PRISM™ 377 DNA Sequencer (Applied Biosystems, Foster City, California, USA) and analyzed

using GENESCAN 3.1 and GENOTYPER 2.5 software (Applied Biosystems, Foster City, California, USA). After confirming the polymorphism of microsatellite loci, primer pairs were ordered where the 5' end of either the forward or reverse primer was tagged with a fluorescent dye (HEX, FAM, or TET). All designed primers were ordered from Integrated DNA Technologies (IDT®).

III. Genotyping Bithynia tentaculata Specimens

Snails were genotyped using 10 different primer pairs thought to amplify 11 different loci (primer pair "GA62" appeared to amplify two different size ranges of products, above and below 240 bp). The following 6 primer pairs were designed at Minnesota State University Mankato using previously described methods: "Bt34", "Bt34(GA)", "GA138", "GA144", "GA154" and "GA62". Primer pairs "Bite03", "Bite16", "Bite29" and "Bite40" were designed by Henningsen *et al.* (2010). 50 ng of template DNA was used in each 15 μL PCR amplification which contained: 12 ng/mL BSA (Roche), 1X DyNAzyme™ II DNA Polymerase buffer (Finnzymes), dNTPs (320 μM each, Promega), 167 nM F primer, 167 nM R primer, and 0.24U DyNAzyme™ II DNA Polymerase (Finnzymes). Reaction mixtures incubated for 3 min at 95°C, followed by 32 cycles of 30 s at 95°C, 30 s at primer-specific annealing temperatures (Table 1) and 72°C for 30 s, followed by 10 min at 72°C in a Mastercycler® gradient thermocycler (Eppendorf). PCR products were electrophoresed on an ABI PRISM™ 377 DNA Sequencer (Applied Biosystems, Foster City, California, USA) and analyzed using

GENESCAN 3.1 and GENOTYPER 2.5 software (Applied Biosystems, Foster City, California, USA).

IV. Statistical Analysis

Null Alleles.

MICRO-CHECKER version 2.2.3 (van Oosterhout et al., 2004) was used to check for errors in scoring and the presence of null alleles. Null alleles were found to exist at all 11 polymorphic loci. Concentrations of DNA templates were verified to be approximately 25 ng/μL and PCR amplifications of DNA for individuals expressing a null allele at a given locus were attempted at least 3 times to ensure that nothing amplified (supporting the presence of a null allele). In addition, the quality of DNA of some individuals that did not amplify was confirmed by electrophoresis in an agarose gel. The original data set was used to determine the frequency of null alleles in each population and at each locus.

Genetic Diversity.

FSTAT version 2.9.3 (Goudet, 2001) was used to determine the number of alleles detected at each locus (N_A), observed and expected heterozygosity (H_O and H_E), and allelic richness (A_R). Allelic richness is useful because it accounts for differences in sample size (Mousadik, Petit, 1996; Petit *et al.*, 1998). The number of individuals with one or more private alleles (N_P) and number of effective alleles (N_E) were calculated for each locus and population using GenAlEx 6.4 (Peakall, Smouse, 2006). Deviations from Hardy-Weinberg Equilibrium (HWE) were tested with GENEPOP version 4.0.10

(Raymond, Rousset, 1995). Default Markov chain parameters for all tests in GENEPOP were used (1000 dememorizations, 100 batches and 1000 iterations per batch). An overall significance level of $P = 0.05$ was maintained for multiple tests of deviation from HWE using Sequential Bonferroni corrections (Rice, 1989).

Population Differentiation.

Populations were differentiated from one another by pairwise F_{ST} values (Weir, Cockerham, 1984) calculated in FSTAT version 2.9.3 (Goudet, 2001). The number of genetically distinct populations (K) was estimated using STRUCTURE version 2.3.3 (Pritchard *et al.*, 2010). Methods described by Evanno *et al.* (2005) were used to estimate ΔK, which gives a more accurate estimate of K because it takes into account the rate of change in and corresponding variances of $L(K)$, the posterior probability that the genotypic data fits a hypothetical number of populations (K), between consecutive values of K. The most probable number of populations (K), or clusters, was estimated by using the admixture model and running simulations with K = 1-10 with 10000 replications. Several modules within PHYLIP version 3.69 (Felsenstein, 2005) were used to create dendograms of relationships between geographical populations based on allele frequencies calculated in GenAlEx 6.4 (Peakall, Smouse, 2006). Bootstrap values were computed in Seqboot by resampling over 1000 replications and branch lengths were calculated in Fitch using matrices of Nei's distance (Nei, 1972) as calculated in Gendist (Felsenstein, 2005). Finally, a consensus tree was made in Consense using the branch lengths calculated in Fitch (Felsenstein, 2005). Trees were viewed using the programs NJplot version 2.3 (Perriere, Gouy, 1996) and TreeViewX version 0.5.0 (Page, 1996).

RESULTS

Snail Collection.

Of the 9 bodies of water sampled, *B. tentaculata* specimens were only found in Lake Shawano, where 50 snails were collected. Another 199 snails from the Lake Butte des Morts area (including Lake Butte des Morts, Lake Poygan and Lake Winneconne) were collected in the summer of 2009. In the fall of 2009, 264 snails from Lake Winnibigoshish were collected. A total of 817 snails were collected during 2008 and 2009. Of these, 304 were collected from Lake Onalaska (Pool 7 of the Upper Mississippi River) in the fall of 2008. Snails from all of these collections were categorized as coming from 4 geographic areas (Lake Butte des Morts, Shawano Lake, Lake Onalaska, and Lake Winnibigoshish; Figure 1).

Detection of Bithynia tentaculata *Microsatellites.*

Of the DNA cloned from an individual snail from Lake Onalaska, a total of 215 DNA fragments produced readable sequences. The sequences of 56 of those fragments were recorded; the mean fragment size was approximately 211 nucleotides, with a standard deviation of 117 nucleotides. Using this information, it is estimated that approximately 45152 to 45387 nucleotides of the *B. tentaculata* genome were sequenced during the screening for microsatellites. 52 of the known 56 sequences recorded appeared to contain microsatellites (Appendix 1). Primers were not designed to amplify all of these loci, either because not enough of the region flanking the repeat was known or the repetitive regions were too small to be classified as microsatellites. Of the 52 sequences

that appeared to contain microsatellites, only 11 were found to be unique microsatellite loci which primers were designed to amplify. Four of these primers did not produce reliable data, either because they were not polymorphic or their products were difficult to interpret on a genotyping gel due to the presence of stutter bands. Seven novel polymorphic microsatellite loci (loci 1-7; Table 1) were identified for *B. tentaculata*. One pair of primers produced a bimodal distribution of allele sizes. These were thought to be at 2 different loci, and are referred to as locus 6 and locus 7 (Table 1). Four additional polymorphic microsatellite loci (loci 8-11), designed by Henningsen *et al*. (2010), were also used to genotype specimens (Table 1). The number of alleles per locus varied from 5 (locus 5, GA154) to 13 (locus 2, Bt34(GA)) across all populations. Allele size ranges include inner flanking regions and sequences complementary to the primers (Table 1).

Null Alleles.

Loci 6, 9, 1, and 7 had the highest prevalence of null alleles, in descending order (Table 2). For the rest of the loci, the frequency of null alleles was less than 15%. The population with the highest null allele frequency was Winnibigoshish (almost 30%) followed by Shawano, Onalaska, and Butte des Morts, in descending order (Table 3). As previously stated, loci with the highest frequency (> 20%) of null alleles were disregarded for the remainder of the data analyses. Therefore, only 4 of the original 11 primer sets designed as a part of this study were used in the final data analyses (2, 3, 4, and 5).

Genetic Diversity.

A total of 161 snails were genotyped at the 7 polymorphic loci used in the study (a maximum of 45 snails amplified from Lake Butte des Morts, 47 from Lake Onalaska, 41 from Shawano Lake, and 26 from Lake Winnibigoshish, see Table 4.1). Sample size (N), number of alleles (N_A), number of private alleles (N_P), number of effective alleles (N_E), and allelic richness (A_R), are shown in Table 4.1 and 4.2. It was found that Butte des Morts had the greatest allelic diversity (*i.e.,* it had the highest number of alleles), followed by Onalaska, Shawano, and Winnibigoshish, respectively. The same trend was found for effective alleles. However, a different trend is shown by allelic richness, with Butte des Morts and Shawano having the highest allelic richness, followed by Onalaska and Winnibigoshish (in descending order). Allelic richness for samples from Winnibigoshish and Onalaska was significantly lower than allelic richness for samples from Butte des Morts and Shawano (Mann-Whitney U-test $P < 0.05$). None of the individuals in Shawano had private alleles, whereas Butte des Morts had the highest number of private alleles, followed by Onalaska and Winnibigoshish.

Observed and expected heterozygosity (H_O and H_E), along with deviations from HWE (F_{IS}, a measure of heterozygote deficiency) according to Weir and Cockerham (1894) are shown in Table 5. With the exception of locus 8 (Bite03) and locus 11 (Bite40), the actual observed heterozygosities (H_o) were less than expected (H_E) for all populations. Significant departures from HWE were observed in 16 of the 28 single-locus exact tests after sequential Bonferroni corrections.

Population Differentiation.

With the exception of Butte des Morts compared to Shawano, which showed little genetic differentiation ($F_{ST} < 0.05$), all populations showed moderate genetic differentiation (F_{ST} between 0.05 and 0.15; Hartl, Clark, 1997) when compared to each other (Table 6). Two genetically distinct populations, or clusters, were identified using STRUCTURE version 2.3.3 (Pritchard, Stephens and Donnelly, 2000; Falush, Stephens and Pritchard, 2003). Figure 2 shows a plot of the mean $L(K)$, or the log probability of there being K genetically distinct populations, and ΔK as a function of K. ΔK is the change in the log probability of K, and has been shown to provide a more accurate estimation of K (Evanno *et al.*, 2005). The mean proportion of membership of each pre-defined population in each of the clusters (Q) is shown in Table 7. Figure 3 is a bar plot that was generated in STRUCTURE version 2.3.3 (Pritchard, Stephens and Donnelly, 2000; Falush, Stephens and Pritchard, 2003), and is a graphical representation of the assignment of individuals to different clusters. Cluster 1 is made up mostly of individuals from Butte des Morts and Shawano, while cluster 2 is made up of mostly individuals from Onalaska and Winnibigoshish. The same pattern is seen in the dendogram of relationships (Figure 4) based on Nei's genetic distance (Nei, 1972).

DISCUSSION

The goals of this study were to compare the genetic diversity of *Bithynia tentaculata* populations in Wisconsin and Minnesota and to determine the geographic pattern of invasion by this species. This required the collection of *B. tentaculata* specimens, identification of microsatellite loci for this species, and their utilization in analyses of genetic diversity and population differentiation.

Snail Collection.

Of the 12 bodies of water that were sampled during this study, *Bithynia tentaculata* specimens were found at only 4 areas (Lake Butte des Morts, Lake Onalaska (Pool 7 of the upper Mississippi River), Shawano Lake, and Lake Winnibigoshish). Since snails were not found at 8 of the locations sampled, it is possible that this species is not yet present at those locations or that it was not detectable using the employed methods. The likelihood of not detecting these snails if they were present is low because a high proportion of samples from the other 4 lakes did contain snails. This argues that if *B. tentaculata* were present in the other 8 water bodies it would have likely been observed.

Microsatellite Detection.

Of the 52 microsatellite loci that were detected through sequencing, 7 were found to amplify microsatellite alleles effectively for most individuals; microsatellite data were not used for 3 of those 7 loci because some individuals failed to amplify at those after three or more attempts. The same samples did, however, amplify at other loci. This

suggests that null alleles were indeed present at those loci (Bt34, GA62.1, GA62.2), which may have contributed to the deficiencies of heterozygotes, or F_{IS} values that were found to deviate significantly from HWE (shown in bold in Table 5). Significant deviations from HWE have also been found in similar studies of another invasive species, the zebra mussel (*Dreissena polymorpha),* and were most likely caused by the presence of null alleles (Astanei *et al.*, 2005). Lake Winnibigoshish, likely the most recently established population, based on the timing of reports of waterfowl mortality, had the highest null allele frequency (Table 3). This is particularly interesting since null alleles are expected to be relatively rare, and they were seen in such high frequencies in the population with the smallest sample size. Null alleles can arise due to a change in the sequence of DNA flanking the microsatellite, including where the primer attaches. Perhaps many of the snails in Lake Winnibigoshish have different sequences in the primer sites than those from Lake Onalaska, which is the snail population that the primers were designed from. This may suggest that *B. tentaculata* snails from locations not sampled in this study played a role in founding the Lake Winnibigoshish population. Previous studies demonstrated that this snail species is also found in southern Quebec, Canada (Hoeve, Scott, 1988). Considering that Lake Winnibigoshish is the most northern of our study populations, there is some reason to believe that snails could have been transported from Canada to this site more frequently than the other sites.

Genetic Characteristics of Snail Populations.

Since allelic richness accounts for differences in sample size, it is a more reliable measure of genetic diversity than simply the number of different alleles that are present in

a population (Mousadik, Petit, 1996; Petit *et al.*, 1998). Lake Butte des Morts and Shawano Lake had the highest allelic richness, followed by Lake Onalaska and Lake Winnibigoshish (Table 4.2). Furthermore, allelic richness for samples from Winnibigoshish and Onalaska was significantly lower than allelic richness for samples from Butte des Morts and Shawano (Mann-Whitney *U*-test $P < 0.05$). This pattern could result from either genetic drift through a founding effect, mutation, or both. If the Lake Onalaska and Lake Winnibigoshish populations descended from either the Shawano Lake or Lake Butte des Morts populations (or both) and gene flow between the ancestor and descendant populations are infrequent, the descendant populations will have received a subset of the allelic diversity in the ancestral populations through immigration. The second process also relies on Winnibigoshish and Onalaska to be descendant populations and if the immigration of those snails had been fairly recent, insufficient time may have passed for mutation to generate new alleles derived from the founding alleles, as could have happened in the older populations in Lake Butte des Morts and Shawano Lake.

It is interesting that no private alleles were found in Shawano Lake. Perhaps individuals from Shawano Lake had a large role in colonizing other populations, thus all of the alleles present in Shawano Lake are present in other populations as well. This conclusion fits well with the dendogram based on Nei's genetic distance (Figure 4), which has strong bootstrap support for the placement of Shawano Lake snails at the basal node. It is more difficult to resolve the reason for the large number of private alleles present in the Lake Butte des Morts snail population. Because Butte des Morts has the highest number of individuals with a private allele, it may be reasonable to assume that it

was one of the earliest established populations and through evolution, new alleles arose that were not passed on to any of the other populations studied. This also suggests one-way gene flow between Butte des Morts and Shawano, with Shawano contributing to Butte des Morts diversity but not the opposite. The fact that the snails that were assigned to the Lake Butte des Morts population were actually collected from Lake Winneconne and Lake Poygan, in addition to Lake Butte des Morts, might also help to explain the large number of private alleles in that sample. If those 3 snail populations are actually more genetically isolated than our grouping of them suggests, we may be inappropriately assigning the private alleles of 3 somewhat distinct populations to just one location.

When compared to each other, Butte des Morts and Shawano displayed little genetic differentiation ($F_{ST} < 0.05$), whereas all other population comparisons showed moderate genetic differentiation (F_{ST} between 0.05 and 0.15; Hartl, Clark, 1997; Table 6). Comparing trends in the mean pairwise F_{ST} values, it is apparent that whenever Butte des Morts is compared with other populations, the mean F_{ST} value across all loci is lower than any of the other 3 comparisons. This means that Butte des Morts has more in common genetically with the other 3 populations than they do with each other. This is a good indication that Lake Butte des Morts, which may have received its alleles from Shawano Lake, played a key role in the colonization of Lake Onalaska and Lake Winnibigoshish.

STRUCTURE version 2.3.3 (Pritchard, Stephens and Donnelly, 2000; Pritchard, Wen and Falush, 2010) assigned individuals to 2 different clusters. The number of clusters was determined using ΔK (Evanno *et al.*, 2005). Figure 3 shows that cluster 1 is

made up of individuals from Butte des Morts and Shawano, whereas cluster 2 is made up of individuals from Winnibigoshish and Onalaska. This is consistent with Butte des Morts and Shawano displaying the least genetic differentiation according to pairwise F_{ST} values (Weir, Cockerham, 1984).

The dendogram created in PHYLIP (Felsenstein, 2005), which is based on Nei's genetic distance (Nei, 1972) suggests that the population of *B. tentaculata* in Shawano Lake was the first of those in the present study to be established, followed by Butte des Morts. While it is not certain whether Winnibigoshish or Onalaska was the next location populated, a comparison of Nei's genetic distance (Nei, 1972; Figure 4) suggests that Shawano played a major role in founding all 3 of the other populations (consistent with its lack of private alleles), most likely through the colonization of Butte des Morts, which then shared alleles with Winnibigoshish and Onalaska. This logic comes from a combination of the dendogram in Figure 4, the F_{ST} values in Table 6, and the fact that no private alleles were found in Shawano Lake. Since Shawano Lake showed little genetic differentiation ($F_{ST} < 0.05$) when compared with Butte des Morts, it follows that Shawano and Butte des Morts are very closely related to one another. Since Shawano Lake showed moderate genetic differentiation with both Winnibigoshish and Onalaska populations, it is fair to interpret this pattern as evidence that these populations are more distantly related. Both Onalaska and Winnibigoshish seem to be more closely related to Butte des Morts than they are to Shawano Lake, or to each other (Table 6).

From a historical perspective, within Wisconsin the presence of *B. tentaculata* was only recorded in Lake Michigan and the Wolf River drainage, including bodies of

water such as Lake Butte des Morts and Shawano Lake, until its presence was noted in Lake Onalaska in 2002 (Sauer *et al.*, 2007). Waterfowl mortalities in the Wisconsin and Minnesota due to parasites associated with *B. tentaculata* were reported in Shawano Lake in 1996, Lake Onalaska in 2002, and Lake Winnibigoshish in 2007 (NWHC, 2011). This mortality pattern follows the population genetic pattern for the snail hosts that are associated with waterfowl mortality in that mortality occurred in Shawano Lake before Lake Onalaska and Lake Winnibigoshish. It is interesting that the genetic evidence obtained in this study suggests that Lake Butte des Morts played a key role in founding *B. tentaculata* populations in Lake Onalaska and Lake Winnibigoshish, while there have been no reports of waterfowl mortality in Lake Butte des Morts. It is typical that the presence of *B. tentaculata* is not acknowledged until there has been noteworthy waterfowl mortality. Perhaps *B. tentaculata* were, indeed, present in Lake Butte des Morts before they were present in Lake Onalaska and Lake Winnibigoshish, as the genetic evidence suggests, but the parasite population has not reached a threshold size that favors notable mortality patterns among waterfowl at Lake Butte des Morts. Perhaps molluscivorous water birds, such as lesser scaup and American coot, do not reach population densities at Lake Butte des Morts that favor transmission of *C. bushiensis* or *S. globulus*.

Although the polymorphic microsatellite loci used in this study were useful in determining the genetic relationships between geographic populations of *B. tentaculata* in Wisconsin and Minnesota, it is still unknown how these snails are able to disperse and invade. It is plausible that they spread via water currents, boat traffic, or even via birds. In

this study, the snail populations in Lake Shawano and Lake Butte des Morts were found to be ancestral to Lake Onalaska and Lake Winnibigoshish. This supports the hypothesis that the *B. tentaculata* populations in eastern Wisconsin lakes are older than and ancestral to the population(s) in Lake Onalaska and Lake Winnibigoshish in northern Minnesota. In other words, the western populations are descendants of the populations in eastern Wisconsin and the alleles in the more recent populations were originally derived from their ancestral populations.

Lake Butte des Morts is connected to Green Bay (Lake Michigan) via Lake Winnebago and the Fox River. This may cause one to believe that Lake Butte des Morts is most likely ancestral to other populations of *B. tentaculata* in the Wisconsin and Minnesota, if they first arrived in the United States in Lake Michigan. However, the Fox River flows North into Green Bay from Lake Butte des Morts, not vice versa. Shawano Lake and Lake Butte des Morts are connected by the Wolf River and a series of other lakes (Lake Poygan and Lake Winneconne), and Shawano Lake drains into Lake Butte des Morts via this series of waterways. This, combined with the genetic data obtained in this study, suggests that *B. tentaculata* snails most likely spread downstream using rivers (as they could have from Lake Shawano to Lake Butte des Morts via the Wolf River) and not upstream (for example, from Lake Michigan to Lake Butte des Morts via the Fox River). The genetic information provided by this study suggests that there is more genetic similarity among populations of *B. tentaculata* that are nearer to each other geographically. Thus, in addition to possibly migrating via water currents, these snails are likely being transported from one body of water to another by land or air. Since Shawano

Lake and Lake Butte des Morts are the only populations in this study that are connected by water, it is unlikely that moving with water currents is the primary method of migration for these snails. Other things these water bodies have in common need to be established, such as sharing the same flyway for waterfowl migration or being visited by the same recreational vessels, which could indicate the primary method of migration of *B. tentaculata.*

Table 1. Details for 11 polymorphic microsatellite loci used to genotype *Bithynia tentaculata* specimens.

Locus	Primer Sequence (5'→3')	T_m (ºC)	Repeat Motif	Size (bp)	*k*
1 (Bt34)	F: GGAGAGAGAGAGACAGAAATACATA	53.0	$(TAGA)_{21}$	193-	
	R: *CGTTTTATGCTCACGCTC	51.5		249	10
2 (Bt34(GA))	F: *TTTCTTCTGCACATCAACG	50.8	$(GA)_{18}$	101-	
	R: TCTATCTATGTATTTCTGTCTCTC	48.9		147	13
3 (GA138)	F: CACCACAGGCACGTTCAG	56.5	$(GATA)_9$	206-	
	R: *TGGGAGGACCACAATATCTCAG	55.9		258	9
4 (GA144)	F: *ATGGGTGACTAAACAATAATGTGATTTC	54.1	$(ACT)_{10}$	310-	
	R: ATTTCTCTACAAAACGTGGGGAAG	55.2		325	6
5 (GA154)	F: GGAGACAAGAGGACGAAGC	55.2	$((GACA)_7/(GATA)_5)$	168-	
	R: *TGGTTATTAGACAGTTTCCAAGGC	55.1		184	5
6 (GA62.1)	F: *TTTGGTCCTTATTGGTGAAGAG	52.6	$(GA)_{45}$	199-	
	R: GAAATACACTGCTTGCAACAG	52.4		237	11
7 (GA62.2)	F: *TTTGGTCCTTATTGGTGAAGAG	52.6	$(GA)_{45}$	241-	
	R: GAAATACACTGCTTGCAACAG	52.4		277	12
8 (Bite03)	F: *AGACCTCCCAATGCTTCAGG	57.0	$(AGT)_{14}$	167-	
	R: GCAACGCTCAAGGCAGTTA	55.9		191	9
9 (Bite 16)	F: GCATCACGAGCAGCCTTTA	55.5	$(GTTT)_6$	261-	
	R: *CCATCCATGTTAGTGGAGCC	55.5		235	10
10 (Bite29)	F: *TGCATCGGTGGGTCTGATTA	56.1	$(GTTT)_8$	208-	
	R: GCTAGCCTCGTATTTCCAGC	55.6		228	6
11 (Bite40)	F: GGCAGCAGCGTTATGTTAGAA	55.5	$(ATC)_7$	248-	
	R: *GAAGTTGGCTCTGTAAGACCG	55.4		269	8

The number of alleles amplified is *k*; size indicates the range of observed alleles in base pairs (including the region flanking the repeat). Primers for loci 2-5 were designed at Minnesota State University Mankato. Primers for loci 8-11 were designed by Henningsen *et al.* (2010).
*5' ends of primers were labeled with either 6-FAM, HEX, or TET.

Table 2. The frequency of null alleles by locus.

Locus	Frequency of null alleles
1 (Bt34)	0.225
2 (Bt34(GA))	0.026
3 (GA138)	0.100
4 (GA144)	0.077
5 (GA154)	0.035
6 (GA62.1)	0.433
7 (GA62.2)	0.215
8 (Bite03)	0.028
9 (Bite16)	0.295
10 (Bite29)	0.064
11 (Bite40)	0.101

Table 3. The frequency of null alleles by population.

Population	Frequency of null alleles
Butte des Morts	0.088
Onalaska	0.088
Shawano	0.126
Winnibigoshish	0.279

Table 4.1. Sample size (N), number of alleles (N_A), number of individuals with one or more private alleles (N_P), number of effective alleles (N_E), and allelic richness (A_R).

Locus/Sample	Butte des Morts	Onalaska	Shawano	Winnibigoshish	Mean/locus
2 (Bt34(GA))					
N	45.000	47.000	41.000	25.000	39.500
N_A	12.000	9.000	7.000	3.000	7.750
N_P	8.000	1.000	0.000	0.000	2.250
N_E	3.004	2.502	2.594	1.279	2.345
A_R	9.799	7.118	6.142	3.000	6.515
3 (GA138)					
N	43.000	46.000	38.000	20.000	36.750
N_A	8.000	3.000	3.000	3.000	4.250
N_P	9.000	2.000	0.000	0.000	2.750
N_E	3.087	2.079	2.208	2.332	2.427
A_R	6.592	2.683	3.000	3.000	3.819
4 (GA144)					
N	44.000	43.000	40.000	23.000	37.500
N_A	4.000	5.000	4.000	4.000	4.250
N_P	1.000	3.000	0.000	0.000	1.000
N_E	1.655	3.546	1.574	1.749	2.131
A_R	3.773	4.778	3.979	4.000	4.133
5 (GA154)					
N	44.000	47.000	39.000	26.000	39.000
N_A	5.000	4.000	4.000	5.000	4.500
N_P	0.000	0.000	0.000	0.000	0.000
N_E	2.251	1.984	1.849	2.522	2.152
A_R	4.581	3.717	3.891	5.000	4.297
8 (Bite03)					
N	45.000	45.000	41.000	26.000	39.250
N_A	8.000	7.000	6.000	4.000	6.250
N_P	0.000	0.000	0.000	0.000	0.000
N_E	4.436	4.655	3.413	1.945	3.612
A_R	6.733	6.391	5.632	4.000	5.689
10 (Bite29)					
N	45.000	46.000	40.000	22.000	38.250
N_A	5.000	5.000	5.000	3.000	4.500
N_P	0.000	0.000	0.000	0.000	0.000
N_E	3.944	2.190	3.682	2.960	3.194
A_R	4.993	4.461	4.798	3.000	4.313
11 (Bite40)					
N	45.000	46.000	37.000	20.000	37.000
N_A	6.000	5.000	7.000	4.000	5.500
N_P	1.000	0.000	0.000	5.000	1.500
N_E	3.127	2.175	3.782	1.294	2.595
A_R	5.271	4.162	6.298	4.000	4.933

Table 4.2. Mean sample size (N), number of alleles (N_A), number of samples with private alleles (N_P), number of effective alleles (N_E), and allelic richness (A_R) across all loci.

Mean (all loci)	N	N_A	N_P	N_E	A_R
Butte des Morts	44.429	6.857	2.714	3.072	5.963
Onalaska	45.714	5.429	0.857	2.733	4.759
Shawano	39.429	5.143	0.000	2.729	4.820
Winnibigoshish	23.142	3.714	0.714	2.012	3.714

Table 5. Expected (H_E) and observed (H_O) heterozygosities, and deviations from HWE (F_{IS}) according to Wier and Cockerham (1984). Values indicated in bold indicate samples that deviate significantly from HWE ($P < 0.05$) after sequential Bonferroni corrections.

Locus/Sample	Butte des Morts	Onalaska	Shawano	Winnibigoshish
2 (Bt34(GA))				
H_E	0.667	0.600	0.615	0.218
H_O	0.489	0.447	0.366	0.160
F_{IS}	**0.278**	**0.266**	**0.415**	0.286
3 (GA138)				
H_E	0.676	0.519	0.547	0.571
H_O	0.535	0.478	0.263	0.150
F_{IS}	**0.220**	0.089	**0.529**	**0.749**
4 (GA144)				
H_E	0.396	0.718	0.365	0.428
H_O	0.250	0.326	0.100	0.304
F_{IS}	**0.378**	**0.555**	**0.732**	0.309
5 (GA154)				
H_E	0.556	0.496	0.459	0.604
H_O	0.318	0.277	0.179	0.308
F_{IS}	**0.437**	**0.451**	**0.617**	**0.505**
8 (Bite03)				
H_E	0.775	0.785	0.707	0.486
H_O	0.889	0.822	0.707	0.500
F_{IS}	-0.137	-0.036	0.012	-0.009
10 (Bite29)				
H_E	0.746	0.543	0.728	0.662
H_O	0.689	0.370	0.400	0.5
F_{IS}	0.088	**0.330**	**0.461**	0.267
11 (Bite40)				
H_E	0.680	0.540	0.736	0.228
H_O	0.778	0.326	0.730	0.250
F_{IS}	-0.132	**0.406**	0.022	-0.073

Table 6. Pairwise F_{ST} values among populations for each locus.

Comparison	Locus 2	Locus 3	Locus 4	Locus 5	Locus 8	Locus 10	Locus 11	Mean
Butte des Morts – Onalaska	0.000	0.020	0.200	0.220	0.060	0.130	0.080	0.100
Butte des Morts – Shawano	0.000	0.060	-0.010	0.000	0.010	0.070	0.060	0.030
Butte des Morts – Winnibigoshish	0.100	0.020	-0.020	0.000	0.180	0.030	0.140	0.060
Onalaska – Shawano	0.010	0.040	0.200	0.260	0.120	0.140	0.060	0.120
Onalaska – Winnibigoshish	0.080	0.060	0.160	0.140	0.170	0.080	0.150	0.120

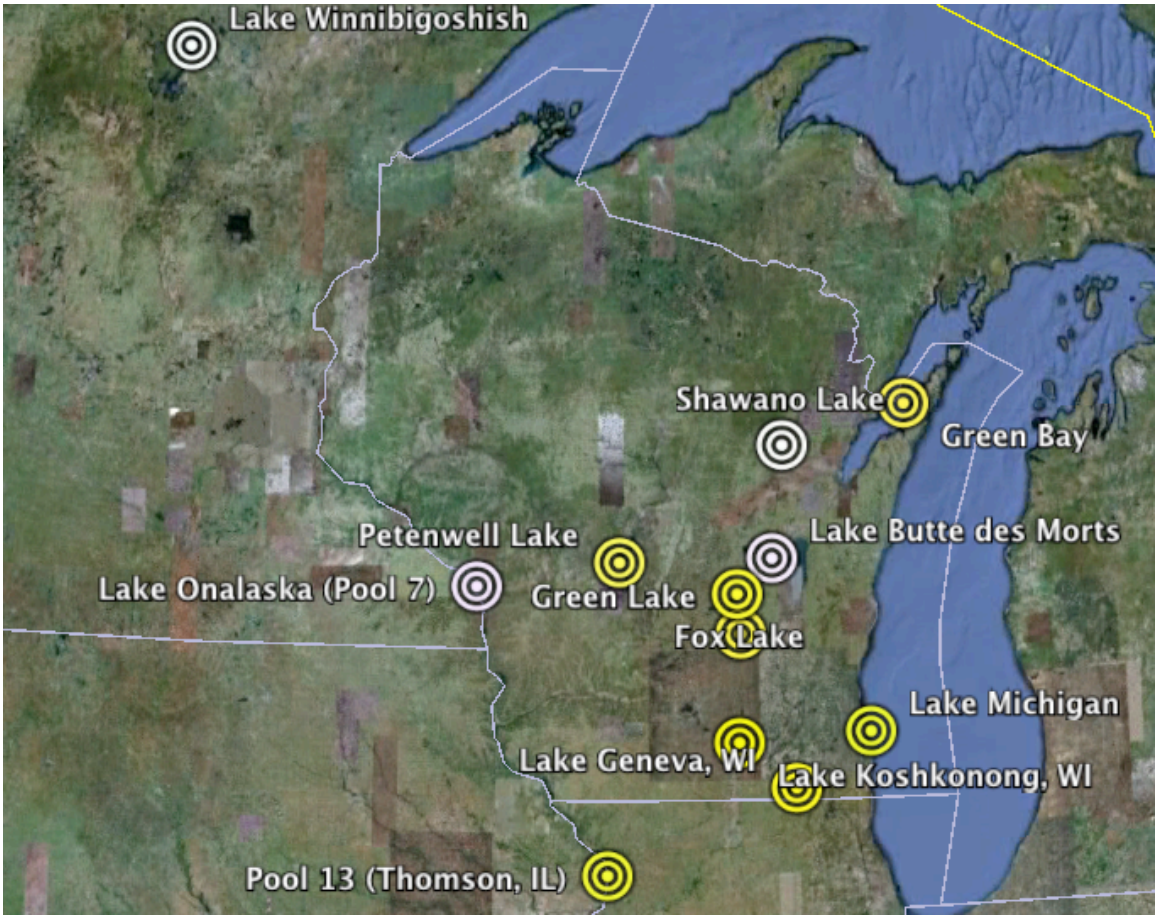

Figure 1. 12 water bodies were sampled for *B. tentaculata* and are represented with a dot: Pool 13 of the Upper Mississippi River, Lake Geneva, Lake Koshkonong, Lake Michigan, Fox Lake, Green Lake, Green Bay, Shawano Lake and Petenwell Lake, Lake Winnibigoshish, and the Lake Butte des Morts areas (including Lake Poygan and Lake Winneconne). Snails were found in 4 areas indicated by a white dot: Lake Butte des Morts (including Lake Poygan and Lake Winneconne), Lake Onalaska, Shawano Lake, and Lake Winnibigoshish. Image from Google Earth (2010).

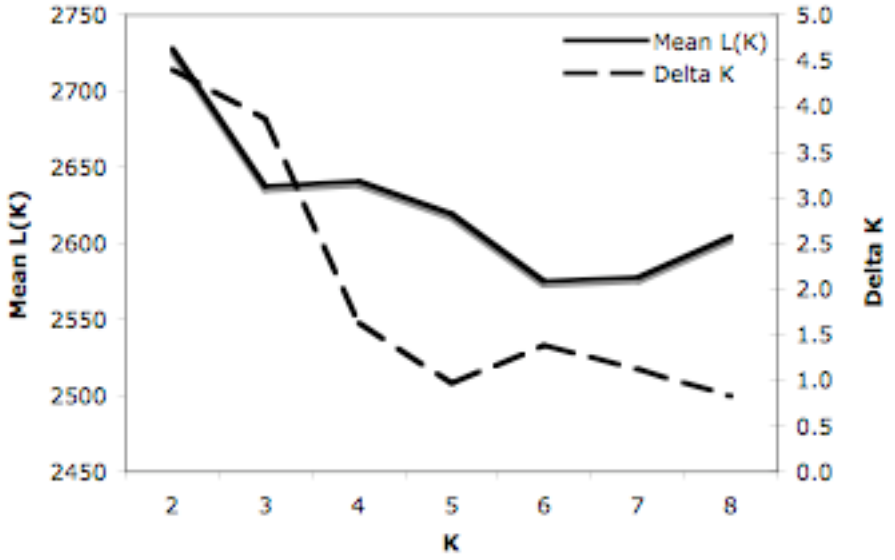

Figure 2. Mean *L(K)*,the mean of the absolute value of the posterior probability that the genotypic data fits a hypothetical number of populations *(K)*, and ΔK as a function of *K*.

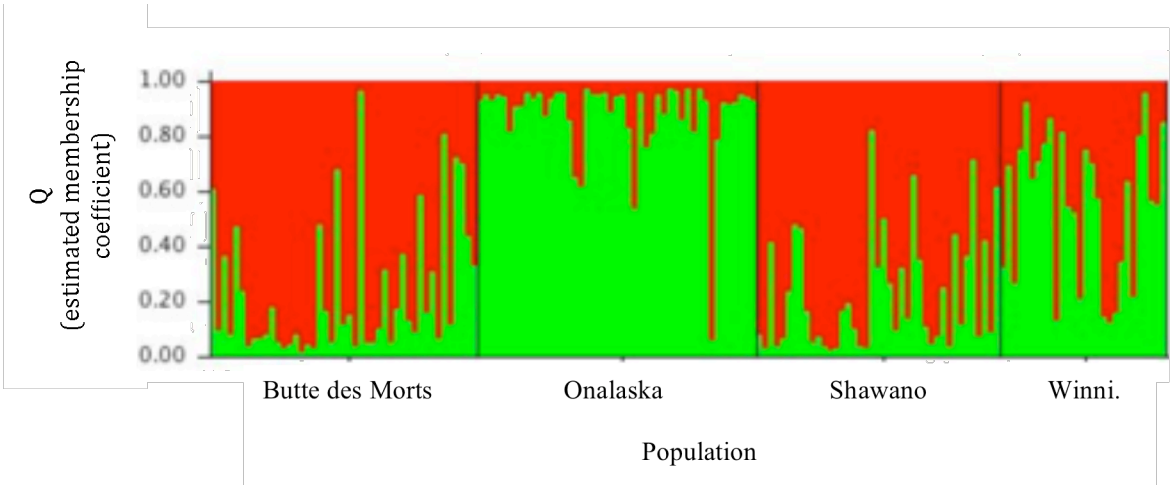

Figure 3. Summary plot of estimates of *Q* (estimated membership coefficients of each individual in each cluster). Each individual is represented by a single vertical line broken into *K* colored segments, with lengths proportional to each of the *K* inferred clusters.

Table 7. The mean proportion of membership (*Q*) of each pre-defined population in each cluster.

Cluster	Butte des Morts	Onalaska	Shawano	Winnibigoshish
1	0.568	0.399	0.570	0.482
2	0.432	0.601	0.430	0.517

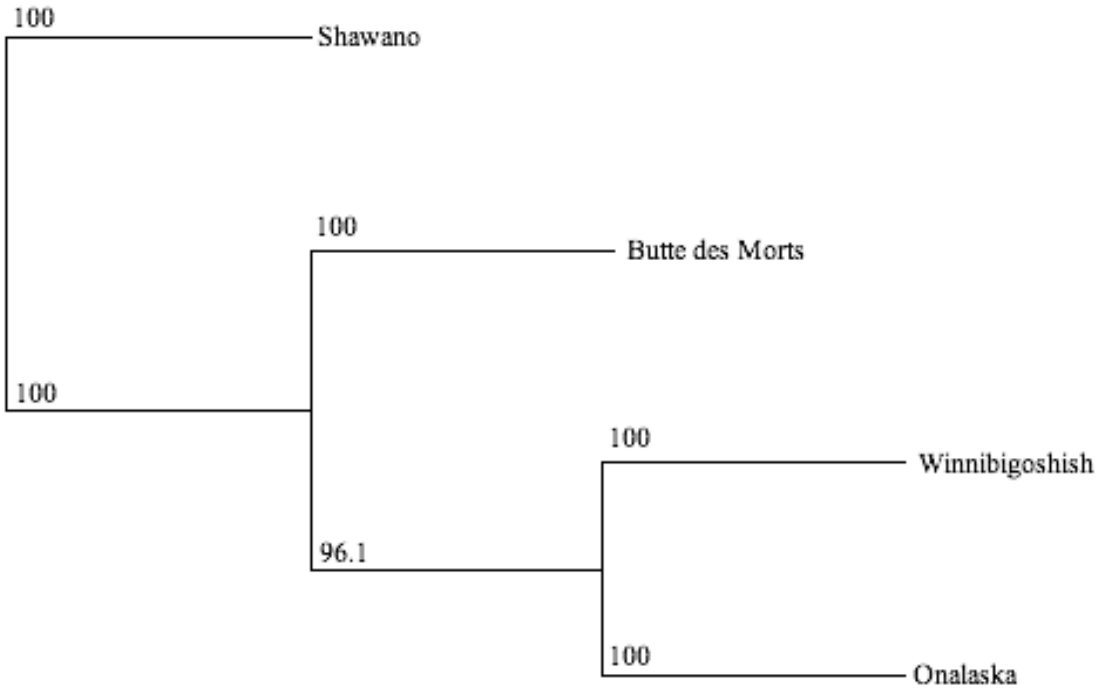

Figure 4. Dendogram of relationships based on Nei's genetic distance (Nei, 1972). Numbers at branch points are bootstrap values computed after repeatedly subsampling (bootstrapping) the data and are given as percentages over 1000 replications.

REFERENCES

(2010) Google Earth.

Astanei I, Gosling E, Wilson J, Powell E (2005) Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymoropha* (Pallas). *Molecular Ecology* **14**, 1655-1666.

Ausubel FM, Brent R, Kingston RE*, et al.* (1999) *Short Protocols in Molecular Biology*, 4 edn. John Wiley & Sons, Inc.

Carr JF, Hiltunen JK (1965) Changes in the bottom fauna of western Lake Erie from 1930 to 1961. *Limnology and Oceanography* **10**, 551-569.

Cole R (2001) Exotic Parasite Causes Large Scale Mortality in American Coots (ed. Interior USDot). U.S. Geological Survey.

Crompton AE, Obbard ME, Peterson SD, Wilson PJ (2008) Population genetic structure in polar bears (*Ursus maritimus*) from Hudson Bay, Canada: Implications of future climate change. *Biological Conservation* **141**, 2528-2539.

Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *TREE* **14**, 17-21.

DeWoody JA (2002) DeWoody's Microsatellite Cloning Protocol (Spring 2002) Based in part on the SNX-linker protocol by Hamilton *et al.* 1999 (*Biotechniques* 27:500) and on Travis Glenn's "Dynabeads2002" protocol.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Faircloth (2008) MSATCOMMANDER ver 0.8.1.

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington, Seattle.

Fisher PJ, HGardner RC, Richardson TE (1996) Single Locus Microsatellites Isolated Using 5' Anchored PCR. *Nucleic Acids Research* **24**, 4369-4371.

Giddings M, Severin J, CVasa S (2006) BaseFinder ver 6.2.3.

Goudet (2001) FSTAT version 2.9.3.

Griekspoor A, Groothuis T (2005) 4Peaks ver 1.7.2.

Hamburger J, Weil M, Pollack J (1987) Detection of *Schistosoma mansoni* DNA in extracts of whole individual snails by dot hybridization. *Parasitology Research* **74**, 97-100.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* **27**, 500-507.

Harman WN (2000) Diminishing species richness of mollusks in Oneida Lake, New York State, USA. *Nautilus* **114**, 120-126.

Hartl DL, Clark AG (1997) *Principles of Population Genetics, Third Edition* Sinauer Associates, Inc., Sunderland, MA.

Henningsen JP, Lance SL, Jones KL*, et al.* (2010) Development and characterization of 17 polymorphic microsatellite loci in the faucet snail, Bithynia tentaculata (Gastropoda: Caenogastropoda: Bithyniidae). *Conservation Genetics Resources* **2010**, 247-250.

Hoeve J, Scott ME (1988) Ecological studies on Cyathocotyle bushiensis (Digenea) and Sphaeridiotrema globulus (Digenea), possible pathogens of dabbling ducks in southern Quebec. *Journal of Wildlife Diseases* **24**, 407-421.

Jokinen E (1992) The Freshwater Snails (Mollusca: Gatropoda) of New York State. University of the State of New York, State Education Dept., The New York State Museum, Albany, New York 12230.

King DG, Soller M, Kashi Y (1997) Evolutionary tuning knobs. *Endeavour* **21**, 36-40.

Kipp RM, Benson A (2010) *Bithynia tentaculata*. U.S. Geological Survey, Gainesville, FL.

Krieger KA (1985) Snail distributions in Lake Erie: the influence of anoxia in the southern central basin nearshore zone. *Ohio Journal of Science* **85**, 230-244.

Lawrence J, Loegering P, Cole R, Cordts S (2008) Scaup and Coot Die-off at Lake Winnibigoshish - 2008 Update. Minnesota Dept. of Natural Resources, Section of Wildlife.

Mills EL, Leach JH, Carlton JT, Secor CL (1993) Exotic species in the Great Lakes: a history of biotic crises and anthropogenic introductions. *Journal of Great Lakes Research* **19**, 1-54.

Minnesota Department of Natural Resources (2009) Faucet Snail (*Bithynia tentaculata*).

Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832-839.

Mullis K, Faloona F, Scharf S*, et al.* (1986) Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **LI**, 263-273.

Nei M (1972) Genetic distance between populations. *American Naturalist* **106**, 283-292.

Neumann P, Elzen PJ (2004) The biology of the small hive beetle (Aethina tumida, Coleoptera: Nitidulidae): Gaps in our knowledge of an invasive species. *Apidologie* **35**, 229-247.

National Wildlife Health Center (2011) Quarterly Mortality Reports. USGS.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

Peakall R, Smouse PE (2006) GenAlEx 6.4.

Perriere G, Gouy M (1996) WWW-Query: An on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364-369.

Petit RJ, Mousadik A, Pons O (1998) Identifying Populations for Conservation on the Basis of Genetic Markers. *Conservation Biology* **12**, 844-855.

Pimental D, Lach L, Zuniga R, Morrison D (1999) Environmental and economic costs of nonindigenous species in the United States. *Bioscience* **50**, 53-65.

Primack RB (2006) *Essentials of Conservation Biology*, 4 edn. Sinauer Associates Inc., Sunderland, Massachusetts.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Populations Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.

Pritchard JK, Wen X, Falush D (2010) Structure.

Raymond M, Rousset F (1995) GENEPOP ver 4.0.10.

Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* **43**, 223-225.

Rozen, Skaletsky (2000) Primer3.

Sauer JS, Cole RA, Nissen JM (2007) Finding the Exotic Faucet Snail (Bithynia tentaculata): Investigation of Waterbird Die-Offs on the Upper Mississippi River National Wildlife and Fish Refuge (ed. Interior USDot), pp. 1-3. U.S. Geological Survey.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371.

Tautz D, Schlötterer C (1994) Simple sequences. *Current Opinion in Genetics & Development* **4**, 832-837.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER ver 2.2.3.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Woxvold IA, Adcock GJ, Mulder RA (2006) Fine-scale genetic structure and dispersal in cooperatively breeding apostlebirds. *Molecular Ecology* **15**, 3139-3146.

Zheng Y, Zhang G, Lin F*, et al.* (2008) Development of microsatellite markers and construction of genetic map in rice blast pathogen *Magnaporthe grisea*. *Fungal Genetics and Biology* **45**, 1340-1347.

APPENDIX 1

DNA Sequences for Isolated Microsatellite Loci

This section describes the nucleotide sequences of microsatellite loci identified from an extract of genomic DNA from *Bithynia tentaculata* snails collected from Lake Onalaska, WI, USA. Titles for sequences have the following format: *sample name-oligonucleotide probe used-colony number sequenced-primer used.* Sequence titles are in bold. "Sample name" indicates the individual snail the DNA came from. For example, "Bt6A" was the title given to the snail to isolate many of the sequences. All of the other sequences were isolated from one other snail from the same site. "Oligonucleotide probe used" refers to the biotinylated microsatellite oligonucleotide that was hybridized to the snail DNA ($(GATA)_7$, $(CATA)_7$, $(GATC)_7$, or $(AC)_{10}$,; Integrated DNA Technologies, Inc. (IDT®)). "Colony number sequenced" is the number that was assigned to the bacterial colony that the DNA was cloned in. "Primer used" indicates whether the T7 or T3 primer for the pBluescript® (pBS) II SK (+) plasmid (Stratagene) was used to obtain the given sequence. If "Consensus" is found at the end of the sequence title, both primers were used in separate sequencing reactions to obtain the final sequence. Text underlined once indicates repetitive sequences (potentially microsatellites). Double-underlined text identifies primer sequences that were used in attempts to amplify that repeat. An "N" in the sequence means that this base was unable to be called with confidence. All sequences are in the 5' to 3' direction for the specified reading frame (T3 or T7).

Bt6A-GATA/C-1-T7
CTGGACGGACGGACGGACGGACGGACGGACACGAAAGAGAGAGAGAGAGACAGAC

Bt6A-GATA/C-3-Consensus
ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTGTTGTCTGTC
TGTCTGTCCGTCCATCTGTCTGTCTATCTATATATCTATCTATCATGTCTGTCATGTTCATATGT
CTATCTATGTATCTATCTATCTGTCTATCTGTCTGTCTATCTGTCTGTCTATCTATCTGTCTATCT
ATCTATTTATATGTCTATCAGTCTGTCTATCTATCTATCTATCTATCTATCTAT

Bt6A-GATA/C-4-T7
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA
TAGATAGATAGATAGATAGATTACAGTAAGAAGCCGTAATCAGGCCTTAGCGCCTCTGTATCT
TAAAATGTTCAATAAAATGTACACCGGGATTACACGTATTAACTAAGACAATACATGTTAATT
ATTACATGATTAAGGTAGTTTATCACCAAT

Bt6A-GATA/C-10-T7
GGGCTANAAGGTAGAGAGATAGAGTGAGGGCTAGAAGGTAGAGAGAGAGAGTGTGGGCTAT
GACGGTAGAGAGTTTGGAGTGTGGGCCAGAGGCAGAGAGAGAGAGTTTGGGCTAGACGGTA
GAGTTTGGAGTGAGGGCTAGAAGGTAGAGAGAGAGAGTGTGGGCTAGAAGGTAGAGAGAGA
GAGAGTGTGGGCTAGAAGGTAGAGAGTTTGGATGTAGGGCTAGAATGGTAGAGAGAGAGAG
TGTGGGCTACAAGGTAGAGAGTTTGGAGTGAGGGCTAGAAGGTAGAGAGAGAGAGAGTGGG
CTATAATGTAGAGAGAGAGAGAGTTGTGGGCTAGAAGTGTAGAGAGAGAGAGTGTGAGTTAG
ATAGTATAGAGAGAGGAGGCTTCTGCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTT
GCTTCCTT

Bt6A-GATA/C-15-T7
CCCTCTACAGAAATGATAGTGTAAGACATCGGTATGCATGCCTTCTTGCTCGCTTTGCTGGGG
CAATTAAGGTAGACCCTTAACCGAAATCGTGTTCACTATACCATATATATGGTCATTTGCATTT
AAGTATAATGGGCTAATAATAACGGCAATAATACATGTCATGTGGGTGTAATGTGTTAGAACA
ATGTGCAAAAAATGTGTGTGATTTTGCCGTGTATACCAACATGTATGTCCATTTCGAATTTAAT
TTGTATCTGTGTAAGTGCAGTTAGAAAATATTGTAGCTTATGGTAAAGTTATGATTTCATGTCT
TATTTGCACTATCTTTTATAAATGTACCAAATTAATTTCTCTTTGGGAGATAATAAAGTAGCCC
AATCTTATTCTTATAGATAGATAGATATATAGATAGATAAATAGATAGACAGACAGACAGAC
AGACAGACAGACAGACAGACAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATG

Bt6A-GATA/C-23-T7
TCTATCTATCGTATCTATCTATCTATCTATCTATCTATCTATTCATCTATCTATCTATCTATCTAT
CTATCTATCTAATCTATCTATCTATCTATATCTATATTTGCGCGCATGGAAAAGCCAATATTTC
GGTTGATGTACWWCATGGATGCATCNGGTCNTT

Bt6A_GATA/C_32d_T7
CTTTCCGACTTCACTCAATCGTCCTTCTCTGCACCTTCCCTCCAGCCAGCCCCGTGATATTTGG
GCCGATTACAGCATGTCATAATATCATACCAAAAGATCGGAGCACTTGCGTTTTATGCTCACG
CTCTTCTCAGCATCTCTTCGCCCCAGAATTCCCCTCCATTTTTCTATCTCTCTTAATGTCCATCT
CTCTTTCTTTCTCTCTTTCTGTCTCTCTCTCAAATCTATCTATCTATCTATCTATCTATCTATCTA
TCTATCTATCTATTCTATTCTATCTATCTATCTATCTATCTATCTATCTATGTATTTCTGTC
TCTCTCTCTCCCTCTCCCTCTCTCTCTCT

Bt6A-GATA/C-34-Consensus
TTTCTTCTGCACATCAACGCGATGGGGTAACGAGCTGCAGAAGGAGAGAGAGAGAGAGAGAG
AGAGAGAGAGAGAGAGAGGGAGAGGGAGAGAGAGAGACAGAAATACATAGATAGATAGAT
AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG
ATAGATAGATTTGAGAGAGAGACAGAAAGAGAGAAAGAAAGAGAGATGGACATTAAGAGAG
ATAGAAAAATGTAGGGGAATTCTGGGGCGAAGAGATGCTGAGAAGAGCGTGAGCATAAAAC
GCAAGTGCTCCGATCTTTTGGTATGATATTATGACATGCTGTAATCGGCCCAAATATCACGGG
GCTGGCTGGAGGGAAGGTGCAGAGAAGGACGATTGAGTGAAGTCGGAAAG

Bt6A-GATA/C-66-T7
TTTGCTTCTGCACATCAACGCGATGGGGTAACTGAGCTGCAGAAGGAGAGAGAGAGAGAGAG
AGAGAGAGAGAGAGA

Bt6A-GATA/C-79-T7
CTTTCCGACTTCACTCAACGTCCTTCTCTGCACCTTCCCTCCAGCCAGCCCCGTGATATTTGGG
CCGATTACAGCATGTCATAATATCATACCAAAAGATCGGAGCACTTGCGTTTTATGCTCACGC
TCTTCTCAGCATCTCTTCGCCCCAGAATTCCCCTCCATTCTTCTATCTCTCTTAATGTCCATCTC
TCTTTCTTTCTCTCTTTCTGTCTCTCTCTCAAATCTATCTATCTATCTATCTATCTATCTATCTAT
CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATGTATTTCTGTCTCT
CTCTCTCCCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTC

Bt6A-$(AC)_{10}$-7-T7
ACTGCAGTGCTTCTCAGTATGTGTGTGTGTGTGTGATTGTGTGTGTGTGATTGTGTGTGTGTGT
GTGTGTGTGA

Bt6A-GATA/C-51-T3
CATGGATCTTGTGAGAGAAGTTGAAAGCATTATCCGTGCCTATTATTACTAGGCTACTACTAC
TACTAACTAGTACTACGTAAATGTAATACTAAGCAGCACTATTTTAAGACTCTCACTCATTCAT
ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTTACTCTATCCATCCATCCATC
CATCCATCTATCCATCTATCTATCTCACCCGTATCTGCGCAGTTGTTAAAGATCGTTCTTTCAG
CCGCCTTCAAGCG

Bt6A-GATA/C-53-T3 (Consensus)
ACCCATTTGTCACATCGTACAACAGACATTCTATCGATCTGTCTCTGTTCTTACTTGCGCTATTT
TACACATTTGTCACATCGTACAACAGACATTCTATCGATCTGTCTCTGTTCTTACTTGCGCTAT
TTTACCCATTTGTCACATCGTACAACAGACATTCTATCGATCTATCTTTACTCTTACTTGCGCTA
TTTTACCCATTTGTCACATCGTACAACAGACATTCGTATCGATCTATCTCTACTCTTACTTGCG
CTATTTTACCCATTTGTCACATCGTACAACAGACATTCTATCGATCTATCTCTACTCTTACTTGC
GCTATTTTACCCATTTGTCACATCGTACAACAGACATTCGA

Bt6A-ID202-T3
TTTGCTTCTGCACATCAACGCGATGGGGTAACGAGCTGCAGAAGGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGGGAGAGGGAGAGAGAGAGACATGAAATACACTAGATAGATA
GATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA
TAGATAGATTTGAGAGAGAGACAGATCAG

Bt6A-ID203-T3
TCTACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCATATCTA

Bt6A-ID79-T3
TTTGCTTCTGCACATCAACGCGATGGGGTAACGAGCTGCAGAAGGAGAGAGAGAGAGGAGAG
AGAGAGAGAGAGAGAGAGAGAG

Bt-CATA4.1-T7
ACGACAAACACTCCACGTTCCCTCGCCAGACCCACGCTGGTCCAACCTTGCCAACTGCGCTAG
TTATAGCCGTATGAAAGGGGTTATTGAGAGGGAGAGAGGTTGAGCGAGAGTTTGTAACAGTC
ATACAGCGAGAGAGAAAGAGAGAGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTATATATGGCGTCGTTTTCTTA

Bt-AC1-T7
GCTAGTGACTAGACTAAAGTTATAACTTTTATGGAGTTGTCTTATAACGAACTGCGTGTGTGT
GTGTGTGTGTGTGTGATGTGTGTGTGTGTGCGTGTGTTGTGTGTGTGTGTGTGTGTA

Bt-AC8.1-T7
GGCTCCAGCCGCTTCCCTGTGAGCAATAAGAAACGACGCCATATATACACGCATACACACAC
ACACACACACACACACACACACATCATTAAACTTTCAATGCTCAAATAGTTTACTTCCAGTCG
GGATTGAGAAGGGGTCAGATGGTTTATG

Bt-CATA7-T7
AAGTATAATAATTGACAATTATTGTGGTGATGATAATCATCCTGTTCTTCATCTTCCTCGTCTG
ACAGGGGCTGCAGATAAATACGTAGTTACCTTTTTATTTCCACTTACGGACTGTCGAACCAGA
GAGATAGCAAACGCCCGTTGGGTGGCTGTCTACAGATAACTTGGATTACGCCCCGTTGCCTTG
CGTGGGAGATAGAAAGAGAATTAGACAGTATGTGTGTGTGTGTTTGTGTGTGTAGGTGCGTGG
GTGTGTCNATTGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

Bt-CATA14-T7
ATGTGCTGCCAAGAGGGTATAAAGTATAAACGGTTCAGGTTATCGATTTCGCGATACGTATTC
CTTGTCACGAACAGGCTAAGATTGGAGCCAGCCTGTAAAAGGGATCCAACATGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGCCTGTGTATGCGCGCTATATTATGGCGT

Bt-CATA21-T7
AACACGGGTATGTGTGTGAATGTGCGTGCGTGTGTGTGTGTTTGCGTGTGTGCGTGTGTGT
GCGTGTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTG

Bt-GATA143-T3
GACACAAGCGATTTAAGTATAAAAACAGAAAGACAGAGAGAGAGAGACAGAGACAGACAGA
CAGAGAGTGATAGAGAGAGAGGATAGAGAAATAGAGAAGAGAGAGGGGGAGAGAGAGAGA
GAGGGGGGGGAGAGA

Bt-AC15-T3
CCGAACTAATGTGAGTGTGTGCATAAGTAGGTGGTGTGTGTATGTGAGCGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTATGCGTGTATATATGGCGTCGTTTCTTATTGCTCAC
AGGGAAGCGGCTGGAGCC

Bt-AC19-T3
AACCAAGACGTGACTGGTCTGTAGCACCAGTCTCCATTCAGAAAGTTGCTGGTTCGTCTGTTG
CGAAATCTAGCAAGGCTCTCCTTCANAGTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTATG
CGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCGGCTGGAGCC

Bt-AC23-T3
GGCTCCAGCCGCTTCCCTGTGAGCAATAAGAAACGACGCCATATATACACGCATACACACAC
ACACACACACACACACACACACACACACACAAGCACTGTTAACGGTCAATAACCCACTGT

Bt-AC30-T3
GTGTGGTGTGGCCATTTTTTTTAGGTAGCCTTGGACACACCCGTGAGTAGTTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTATGCGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGC
GGCTGGAGC

Bt-AC31-T3
CATCTGCATTGAGTAATACACACAATGACACAAGTCTGTTTTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTATGCGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCGCTGGA
GCC

Bt-AC34-T3
ATCACAACTCTGCCTCCCGGTGGTGGCGCACCAGGGTCGGCCGCACCTCCCCCTAGCCAACCG
CACAACTGCCCTTTTCTGTCTTTCTTTTTCCACCATGGATTCCGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTATGCGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCGGCTGGA
GCC

Bt-AC62-T3
AACCAAGACGTGACTGGTCTGTAGCACCAGTCTCCtATTCAGAAAGTTGCTGGTTCGTCTGTTG
CGAAATCTAGCAAGGCTCTCCTTCAnAGTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTATG
CGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCGGCTGGAGCC

Bt-AC65-T3
ACTTGCTGAGGTTGCTCGCTCTTGACGGGATGTTGCTGGCACAGAGGGGTAGTCTGTCCTGAA
AGTGCCACCAGACACAAACATTNGTATTTTTATCGTCCATTAATTATTTTTGTGCGAGTGTTTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCG
GCTGGAGCC

Bt-CATA25-T7
TGAATATGATGTAGCAAATAATACCCCCTGAACAATGAACCGATTTAGCCGGAaACAAAAAA
GTCTGGACACCATAAATAGGGACACTAAAAAGCCGCGAGGCTTAATCAAGTCCCTCTGAACT
GACCTGgATGACTGTCATGTAATTATTCTCTGTCTACCTGTCTGTGCTTTGCATAAAgGGTTTTG
TGTGTGTGTGTGTGTGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGCgTGTATAT
ATGGCGTCGtTTCtTAtTGCTCACAgGgaAgCgGCTGGAgCc

Bt-CATA26-Consensus (T7 reading frame)
TCCTCGTCCCTTTCCACCGACGCAGGATTCCTTACCCTAACAGAATCGCNTAAACAAAACAAA
AAAACAAAAATCCGTTGGGATTTCGGCGATGTATTTGTGTGCGTTTGTGGTTGATTGTGTGTAT
CTGTTTATTTTATGTGTGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGCGTGTATATAT
GGCGTCGTTTCTTATTGCTCACAGGGAAGCGGCTGGAGCC

Bt-CATA27-T7
TCGGCCTTTAAATGCTTAATAAACGAAGTGTGAGGTATCCTAAGAAATTAGTGTGCCAGTGTG
TGTGTGTGTGTGGTGTGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAATAT

Bt-CATA39-Consensus (from T3 reading frame)
GGCTCCAGCCGCTTCCCTGTGAGCAATAAGAAACGACGCCATATATACACGCATACACACGC
ATACACACACACACACACACGCACACACTCGCAATCGACACACCCACGCACCTACACACACA
AACACACACACACACATACTGCTTAATTCTCTTTCTATCTCCCACGCAAGGCAACGGGGCGTAAT
CCAAGTTATCTGTAGACAGCCACCCANCGGGCGTTTGCTATCTCTCTGGTTCGACAGTCCGTA
AGTGGAAATAAAAAGGTAACTACGTATTTATCTGCAGCCCCTGTCAGACGAGGAAGATGAAG
AACAGGATGATTATCATCACCACAATAATTGTCAATTATTATACTT

Bt-CATA41-T3
TGTGTATGTAGGTGTGTGTGTGTGTAATGCAAGCCATGCCCAAGTTTAAACAAAACCATGAAA
ACCAAAGCAAATAATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGC

Bt-CATA45-T7 (Consensus of T3 &T7; read in T7 frame)
CGGACCGATGACGGACGCTGGCTGGGTGTCGTCCATGGCTTGATTTATTCTAAATTTTTAGTTT
GTTTTCAGGTGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGCGTGTATATATGGCGTCG
TTTCTTATTGCTCACAGGGAAGCGGCTGGAGCc

Bt-CATA67-Consensus (read in T7 frame)
GAGTGTTAGTGAGTCCTTTCAAGTAATTATTGTTTCTTTATCGTTGTAATTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGCGTGTATATATGGCGTCGTTT
CTTATTGCTCACAGGGAAGCGGCTGGAGCC

Bt-CATA69-T3
NCGAAGAGAGCNCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTA

Bt-GATC31-T3
ACAGGCACCGTCGGTAGATGGCACCACTGGTCTTGCATTCGTTACCAGTCTGCTGTTTCGGGC
AGGAACAGACATCACAGTGCCATAATGACCAAAACCTAATGCGGAACGAGGAAGACAACNN
AATACAGAAAGAAGACGAAGAGATAGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGAGAGAGAGAGAGGGAAGTGGAAAGGATGGGAGAGAGGGAGGGAGAAGAGAA

Bt-GATA144-T3
ATTTCCATTACACAGTCATGGGTGACTAAACAATAATGTGATTTCTTCCCACGAAACTTTCTGT
AAGTCTCGTAGGTATACCCATAAGGGACATTCTGAGGTCTCGTAGAGGTGTGTTAGTGTCAGC
ATCGGCGTATCAGCATCCTATAATTTAATTGCTACTACTACTATTACCACTACTACTGCTATGA
CATCACTACTACTACTACTACTACTACTACTAAAGGGGATATCCACACTAACAAACCCTA
TTAGACCTCAGAATTAGATCTCAGAATATACCTTTGAGACCGAGAGACTATCTGAGAATACTC
ATCNCCACGTTTGTAGAGAGATCCGTAGTTATTCTGTCTCTGTTG

Bt-GATA-P138-T3
GCGGCCCATTGCTGTTCACCACAGGCACGTTCAGTTCATGCAATTTACATTCAGTAAATTTTTCC
TATCTATCTATCTATCTATCTATCTATCTATCCCATAGTCCTGCAGACCGTTGGGGGCAC
CACCAACAACTCGACAAACAACTTCCTCCATTCCACTCTGTGTTCTTTGACATTCTTAGTGCAT
GGCTCAGTTTCAGTCCTGCTGACCCTGAGATATTGTGGTCCTCCCAATTTTTCTCCCTCtGTCTG
CGTCCCCTTCGCCCCTT

Bt-GATA62-T3 Consensus
ATTTCGGTCAAACTTAAAGTATTCTTCAAACTCATTAAACACATTTCCTAAAATTTGGTCCTTA
TTGGTGAAGAGAGACAAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT
TGGGGTTAGAATTAAAGAGGACAAGAGAGAGAGAGAGAGAGTTGGGGTTAGAATTAAAGAG
GACAAGAGAGAGAGAGAGAGAGAGTTGGGGTTAGAATTAAAGAGGACAAGAGAGAGAG
AGAGAGAGAGAGTTGGGGTTAGAATTAAAGAGGACAAAACAAACACATATTAAAAAATGTTT
TGCTGTTGCAAGCAGTGTATTTCCTCTGACAGGCAATGGAAACTGT

Bt-GATA-P200-T3
TGCCAAATCATCAGTGGTGCCCCAACGGCCGACAAGTCTACGGGATAGATAGATGGATAGAT
ATGATAGATAGATAGATGGATAGATAGATAGATAGATAGATATGATGGATATGATAGAATTA
AAGCGAATGTTCCCATACTACTCGTTACCTAGCcCCCAGTCCGCTAT

Bt-GATA-P205-T3
AGCCCCTGGTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCATGATCCGCGGACTC
CTGTAAAGTAAGGATTTAATTTACGGAATGAACGGCATTTCTAGAACAGTAAATTTATCAATA
CAGAAATGCCGTTCA

Bt-GATA-P206-T3
TCTCTCTCTTTCTCTCTCTCTCTCTCGCTCTCTCTCTCTCNGGGGGACGCACAGAGCGCAAATG
NAATGGCACCAGTGTTCAAAATGGGCTCTGTATATTTTTGTTAATATCTGCTGGATTGGAGACT
TTGTTTGTGTTGAAATTTGACAAGGATCTTTGTTAGGTC

Bt-GATA-P217-T3
GGTCATAACGCTTTCCTCTTTTGACCTTAGCTGTGGTGTTAATAGATAGATAGATAGATAGATA
GATAGATAGATAGATAGAA

Bt-GATA-P150-T3 Consensus
AAGGTCGTCTAAAGGTCACCGGCTTAGCACTCTGGCGGCGACAGTGGCTGTTGGAACCGTCTC
AAACGCATTGTTTTATTAACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCGTGTGTAT
GCGTGTATATATGGCGTCGTTTCTTATTGCTCACAGGGAAGCGGCTGGATGCC

Bt-GATA-L154-T3
CGGAGACAAGAGGACGAAGCCCGATATAGAGACAGAGGCAGAGAGAGACAGAGAGACAGA
GACAGAGTGAGAGAGACAGACAGACAGACAGACAGACAGAGAGATAGATAGATAGATAGAT
AGATAGCTCTAACTCGAACTTTGTCAGAGAACAATAGCCTTGGAAACTGTCTAATAACCAGTC
CCATTTCCTAGCAAAGCAGTTTGGACAAAGGCACCC

Bt-GATA-L78-T3 Consensus
ATGTGTTAGTAAAGTAAACAATAAACAAGCAGATGAAACCATTATTGTTCCGTTGGTCCCTTT
CCATTGGACATAACCAAAACAAACAAAAAAAGAGTGAGAGAGAGAGAGAGAGTCATTTCCG
TGTTTGGGTGTCATGGGACTATCTTGCTTAGTGGTCAAGGTCACAGATTATATGTGCTGTCAAG
GTTATAACTATTTTAGCCCTTTCCGCACC

Bt-GATA-L114-T3 Consensus
TCAGCNGCTCCTGTACATAGAACGACGCTATATACACGCATACACAGCACACACACACACACAC
ACACACACACACANNTGTATAACTATGCGTTGAGACGTCCACAGCCACTGTCGCCAGAGTGCT
AAGCCGGTGACCTTTAGACGACCTTGCTTCTGCTAGCAGAAGCATCCGCAGCGGCATCAAACA
GGGCTGTGTGCNTGCGCCGACGCTTTTGGCATCTTCTTCGCAGTCCTGCTGAAGCACGCCTTTG
GGACTTCAGCAGAAGGTGTCTACCTCCACACACGGTCAGATGGTAGACTGTATAACCTGGCTC
GACTGAAGGCTAAGAGCAAAGTACGACAAGCGACC

Bt-GATA-L123-T3 Consensus
GATCCTGAACACCTGTGGGCGATTTCTGACATTAATTGTAACCAGTCCTTACGGTGTAAGCTT
AGCAACTTAGTGTGCGTGTGTGTGTCTGAGAGAGATAGAAAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGTCTGTGTGTGTGTGTGTGTGTGCGTGTGTGTGTGTGAGAGTTAGAGGA
GTGAAGGGTGGATTAGTAATGATAGCTGTGTGAGGAGACTGCGTGTGCTTCTAGCTAGCAAA
AGCAAGCCTGCATTCGATGGATGCGTCACAGACTTTGTTCTTGAAGGGTTGCGCTAACCCGTT
CTCCATCCCAGGCTTTTAGTTGGTACTGAAACGTCTGCCTTCC