

FINAL PROJECT REPORT

Great Lakes Fish and Wildlife Restoration Act

- **Project title:** Strain Composition of Lake Trout at Lake Michigan's Mid-Lake Reef Complex
- **Project sponsor:** U.S. Fish and Wildlife Service
- **FWS agreement number:** 30181-8-G019
- **Principal investigators/Project leaders:** Brian L. Sloss and John A. Janssen
- **Final report authors:** Brian L. Sloss and Meaghan Proctor
- **Study objectives:** Identify the genetic heritage of fertilized eggs and hatched fry at two sites on Lake Michigan's Mid-lake Reef Complex.
- **Description of tasks:**
 - Development of a suite of lake trout-specific microsatellite genetic markers.
 - Completed. See attached thesis, Chapter 2 for full details.
 - Simulation of mixed-strain hybrid genotypes for determining efficacy of hybrid identification methods.
 - Completed. See attached thesis, Chapter 1 for full details and results.
 - Assessment of genetic heritage of lake trout embryo and fry.
 - Not completed due to findings of simulation study showing no power or accuracy for detecting genetic heritage. Samples are archived and ready for analysis when appropriate tools for doing so are identified.
- **Major findings and accomplishments**

The major findings and accomplishments of this research have been compiled into a comprehensive Masters of Science thesis by the graduate student funded on this project, Meaghan Proctor. The thesis is attached to this report and highlights/accomplishments will be spelled out here. This research was analytically intense and the accompanying thesis represents an in-depth assessment of the available genetic reference data's ability to accurately perform genetic heritage estimates of naturally produced lake trout in Lake Michigan.

The first objective was to determine the efficacy of statistical algorithms to predict genetic heritage of sampled lake trout given the likelihood of cross-strain products occurring on the reef complex. Assuming strains will spawn only with the same strain is unreasonable and, thus, any genetic heritage estimate must be capable of considering at least interstrain F_1 products. The specific objective was to determine if established genetic-based individual assignment and/or hybrid-admixture detection methods were unbiased and showed sufficient accuracy to identify interstrain lake trout hybrids.

This objective was accomplished by constructing multiple simulated lake trout populations based on the genetic data from Page (2001) and Page et al. (2003) and assuming linkage equilibrium among the sampled microsatellite loci and Hardy-Weinberg equilibrium at each locus. These simulated populations comprised pure strain populations (we focused on only four primary strains: Green Lake [GLW], Lewis Lake [LLW], Marquette strain [SMD], and Seneca Lake [SLW]), all six possible F₁ combinations, 12 possible backcross (B_x) combinations, and 21 possible F₂ combinations (Table 1). Reference data was simulated using the published allele frequency data of Page (2001) and Page et al. (2003). This simulated data was used to assess the ability of six primary statistical packages using similar but different underlying assignment algorithms to assign individuals of known genetic origin to their correct groups (Table 2).

Results showed that no single package/algorithm showed acceptable accuracy in predicting the genetic heritage of naturally produced lake trout. Some evidence of a stepwise approach appeared promising but results were hampered by a lack of enough genetic data. Personal communication with Wendy Stott (USGS, Great Lakes Science Center) and Kim Scribner (Michigan State University) suggested similar conclusions were being drawn and further markers were being explored for lake trout genetic heritage assessments. Therefore, the initial second objective of our research, predict genetic heritage of lake trout produced on the reef complex, was not completed.

In lieu of the stated second objective, we developed further lake trout microsatellite primers based on what we believed to be the first lake trout-specific microsatellite-enriched lake trout subgenomic DNA library. We developed 11 additional lake trout microsatellite loci and optimized multiplex PCR reactions where all 11 loci could be screened using only three PCR amplifications (Table 3). These loci were polymorphic in tests of hatchery broodfish from the Codrington Research Hatchery, Ontario, CN (number of alleles range from 2-22 and expected heterozygosity ranged from 0.067-0.921; Table 3). Subsequent to our development of these primers, an additional suite of 12 microsatellite markers were developed (Rollins et al. 2009). Therefore, we have contributed to a growing need for genetic markers and data for lake trout management and restoration.

- **Management implications of your work**

The specific management implications of our work are related to the ability of management agencies to accurately assess the overall efficiency of lake trout restoration efforts in Lake Michigan. Previous research assessing survival of stocked strains provide an initial estimate of restoration success and strain-specific survival but to be ultimately successful these strains have to produce future generations of lake trout without the assistance of human activities. Therefore, the development and use of genetic-based monitoring techniques such as hybrid estimation of lake trout are imperative to provide short- and long-term assessments of management activities and allow adaptive changes in the program to maximize success. Our work shows that the current state of knowledge and current data level are insufficient to provide this insight. However, the development of new markers and separate but ongoing research by numerous labs to develop new and innovative analysis approaches will provide a

foundation for eventual assessment of the genetic heritage of archived larval lake trout from this study.

- **Any relevant pictures or images associated with the project**

No relevant pictures or images to my knowledge on this aspect of the project.

- **Additional restoration work needed and/or areas for future research**

As with any restoration project, further research and efforts exist for the genetic aspects of lake trout restoration work. The developed microsatellite primers from this project and from the work of Rollins et al. (2009) should be added to the reference data from Page (2001) and Page et al. (2003) to provide a more comprehensive database for lake trout restoration assessments. Further, we would suggest a US-Canadian Cooperative be explored to standardize lake trout genetic data collection similar to the efforts currently being undertaken for lake sturgeon and brook trout. This would maximize the impact of individual research objectives and provide the framework for the participating nation's fisheries programs to maximize their investments to aid in lake trout restoration. Informal efforts are currently being undertaken to do this but a more formal framework would ensure all researchers are aware of each other's efforts and more effective funding disbursement.

- **List of presentations delivered and outreach activities**

Several research presentations have occurred based on this research. In addition, the graduate student on this project, Meaghan Proctor, received the Norman S. Baldwin Scholarship from the International Association of Great Lakes Research (IAGLR) in recognition of an outstanding graduate student in Great Lakes Research. The specific presentations were:

Proctor, M.E., B.L. Sloss, J.J. Janssen, and M. Rise. 2009. Assessing our Abilities to Distinguish among Lake Trout Hatchery Strains and their Potential Hybrid Offspring on Lake Michigan's Mid-Lake Reef complex. Annual Meeting of the International Association of Great Lakes Research. Toledo, OH. INVITED.

Proctor, M.E., B.L. Sloss, J.J. Janssen, and M.L. Rise. 2008. Efficiency of genetic algorithms in identifying pure strain and hybrid lake trout in Lake Michigan's lake trout restoration efforts. Wisconsin Chapter of the American Fisheries Society Annual Meeting. Wausau, WI.

Proctor, M.E., B.L. Sloss, J.J. Janssen, and M.L. Rise. 2007. Estimating strain contribution of lake trout naturally produced on Lake Michigan's mid-lake reef complex. The 2007 Midwest Fish and Wildlife Conference, Madison, WI.

- **List of reports and peer-reviewed papers completed or in-progress**

To date, no peer-reviewed papers have been published but two papers are in preparation. One outlining the efficacy of the algorithms and one describing the new genetic loci. All

future papers will be shared with the USFWS prior to publication as part of the USGS FSP review process.

Table 1. Potential pure strain and hybrid crosses on the mid-lake reef complex. SMD = Marquette strain, LLW = Lewis Lake strain, GLW = Green Lake strain, SLW = Seneca Lake strain, F₁ = Pure x Pure, B_x = F₁ x Pure, F₂ = F₁ x F₁.

Pure strain	F ₁	B _x	F ₂
SMD	SMDxLLW	SMD-LLWF1xSMD	SMD-LLWF1xSMD-LLWF1
LLW	SMDxGLW	SMD-LLWF1xLLW	SMD-LLWF1xSMD-GLWF1
GLW	SMDxSLW	SMD-GLWF1xSMD	SMD-LLWF1xSMD-SLWF1
SLW	LLWxGLW	SMD-GLWF1xGLW	SMD-LLWF1xLLW-GLWF1
	LLWxSLW	SMD-SLWF1xSMD	SMD-LLWF1xLLW-SLWF1
	GLWxSLW	SMD-SLWF1xSLW	SMD-LLWF1xGLW-SLWF1
		LLW-GLWF1xLLW	SMD-GLWF1xSMD-GLWF1
		LLW-GLWF1xGLW	SMD-GLWF1xSMD-SLWF1
		LLW-SLWF1xLLW	SMD-GLWF1xLLW-GLWF1
		LLW-SLWF1xSLW	SMD-GLWF1xLLW-SLWF1
		GLW-SLWF1xGLW	SMD-GLWF1xGLW-SLWF1
		GLW-SLWF1xSLW	SMD-SLWF1xSMD-SLWF1
			SMD-SLWF1xLLW-GLWF1
			SMD-SLWF1xLLW-SLWF1
			SMD-SLWF1xGLW-SLWF1
			LLW-GLWF1xLLW-GLWF1
			LLW-GLWF1xLLW-SLWF1
			LLW-GLWF1xGLW-SLWF1
			LLW-SLWF1xLLW-SLWF1
			LLW-SLWF1xGLW-SLWF1
			GLW-SLWF1xGLW-SLWF1

Table 2. List of genetic programs tested for accuracy and bias in assessing the genetic heritage of simulated lake trout.

Program	Method	Reference
GeneClass2	ML-based assignment testing	Piry et al. 2004
GMA	ML-based assignment testing	Kalinowski 2003
WHICHRUN	ML-based assignment testing	Banks and Eichert 2000
NewHybrids	Admixture using Bayesian clustering	Anderson and Thompson 2002
BAPS	Admixture using Bayesian clustering	Corander et al. 2004
STRUCTURE	Admixture using Bayesian clustering	Pritchard et al. 2000

Table 3. Summary data for 11 microsatellite primer pairs developed for lake trout including the number of alleles at each locus (A), the size range of observed alleles (Range), and observed and expected heterozygosity (H_o and H_e , respectively). Dye labels used for genotyping are located at the beginning of the forward primer in all cases except for one located on the reverse primer.

Locus ID	Primer sequence (5' - 3')	Repeat motif	A	Range	H_o	H_e
Sna 2A ^{+#^}	F: <i>6FAM</i> - CTATCGTGCGCCATGAAAAC R: GATTCAACCACCGATTCAAC	(AGTGT) ₇	7	138 - 182	0.38636	0.66614
Sna 13Y	F: <i>NED</i> - AAACCCCCTTTCAGTTCACC R: CAGTGTGAGAACAAGCAGAG	(CA) ₉	2	156 - 164	0.22727	0.23824
Sna 15E ^{+#^}	F: <i>HEX</i> - TTGGAAATATCTGCTGTAGCC R: AGGAAAGGAAAGTGCTTGTG	(AC) ₁₁ CG(CA) ₂	22	248 - 357	0.80952	0.92083
Sna 19A	F: <i>HEX</i> - GGCCGATGCACTCCTGAC R: TGCTGTAGGCCACCAAATAC	(CT) ₉	2	81 - 85	0.02273	0.06661
Sna 40V	F: <i>6FAM</i> - GTGTCTGCATAAAGCCTTGC R: GAGGCAGAACCGACTCTCTG	(AC) ₈	3	236 - 242	0.06818	0.06714
Sna 44Eb [#]	F: GCAATCACCCCTAACTCAAGC R: <i>HEX</i> - TCCAAGTTGGCTCACTTTAAC	(TC) ₁₂ G(CT) ₈	5	147 - 157	0.65909	0.70272
Sna 48A	F: <i>NED</i> - TGATTTTGATGCGAAGTGGA R: CGGGGAAAGTGCTGGATT	*	5	119 - 155	0.54545	0.51959
Sna 63Y	F: <i>NED</i> - GCACAAGTGTACCGCTTC R: ATCCATCCGTGTTCTCAACC	(GA) ₁₂	2	195 - 199	0.36364	0.50470
Sna 64A	F: <i>6FAM</i> - CACTTCTCCCTTCATCATTTCC R: AGTGGCTGAAACGTCAAACC	(TC) ₈	2	194 - 198	0.15909	0.14812
Sna 79A	F: <i>HEX</i> - AGCTAACTGTCTCTCAAACCTC R: TTTGGTTACTACATGATTCC	(AC) ₁₁	4	113 - 119	0.50000	0.51541
Sna 82Y ^{+#^}	F: <i>6FAM</i> - GAGCGTGTGCGCTTCAGT R: AACACAAATAGTAGGGAGGCAAG	(GA) ₁₆	5	106 - 124	0.52273	0.68391

* Complex microsatellite motif:(GT)₃(GC)₂GATT(GT)₅TA(TG)₅TTA(TG)₉

⁺ Hardy-Weinberg disequilibrium

[#] Linkage disequilibrium

[^] Possible null allele

- **Literature Cited**

Anderson, E.C. and E.A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217 – 1229.

Banks, M.A., and W. Eichert. 2000. WHICHRUN (version 3.2) a computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* 91:87 – 89.

Corander, J., P. Waldmann, P. Marttinen, and M.J. Sillanpää. 2004. BAPS 2: Enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* 20: 2363–2369.

Kalinowski, S.T. 2003. Genetic Mixture Analysis 1.0. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from <http://www.montana.edu/kalinowski>.

Page, K.S. 2001. Genetic diversity and interrelationships of wild and hatchery lake trout in the upper Great Lakes: inferences for broodstock management and development of restoration strategies. Master's thesis. Michigan State University, East Lansing.

Page, K.S., K.T. Scribner, K.R. Bennett, L.M. Garzel, and M.K. Burnham-Curtis. 2003. Genetic assessment of strain-specific sources of lake trout recruitment in the Great Lakes. *Transactions of the American Fisheries Society* 132:877 – 894.

Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95:536 – 539.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945 – 959.

Rollins, M.F., N.V. Vu, I.B. Spies, and S.T. Kalinowski. 2009. Twelve microsatellite loci for lake trout (*Salvelinus namaycush*). *Molecular Ecology Resources* 9: 871-873.

- **Attachment**

Masters of Science Thesis of Meaghan Proctor.

GENETIC IDENTIFICATION OF SIMULATED LAKE TROUT FROM LAKE
MICHIGAN: EFFICACY OF GENETIC ALGORITHMS TO DETERMINE CROSS-
STRAIN HYBRIDS BASED ON AVAILABLE GENETIC DATA

by

Meaghan E. Proctor

A Thesis

submitted in partial fulfillment of the

requirements of the degree of

MASTER OF SCIENCE

IN

NATURAL RESOURCES (FISHERIES)

College of Natural Resources

UNIVERSITY OF WISCONSIN

Stevens Point, Wisconsin

July 24, 2009

APPROVED BY THE GRADUATE COMMITTEE OF

Dr. Brian L. Sloss, Committee Chair
Professor of Fisheries
Wisconsin Cooperative Fishery Research Unit

Dr. Eric M. Anderson
Professor of Wildlife
University of Wisconsin – Stevens Point

Dr. Michael J. Hansen
Professor of Fisheries
University of Wisconsin – Stevens Point

ABSTRACT

In the 1950s, lake trout (*Salvelinus namaycush*) were functionally extirpated from Lake Michigan from a combination of overharvest and introduced species. Since the 1960s, restoration efforts to return lake trout to Lake Michigan have involved annual stocking of millions of yearling lake trout from four strains into the lake. However, restoration of self-sustaining lake trout populations has not been successful. If lake trout restoration is to be successful, fisheries managers will need to focus on stocking lake trout strains that are either adapted to their stocking locations or have adequate genetic diversity and shown reproductive success in past efforts. One of the best ways to assess strain fitness is to determine reproductive success and survival of different strains through genetic data analyses. While genetic data analyses are essential for determining which strains are most appropriate for stocking, questions remain if current genetic markers and statistical approaches are able to accurately determine the origin strain of an individual lake trout that is the result of hybridization between individuals of two different pure strains. Previous studies have found genetic markers to be useful for studying the origin of pure strain lake trout. But as lake trout aggregates of multiple broodstock origins congregate on spawning reefs, spawning between lake trout of different hatchery origins may create interstrain hybrid offspring. Many statistical computer algorithms have been written for identifying individuals to populations, but as of yet, most of these models are untested for identifying individuals of hybrid origin. If fishery biologists rely on results from these models when making management decisions, models must be both accurate and unbiased, even for hybrid individuals. I tested six individual assignment or hybrid/admixture algorithms (i.e., BAPS v4.14, NewHybrids v1.1, STRUCTURE v2.2,

GeneClass2, GMA, and WhichRun) to determine their accuracy and bias for assigning parentage to individuals from a simulated population of pure strain and interstrain hybrid offspring. These algorithms were moderately successful for assigning pure strain individuals, but less successful for assigning interstrain hybrid offspring. Success in assigning pure strain individuals ranged from 8.9% to 40.0% in the maximum likelihood (ML) algorithms and from 36.8% to 99.6% in the Bayesian clustering algorithms. For F₁ hybrid individuals, assignment success ranged from 4.3% to 46.2% in the ML algorithms and from 7.0% to 95.7% in Bayesian clustering algorithms. Assignment success of advanced hybrid individuals (i.e., B_x and F₂) only exceeded 65% in one algorithm (i.e., NewHybrids). While NewHybrids consistently had the highest assignment success, its use in studies of Lake Michigan's lake trout may be limited by the algorithm's limitation of a maximum of two reference populations during assignment. As a result of this poor performance, none of these methods were sufficient to be used for the assignment of lake trout of unknown origin with currently published genetic markers. Therefore, I was unable to determine the origins of lake trout eggs and fry collected from the mid-lake reef complex. Developing additional lake trout microsatellite genetic markers may provide tools necessary to distinguish among stocked strains and their potential hybrid offspring. Eleven lake trout specific microsatellite loci were developed and grouped into three multiplex reactions. These loci had an average heterozygosity of 0.450 while individual locus heterozygosity ranged from 0.023 to 0.810. The number of alleles per locus ranged from 2 – 22. These reactions were tested for performance with four other salmonid species with varying success. Between 2 and 7 loci were successfully amplified for each salmonid, and a subset of each were polymorphic.

ACKNOWLEDGEMENTS

I would like to thank everyone who has contributed to the success of this project and encouraged me through this long process. First, I would like to give special thanks to Dr. Brian Sloss who took me on as a student even though I had no prior genetics or fisheries experience. By doing this, he challenged me to expand my knowledge base and become a better-rounded biologist able to face any challenges that present themselves over the course of my natural resources career. I would like to thank Dr. Eric Anderson and Dr. Michael Hansen for their interest in my research and their advice on my thesis. Next, thanks to Dr. Michael Bozek for challenging my opinions and forcing me to look at the world from a different perspective. To Ryan Franckowiak, thank you for not only your assistance in the lab, but for the many interesting conversations we shared about science, music, and politics. Thanks to the other graduate students for pushing me to drive myself harder. Thanks to Andrea Musch whose knowledge of UWSP is unsurpassed. I would like to thank the Great Lakes Fishery Commission, the USGS-Wisconsin Cooperative Fishery Research Unit, and UWSP for the technical and financial support necessary for the completion of this project. Thank you to the International Association for Great Lakes Research and Salmon Unlimited of Wisconsin for providing me with scholarships. I would like to acknowledge my parents, Edwin and Deborah Proctor, and my siblings, Rachel and Jason, who always took an interest in my work even if they didn't always understand it. I would especially like to thank Steven Cibarich for his love and support. Without you I may not have made it through to the end. And last, but definitely not least, I would like to thank Monkey Pox, my beloved pet prairie dog, who was very supportive with all of her affection. You are dearly missed.

TABLE OF CONTENTS

TITLE PAGE	i
COMMITTEE SIGNATURES	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	xi
LIST OF APPENDICIES	xii
INTRODUCTION	1
<i>Lake Trout in the Great Lakes</i>	3
LITERATURE CITED	19
CHAPTER 1: EFFICACY OF GENETIC ALGORITHMS IN IDENTIFYING SIMULATED PURE STRAIN AND HYBRID LAKE TROUT IN LAKE MICHIGAN'S LAKE TROUT RESTORATION EFFORTS	
ABSTRACT	28
INTRODUCTION	30
MATERIALS AND METHODS	32
<i>Experimental Approach</i>	32
<i>Data simulation</i>	34
<i>Population differentiation</i>	36
<i>Bayesian clustering</i>	37
BAPS v4.14	39
NewHybrids v1.1	41
STRUCTURE v2.2	42
<i>Maximum-likelihood assignment</i>	44
GeneClass2	44

GMA	46
WhichRun	47
<i>Alternative approaches to assignment</i>	48
<i>Bias assessment</i>	52
RESULTS	53
<i>Population differentiation</i>	53
<i>Self-assignment/pure strain assignment</i>	53
<i>F₁ assignment</i>	54
<i>B_x assignment</i>	55
<i>F₂ assignment</i>	56
<i>Alternative approaches to assignment</i>	57
<i>Bias assessment</i>	59
DISCUSSION	61
REFERENCES	73

CHAPTER 2: DEVELOPMENT OF NEW MICROSATELLITE LOCI AND
MULTIPLEX REACTIONS FOR LAKE TROUT AND THEIR USEFULNESS IN
OTHER SALMONID SPECIES

ABSTRACT	99
INTRODUCTION	100
METHODS	102
RESULTS	106
DISCUSSION	108
LITERATURE CITED	111

LIST OF TABLES

CHAPTER 1

Table 1.1.	Genetic loci and allele frequencies used in the creation of a theoretical population of pure strain lake trout on the mid-lake reef complex from Page et al. 2003. SMD = Marquette strain, LLW = Lewis Lake strain, GLW = Green Lake strain, SLW = Seneca Lake strain.....	80
Table 1.2.	Potential pure strain and hybrid crosses on the mid-lake reef complex. SMD = Marquette strain, LLW = Lewis Lake strain, GLW = Green Lake strain, SLW = Seneca Lake strain, F_1 = Pure x Pure, B_x = F_1 x Pure, F_2 = F_1 x F_1	81
Table 1.3.	List of programs tested for accuracy and bias	82
Table 1.4.	Summary data for pairwise F_{st} score among four pure strains and three hybrid categories	83
Table 1.5.	Percentage of individuals in four categories assigned ($p \geq 0.80$ or $LOD \geq 2$) to an origin strain ¹ or to an incorrect strain during assignment in three maximum likelihood algorithms and two Bayesian clustering algorithms	84
Table 1.6.	Percentage of individuals in four categories assigned ($p \geq 0.80$) in NewHybrids and the breakdown of correct and incorrect assignment based on the number of true origin strains used as a prior	85
Table 1.7.	Comparison of the number of individuals assigned ($p \geq 0.80$) in GeneClass2 when using 100 or 100,000 simulated individuals. Significant differences are marked by an *.....	86
Table 1.8.	The number of individuals assigned to origin and non-origin strains in WhichRun with a relaxed assignment threshold. Confidence based on $LOD \geq 2$ and $LOD \geq 1.5$ are presented. Significant differences are marked by an *	87
Table 1.9.	Percentage of pure strain individuals correctly and incorrectly assigned in three maximum likelihood algorithms. Data for $p \geq 0.80$ and $p \geq 0.50$ assignment thresholds (GeneClass2 and GMA), $LOD \geq 2$ and $LOD \geq 0.5$ (WhichRun), and pure assignment are presented. P-values represent the significance found in ANOVA/Kruskal-Wallis tests. Significant tests p-values are marked by ⁺ . Superscripts indicate which threshold levels are significantly different as found in LSD post hoc tests/Mann-Whitney U-tests.....	88

Table 1.10.	Percentage of F ₁ hybrid individuals assigned with pure assignment in three maximum likelihood algorithms based on the most likely and second most likely origin populations	89
Table 1.11.	Number of pure strain and F ₁ individuals assigned (pure assignment) to an origin strain and incorrectly assigned in three maximum likelihood algorithms when the baseline file contained pure strain and F ₁ populations.....	90
Table 1.12.	Performance capabilities of two Bayesian clustering algorithms when clustering of pure strain and F ₁ individuals was performed with pure strain and F ₁ populations in the baseline prior file. Assignment based on a $p \geq 0.50$ threshold.....	91
Table 1.13.	Assignment success ($p \geq 0.80$) of NewHybrids when the unknown populations were created based on the individual's assignment in GMA	92
Table 1.14.	Results of bias tests for each pure strain in five assignment algorithms	93
Table 1.15.	Results of bias tests for each F ₁ hybrid in five assignment algorithms	94
Table 1.16.	Results of each bias test performed for NewHybrids where the bias occurs	95

CHAPTER 2

Table 2.1:	Strains used for identifying lake trout microsatellite loci.....	113
Table 2.2:	Summary data for 11 microsatellite primer pairs developed for lake trout including the number of alleles at each locus (A), the size range of observed alleles (Range), and observed and expected heterozygosity (H _o and H _e , respectively). Dye labels used for genotyping are located at the beginning of the forward primer in all cases except for one located on the reverse primer	114
Table 2.3:	Allele frequencies at each of the 11 microsatellite loci developed for lake trout in this study.....	115
Table 2.4:	Tests for heterozygote deficiency at 11 lake trout microsatellite loci and their significant levels following Bonferroni correction	116

Table 2.5:	Multiplex reaction conditions for lake trout microsatellite loci developed in this study. Conditions include the concentration of the various PCR cocktail reagents and the multiplex-specific annealing temperature (T_a). All PCRs were 10 μ L total volume with ~60 ng DNA/reaction.....	117
Table 2.6:	Lake trout microsatellite loci that successfully amplified four other salmonid species' DNA. The number of alleles (A), the size range of observed alleles (Range) and the observed and expected heterozygosities (H_o and H_e , respectively) are presented	118

LIST OF FIGURES

GENERAL INTRODUCTION

Figure 1.	Location of the Northern Refuge and the Southern Refuge which includes the mid-lake reef complex (1 = Sheboygan Reef, 2 = Northeast Reef, 3 = East Reef, and 4 = Milwaukee Reef) in Lake Michigan (adapted from Bronte et al. 2007)27
-----------	---

LIST OF APPENDICIES

CHAPTER 1

- Appendix 1.1: Results of t-tests/Mann-Whitney U-tests for comparisons between using 100 and 100,000 simulated individuals for assignment in GeneClass296
- Appendix 1.2: Results of Mann-Whitney U-tests for comparisons between using $LOD \geq 2$ and $LOD \geq 1.5$ for assignment in WhichRun97
- Appendix 1.3: Table of ANOVA results comparing the assignment capabilities and error rates of three maximum likelihood algorithms at three assignment thresholds (GeneClass2 and GMA: $p \geq 0.80$, $p \geq 0.50$, and pure assignment, WhichRun: $LOD \geq 2$, $LOD \geq 0.5$, and pure assignment). Significant results are marked with an *98

CHAPTER 2

- Appendix 2.1: List of lake trout microsatellite loci identified but not finalized for use120

INTRODUCTION

While genetic data analyses are essential for determining which strains of lake trout (*Salvelinus namaycush*) currently used for stocking Lake Michigan are most appropriate (Perkins et al. 1995; Page et al. 2003; DeKoning et al. 2006), the efficacy of strain identification methods after stocking has not been evaluated. A key issue in genetic identification of current strains is the efficacy of genetic markers and statistical methods to accurately determine the strain of origin of an individual lake trout that is the result of hybridization between individuals of two different pure strains (O’Gorman et al. 1998). Previous studies have found genetic markers to be useful for studying the origin of individual lake trout of pure-strain origin (Burnham-Curtis et al. 1995; Page 2001; Page et al. 2003). Page et al. (2003) were able to accurately classify each lake trout hatchery strains by using seven microsatellite loci while DeKoning et al. (2006) were able to document unequal spawning success among strains in the wild. However, as lake trout aggregates of multiple broodstock origins congregate on spawning reefs, spawning may occur between lake trout of different hatchery origins.

Inter-strain crossing of genes results from independent assortment of genes from different broodstocks (i.e., Mendel’s Law of Independent Assortment) and can create ambiguity in determining the strain of resulting hybrid offspring. After a few years of F_1 crosses, the ability to determine corresponding strains responsible for natural reproduction could be difficult. Combine the presence of F_1 individuals with the potential for crosses between two hybrids, or a hybrid backcrossed with a parental strain (cumulatively called F_x hybrids), and the ability to determine the strain of origin of a particular fish could be all but impossible. Many statistical computer algorithms have

been written for identifying individuals to populations, but as of yet, most of these models are untested in situations where the individuals may be of hybrid origin. If fishery biologists rely on results from these models to make management decisions, models must be both accurate and unbiased, even for hybrid individuals. Accuracy, or the freedom from error, is the ability of the algorithm to correctly assign an individual to its origin. Bias is a consistent tendency of an algorithm to incorrectly assign an individual's origin. Accuracy depends on the number of contributing strains of lake trout and the genetic differentiation among the strains (Page et al. 2003). Bias will be greatest when strains are genetically similar but differ in abundance, a distinct likelihood in Lake Michigan lake trout because of differences in stocking among strains (Holey et al. 1995; Page et al. 2003).

The goal of my study is to determine if current statistical algorithms are able to differentiate among four pure lake trout strains currently stocked on Lake Michigan's mid-lake reef complex, and each potential hybrid cross up to the F_2 that would be expected to occur if each strain is reproducing. If a method is found to have high assignment success, then that method will be used to assign lake trout eggs and fry collected from the mid-lake reef complex to their strain of origin. However, if no algorithm is found to have sufficient performance capabilities to distinguish among pure strains and hybrid crosses, then additional lake trout specific genetic markers will be developed to provide increased discriminatory ability in future lake trout studies. The three project objectives for this study are:

- 1) To determine if established genetic-based individual assignment and/or hybrid-admixture detection methods were unbiased and showed sufficient accuracy to identify interstrain lake trout hybrids.
- 2) To determine the likely strain of origin for lake trout eggs and/or larval fish collected in the mid-lake reef complex of Lake Michigan.
- 3) To develop, if necessary, additional genetic markers specifically designed for lake trout to provide increased discriminatory ability in future studies of lake trout reproduction.

Lake Trout in the Great Lakes

Lake trout (*Salvelinus namaycush*) are a morphologically diverse piscivorous fish endemic to northern North America that historically sat at the apex of simple predator-prey ecosystems (Krueger and Ihssen 1995; Page et al. 2003). The variety of physical forms exhibited by lake trout allowed for use of such different habitat types as rivers and deep lakes (Page et al. 2003). Three morphotypes (lean, siscowet, and humpers) have been delineated on the basis of facial characteristics, body fat content, body shape, and unique habitat preferences (Burnham-Curtis et al. 1995; Krueger and Ihssen 1995; Page et al. 2003). Lean lake trout are the most common of the three morphotypes and inhabit and spawn in shallow water (Burnham-Curtis et al. 1995; Krueger and Ihssen 1995). They have a fusiform body with a long pointed snout, small eyes and fins, and low fat levels (Sitar et al. 2007). Because lean lake trout do not begin to mature until an adult size is reached, beginning at age 5 or 6, they are at an increased risk of sea lamprey (*Petromyzon marinus*) predation or fishing related mortality before their first spawning

event compared to the siscowet and humper strains that spawns at a smaller size (Burnham-Curtis et al. 1995); both sea lamprey predation and fishery operations tend to favor larger fish (Burnham-Curtis et al. 1995). Siscowet lake trout have a more stout body than lean lake trout with a convex snout, and large eyes and fins (Sitar et al. 2007). Siscowets have the highest fat content of the three morphotypes (Krueger and Ihssen 1995). They inhabit deep water, often as deep as 100 m (Krueger and Ihssen 1995; Sitar et al. 2007), and vary greatly in spawning times among populations (Eschmeyer 1957; Bronte 1993; Burnham-Curtis et al. 1995). Humper lake trout are the least common of the three morphotypes (Sitar et al. 2008). They are an intermediate form with large eyes, a thin ventral body wall, and grow to a smaller adult size (Rahrer 1965; Burnham-Curtis et al. 1995; Krueger and Ihssen 1995). Humper lake trout could be the result of introgression between individuals of lean and siscowet morphotypes (Zimmerman et al. 2007). All three morphotypes exist concurrently in individual lakes throughout their range (Smith and Snell 1891; Bronte and Moore 2007) but exhibit significant interstrain and intrastrain (Bronte and Moore 2007) genetic diversity (Burnham-Curtis et al. 1995) as a result of historical isolation of the different stocks throughout the heterogeneous habitat and large size of the Laurentian Great Lakes (Burnham-Curtis 1993; Eshenroder et al. 1995b).

Historically, the lake trout was one of the most important freshwater fish in both commercial and recreational fisheries of North America (Martin and Olver 1980). Prior to establishment of the sea lamprey in the Great Lakes, long-term average catches of lake trout ranged between 0.24 and 0.49 kg/ha (Martin and Lover 1980). Lake Michigan supported the largest commercial lake trout fishery and sustained higher annual catches

than any other Great Lake, despite being only the third largest of the five lakes (Holey et al. 1995). From the 1890s to the mid-1940s, commercial fishers in Lake Michigan landed ~2.7 million kg of lake trout annually (Wells and McLain 1973; Ward et al. 2000). In the 1950s, lake trout populations collapsed and became functionally extirpated (Holey et al. 1995). Large-scale reintroduction efforts were begun in 1965 with the goal of reestablishing a self-sustaining lake trout population capable of supporting a commercial fishery (Holey et al. 1995). In spite of these reintroduction efforts, lake trout populations in Lake Michigan have yet to recover.

Lake trout populations have shown susceptibility to over-exploitation (Martin and Olver 1980; Eshenroder et al. 1995b). Van Oosten (1949) and Wells and McLain (1972) felt that the decline in lake trout abundance during 1893 – 1938 in Lake Michigan was because of high exploitation rates (Martin and Olver 1980). Christie (1972) suggested that the collapse of the lake trout fishery in Lake Ontario at the beginning of the 1900s was the result of overfishing (Martin and Olver 1980). Fishing pressure has also been credited with the lake trout population decline in Lake Erie (Hartman 1972; Martin and Olver 1980).

Over-fishing alone may not have been the cause of the lake trout population collapse. Instead, over-fishing may have reduced the lake trout population to a density at which it was vulnerable to sea lamprey predation and habitat degradation (Martin and Olver 1980; Fitzsimons 1995; Fitzsimons et al. 1995). These effects combined to cause a complete extirpation of lake trout from Lake Michigan by the mid-1950s (Eschmeyer 1957; Holey et al. 1995). However, lake trout extirpation was not limited to Lake Michigan. By the 1960s, native lake trout had also been extirpated from Lakes Ontario

and Erie, and most of Lake Huron (Krueger and Ihssen 1995). Within each lake basin, the collapse of lake trout occurred over the course of approximately one generation (~5 years in Lake Michigan and ~10 years in Lakes Huron and Superior; Hansen 1999; Guinand et al. 2003).

In 1955, the Great Lakes Fishery Commission (GLFC) was formed with the purpose of developing and coordinating fishery research programs, advising U.S. and Canadian governments on measures to improve fisheries, and implementing programs to control sea lamprey (Fetterolf 1980; Holey et al. 1995). The GLFC emphasized the development and maintenance of hatchery lake trout broodstocks for stocking of U.S. waters of the upper Great Lakes (Fetterolf 1980; Page et al. 2005; GLFC 2006). This stocking program, managed by state and federal agencies, focused on restoration of inshore lean lake trout populations (Hatch 1984; Janssen et al. 2007), but also recommended using broodstock sources closely matched to environmental conditions (Elrod et al. 1996) and developing hatchery offspring from multiple parental crosses to maximize genetic diversity (Schneider et al. 1983; Burnham-Curtis et al. 1995). These broodstocks were separated into deep-water and shallow water forms, and included the three morphotypes present in the early 1900s (Holey et al. 1995).

In the Great Lakes, eight lake trout strains have been used for stocking. These strains were selected because they originated from endemic Great Lakes sources and matched ecological and habitat characteristics of regions to be stocked (Page et al. 2003). Use of multiple, divergent broodstocks helped to capture the genetic and ecological diversity present in the remaining lake trout populations across the upper Great Lake basin (Krueger et al. 1983; Kincaid et al. 1993; Krueger and Ihssen 1995; Page et al.

2005). In Lake Michigan, six strains were used for stocking, four of which (Seneca Lake, Lewis Lake, Green Lake, and Marquette) are thought to be potential contributors to natural spawning on the mid-lake reef complex (DeKoning et al. 2006).

In preparation for lake trout stocking, toxicants specific to sea lamprey (i.e., lampricides) were applied to Lake Michigan tributary streams in which sea lamprey were spawning (Smith and Tibbles 1980; Lavis et al. 2003). Following the application of lampricides and subsequent reduction of sea lamprey populations by 80 – 90% (Lavis et al. 2003), lake trout stocking began in Lake Michigan in 1965 with the release of 1.1 million hatchery-reared yearlings across the northern third of the lake (Holey et al. 1995; McKee et al. 2004). Shortly thereafter, the Lake Michigan Study Group (LMSG) was formed to evaluate effectiveness of lamprey control and large-scale stocking of hatchery-reared fish (Holey et al. 1995). Stocking levels eventually reached an average of ~2.4 million yearling lake trout per year (Holey et al. 1995; McKee et al. 2004) but this number was only half of that recommended for restoration (Bronte et al. 2003).

Work towards a new management plan was begun when successful reproduction of lake trout was not observed by the late 1970s and led to the creation of the Lake Wide Management Plan (1985), a multi-agency coordinated focus on rehabilitation efforts specific to Lake Michigan (Holey et al. 1995; McKee et al. 2004). This plan had five goals, including: 1) achieve a self-sustaining lake trout population able to yield annual harvests of 500,000 – 700,000 fish (~1,000 metric tons), 2) annual stocking of 3.534 million lake trout throughout Lake Michigan, 3) a maximum annual total mortality of 40%, 4) delineation of four rehabilitation zones with specific stocking rates for each zone, and 5) evaluation of multiple lake trout strains used for stocking (LMLTTC 1985;

Holey et al. 1995; O’Gorman et al. 1998; McKee et al. 2004; DeKoning et al. 2006). The Lake Wide Management Plan also included the creation of two refuges in Lake Michigan, one in the north and one in the south, where most lake trout stocking would occur (LMLTTC 1985). These two refuge zones historically provided ~36% of the total lake trout harvest despite accounting for ~10% of Lake Michigan’s surface area (Dawson et al. 1997). The southern refuge was created over a complex of four large, deep-water (45 – 80 m) spawning reefs (Sheboygan, Northeast, East, and Milwaukee) and is 2,859 km² in size (Figure 1). Historically, these reefs, known as the mid-lake reef complex, were one of the most productive lake trout spawning areas in Lake Michigan. The original stocking plan called for 750,000 lake trout from the Marquette, Seneca Lake, and Green Lake or Lewis Lake strains to be stocked annually in equal proportions. Stocked fish received a coded wire tag in the snout and an adipose fin clip to identify strain, year-class, and location of stocking (Holey et al. 1995; Bronte et al. 2007). Despite the goal of equal strain contribution to stocking, the Marquette strain composed 65% of the 7.1 million fish stocked into the southern refuge during 1980 – 1992 because a viral disease infected multiple state and federal hatcheries and all exposed fish had to be destroyed (Holey et al. 1995).

The four strains that were selected for stocking were chosen because they represented either alternative morphotypes or differed ecologically and could provide fisheries managers with a way to infer adaptability of strains to contemporary Lake Michigan conditions through interstrain variations in survival and reproduction (Holey et al. 1995). The Seneca Lake strain is a deep-water, early-fall spawning population of the lean morphotype originating in Seneca Lake of the Finger Lakes system in upstate New

York (Royce 1951; Krueger and Ihssen 1995). The Lewis Lake strain originated from northern Lake Michigan and was introduced into Lewis Lake, Wyoming before 1900 (Krueger et al. 1983; Krueger and Ihssen 1995). This strain may have originated from the siscowet morphotype (Smith and Snell 1891; Krueger and Ihssen 1995). The original Green Lake strain was a deep-water strain developed in the late 1800s with lake trout from southern Lake Michigan, including the mid-lake reef complex, that were stocked into Green Lake, Wisconsin (Hacker 1957; Burnham-Curtis et al. 1995; Holey et al. 1995; Page et al. 2003). A management decision led to discontinuation of the Green Lake strain in the lake trout propagation program in 1976 (Krueger et al. 1983). However, the Green Lake strain was one of the last known sources of Lake Michigan origin lake trout (Kincaid et al. 1993; Krueger and Ihssen 1995), so fisheries managers later worked to rejuvenate the strain from the few remaining fish in the hatchery system and gametes collected from feral lake trout of Green Lake origin collected on southern Lake Michigan reefs between 1986 and 1989 (Kincaid et al. 1993; Page et al. 2003). The Marquette strain was founded in the late 1940s with lake trout collected from in-shore populations on the southern shore of Lake Superior near Marquette, Michigan (Krueger et al. 1983; Krueger and Ihssen 1995; Holey et al. 1995; Guinand et al. 2003). Because the Marquette strain used a different habitat than the other strains (inhabits and spawns in shallow water), it was designated as a control group for survival comparisons with other strains (Krueger et al. 1983; McKee et al. 2004). These strains continue to be reared in hatcheries and stocked annually.

After more than 20 years of implementing the Lake Wide Management Plan, lake trout rehabilitation efforts face many impediments that can be summarized into three

broad categories: low lake-wide population sizes, spawning aggregations too diffuse and in inappropriate locations, and low survival of early life stages (Perkins and Krueger 1995; Bronte et al. 2003). Holey et al. (1995) suggested that the number of lake trout stocked annually (~2.5 million yearlings) was not sufficient to establish a self-sustaining population because it is well below the estimated average number of young produced each year (~10 million yearlings) by historical populations. This smaller population of adults was not producing enough eggs and fry to survive early life history mortality and reach adulthood (Holey et al. 1995; Claramunt et al. 2005).

The apparent lack of stocking success has been attributed, in part, to strategic decisions associated with propagation of lake trout. Until the mid-1980s, stocking was accomplished at shoreline sites that were easily accessible to stocking trucks (Holey et al. 1995; Bronte et al. 2007). Thoughts at that time were that these fish would find suitable habitat for spawning, but many of the fish returned to their stocking sites for spawning, even if it was inappropriate habitat (Dawson et al. 1997; Bronte et al. 2007; Janssen et al. 2007). Subsequently, stocking efforts were shifted towards stocking large numbers of lake trout in appropriate spawning habitat, thereby resulting in large concentrations of adult lake trout being observed in spawning condition in more appropriate habitat (Edsall et al. 1995; Janssen et al. 2006; Bronte et al. 2007). However, production of only a small number of fry, deemed insufficient for the establishment of a self-sustaining population, has been observed to date (Holey et al. 1995; Edsall and Kennedy 1995, Janssen et al. 2006). This could be the result of lower reproductive success by hatchery origin lake trout compared to naturally produced lake trout or lower viability of eggs from

individuals in their first year of spawning (Anderson and Collins 1995; Perkins and Krueger 1995; Fleming and Petersson 2001; Madenjian et al. 2004).

Use of suboptimal spawning habitats by introduced lake trout continues to be an impediment to lake trout restoration. The highest concentrations of eggs are found in suboptimal shallow, in-shore areas (<6 m) (Schreiner et al. 1995; Marsden and Janssen 1997). These shallow water areas tend to have high concentrations of exotic egg predators such as the rusty crayfish (*Orconectes rusticus*) and round goby (*Neogobius melanostomus*) (Claramunt et al. 2005; Jonas et al. 2005), fry predators such as alewives (*Alosa pseudoharengus*) (Holey et al. 1995; Krueger et al. 1995b; Marsden and Janssen 1997; Madenjian et al. 2008), and extensive wave action that can damage or dislodge eggs (Eshenroder et al. 1995a; Marsden et al. 1995; Perkins and Krueger 1995; Fitzsimons et al. 2005). Shallow water areas are also susceptible to colonization by zebra mussels (*Dreissena polymorpha*) that have a strong negative effect on lake trout spawning activity and egg survival (Marsden and Chotkowski 2001). Zebra mussels can occlude interstitial spaces where lake trout eggs incubate and can create areas of poor water quality because of the mussels' oxygen consumption and their waste product sedimentation (Marsden and Chotkowski 2001). This shallow spawning tendency could be a consequence of heavy dependence on lean morphotypes of lake trout (Janssen et al. 2007; Sitar et al. 2007), which prefer a shallower depth, for stocking from the beginning of the stocking program until the late 1980s (Holey et al. 1995) and the loss of the deep-dwelling morphotypes (Gunn 1995; Ward et al. 2000), possibly combined with strain acclimation to the shallow water of the hatchery raceways.

Another significant challenge to lake trout restoration is that the ecology of Lake Michigan has changed significantly since lake trout were first extirpated (Miller and Holey 1992; Marsden et al. 1995). The introduction and continued propagation and recruitment of Pacific salmonids (*Oncorhynchus* spp.) and other native and non-native trout (brook trout, *S. fontinalis*, brown trout, *Salmo trutta*, and splake, *S. fontinalis* x *S. namaycush*) likely impede successful lake trout reproduction through increased interspecific competition and hybridization (Evans and Olver 1995; Gunn 1995; Holey et al. 1995; Noakes and Curry 1995). In addition to direct competition, the prey base of Lake Michigan has changed dramatically since lake trout extirpation. A shift has occurred in Lake Michigan in the availability of potential prey from predominately cisco (*Coregonus artedii*) to a mixture of alewife and rainbow smelt (*Osmerus mordax*) (Miller and Holey 1992; Ward et al. 2000; Fitzsimons et al. 2007). Adult lake trout with diets high in alewives and rainbow smelt suffer from thiamine deficiency leading to Early Mortality Syndrome (EMS) in their eggs and fry (Fitzsimons et al. 1995; Honeyfield et al. 2005; Bronte et al. 2007; Fitzsimons et al. 2007; Madenjian et al. 2008). The continued presence of sea lamprey causes injuries to lake trout of all morphotypes (O’Gorman et al. 1998; Sitar et al. 2007) and account for an average of 7% of lake trout mortality in Lake Michigan (Fetterolf 1980; Holey et al. 1995).

Genetic concerns in relation to re-establishment of lake trout in Lake Michigan are related to loss of genetic diversity and comparatively low available genetic diversity for subsequent rehabilitation efforts (Burnham-Curtis et al. 1995; Krueger et al. 1995a; Meffe 1995). Because most modern genetic techniques were developed after many lake trout stocks were already extinct, the original genetic diversity of lake trout can only be

inferred from historical observations and genetic analyses of native stocks that remained in Lake Superior (Krueger and Ihssen 1995). Comparisons of hatchery broodstocks to extant Lake Superior populations have shown that genetic diversity of extant populations is higher than in hatchery populations (Grewe and Hebert 1988; Ihssen et al. 1988; Grewe et al. 1994; Burnham-Curtis et al. 1995). This lower genetic diversity in hatchery broodstocks has been implicated as a constraint on lake trout rehabilitation (Loftus 1976; Krueger et al. 1981; Ihssen 1984) because multiple alleles with selective differences among them are needed if natural selection is to cull the least adaptive alleles from the population. Now-extinct stocks probably contained locally adapted traits, which developed over thousands of years, for specific habitats or regions of the Great Lakes that provided for better survival and reproduction within those habitats versus other habitats (Krueger and Ihssen 1995; Meffe 1995; Dawson et al. 1997; Morbey et al. 2008). As hatchery source lake trout and their offspring successfully reproduce in Lake Michigan, adaptations to contemporary conditions could develop given sufficient genetic diversity. Finally, documented progeny from natural spawning has shown a reduction in genetic diversity compared to the genetic diversity in hatchery strains (Page et al. 2003). This suggests only a subset of hatchery-reared fish spawn in the wild (DeKoning et al. 2006).

Hatchery selection and behavioral alterations as a consequence of adaptation to the hatchery environment is a concern with lake trout rehabilitation (Gunn 1995; Einum and Fleming 2001; Fleming and Petersson 2001; Ford 2002). While hatcheries have been essential in preserving remnant lake trout genetic diversity, including ecological, physiological, and phenotypic traits (Meffe 1995; Page et al. 2005), source broodstocks may no longer be adapted to life in wild conditions (Einum and Fleming 2001). Over

time, reduced selection pressures of a hatchery environment could allow for survival of fish that would not have survived under natural conditions (Einum and Fleming 2001; Ford 2002). As these fish that are maladapted to wild conditions contribute their genes to the hatchery population through spawning, the entire population could lose adaptations that would allow for greater survival of particular strains at specific habitats within Lake Michigan (Ford 2002). Loss of these locally adapted gene complexes, as a result of selection for the hatchery environment, could negatively impact fitness of hatchery-origin lake trout after release (Lynch and O’Hely 2001; Wang and Ryman 2001; Page et al. 2005; McDermid et al. 2007). Additionally, changes in genetic diversity as a result of natural selection can be exacerbated by genetic drift over the course of multiple generations in captivity because of a low effective population size (N_e) as a result of a small parental population (Allendorf and Pheleps 1980; Page et al. 2005). Even with large numbers of adults in the hatchery program, not all adults will successfully produce offspring (Page et al. 2005).

If lake trout restoration efforts are to be successful, fisheries managers will need to focus on stocking lake trout strains that are either adapted to their stocking locations (Burnham-Curtis et al. 1995; Morbey et al. 2008) or have adequate genetic diversity, and have shown reproductive success in past efforts (Krueger and Ihssen 1995; Perkins et al. 1995; O’Gorman et al. 1998; DeKoning et al. 2006). One of the best ways to assess strain fitness is to determine reproductive success and survival of different strains (Perkins et al. 1995; Krueger et al. 1989; O’Gorman et al. 1998; Page et al. 2003). Many commonly used management techniques are unable to determine contributions of different strains to reproduction. First, presence of a strain on a spawning reef,

identifiable by a fin clip during a catch-per-unit-effort survey, does not indicate contribution to reproduction and spawning success (Page et al. 2003; DeKoning et al. 2006). Furthermore, presence does not accurately reflect recruitment potential, because a few adults or strains present in an aggregate could contribute disproportionately to spawning and recruitment (Perkins et al. 1995; Page et al. 2003; DeKoning 2006). Second, coded-wire tags (CWT), useful for studies involving life history aspects of individual fish, cannot be used to determine reproductive success (Utter and Ryman 1993) and juvenile production of different strains (Page et al. 2003; DeKoning et al. 2006). Additionally, retrieval of CWTs requires sacrifice of individuals, a concern for a species that spawns in multiple years and is already experiencing reproductive challenges.

The best way to determine lake trout strain contribution to natural reproduction is through the use of genetic analysis (Eshenroder et al. 1995b; Page et al. 2003, DeKoning et al. 2006). Genetic techniques are able to address issues relating to strain survival and recruitment, as well as habitat use of different strains, and knowledge that genetic data allows for a better understanding of heritable traits and adaptive fitness of populations (Burnham-Curtis et al. 1995). Of the many genetic markers available, microsatellite loci are an appropriate choice for strain identification because of a high amount of genetic variation and subsequent increased discriminatory power (Page et al. 2003). Microsatellites have also shown high rates of polymorphism in lake trout (Angers and Bernatchez 1996; Page 2001), which provides more discriminatory power when multiple strains are present.

While genetic data analyses are essential for determining which strains are most appropriate for stocking (Perkins et al. 1995; Page et al. 2003; DeKoning et al. 2006), efficacy of such analyses for strain identification have not been evaluated. Do current genetic markers and statistical approaches accurately determine the strain of origin of an individual lake trout that is the result of hybridization between individuals of two different pure strains (O’Gorman et al. 1998)? Previous studies have found genetic markers to be useful for studying the origin of individual lake trout of pure-strain origin (Burnham-Curtis et al. 1995; Page 2001; Page et al. 2003). Page et al. (2003) accurately classified each lake trout hatchery strain by using seven microsatellite loci, and DeKoning et al. (2006) documented unequal spawning success among strains in the wild. But as lake trout aggregates of multiple broodstock origins congregate on spawning reefs, spawning will likely occur between lake trout of different hatchery origins.

Inter-strain crossing of genes which results from independent assortment of genes from different broodstocks (i.e., Mendel’s Law of Independent Assortment) can create ambiguity in determining the strain of resulting hybrid offspring. After a few years of F_1 crosses, the ability to determine corresponding strains responsible for natural reproduction could be difficult. Combine the presence of F_1 individuals with the potential for crosses between two hybrids, or a hybrid backcrossed with a parental strain (cumulatively called F_x hybrids), and the ability to determine the strain of origin of a particular fish could be all but impossible. Many statistical computer algorithms have been written for identifying individuals to populations, but most of these models are untested for identifying individuals of hybrid origin. If fishery biologists are going to rely on results from genetic analyses these models to make management decisions, the

models must be accurate and unbiased, even when identifying hybrid individuals.

Accuracy, or the freedom from error, is the ability of the algorithm to correctly assign an individual to its origin. Bias is a consistent tendency of an algorithm to incorrectly assign an individual's origin. Accuracy depends on the number of contributing strains of lake trout and the genetic differentiation among the strains (Page et al. 2003). Bias will be greatest when strains are genetically similar but differ in abundance, a distinct likelihood in Lake Michigan lake trout because of differences in stocking among strains (Holey et al. 1995; Page et al. 2003).

The goal of my study was to determine if current statistical algorithms were able to differentiate among the four pure lake trout strains currently stocked on Lake Michigan's mid-lake reef complex, and each potential hybrid cross up to the F_2 that would be expected to occur if each of the strains was reproducing. If a method was found to have high assignment success, then that method would be used to assign lake trout eggs and fry collected from the mid-lake reef complex to their strain(s) of origin. But, if no algorithm was found to have sufficient performance capabilities to distinguish among the pure strains and hybrid crosses, then additional lake trout specific genetic markers would be developed to provide increased discriminatory ability in future lake trout studies. The three project objectives for this study were:

- 1) To determine if established genetic-based individual assignment and/or hybrid-admixture detection methods were unbiased and showed sufficient accuracy to identify interstrain lake trout hybrids.

- 2) To determine the likely strain of origin for lake trout eggs and/or larval fish collected in the mid-lake reef complex of Lake Michigan.
- 3) To develop, if necessary, additional genetic markers specifically designed for lake trout to provide increased discriminatory ability in future studies of lake trout reproduction.

LITERATURE CITED

- Allendorf, F.W. and S.R. Phelps. 1980. Loss of genetic variation in a hatchery stock of cutthroat trout. *Transactions of the American Fisheries Society* 109:537 – 543.
- Anderson, D.M. and J.J. Collins. 1995. Natural reproduction by stocked lake trout (*Salvelinus namaycush*) and hybrid (backcross) lake trout in South Bay, Lake Huron. *Journal of Great Lakes Research* 21(Supplement 1):260 – 266.
- Angers, B. and L. Bernatchez. 1996. Usefulness of heterologous microsatellites obtained from brook charr, *Salvelinus fontinalis* Mitchill, in other *Salvelinus* species. *Molecular Ecology* 5:317 – 319
- Bronte, C.R. 1993. Evidence of spring spawning lake trout in Lake Superior. *Journal of Great Lakes Research* 19:625 – 629.
- Bronte, C. R., M.E. Holey, C.P. Madenjian, J.L. Jonas, R.M. Claramunt, P.C. McKee, M.L. Toney, M.P. Ebener, B. Breidert, G.W. Fleischer, R.H.A.W. Martell Jr., and E.J. Olsen. 2007. Relative abundance, site fidelity, and survival of adult lake trout in Lake Michigan from 1999 to 2001: Implication for future restoration strategies.
- Bronte, C.R., J. Jonas, M.E. Holey, R.L. Eshenroder, M.L. Toney, P. McKee, B. Breidert, R.M. Claramunt, M.P. Ebener, C.C. Krueger, G. Wright, and R. Hess. 2003. Possible impediments to lake trout restoration in Lake Michigan. Great Lakes Fishery Commission, Ann Arbor, Michigan. Available: <http://www.glfc.org/lakecom/lmc/ltrestore.pdf>. (June 2007).
- Bronte, C.R. and S.A. Moore. 2007. Morphological variation of siscowet lake trout in Lake Superior. *Transactions of the American Fisheries Society* 136:509 – 517.
- Burnham-Curtis, M.K., C.C. Krueger, D.R. Schreiner, J.E. Johnson, T.J. Stewart, R.M. Horrall, W.R. MacCallum, R. Kenyon, and R.E. Lange. 1995. Genetic strategies for lake trout rehabilitation: a synthesis. *Journal of Great Lakes Research* 21(Supplement 1):477 – 486.
- Christie, U.J. 1972. Lake Ontario: effects of exploitation, introductions, and eutrophication on the salmonids community. *Journal of the Fisheries Research Board of Canada* 29:913 – 929.
- Claramunt, R.M., J.L. Jonas, J.D. Fitzsimons, and J.E. Marsden. 2005. Influences of spawning habitat characteristics and interstitial predators on lake trout egg deposition and mortality. *Transactions of the American Fisheries Society* 134:1048 – 1057.

- Dawson, K.A., R.L. Eshenroder, M.E. Holey, and C. Ward. 1997. Quantification of historic lake trout (*Salvelinus namaycush*) spawning aggregations in Lake Michigan. *Canadian Journal of Fisheries and Aquatic Science* 54:2290 – 2302.
- DeKoning, J., K. Keatley, R. Phillips, J. Rhydderch, J. Janssen, and M. Noakes. 2006. Genetic analysis of wild lake trout embryos recovered from Lake Michigan. *Transactions of the American Fisheries Society* 135:399 – 407.
- Edsall, T.A., M.E. Holey, B.A. Manny, and G.W. Kennedy. 1995. An evaluation of lake trout reproductive habitat on Clay Banks Reef, northwestern Lake Michigan. *Journal of Great Lakes Research* 21(Supplement 1):418 – 432.
- Edsall, T.A. and G.W. Kennedy. 1995. Availability of lake trout reproductive habitat in the Great Lakes. *Journal of Great Lakes Research* 21(Supplement 1):290 – 301.
- Edsall, T.A., G.W. Kennedy, and W.H. Horns. 1996. Potential spawning habitat for lake trout on Julian's Reef, Lake Michigan. *Journal of Great Lakes Research* 22:83 – 88.
- Einum, S. and I.A. Fleming. 2001. Implications of stocking: ecological interactions between wild and released salmonids. *Nordic Journal of Freshwater Research* 75:56 – 70.
- Elrod, J.H., R. O'Gorman, C.P. Schneider, and T. Schaner. 1996. Geographical distributions of lake trout strains stocked in Lake Ontario. *Journal of Great Lakes Research* 22:871 – 883.
- Eschmeyer, P.H. 1957. The near extinction of lake trout in Lake Michigan. *Transactions of the American Fisheries Society* 85:102 – 119.
- Eshenroder, R.L., C.R. Bronte, and J.W. Peck. 1995a. Comparison of lake trout-egg survival at inshore and offshore and shallow-water and deepwater sites in Lake Superior. *Journal of Great Lakes Research* 21(Supplement 1):313 – 322.
- Eshenroder, R.L., E.J. Crossman, G.K. Meffe, C.H. Olver, and E.P. Pister. 1995b. Lake trout rehabilitation in the Great Lakes: an evolutionary, ecological, and ethical perspective. *Journal of Great Lakes Research* 21(Supplement 1):518 – 529.
- Evans, D.O. and C.H. Olver. 1995. Introduction of lake trout (*Salvelinus namaycush*) to inland lakes of Ontario, Canada: factors contributing to successful colonization. *Journal of Great Lakes Research* 21(Supplement 1):30 – 53.
- Fetterolf, C.M. Jr. 1980. Why a Great Lakes Fishery Commission and why a sea lamprey international symposium. *Canadian Journal of Fisheries and Aquatic Sciences* 37:1588 – 1593.

- Fleming, I.A. and E. Petersson. 2001. The ability of released, hatchery salmonids to breed and contribute to the natural productivity of wild populations. *Nordic Journal of Freshwater Research* 75:71 – 98.
- Fitzsimons, J.D. 1995. A critical review of the effects of contaminants on early life state (ELS) mortality of lake trout in the Great Lakes. *Journal of Great Lakes Research* 21(Supplement 1): 267 – 276.
- Fitzsimons, J., G. Fodor, B. Williston, H. Don, B. Gray, M. Benner, T. Breedon, and D. Gilroy. 2005. Deepwater spawning by lake trout (*Salvelinus namaycush*) in Keuka Lake, New York. *Journal of Great Lakes Research* 31:1 – 10.
- Fitzsimons, J.D., S. Huestis, and B. Williston. 1995. Occurrence of a swim-up syndrome in Lake Ontario lake trout in relation to contaminants and cultural practices. *Journal of Great Lakes Research* 21(Supplement 1):277 – 285.
- Fitzsimons, J.D., B. Williston, G. Williston, L. Brown, A. El-Shaarawi, L. Vandendyck, D. Honeyfield, D. Tillitt, M. Wolgamood, and S.B. Brown. 2007. Egg thiamine status of Lake Ontario salmonines 1995 – 2004 with emphasis on lake trout. *Journal of Great Lakes Research* 33:93 – 103.
- Ford, M.J. 2002. Selection in captivity during supportive breeding may reduce fitness in the wild. *Conservation Biology* 16:815 – 825.
- GLFC (Great Lakes Fishery Commission). 2006. Fishery Research Priorities for the Great Lakes. Great Lakes Fishery Commission, Ann Arbor, Michigan, USA. Available: <http://www.glfc.org/research/Priorities.pdf>. (August 2007).
- Grewe, P.E. and P.D.N. Hebert. 1988. Mitochondrial DNA diversity among brood stocks of the lake trout *Salvelinus namaycush*. *Canadian Journal of Fisheries and Aquatic Science* 45:2114 – 2122.
- Grewe, P.E., C.C. Krueger, J.E. Marsden, C.F. Aquadro, and B. May. 1994. Hatchery origins of naturally produced lake trout fry captured in Lake Ontario: temporal and spatial variability based on allozyme and mitochondrial DNA data. *Transactions of the American Fisheries Society* 123:309 – 320.
- Guinand, B., K.T. Scribner, K.S. Page, and M.K. Burnham-Curtis. 2003. Genetic variation over space and time: analyses of extinct and remnant lake trout populations in the Upper Great Lakes. *The Proceedings of the Royal Society of London* 270:425 – 433.
- Gunn, J.M. 1995. Spawning behavior of lake trout: effects on colonization ability. *Journal of Great Lakes Research* 21(Supplement 1):323 – 329.

- Hacker, V.A. 1957. Biology and management of lake trout in Green Lake, Wisconsin. *Transactions of the American Fisheries Society* 86:71 – 83.
- Hansen, M.J. 1999. Lake trout in the Great Lakes: basinwide stock collapse and binational restoration. In *Great lakes fisheries policy and management: a binational perspective* (ed. W. W. Taylor and C. P. Ferreri). East Lansing, MI: Michigan State University Press.
- Hartman, W.L. 1972. Lake Erie: effects of exploitation, environmental changes, and new species on the fishery resources. *Journal of the Fisheries Research Board of Canada* 29:899 – 912.
- Hatch, R.W. 1984. Population dynamics and species interactions. Pages 4-9 in R. L. Eshenroder, T. P. Poe, and C. H. Olver, editors. *Strategies for rehabilitation of lake trout in the Great Lakes: Proceedings of a conference on lake trout research, August 1983*. Great Lakes Fisheries Committee Technical Report No. 40.
- Holey, M.E., R.W. Rybicki, G.W. Eck, E.H. Brown Jr., J.E. Marsden, D.S. Lavis, M.L. Toney, T.N. Trudeau, and R.M. Horrall. 1995. Progress toward lake trout restoration in Lake Michigan. *Journal of Great Lakes Research* 21(Supplement 1):128 – 151.
- Honeyfield, D.C., J.P. Hinterkopf, J.D. Fitzsimons, D.E. Tillitt, J.L. Zajicek, and S.B. Brown. 2005. Development of thiamine deficiencies and early mortality syndrome in lake trout by feeding experimental and feral fish diets containing thiaminase. *Journal of Aquatic Animal Health* 17:4 – 12.
- Ihssen, P.E. 1984. Genetics. Pages 15-21 in R.L. Eshenroder, T.P. Poe, and C.H. Olver, editors. *Strategies for rehabilitation of lake trout in the Great Lakes: Proceedings of a conference on lake trout research, August 1983*. Great Lakes Fisheries Committee Technical Report No. 40.
- Ihssen, P.E., J.M. Casselman, G.W. Martin and R.B. Phillips. 1988. Biochemical genetic differentiation of lake trout (*Salvelinus namaycush*) stocks of the Great Lakes Region. *Canadian Journal of Fisheries and Aquatic Sciences* 45:1018 – 1029.
- Janssen, J., D.J. Jude, T.A. Edsall, R.W. Paddock, N. Wattrus, M. Toney, and P. McKee. 2006. Evidence of lake trout reproduction at Lake Michigan's mid-lake reef complex. *Journal of Great Lakes Research* 32:749 – 763.
- Janssen, J., J.E. Marsden, C.R. Bronte, D.J. Jude, S.P. Sitar, and F.W. Goetz. 2007. Challenges to deep-water reproduction by lake trout: pertinence to restoration in Lake Michigan. *Journal of Great Lakes Research* 33(Supplement 1):59 – 74.

- Jonas, J.L., R.M. Claramunt, J.D. Fitzsimons, J.E. Marsden, and B.J. Ellrott. 2005. Estimates of egg deposition and effects of lake trout (*Salvelinus namaycush*) egg predators in three regions of the Great Lakes. *Canadian Journal of Fisheries and Aquatic Science* 62:2254 – 2264.
- Kincaid, H.L., C.C. Krueger, and B. May. 1993. Preservation of genetic variation in the Green Lake strain lake trout derived from remnant domestic and feral populations. *North American Journal of Fisheries Management* 13:318 – 325.
- Krueger, C.C., A.J. Gharrett, T.R. Dehring, and F.W. Allendorf. 1981. Genetic aspects of fisheries rehabilitation programs. *Canadian Journal of Fisheries and Aquatic Science* 38:1877 – 1881.
- Krueger, C.C., R.M. Horrall, and H. Gruenthal. 1983. Strategy for the use of lake trout strains in Lake Michigan. Report of the genetics subcommittee to the Lake Trout Technical Committee for Lake Michigan of the Great Lakes Fishery Commission, Ann Arbor, Michigan.
- Krueger, C.C. and P.E. Ihssen. 1995. Review of genetics of lake trout in the Great Lakes: history, molecular genetics, physiology, strain comparisons, and restoration management. *Journal of Great Lakes Research* 21(Supplement 1):348 – 363.
- Krueger, C.C., M.L. Jones, and W.W. Taylor. 1995a. Restoration of lake trout in the Great Lakes: challenges and strategies for future management. *Journal of Great Lakes Research* 21(Supplement 1):547 – 558.
- Krueger, C.C., J.E. Marsden, H.L. Kincaid, and B. May. 1989. Genetic differentiation among lake trout strains stocked into Lake Ontario. *Transactions of the American Fisheries Society* 118:317 – 330.
- Krueger, C.C., D.L. Perkins, E.L. Mills, and J.E. Marsden. 1995b. Predation by alewives on lake trout fry in Lake Ontario: role of an exotic species in preventing restoration of a native species. *Journal of Great Lakes Research* 21(Supplement 1):458 – 469.
- Lavis, D.S., M.P. Henson, D.A. Johnson, E.M. Koon, and D.J. Ollila. 2003. A case history of sea lamprey control in Lake Michigan: 1979 to 1999. *Journal of Great Lakes Research* 29(Supplement 1):584 – 598.
- LMLTTC (Lake Michigan Lake Trout Technical Committee). 1985. A draft lakewide plan for lake trout restoration in Lake Michigan. In *Minutes of the Lake Michigan Committee, 1985 annual meeting*, Great Lakes Fisheries Commission, Ann Arbor, Michigan.

- Loftus, K.H. 1976. Science for Canada's fisheries rehabilitation needs. *Journal of the Fisheries Research Board of Canada* 33:1822 – 1857.
- Lynch, M. and M. O'Heley. 2001. Captive breeding and the genetic fitness of natural populations. *Conservation Genetics* 2:363 – 378.
- Madenjian, C.P., T.J DeSorcie, J.R. McClain, A.P. Woldt, J.D. Holuszko, and C.A. Bowen II. 2004. Status of lake trout rehabilitation on Six Fathom Bank and Yankee Reef in Lake Huron. *North American Journal of Fisheries Management* 24:1003 – 1016.
- Madenjian, C.P., R. O'Gorman, D.B. Bunnell, R.L. Argyle, E.F. Roseman, D.M. Warner, J.D. Stockwell, and M.A. Stapanian. 2008. Adverse effects of alewives on Laurentian Great Lakes fish communities. *North American Journal of Fisheries Management* 28:263 – 282.
- Marsden, J.E., J.M. Casselman, T.A. Edsall, R.F. Elliott, J.D. Fitzsimons, W.H. Horns, B.A. Manny, S.C. McAughey, P.G. Sly, and B.L. Swanson. 1995. Lake trout spawning habitat in the Great Lakes – a review of current knowledge. *Journal of Great Lakes Research* 21(Supplement 1):487 – 497.
- Marsden, J.E. and M.A. Chotkowski. 2001. Lake trout spawning on artificial reefs and the effect of zebra mussels: fatal attraction? *Journal of Great Lakes Research* 27:33 – 43.
- Marsden J.E. and J. Janssen. 1997. Evidence of lake trout spawning on a deep reef in Lake Michigan using an ROV-based egg collector. *Journal of Great Lakes Research* 23:450 – 457.
- Martin, N.V. and C.H. Olver. 1980. The lake charr, *Salvelinus namaycush*. Pages 205–277 in E.K. Balon, editor. *Charr: Salmonid Fishes of the Genus Salvelinus*. Dr. W. Junk bv Publishers, The Hauge, The Netherlands.
- McDermid, J.L., P.E. Ihssen, W.N. Sloan, and B.J. Shuter. 2007. Genetic and environmental influences on life history traits in lake trout. *Transactions of the American Fisheries Society* 136:1018 – 1029.
- McKee, P.C., M.L. Toney, M.J. Hansen, and M.E. Holey. 2004. Performance of two strains of lake trout stocked in the Midlake Refuge of Lake Michigan. *North American Journal of Fisheries Management* 24:1101 – 1111.
- Meffe, G.K. 1995. Genetic and ecological guidelines for species reintroduction programs: application to Great Lakes fishes. *Journal of Great Lakes Research* 21(Supplement 1):3 – 9.

- Miller, M.A. and M.E. Holey. 1992. Diets of lake trout inhabiting nearshore and offshore Lake Michigan environments. *Journal of Great Lakes Research* 18:51 – 60.
- Morbey, Y.E., D.M. Anderson, and B.A. Henderson. 2008. Progress toward the rehabilitation of lake trout (*Salvelinus namaycush*) in South Bay, Lake Huron. *Journal of Great Lakes Research* 34:287 – 300.
- Noakes, D.L.G. and R.A. Curry. 1995. Lessons to be learned from attempts to restore *Salvelinus* species other than *S. namaycush*: a review of reproductive behavior. *Journal of Great Lakes Research* 21(Supplement 1):54 – 64.
- O’Gorman, R., J.H. Elrod, and C.P. Schneider. 1998. Reproductive potential and fecundity of lake trout strain in southern and eastern waters of Lake Ontario, 1977 – 1994. *Journal of Great Lakes Research* 24:131 – 144.
- Page, K.S. 2001. Genetic diversity and interrelationships of wild and hatchery lake trout in the upper Great Lakes: inferences for broodstock management and development of restoration strategies. Master’s thesis. Michigan State University, East Lansing.
- Page, K.S., K.T. Scribner, D. Bast, M.E. Holey, and M.K. Burnham-Curtis. 2005. Genetic evaluation of a Great Lakes lake trout hatchery program. *Transactions of the American Fisheries Society* 134:872 – 891.
- Page, K.S., K.T. Scribner, K.R. Bennett, L.M. Garzel, and M.K. Burnham-Curtis. 2003. Genetic assessment of strain-specific sources of lake trout recruitment in the Great Lakes. *Transactions of the American Fisheries Society* 132:877 – 894.
- Perkins, D.L., J.D. Fitzsimons, J.E. Marsden, C.C. Krueger, and B. May. 1995. Differences in reproduction among hatchery strains of lake trout at eight spawning areas in Lake Ontario: genetic evidence from mixed-stock analysis. *Journal of Great Lakes Research* 21(Supplement 1):364 – 364.
- Perkins, D.L. and C.C. Krueger. 1995. Dynamics of reproduction by hatchery-origin lake trout (*Salvelinus namaycush*) at Stony Island Reef, Lake Ontario. *Journal of Great Lakes Research* 21(Supplement 1):400 – 417.
- Rahrer, J.H. 1965. Age, growth, maturity, and fecundity of “humper” lake trout, Isle Royale, Lake Superior. *Transactions of the American Fisheries Society* 94:75 – 83.
- Royce, W.F. 1951. Breeding habits of lake trout in New York. U.S. Fish and Wildlife Service Fisheries Bulletin 52:59 – 76.

- Schneider, C.P., D.P. Kolenosky, and D.B. Goldthwaite. 1983 *A joint plan for the rehabilitation of lake trout in Lake Ontario*. Ann Arbor, MI: Great Lakes Fishery Commission, Lake Ontario Committee, Special Publication.
- Schreiner, D.R., C.R. Bronte, N.R. Payne, J.D. Fitzsimons, and J.M. Casselman. 1995. Use of egg traps to investigate lake trout spawning in the Great Lakes. *Journal of Great Lakes Research* 21(Supplement 1):433 – 444.
- Sitar, S.P., H.M. Morales, M.T. Mata, B.B. Bastar, D.M. Dupras, G.D. Kleaver, and K.D. Rathbun. 2008. Survey of siscowet lake trout at their maximum depth in Lake Superior. *Journal of Great Lakes Research* 34:276 – 286.
- Smith, B.R. and J.J. Tibbles. 1980. Sea lamprey (*Petromyzon marinus*) in Lakes Huron, Michigan, and Superior: history of invasion and control, 1936 – 78. *Canadian Journal of Fisheries and Aquatic Sciences* 37:1780 – 1801.
- Smith, H.M. and M.M. Snell. 1891. The fisheries of Lake Michigan. In *Report of the Commissioner for 1887*, eds. H. M. Smith, M. –M. Snell, and J. W. Collins, pp. 71 – 204. U.S. Committee of Fish and Fisheries, Part XV. Washington, D.C.: Government Printing Office.
- Van Oosten, J. 1949. A definition of depletion of fish stockes. *Transactions of the American Fisheries Society* 76:283 – 289.
- Utter, F. and N. Ryman. 1993. Genetic markers and mixed stock fisheries. *Fisheries* 18:11 – 21.
- Wang, J. and N. Ryman. 2001. Genetic effects of multiple generations of supportive breeding. *Conservation Biology* 15:1619 – 1631.
- Ward, C., R.L. Eshenroder, and J.R. Bence. 2000. Relative abundance of lake trout and burbot in the main basin of Lake Michigan in the early 1930s. *Transactions of the American Fisheries Society* 129:282 – 295.
- Wells, L. and A.L. McLain. 1972. Lake Michigan: effects of exploitation, introductions, and eutrophication on the salmonids community. *Journal of the Fisheries Research Board of Canada* 29:889 – 898.
- Wells, L. and A.L. McLain. 1973. Lake Michigan: man's effects on native fish stocks and other biota. Great Lakes Fishery Commission, Technical Report 20.
- Zimmerman, M.S., C.C. Krueger, and R.L. Eshenroder. 2007. Morphological and ecological differences between shallow- and deep-water lake trout in Lake Mistassini, Quebec. *Journal of Great Lakes Research* 33:156 – 169.

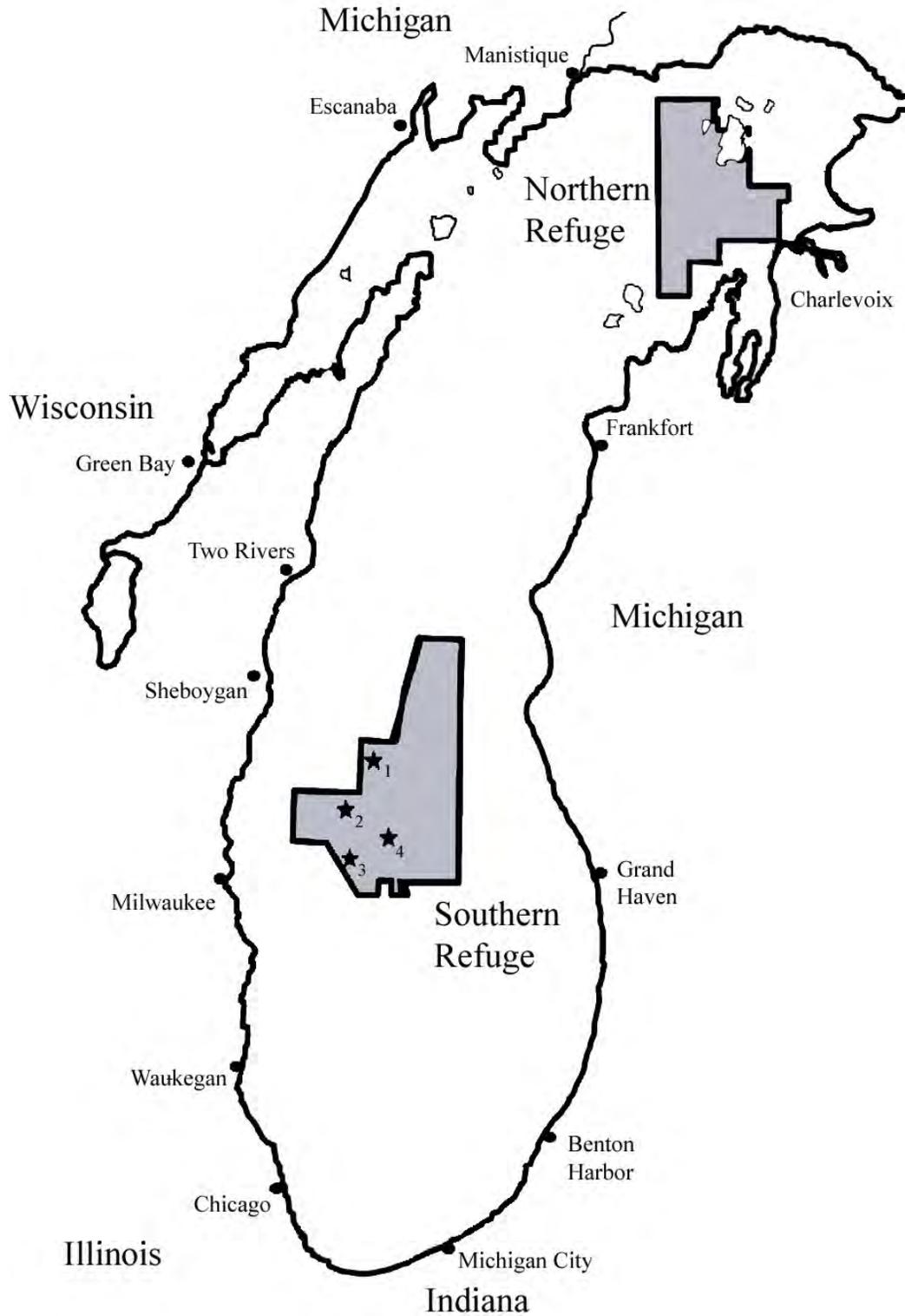


Figure 1. Location of the Northern Refuge and the Southern Refuge which includes the mid-lake reef complex (1 = Sheboygan Reef, 2 = Northeast Reef, 3 = East Reef, and 4 = Milwaukee Reef) in Lake Michigan; adapted from Bronte et al. 2007).

CHAPTER 1

EFFICACY OF GENETIC ALGORITHMS IN IDENTIFYING SIMULATED PURE STRAIN AND HYBRID LAKE TROUT IN LAKE MICHIGAN'S LAKE TROUT RESTORATION EFFORTS

ABSTRACT

Current Lake Michigan lake trout (*Salvelinus namaycush*) management plans include the propagation of four hatchery strains to recolonize the Lake Michigan basin and determine the most appropriate strains for continued stocking. By determining which strains are reproducing in contemporary conditions, fishery managers will be able to focus subsequent stocking on successful strains. Ambiguity in determining which strains are reproducing can occur because any offspring could be the result of mating between fish of different hatchery strains (interstrain hybrids). The objective of this study was to determine if established genetic-based individual assignment and/or hybrid-admixture detection methods were sufficiently accurate and unbiased to identify interstrain lake trout hybrids using available reference data for all four strains. Simulated pure-strain offspring and various interstrain crosses were tested to determine accuracy and bias of six algorithms (i.e., BAPS v4.14, NewHybrids v1.1, STRUCTURE v2.2, GeneClass2, GMA, and WhichRun). All algorithms were moderately successful for assigning pure-strain individuals to their origin strain, but had limited success in assigning interstrain hybrid offspring to their strains of origin. Success in assigning pure-strain individuals ranged from 8.9% to 40.0% for maximum likelihood algorithms and from 36.8% to 99.6% for Bayesian clustering algorithms. For F₁ hybrid individuals, assignment success ranged from 4.3% to 46.2% for maximum likelihood algorithms and from 7.0% to 95.7% for Bayesian clustering algorithms. Assignment success of advanced hybrid individuals (i.e.,

B_x and F_2) only exceeded 65% for the New Hybrids algorithm. While NewHybrids consistently had the highest assignment success, its use in studies of Lake Michigan's lake trout may be limited by the algorithm's limitation of a maximum of two reference populations during assignment. Following initial trials, assignment thresholds and parameters were relaxed to determine if original assignment thresholds were too stringent. These additional trials also had limited success in assigning hybrid individuals to their origin strains. Overall, no method was found to have high enough performance capabilities to be used for assigning individuals of unknown origin to their genetic strains using available reference data for four strains propagated in the Lake Michigan lake trout restoration program.

INTRODUCTION

Following the extirpation of lake trout (*Salvelinus namaycush*) from Lake Michigan in the mid-1950s (Eschmeyer 1957), a propagation program was begun (Fetterolf 1980; Holey et al. 1995). In 1985, stocking efforts were refined through the creation of the Lake Wide Management Plan (LMLTTC 1985), which included stocking multiple strains of lake trout of different morphotypes and ecological preferences (Holey et al. 1995). Differential survival among strains would provide fisheries managers with a way to infer adaptability of strains to contemporary Lake Michigan conditions through interstrain variation in survival and reproduction (Holey et al. 1995). The four strains currently used for restoration efforts are Green Lake, Lewis Lake, Seneca Lake, and Marquette. Ancestors of fish comprising the Green Lake strain were collected in southern Lake Michigan and released into Green Lake, Wisconsin (Burnham-Curtis et al. 1995). The Lewis Lake strain, possibly a siscowet morphotype (Smith and Snell 1891; Krueger and Ihssen 1995), originated from northern Lake Michigan and was introduced into Lewis Lake, Wyoming before 1900 (Krueger et al. 1983; Krueger and Ihssen 1995). The Seneca Lake strain is a deep-water, early-fall spawning population of the lean morphotype originating in Seneca Lake of the Finger Lakes system in upstate New York (Royce 1951; Krueger and Ihssen 1995). The Marquette strain was founded in the late 1940s from lake trout collected from in-shore populations on the southern shore of Lake Superior near Marquette, Michigan (Krueger et al. 1983; Krueger and Ihssen 1995; Holey et al. 1995; Guinand et al. 2003).

In situations attempting to reintroduce a species, such as the lake trout, through the release of multiple strains of a species, monitoring success by assigning individuals to

their strain of origin is essential to assess reproductive success, a precursor to self-sustainability, of specific strains. Differential reproductive success among lake trout strains is possible (DeKoning et al. 2006), and interstrain hybridization may occur if more than one strain successfully reproduces. When managers are unable to directly infer origin of individuals through visual observations of specific strains participating in spawning events, genetic data collected from eggs and fry must be used with statistical assignment methods to determine reproductively successful strains (Paetkau et al. 1995; Pritchard et al. 2000; Cegelski et al. 2003). Most assignment methods are untested in these types of situations so caution should be used when basing management decisions on results obtained by assignment methods until they are tested for their accuracy and bias in assignment of interstrain hybrid offspring.

The goal of this research was to evaluate available genetic data for four strains of lake trout stocked into Lake Michigan (Page et al. 2003) for its ability to assign individuals of various crosses. Specific objectives were 1) to determine if established genetic-based individual assignment and/or hybrid-admixture detection methods were unbiased and showed sufficient accuracy to identify interstrain lake trout hybrids using the reference data currently in place for lake trout strains in Lake Michigan and, if possible, 2) to determine the likely strain of origin for lake trout eggs and/or larval fish collected on the mid-lake reef complex of Lake Michigan.

MATERIALS AND METHODS

Experimental Approach

When assignment methods are used to make decisions regarding which lake trout strains to stock on the mid-lake reef complex, the major concern is not the minor genetic differences among the strains, but rather the ability to detect those strains that reproduce best under current ecological conditions (Marsden et al. 1989). Strain reproductive success will need to be assessed with assignment of eggs and fry collected from the mid-lake reef complex to their most probable strains of origin by an assignment method found to have the highest performance capabilities.

I used simulated populations of lake trout, that included the four pure strains stocked on Lake Michigan's mid-lake reef complex and each potential hybrid cross up to the F₂ generation that could occur from the pure strains, to test performance (i.e., accuracy and bias) of six assignment methods that used either Bayesian clustering or maximum likelihood estimation. These six statistical algorithms were chosen because they are frequently used in other peer-reviewed studies to assign individuals to populations, and therefore are likely methods to be used for future lake trout studies. The benefit to simulating the population of individuals, represented as multilocus genotypes, for testing assignment accuracy was that an almost unlimited number of individuals could be created in the population (Roques et al. 1999). These individuals were of known origin, but were treated as unknowns during assignment. Genotypic data for all simulated populations (details to follow) consisted of the seven microsatellite loci sampled by Page et al. (2003). Microsatellites are highly variable non-coding regions of nuclear DNA, consisting of repeated simple sequence motifs (Hansen et al. 2001). The

seven loci used were *Sfo1* (3 alleles), *Sfo112* (5 alleles), *Sfo18* (9 alleles), *Oneμ10* (3 alleles), *Ogo1a* (6 alleles), *Scoμ19* (11 alleles), and *Ssa85* (6 alleles) (Table 1.1) (Page et al. 2003; DeKoning et al. 2006). For each algorithm, accuracy was measured as the percentage of individuals assigned to their correct origin strain (Berry et al. 2004). Bias was calculated as the percentage of individuals assigned to a non-origin strain (Blott et al. 1999). If an individual's assignment confidence did not meet or exceed the chosen assignment threshold for any population, it was considered to be of ambiguous ancestry (Cegelski et al. 2003).

Assignment tests use genetic information to determine the population from which an individual most likely originated based on the multilocus genotype of the individual and the probability of that genotype occurring in each potential source population (Cornuet et al. 1999; Vázquez-Domínguez et al. 2001; Berry et al. 2004; Pearse and Crandall 2004; Manel et al. 2005). This process can be used to determine if an individual belongs to the population from which it was sampled or to assign unknown individuals to a set of baseline samples from known populations (Hansen et al. 2001). After allele frequencies from potential source populations were calculated, unknown individuals were assigned to a known population from which the frequency of its genotype was the greatest (Cornuet et al. 1999; Luikart and England 1999; Koljonen et al. 2005).

Assignment methods tend to use either classification or clustering. During classification, individuals are assigned to predefined categories and bias results if the true origin population for an individual was not sampled (Manel et al. 2005). Clustering methods construct categories from data rather than using predefined categories (Manel et al. 2005). This process relies on presence of linkage disequilibrium within a mixture sample

to create groups of individuals that minimize the overall amount of linkage disequilibrium (Manel et al. 2005). Individuals are then assigned to the cluster in which it has the highest probability of belonging (Manel et al. 2005). Clustering methods are particularly useful when genetic data for potential source populations are not available, population boundaries are uncertain, or some potential sources have not been sampled (Manel et al. 2005).

Both classification and clustering methods assume that source populations are in Hardy-Weinberg (HWE) and linkage equilibrium (LE), all potential source populations have been defined, random sampling, and adequate samples have been collected for each potential source population, and allele frequencies are known without error (Manel et al. 2005; Guinand et al. 2006; Hauser et al. 2006). Many assignment tests designate a single population as the probable source for the individual being assigned (Cornuet et al. 1999), but some are able to detect immigrants into a population, or offspring of immigrants within a population (Eldridge et al. 2001). The percentage of individuals correctly assigned to their source population and estimates of the level of genetic differentiation between populations (i.e., F_{st}) are useful predictors of assignment method performance (Cornuet et al. 1999; Berry et al. 2004).

Data Simulation

Simulated populations of pure strain, F_1 (pure strain x pure strain), F_2 (F_1 x F_1), and first generation backcrossed (B_x ; pure strain x F_1) individuals were constructed from observed allele frequencies (Table 1.1) for each of four pure hatchery lake trout strains (Green Lake [GLW], Lewis Lake [LLW], Marquette [SMD], and Seneca Lake [SLW])

by Page et al. (2003). Simulations assumed that allele frequencies of the pure strain populations were equal to posterior mean estimates found by Page et al. (2003) (Anderson and Thompson 2002). For each pure strain, 1,000 individual multilocus genotypes were created in Excel by randomly sampling without replacement one allele from each of two pure strain allele frequencies (Blott et al. 1999). These pure strain genotypes were simulated assuming HWE and LE. All hybrid offspring were simulated in HybridLab (Nielsen et al. 2006), a computer program designed to generate artificial hybrids between pairs of potential parental populations (Nielsen et al. 2001). This method mimics an instantaneous population expansion in which pure strain individuals spawn F_1 hybrid populations that subsequently spawn B_x hybrid populations and F_2 hybrid populations (Choisy et al. 2004; Latch et al. 2006). This method assumes that no genetic drift, selection, or mutation occurs (Choisy et al. 2004; Latch et al. 2006).

To generate hybrid individuals, HybridLab first estimates allele frequencies at each locus in parental populations from individual multilocus genotypes within each population (Nielsen et al. 2006). Next, HybridLab randomly draws one allele from each parental population at each locus as a function of their calculated frequency distributions (Nielsen et al. 2001; Nielsen et al. 2006). Different levels of hybridization are created by use of different parental populations as the input file. For example, input files used to generate F_1 hybrid individuals are two pure strain populations. These F_1 populations can then be used as input files to generate B_x and F_2 individuals by inputting an F_1 and a pure strain population or two F_1 populations, respectively (Nielsen et al. 2006). For each potential F_1 , F_2 , and B_x cross (Table 1.2), 1,000 individuals were generated for use in testing performance of Bayesian clustering and maximum likelihood (ML) assignment

methods used in six computer algorithms (Table 1.3). Each simulated individual was given a unique identification code to allow for recognition among trials. All simulated genotype population data were converted into GenePop (Raymond and Rousset 1995) or other appropriate formats through either Microsatellite Toolkit in Excel (Park 2001) or Formatomatic (Manoukis 2007). Input files were used in six Bayesian clustering or ML algorithms (Table 1.3) to assess performance (i.e., bias and accuracy) of each algorithm for assigning pure strain, F₁ hybrid and B_x hybrid lake trout individuals to their strain of origin. Bias was defined as a consistent tendency of an algorithm to incorrectly assign an individual's origin. Accuracy was defined as the ability of an algorithm to correctly assign an individual to its true origin.

Population Differentiation

Population differentiation among the four pure strain and F₁ hybrids, B_x hybrids, and F₂ hybrids were assessed by calculating pairwise F_{st} scores in Arlequin v3.1 (Excoffier et al. 2006). F_{st} is a measure of genetic differentiation first proposed by Wright (1951) that compares subpopulation genetic diversity, *s*, to total population genetic diversity, *t*, (Balloux et al. 2002; Neigel 2002; Hedrick 2005):

$$F_{st} = \frac{\sigma_a^2}{\sigma_T^2}$$

where σ_a^2 is covariance among populations and σ_T^2 is total covariance (Excoffier et al. 2006). The F_{st} scores range from 0 to 1 with a score of 0 meaning that the same allele frequencies are present in all subpopulations and a score of 1 meaning that subpopulations are fixed for different alleles (Hedrick 2005). Populations that share more alleles (i.e., less genetically distinct) have F_{st} scores closer to 0 whereas populations

that share fewer alleles (i.e., more genetically distinct) have scores closer to 1. The null hypothesis of this test is that two population's allele frequencies' do not differ (Excoffier et al. 2006).

Pairwise F_{st} estimates were calculated in Arlequin with 25,000 permutations to test for significance of the value compared to zero (i.e., no difference). Input files for comparisons among pure strain and F_1 hybrid crosses and pure strain and B_x hybrid crosses consisted of 1,000 individuals from each of four pure strains and either 1,000 individuals from each F_1 cross or 1,000 individuals from each B_x cross. Input files for comparisons among pure strain and F_2 crosses were abbreviated to 250 individuals from each pure strain and F_2 hybrid cross because program limitations would not accept the large files that would have been necessary for an entire pure strain and F_2 comparison.

Bayesian Clustering

Three Bayesian-based computer algorithms, commonly used in other fishery studies throughout peer-reviewed literature, were assessed for accuracy and bias of assignment of origin and admixture detection. Bayesian assignment methods calculate the probability of an individual's multilocus genotype occurring in a cluster, given the allele frequencies at different loci in each cluster (Beaumont and Rannala 2004) and the assumption of HWE and LE within each randomly mating subpopulation (Mank and Avise 2004). Each cluster fits genetic criteria that define it as a group, with individuals from one cluster possessing multilocus genotypes more similar to each other than to multilocus genotypes contained in other clusters (Cornuet et al. 1999). This allows for genetic similarity (i.e., the proportion of shared alleles across loci) to act as a proxy for

genetic ancestry (Davies et al. 1999; Mank and Avise 2004). Cluster analysis is subject to considerable uncertainty unless true populations are strongly divergent (Manel et al. 2005).

Bayesian analysis is based on Reverend Thomas Bayes method for estimating an unknown quantity based on information available about the unknown prior to sampling (Pella and Masuda 2001). This available information is in the form of a prior probability distribution (Pella and Masuda 2001). Bayes' theorem states that the posterior probability is proportional to the product of the prior and the likelihood of the sample:

$$p(\text{parameter} \mid \text{data}) = \frac{p(\text{parameter})p(\text{data} \mid \text{parameter})}{p(\text{data})}$$

where $p(\text{parameter} \mid \text{data})$ is the posterior probability, $p(\text{parameter})$ is the prior distribution, $p(\text{data} \mid \text{parameter})$ is the conditional probability, and $p(\text{data})$ is the marginal probability (Shoemaker et al. 1999). This analysis makes probability statements about the parameters and calculates the posterior probability of each parameter given the data (Shoemaker et al. 1999). Bayesian analysis allows for simultaneous estimation of many interdependent parameters in complex models (Latch et al. 2006) through the use of Markov Chain Monte Carlo (MCMC) to repeatedly sample from a simulated distribution that is expected to contain the true posterior distribution (Cowles and Carlin 1996; Shoemaker et al. 1999).

Markov Chain Monte Carlo is a method that simulates a stochastic process to study complicated probability distributions that cannot be easily studied using other methods (Beaumont and Rannala 2004). In MCMC, the properties of random variables are studied by simulating many instances of a variable, sampling the distribution, and determining the posterior probability of a result (Cowles and Carlin 1996; Beaumont and

Rannala 2004). In any iteration, the starting point of the Markov chain is based on the output of the previous iteration so future states of the chain are completely determined by the current state of the chain (Beaumont and Rannala 2004). The Markov chain is expected to converge on the true posterior probability distribution over many thousands of iterations (Cowles and Carlin 1996; Beaumont and Rannala 2004). Convergence is not determined by a set rule, so length of the chain depends on the data (Cowles and Carlin 1996).

Oftentimes, Bayesian priors are used to give the Markov chain a starting point for convergence. Without priors the Markov chain may not reach convergence even if allowed to run for an extremely large number of iterations (Mank and Avise 2004). These priors allow the user to take into account all relevant data representing the identity of the tested sample (i.e., sampling location, suspected population membership, or population allele frequency estimates) (Baudouin et al. 2004; Beaumont and Rannala 2004; Mank and Avise 2004). Priors may also introduce biases or circular reasoning in which the individual assignment merely recovers information provided as a prior (Mank and Avise 2004). If no priors are used, then all populations are considered to be equally possible (Baudouin et al. 2004). Collection of accurate baseline data for use as priors is often impeded by sampling effort required to obtain an accurate allele frequency distribution and inherent inaccuracy of allele frequency estimation process (Guinand et al. 2004; Hauser et al. 2006).

Bayesian Analysis of Population Structure (BAPS) v4.14.—BAPS v4.14 uses Bayesian inference to determine the genetic structure of a population (Corander and Marttinen 2006). The algorithm uses “trained clustering” in which individuals of known

origin are used to find the best clustering for individuals of unknown origin (Corander and Marttinen 2006). BAPS treats allele frequencies of molecular markers and the number of genetically distinct populations as random variables (Corander and Marttinen 2006) and assumes that genetic loci are unlinked and source populations are in HWE (Corander and Marttinen 2006). Partial or complete baseline data from presumed source populations can also be incorporated into priors for cluster “training” (Corander and Marttinen 2006). This version of BAPS uses a stochastic optimization to infer the posterior mode of the genetic structure including the number of populations in the data (Corander and Marttinen 2006). Once the number of populations is determined, individual admixtures are assessed using a Markov chain-based algorithm (Corander and Marttinen 2006). In this step, any individual can be assigned to an admixture ancestry that includes any of the k source populations (Corander and Marttinen 2006).

BAPS v4.14 was used to examine simulated data using the “admixture based on pre-defined clustering” option. Data files containing 1,000 known individuals of each of the four pure strains and 1,000 unknown hybrid offspring of a single cross were input into BAPS in the GenePop format. In this format, each of the four pure strain populations and the hybrid population were considered separate populations. Each pure strain population was assigned a discrete code in the input file which allowed the algorithm to use the population’s data for trained clustering. The hybrid population was assigned a code (-1) to identify hybrid individuals as unknown origin (not preassigned to any particular population). This hybrid population was used to define one more cluster with respect to which admixture proportions to be estimated (Corander and Marttinen 2006). Each run included a minimum population size of 750 individuals during admixture estimation, 100

iterations for estimation of admixture coefficients for each individual, 500 reference individuals from each population, and 25 iterations to estimate admixture coefficients for reference individuals. These conditions were more than adequate according to Corander and Marttinen (2006).

Assignment using BAPS was performed for each of the 39 hybrid crosses. An additional pure strain self-assignment was performed in which the four pure strains were used as the known reference file and as an unknown input file. Results obtained from 40 runs were exported to Excel to compare how many individuals from each cross met or exceeded the assignment threshold ($p \geq 0.80$) for belonging to each cluster (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008). The threshold of $p \geq 0.80$ represents a balance between a high assignment confidence necessary for the basis of management decisions and a high rate of confident assignments.

NewHybrids v1.1.—NewHybrids v1.1 (Anderson and Thompson 2002) models hybridized populations as a mixture with unknown proportions of individuals from six different (pure strain A, pure strain B, F₁, F₂, backcross to A, backcross to B) genotype frequency classes (Anderson and Thompson 2002). The model assumes that members of the same hybrid category have the same expected proportion of their genes originating from each pure strain (Anderson and Thompson 2002). Only two input populations per run are supported by NewHybrids, a drawback under the current lake trout propagation scheme which uses four hatchery strains. Markov Chain Monte Carlo sampling is used to determine the posterior probability of an individual belonging to each of the two pure strains or four hybrid classes that could occur as a result of the two pure strains hybridizing (Gunnell et al. 2007). Assumptions of the model are: 1) individual samples

are from potentially hybridizing populations; 2) genetic loci are unlinked; 3) individuals in the origin populations n generations prior to sampling were in HWE and LE; and 4) linkage and Hardy-Weinberg disequilibrium in the mixed population are entirely the result of mixing and admixing of population A and population B gene pools (Anderson and Thompson 2002).

Several features distinguish NewHybrids from other methods. First, NewHybrids does not require that pure strain individuals be sampled separately, but can include this information as priors if available (Anderson and Thompson 2002). Second, NewHybrids does not require loci to be diagnostic or species to possess unique alleles to distinguish between hybrid categories (Anderson and Thompson 2002). Third, uncertainty associated with allele frequency estimation is incorporated into the model (Anderson and Thompson 2002). Finally, NewHybrids compares allele frequencies among individuals to determine the probability that they belong to one of two pure lineages or from a hybrid mixture (Wares et al 2004, Gunnell et al. 2007).

NewHybrids analyses were performed in the non-graphical version with Jeffrey's prior distribution at each locus, 45,000 sweeps for a burn-in, and 45,000 sweeps for posterior probability calculations. Each pure strain and hybrid cross was input as an unknown population and compared to each possible pure strain or unknown pure strain input file and known pure strain or pure strain input combination. Allele frequencies of the two pure strains were used as priors. Individuals were considered unambiguously assigned to a pure strain or hybrid category if their assignment score was ≥ 0.80 (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008).

STRUCTURE v2.2.—STRUCTURE v2.2 (Pritchard et al. 2000) is a fully Bayesian method for evaluating genetic admixture in samples of known origin or of unknown origin, and thereby provides a framework for incorporating uncertainty of parameter estimation into an inference procedure (Manel et al. 2002). A MCMC algorithm is used to cluster individuals into groups that maximize HWE and linkage equilibrium between loci on the basis of multilocus genotypic data (Pritchard et al. 2000; Falush et al. 2007; Barilani et al. 2007). A posterior probability of an individual's membership in a population within all sampled populations is computed and directly interpreted as the probability that each individual originated from each sampled population (Manel et al. 2002; Wallace 2006). STRUCTURE assumes that all genetic material of the sampled individuals comes from one or more of k unobserved populations characterized by a set of allele frequencies at each locus (Falush et al. 2007). The admixture model in STRUCTURE allows for the possibility that individuals may have mixed ancestry in more than one of k populations (Falush et al. 2007). The entire process is completed by updating subsets of parameters to new values conditional on the data and current values of parameters at the completions of each iteration until the algorithm converges on the posterior distribution of all parameters given the data (Falush et al. 2007).

Admixture analysis in STRUCTURE was conducted using the admixture model with a burn-in and run length of 50,000 iterations each. The input file for each run consisted of 1,000 pure strain individuals from each of four strains with USEPOPINFO flag set to 1 (to use the provided population information) and 1,000 individuals from a single pure strain or F₁ hybrid class with the USEPOPINFO set to 0 (do not use provided

population information). Runs with the pure strain unknown population provided self-assignment tests whereas runs with hybrid classes provided tests of hybrid identification. An *a priori* assignment threshold of $p \geq 0.80$ (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008) was used to determine how many individuals from each run were confidently identified to each cluster.

Maximum Likelihood Assignment

Maximum likelihood estimation was developed by R.A. Fisher in the 1920s and is based on the principle that the desired probability distribution is the one that makes the observed data “most likely,” thereby testing the data against a given parameter set [L(data | parameters)] (Myung 2003, Beaumont and Rannala 2004). These methods produce point estimates of model parameters that maximize the likelihood function of the parameters for a fixed set of data (Beaumont and Rannala 2004; Manel et al. 2005). Three computer algorithms based on maximum likelihood assignment were assessed for accuracy and bias.

GeneClass2.—GeneClass2 (Piry et al. 2004) uses predetermined sets of reference populations to assign or exclude a set of unknown individuals to or from populations based on their likelihood of occurring in reference populations (Baudouin et al. 2004; Piry et al. 2004). The maximum likelihood of an individual occurring in each reference population is calculated assuming Hardy-Weinberg and linkage equilibrium, and the individual is assigned to the population in which the maximum likelihood estimate is the highest (Piry et al. 2004).

Two approaches can be employed in GeneClass2 based on criteria for assignment: a threshold assignment or a simulation assignment. Threshold assignment allows an *a priori* minimum threshold to be met before a population is considered to be a likely source candidate (Cornuet et al. 1999) and allows for populations to be excluded from consideration as an individual's source (Pearse and Crandall 2004). This is an important feature because in some situations the true source population may not have been sampled, and rather than selecting the most likely of all likelihoods, all populations with low likelihoods can be excluded from consideration (Cornuet et al. 1999; Pearse and Crandall 2004). In a reference file of k populations, the score of an individual i in a population l is computed according to the equation:

$$likelihood_{i,l} = \frac{L_{i,l}}{\sum_{j=1}^k L_{i,j}}$$

with $L_{i,l}$ the likelihood value of the individual i in the population l (Piry et al. 2004).

The simulation method uses a partial Bayesian method based on Rannala and Mountain (1997) (Maudet et al. 2002). Allele frequency estimates for each reference population are used to simulate a given number of multilocus genotypes within the genotype distribution (Cornuet et al. 1999; Manel et al. 2002). The likelihood of an individual occurring in each reference population is calculated through the leave-one-out method (Efron 1983), which compares the observed frequency of the tested individual's genotype occurring within the simulated genotypic distribution to the candidate population's genotypic distribution (Manel et al. 2002; Maudet et al. 2002). The leave-one-out method sequentially removes the individual being assigned when calculating the allele frequencies in each population to reduce bias that would be introduced by having a

multilocus genotype in one population perfectly match the genotype under consideration (Hansen et al. 2001; Manel et al. 2002; Berry et al. 2004).

The threshold method was used to assign individuals at a minimum threshold level of 0.05 based on the Rannala and Mountain (1997) Bayesian criteria for computation. This criterion performed better than other assignment methods under a variety of conditions (Cornuet et al. 1999; Nielsen et al. 2001). Reference populations for each run consisted of 1,000 pure strain individuals from each strain. Self-assignment success was assessed by using the reference population input file as the unknown input file. Each hybrid assessment trial was run by inputting 1,000 individuals from a single hybrid cross as unknown individuals to be assigned in each run. Hybrid assessment trials were run 39 times, or once with each hybrid cross. Simulation assignment was completed using the above settings and input files with the addition of Monte Carlo resampling based on Rannala and Mountain (1997) and the use of 100 simulated individuals. Individuals were considered unambiguously assigned to a pure strain or hybrid category if their assignment score was greater than or equal to 0.80 (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008). Confidence in assignment for both the threshold and the simulation method was assessed through bootstrap resampling (Piry et al. 2004), a process that repeatedly resamples the data with replacement to determine how often a particular outcome would be expected to occur (Beaumont and Rannala 2004). For all iterations, 500 bootstraps were performed.

Genetic Mixture Analysis (GMA).—GMA uses baseline genetic data to estimate the composition of a sample containing individuals from populations included in baseline files (Kalinowski 2003). Estimates of mixture proportions of populations included in the

sample are incorporated into the algorithm while performing assignment tests (Kalinowski 2003). The method of Rannala and Mountain (1997) is used to estimate the probability of observing a genotype in a particular baseline population based on Bayes' rule and the estimated contribution of that baseline population in the mixture (Kalinowski 2003). Individuals are assigned to the population in which they have the highest probability of belonging (Kalinowski 2003) assuming that all possible populations of origin are represented in the baseline genetic data (Koljonen et al. 2005). Individual probabilities of belonging are calculated based on the formula:

$$P_{ij} = \frac{m_j f_{ij}}{\sum_j m_j f_{ij}}$$

where P_{ij} represents the probability that an individual i of unknown origin belongs to population j in the baseline file, f_{ij} represents the frequency of the i^{th} fish's genotype in the j^{th} population of the baseline, and m_j represents the estimated stock composition of the sample (Kalinowski 2008). GMA was updated and expanded into ONCOR in 2008. A sample of tests was re-run in ONCOR and results were compared between algorithms to confirm compatibility of results. While the assignment probabilities found for particular individuals were not always exactly the same between the two versions of the algorithm, the assignment probabilities were found to be very consistent with differences generally being less than 1%.

Assignment testing under the "estimation" method was completed in GMA with baseline file of 1,000 individuals from each of four pure strains and 1,000 individuals from a single hybrid cross as the mixture file. A single test was performed for each of the 39 hybrid crosses. Individuals were considered unambiguously assigned to a pure strain

or hybrid category if their assignment score was greater than or equal to 0.80 (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008).

WhichRun.—WhichRun (Banks and Eichert 2000) calculates the likelihood of a given individual's multilocus genotype originating from each of two or more candidate populations (Pearse and Crandall 2004). Individuals are assigned to the population in which their likelihood of belonging is highest (Banks and Eichert 2000). Jackknifing is used to evaluate baseline data and assess the chance of correct allocation of individuals of unknown origin (Banks and Eichert 2000). A Log of Odds (LOD) score is calculated for each unknown individual. This LOD score is a ratio between the likelihoods of the most likely population and the second most likely population (Roques et al. 1999, Maudet et al. 2002; Cegelski et al. 2003). On this scale, a LOD of 1 is equivalent to an individual being 10 times more likely to occur in one population than another while a LOD of 2 is equivalent to an individual being 100 times more likely to occur in one population than another (Roques et al. 1999; Cegelski et al. 2003). WhichRun assumes populations are in HWE and LE and that an individual with an unknown origin is equally likely to have originated from each baseline population.

WhichRun assignment was performed on simulated baseline files containing 1,000 individuals from each of four pure strains. Each hybrid cross was tested with 1,000 simulated hybrid individuals designated as unknown individuals to be assigned to one of four baseline populations. This testing procedure was repeated 39 times, once for each hybrid cross. Pure strain self-assignment was tested by using the baseline input file as the unknown individual input file. Individuals were considered unambiguously assigned to a pure strain or hybrid category if their assignment score was greater than or equal to a

LOD 2. While this criterion was somewhat arbitrary, it represents a strong confidence (100 times more likely) in assignment to one strain over another.

Alternative Approaches to Assignment

Additional trials were used to determine if assignment success could be increased through additional steps or modifications of the approach. The first group of trials involved the relaxation of some of the initial run parameters for all genetic categories. First, GeneClass2 was performed with the number of simulated individuals increased from 100 to 100,000, but all other conditions (i.e., data inputs, assignment threshold of $p \geq 0.80$) of the run the same as in other trials. Second, assignments performed by WhichRun in the initial trials were analyzed with an assignment threshold of $LOD \geq 1.5$ relaxed from the initial $LOD \geq 2$. The new LOD score corresponds to an assignment being ~32 times more likely than the second most likely assignment. Significance ($p \leq 0.05$) of any observed increases in assignment success was calculated for trials in both algorithms with t-tests (SPSS v.14 2005). Normality of the data was tested with a Kolmogorov-Smirnov test while equal variance was tested with a Levene's test for equal variance. If the data did not meet assumptions of the parametric test (i.e., normality and equal variance), Mann-Whitney U-tests were used instead.

The second group of trials was conducted only for pure strain and F_1 populations. First, the assignment threshold in the three ML algorithms was reduced from a p-value of $p \geq 0.80$ to $p \geq 0.50$ ($LOD \geq 0.5$, or assignments ~ 3 times more likely, in WhichRun) and again reduced to a *pure assignment*. In pure assignment, each individual was accepted as belonging to the population deemed most likely by the algorithm, with no concern for

how much more likely the assignment was considered to be. Significant ($p \leq 0.05$) increases in assignment success was tested using ANOVAs comparing the three assignment thresholds (i.e., $p \geq 0.80$, $p \geq 0.50$, and pure assignment) and the LSD *post hoc* test. A Kolmogorov-Smirnov test was used to determine if the data met the assumption of normality while a Levene's test of equal variance was used to determine if the data had equal variance. If the data did not meet the assumptions for the ANOVA (i.e., normality and equal variance), a Kruskal-Wallis test was performed. If the Kruskal-Wallis test was found to be significant ($p \leq 0.05$), then a Mann-Whitney U-test was used to determine which groups differed significantly.

In the third set of trials, assignment was assessed by considering which two pure strain populations were the most likely origin populations for each F_1 hybrid individual based on a pure assignment threshold. This trial was only performed for the ML algorithms, and GeneClass2 was conducted without MCMC simulations. If the algorithm considered only a single pure strain to be likely for a given individual based on pure assignment, that individual was assumed to belong to that one pure strain. If the algorithm found two or more pure strains to be likely for an individual using pure assignment, the individual was assumed to be an F_1 hybrid of the two most likely populations. Assignment success for each of the three ML algorithms was qualitatively assessed based on the number of origin and non-origin strains each F_1 hybrid individual was assigned.

The fourth set of trials used a baseline file that contained both pure strain and F_1 hybrid individuals to determine if the algorithms would assign F_1 hybrid individuals to the correct hybrid category if given the option. All algorithms except NewHybrids were

tested in this trial with a baseline input file that contained 1,000 individuals from each of the four pure strains and 1,000 individuals from each of the six F₁ hybrid populations.

Input files of unknown individuals were the same as those used in the other trials.

Assignment thresholds used for these trials were pure assignment for the ML algorithms or $p \geq 0.50$ for Bayesian algorithms. Assignment success was measured as the percentage of individuals correctly assigned to their origin population.

Finally, a hierarchical approach was tested where outputs from F₁ trials in GMA were used as input files for NewHybrids. GMA was selected for the first step of this process because it had a very low error rate when assigning F₁ individuals, and had one of the highest rates of confident assignment to an origin strain. NewHybrids was selected for the second step of this process because it was the only algorithm that was able to correctly assign hybrid individuals to their true hybrid category, and did so with a high success rate when using the individuals' two origin strains as baseline files. For this method, the GMA output of a single hybrid cross was sorted based on the first and second most likely populations for a particular individual. Individuals were grouped with other individuals that had the same two most likely populations of origin. For example, if individual 1 assigned to GLW (most likely) and SLW (second most likely), it would be grouped with individual 2 that assigned to SLW (most likely) and GLW (second most likely). Individuals that were assigned to only one strain were grouped with any individual that shared the most likely or second most likely strain [i.e., individual 3 was assigned only to GLW and was grouped with individual 4 (GLW most likely, SMD second most likely) and also with individual 5 (LLW most likely, GLW second most likely)]. These new composite groups were input into the non-graphical version of

NewHybrids using Jeffrey's prior distribution at each locus, a 45,000 sweep burn-in, and 45,000 sweeps for posterior probability calculations. The prior for each run was allele frequencies of the two strains to which the group was assigned by GMA rather than the true strains to which the individuals belonged. Individuals were again considered unambiguously assigned to a pure strain or hybrid category if their assignment score was greater than or equal to 0.80 (Pierpaoli et al. 2003; Adams et al. 2007; Oliveira et al. 2008).

Bias Assessment

Assignment bias for pure strain and F_1 individuals was determined for each of the six algorithms. Output files from trials for each algorithm were analyzed to determine how many individuals were incorrectly assigned ($p < 0.80$) and to which incorrect strain the individual was assigned. Chi-square tests of association were used to determine if errors were randomly distributed across all potential categories or biased towards certain outcomes. For each algorithm, except for NewHybrids, the expected error for pure strain assignment was the total number of individuals from one strain incorrectly assigned divided by the three potential erroneous strains. For F_1 hybrid assignments, the expected error was the total number of individuals incorrectly assigned divided by the four pure strains. In NewHybrids, the expected error for pure strains was calculated as the total number of individuals erroneously assigned divided by the five erroneous assignment categories. Bias for F_1 individuals was calculated based on the type of strains included in the assignment prior: the three categories used were two origin strains, an origin and an incorrect strain, and two incorrect strains. When assignment was to two origin strains,

expected error for F_1 individuals was calculated as the total number of individuals erroneously assigned divided by the five error categories. For the other two categories, the expected error was calculated as the total number of individuals erroneously assigned divided by the six error categories.

RESULTS

Population Differentiation

Generally, pairwise F_{st} values decreased as comparisons included populations with an advanced level of hybridization (Table 1.4). All pairwise comparisons between two pure strains met or exceeded a threshold of 0.03. However, the percentage of comparisons below this threshold increased up to 48.8% when the pairwise comparison included a strain with increasingly advanced hybrid ancestry. In addition, the minimum F_{st} , maximum F_{st} , and mean F_{st} all decreased as increasingly hybridized populations were compared to pure strains. A similar pattern was observed when pairwise F_{st} comparisons were within a single hybrid category. Comparisons with the highest level of population differentiation tended to occur among two pure strains whereas the lowest level of population differentiation occurred among two F_2 strains. Additionally, the percentage of comparisons below an F_{st} of 0.03 increased dramatically as the level of hybridization increased.

Self Assignment/Pure Strain Assignment

Most (56.0% in GeneClass2 to 83.7% in WhichRun) pure strain individuals were ambiguously assigned by ML algorithms (Table 1.5), which confidently assigned fewer than half (range: 8.9% in GeneClass2 with 100 simulated individuals to 40.0% in GeneClass2 with no simulated individuals) of pure strain individuals. An undetectable or low to moderate error rate (range: 0.0% in WhichRun to 35.1% in GeneClass2) was observed.

Bayesian methods were less confident in assignment than ML algorithms. Clusters formed by BAPS and STRUCTURE tended to include individuals of multiple ancestries instead of a single ancestry. Assuming the cluster with the highest number of individuals from a particular strain was the true cluster, 52.9% and 33.8% of pure strain individuals were correctly assigned by BAPS and STRUCTURE, respectively. Error rates for these two algorithms ranged from 1.9% in STRUCTURE to 11.2% in BAPS, and ambiguity ranged from 36.2% in BAPS to 61.2% in STRUCTURE.

NewHybrids confidently assigned a high percentage (99.6%) of pure strain individuals to their strain of origin for three of four strains (i.e., GLW, LLW, SMD) when the prior included the true origin strain and an unknown strain (Table 1.6). No pure strain individuals were incorrectly assigned when using this type of self-assignment. The SLW strain could not be assigned because of an error within the program that could not be resolved. When using prior information that included either the correct strain and an incorrect strain or two incorrect strains, the percentage of confident assignments decreased (81.2% and 77.7%, respectively). In the first of these two cases, 100% of individuals confidently assigned were correctly assigned to their origin strain, whereas 100% of individuals confidently assigned were in error with two non-origin strains in the prior. Additionally, ambiguous assignment rates increased as the number of non-origin strains in the prior increased up to the maximum of 2 (range: 0.4% to 22.3%).

F₁ Assignment

Assignment of F₁ hybrids showed lower confidence, higher error rates, and higher ambiguity than pure strain assignments. None of the ML algorithms were able to assign

F₁ hybrid individuals to their two origin strains. In some cases, algorithms assigned F₁ individuals to one of the individual's two origin strains (range = 4.3% in WhichRun to 46.2% in GMA). Assignment error rate was low (0.0% in WhichRun to 13.7% in GeneClass2) for ML algorithms with few F₁ individuals being assigned to non-origin pure strains. Most assignments were ambiguous (53.1% in GMA to 95.7% in WhichRun).

Of the Bayesian methods, only NewHybrids successfully assigned F₁ individuals to their two origin strains (Tables 1.6). NewHybrids confidently assigned over 80% of all F₁ individuals no matter what strains were used in the prior, but error rates jumped from 0.0% when two origin strains were used to 100% if at least one of two strains used in the prior was a non-origin strain. When the prior contained an origin and non-origin strain, 40.8% of confidently assigned individuals were assigned to their true origin strain.

Compared to NewHybrids, BAPS and STRUCTURE performed poorly when assigning individuals of hybrid origin (Table 1.5). For F₁ hybrid individuals, confident assignment ranged from 7.0% in STRUCTURE to 39.6% in BAPS. Assignment to one of two origin strains ranged from 6.1% in STRUCTURE to 23.8% in BAPS. Error rate was 15.9% for BAPS and 0.9% for STRUCTURE but ambiguous assignment rates were 60.4% for BAPS and 93.0% for STRUCTURE.

B_x Assignment

ML algorithms were unable to confidently assign a high number of B_x individuals to their correct origin (Table 1.5). Confident assignment ranged from 8.1% in WhichRun to 69.9% in GMA, but none of these assignments were to a B_x category. Instead, B_x

individuals were assigned to pure strains, with some individuals assigned to one origin strain (range: 8.1% in WhichRun to 64.0% in GMA). Error rates ranged from 0.0% in WhichRun to 28.3% in GeneClass2 when run with no MCMC simulations. Ambiguous assignment of B_x individuals ranged from 30.1% in GMA to 91.9% in WhichRun.

NewHybrids was able to confidently assign most B_x individuals no matter what strains were used in the prior, but accuracy of assignments varied substantially. When two origin strains were used in the prior, 94.8% of B_x individuals were confidently assigned, of which 100% were correctly assigned. Confident assignment when at least one non-origin strain was used in the prior was 88.2% with one non-origin strain and 63.4% with two non-origin strains, and in each case, 100% of assignments were incorrect. A large percentage (84.6%) of erroneous assignments was to an origin strain if one was included in the prior. Ambiguous assignments ranged from 5.2% when two origin strains were used as the prior to 36.6% when two non-origin strains were used as the prior.

For BAPS, fewer than half (48.3%) of all B_x individuals were confidently assigned and only 19.9% were assigned to an origin strain. The error rate in BAPS was 28.5%, but represented over half (58.9%) of confident assignments. The remaining 51.7% were ambiguously assigned by BAPS.

F₂ Assignment

For ML algorithms, over half of all F_2 individuals were ambiguously assigned (range: 58.8% in GMA to 95.4% in WhichRun; Table 1.5). Confident assignment ranged

from 4.6% in WhichRun to 41.2% in GMA. Error ranged from 0.0% in WhichRun to 7.7% in GeneClass2 when run with 100 simulated individuals.

Assignment success in NewHybrids depended upon the number of origin strains included in the prior (Table 1.6). Confident assignment ranged from 67.5% when the prior included two origin strains to 90.2% when two non-origin strains were used in the prior. Similar to previous trials, 100% of individuals were erroneously assigned if at least one of the two strains in the prior was not from the individual's genetic ancestry. A high percentage (71.1%) of F₂ individuals was erroneously assigned when two origin strains were used in the prior, and some individuals were assigned as a pure origin strain when at least one origin strain was used in the prior. Pure strain assignments were 13.5% with two origin strains and 82.7% with one origin strain in the prior. Ambiguous assignments ranged from 9.8% with two non-origin strains in the prior to 32.5% with two origin strains in the prior.

For BAPS only 40.0% of F₂ individuals were confidently assigned to their strains of origin (Table 1.6). Confident assignment was 26.6% and error was 13.4%. The remaining 60.0% of F₂ hybrid individuals were ambiguously assigned by BAPS.

Alternative approaches to assignment

Assignment success did not increase when input settings were relaxed for algorithms. First, in GeneClass2 increasing the number of simulated individuals used for assignment from 100 to 100,000 did not significantly change the number of individuals confidently assigned or the error rate (Table 1.7). Second, in WhichRun relaxation of the assignment threshold from $\text{LOD} \geq 2$ (assignment 100 times more likely than the second

most likely assignment) to $\text{LOD} \geq 1.5$ (assignment ~ 32 times likely than the second most likely assignment) resulted in significantly more hybrid individuals being assigned as F_1 and F_2 individuals (9.6% for F_1 , 10.0% for F_2) (Table 1.8), but error rates in each hybrid category significantly increased with the relaxation of the assignment threshold (range: 0.2% to 0.3%). Third, assignment thresholds were relaxed in each of the ML algorithms, first to ≥ 0.50 ($\text{LOD} \geq 0.5$ in WhichRun) and then to acceptance of the “most likely population” as the individual’s origin. In the first step, no ML algorithm correctly assigned significantly more individuals, but the error rate increased significantly in WhichRun (Table 1.9) to 5.1%. In the second step, only WhichRun correctly assigned significantly more pure strain individuals (74.8%), but each ML algorithm had a significant increase in the error rate (range: 25.2% in WhichRun to 26.9% in GMA). Fourth, assignment was made by accepting the two most likely populations as the individual’s origin (Table 1.10). With this method, GMA assigned 71.6% of F_1 individuals to only origin strains while GeneClass2 and WhichRun assigned 35.5% and 36.6% of F_1 hybrid individuals to only origin strains. Fifth, F_1 populations were included with pure strain populations in the baseline file, and assignment of pure strain and F_1 individuals was based on the “most likely population” (Table 1.11). This method had lower assignment success than trials in which baseline files contained only four pure strains. Success in pure strain assignment ranged from 42.8% in GMA to 55.3% in both GeneClass2 and WhichRun whereas success in F_1 individual assignment ranged from 23.7% in WhichRun to 30.3% in GMA. A similar method was attempted in BAPS and STRUCTURE with an assignment threshold of ≥ 0.50 (Table 1.12). BAPS was able to confidently assign 82.6% of pure strain individuals, compared to 13.5% confident

assignment in STRUCTURE, but these assignments were scattered throughout the ten populations included in the prior, and each cluster contained individuals from multiple pure strains. Assignment success of the F₁ individuals was lower than for pure strain individuals when using pure strain and F₁ populations in the baseline. BAPS confidently assigned 69.5% of F₁ individuals compared to 9.0% assigned in STRUCTURE. Again, assignments were scattered throughout baseline populations, and clusters formed by algorithms included individuals from multiple hybrid populations. Finally, the hierarchical approach, where pure assignment output from GMA was used as an unknown input file in NewHybrids, also had limited success (Table 1.13). When GMA correctly selected two origin strains for an F₁ individual, then NewHybrids correctly assigned 62.6% of F₁ hybrid individuals with a 0.0% error rate. However, when at least one non-origin strain was selected by GMA as one of the two most likely origin populations, then NewHybrids correctly assigned 0.0% and incorrectly assigned 67.5% of the F₁ hybrid individuals.

Bias Assessment

Each algorithm was biased and the magnitude of bias depended upon the genetic category and origin strains of individuals being assigned (Tables 1.14 and 1.15). GeneClass2 was unbiased in assignment of pure strain individuals, but each F₁ population was biased towards one origin strain. GMA was unbiased for pure strains, but assignment of F₁ populations was biased to both origin strains. WhichRun was biased in the assignment of the pure strains and each F₁ population. For F₁ populations, bias was always towards one of two origin strains. BAPS was strongly biased in the assignment of

both pure strain and F_1 populations, but was not consistently biased towards hybrid origins. STRUCTURE was biased for two pure strain populations and each F_1 population. NewHybrids was biased when the assignment prior contained one origin strain and one incorrect strain, or two incorrect strains (Table 1.16), and for each F_1 hybrid assignment.

DISCUSSION

Overall, no assignment method performed well enough to determine which lake trout hatchery strains were reproducing under natural conditions. BAPS and STRUCTURE did not assign pure strain or hybrid categories to their origin strains, perhaps because F_{st} values were low among pure strain and hybrid populations. For example BAPS and STRUCTURE begin to lose their assignment abilities at $F_{st} \leq 0.03$ (Latch et al. 2006). Below this point, neither algorithm was able to discern a clear pattern in the data, yet displayed false certainty about the number of populations present and confidence of individual assignment (Latch et al. 2006). In contrast, if F_{st} values reached 0.39, both BAPS and STRUCTURE were able to correctly assign 97% of individuals in the study (Latch et al. 2006). If populations that are below F_{st} values of 0.03 are considered to be indistinguishable from each other by BAPS and STRUCTURE (Latch et al. 2006), then 1/3 to 1/2 of populations in each hybrid category could not be discerned from pure strain populations.

When F_1 hybrid populations were included in the baseline file, MCMC can be problematic. For example, MCMC methods have a difficult time effectively exploring the complex topologies that occur when high numbers of populations are assessed (Mark and Avise 2004). Further, the MCMC may not have mixed effectively, which could cause it to become stuck at a likelihood maximum resulting in a poor estimate of probabilities (Choisy et al. 2004). Choisy et al. (2004) found this to occur in ~10% of cases when using MCMC, and observed this problem to occur more frequently when a high number of loci or a large sample size were used. The sample sizes used for each population in this simulated study were much larger than would realistically be used in

empirical studies because of economic and time constraints, as well as the feasibility and pragmatism of sampling populations of this size. The issue of poor mixing could have been addressed by increasing the number of iterations for each run, but as MCMC runs take the most time of all methods tested this may not be a feasible option.

BAPS and STRUCTURE were consistently biased in assignment for both pure strain and F_1 hybrid individuals. For pure strains, BAPS was strongly biased towards the strain that was least differentiated from the strain under assignment. In only one case (SLW biased towards SMD) was BAPS biased towards the most differentiated strain instead of the least differentiated strain. In this case, population differentiation was relatively high for each comparison (F_{st} ranged from 0.116 to 0.129). For F_1 hybrid populations, BAPS was biased toward incorrect strains in half of the trials, which makes BAPS a problematic method for assignment of individuals of unknown origin because managers could mistakenly stock strains that are not contributing to reproduction. STRUCTURE only showed bias for strains with an F_{st} score that was borderline for assignment success (GLW, GMD: $F_{st} = 0.03$). Because performance of STRUCTURE degrades when F_{st} scores are low, STRUCTURE would not be a suitable method to choose for this research and management issue.

NewHybrids successfully assigned pure strain individuals to their true origin when the baseline prior file included the origin strain. However, when the true origin strain was not included in baseline prior file, NewHybrids confidently but erroneously assigned a large percentage of pure strain individuals. Previous studies have suggested that incorrect *a priori* population groupings can diminish the power of assignment tests to elucidate biological reality, which could lead to unsuitable conservation or management

strategies (Pearse and Crandall 2004; Latch et al. 2006). A similar pattern was observed with the assignment of F_1 and B_x hybrid populations. In each case, confident assignment rates exceeded 90% and no error occurred when two origin strains were used for assignment, but as the number of non-origin strains in the baseline prior file increased, confident assignments remained high but error increased to 100%. When F_2 hybrid populations were considered, the error rate remained high no matter how many origin strains were included in the baseline prior file. This was expected because genetic mixing prior to the generation of F_2 hybrid individuals created a complex situation with a large number of possible crosses and high genetic similarity among them (Marsden et al. 1989).

In general, NewHybrids worked well when origin strains were included in the baseline prior file. However, four pure strains are currently used for stocking and NewHybrids is limited to two populations in the baseline prior file. If NewHybrids is to be used successfully there needs to be some method of confidently and correctly narrowing down the strains that will be used as a prior to those that have a strong likelihood of being the ancestor strains of individuals under consideration. This suggests that NewHybrids has potential to be used in the second step of a two-step hierarchical assignment process. An important benefit of using a hierarchical method of assignment testing is increased confidence of assignment resulting from concordance among tests (Cegelski et al. 2003). Past studies have cautioned against basing conclusions and management recommendations on a single assignment test (Manuel et al. 2002; Cegelski et al. 2003; Latch et al. 2006). Therefore, a hierarchical assignment method would help quell concerns about the accuracy of an individual assignment.

NewHybrids was biased when reference files for assignment of hybrid individuals included incorrect strains. When one reference population was a parental strain and the second reference file was incorrect, NewHybrids tended to assign individuals as either the parental strain or as an incorrect F₁ hybrid. In both cases, the manager would continue stocking at least one correct and reproductive strain. But when two incorrect strains were included as reference populations in the prior, assignment tended to be biased towards advanced hybrid categories. This could lead managers to falsely think that recruitment to adulthood was occurring in small numbers. If this were the case, success of reintroduction could be overestimated.

GeneClass2 was consistently average for assigning individuals to their strains of origin. The threshold assignment method that did not use MCMC simulations tended to have higher correct assignment rates and lower error rates than the simulation-exclusion method when run with 100 simulated individuals, but these rates were not high enough to consider using GeneClass2 with the current suite of microsatellite markers and wild-origin individuals. Even when the number of simulated individuals was increased from 100 to 100,000, assignment rates did not significantly change. Unlike Eldridge et al. (2001), who found that GeneClass (Cornuet et al. 1999) performed well in the absence of a significant assignment threshold, relaxing the assignment threshold from $p \geq 0.80$ to pure assignment in my study did not increase assignment success, but rather increased the error rate of the algorithm. While confident assignment is important in research and forensics, management decisions about stocking may require determining the origin of as many individuals as possible through pure assignment rather than fewer individuals with a higher assignment confidence (Hauser et al. 2006). Assignment accuracy could

potentially be improved if the number of microsatellite markers were increased from the seven currently in use (Eldridge et al. 2001; Berry et al. 2004; Pearse and Crandall 2004). GeneClass2 was not biased in the assignment of pure strain individuals, though assignment of F₁ hybrid individuals was strongly biased towards accurate parental strains. This could be important if assignment accuracy of the method increases in the future.

While GMA had moderate rates of confident assignment for each genetic category, the error rate was promising. The highest error rate was 5.9% for B_x hybrid individuals and in each other case was less than 2%. Pure assignment methods had high assignment success for pure strain individuals (73.1%), but the error rate of 30% limits the method's usefulness for fishery managers. For F₁ hybrid individuals, the method of accepting up to two populations that were found to be the most likely origin for that individual had promise, but again, the high error rate found with the current suite of microsatellite loci was unacceptable for the basis of management decisions. With this method, GMA correctly assigned 60.6% of the F₁ hybrid individuals to their two origin strains, and an additional 11% were assigned as one of the two origin pure strains. While 11% of the assignments are not completely accurate, they should also not be considered in complete error. A fisheries manager faced with these assignments would likely continue stocking a pure strain that was contributing to reproduction. Overall, GMA has potential to be used as the first step in a hierarchical approach, but rates of correct, confident assignments would need to improve. Increasing the number of microsatellite loci used for assignment could help to improve assignment rates in GMA (Eldridge et al. 2001; Berry et al. 2004; Pearse and Crandall 2004). Like GeneClass2, GMA was not biased in pure strain assignment, but for F₁ hybrid assignments, GMA was biased toward

the two parental strains for the individual under consideration. This suggests that if a method is found to decrease the rate of ambiguous assignments in GMA, then assignment could be based on strains found to be the first and second most likely strains of origin.

Similar to GMA, WhichRun also displayed very low error rates but only moderate assignment success. For all genetic categories, the error rate of WhichRun was 0%, while assignment success ranged from 4% to 16%. This success rate was too low to be useful for management decisions. Relaxing the assignment threshold from $\text{LOD} \geq 2$ to $\text{LOD} \geq 1.5$ significantly increased assignment success of F_1 and F_2 hybrid individuals, but not enough to be acceptable. While the error rate for each hybrid category under the relaxed assignment threshold remained under 1%, this increase was significant because of the few number of individuals confidently assigned to any origin. Further relaxation to pure assignment from the higher LOD thresholds significantly increased assignment success of pure strain individuals to 74.8%, but error increased to 25.2%. Overall, WhichRun has potential to be used in a hierarchical approach, but assignment success would need to increase dramatically. This could potentially be accomplished through the use of additional loci (Eldridge et al. 2001; Berry et al. 2004; Pearse and Crandall 2004). WhichRun was significantly biased for the assignment of GLW and SMD pure strains. But considering the low number of erroneously assigned individuals during assignment of these two pure strains (3 and 11 individuals respectively) this bias does not seem to be a large issue. In the assignment of F_1 hybrid individuals, WhichRun was biased to either one or two of the individual's parental strains. Based on these assignments, reproductively successful strains would likely be continued to be stocked.

The success found in this study for assigning pure strain lake trout to their origin strain was much lower than the success found by Page et al. (2003) and DeKoning et al. (2006). In these two studies, assignment success rates ranged from 76.5% to 97.6% and 92.7% to 99.8%, respectively. One possible reason for this lack of reproducibility could be the result of violations in the assumptions of linkage equilibrium and Hardy-Weinberg equilibrium in the empirical data. In the case of linkage, it may have been slight enough to be statistically insignificant, but still strong enough to cause a violation in the assumption of linkage equilibrium. By using a simulation approach, in which each individual's alleles were randomly selected from the allelic distribution, the potential linkage would have been removed from the data, which could have caused a reduction in assignment success. Similarly, the empirical populations may have been out of HWE by a statistically insignificant amount, and therefore not actually a randomly mating population. Within the hatchery populations studied by Page et al. (2003) this could result from a subset of individuals spawning successfully (Allendorf and Phelps 1980). This would cause individuals from within a single hatchery to be slightly more similar than would be expected among the hatchery populations. Again, using simulated populations that were created randomly from the allelic frequencies of each hatchery strain would disrupt any slight inbreeding that could be occurring, and bring the populations back into accordance with the assumption of HWE.

Use of a simulated population to test algorithms has two potential drawbacks. First, simulation is based on allele frequencies of source populations, which are an estimate (Davies et al. 1999; Roques et al. 1999). Therefore, simulated genotypes are not exactly the same as source populations (Roques et al. 1999). Second, success in

assigning simulated individuals to their origins is likely to be overestimated (Roques et al. 1999). For that reason, performance of assignment methods in this study occurred under ideal conditions, and could be lower with empirical data where origin of each individual is not known (Roques et al. 1999). Overall, benefits of using a simulated population of individuals for testing performance of each assignment method outweigh the drawbacks. Fisheries managers will need to have confidence in the method chosen and the results obtained by that method. Further, fisheries managers will need to know the strengths and weaknesses of each assignment method, which can only be obtained through simulations and evaluation of assignments for individuals with known origins (Hansen et al. 2001; Cegelski et al. 2003)

Currently, low levels of genetic differentiation among pure strain populations and hybrid populations seem to be the largest impediment to increased assignment success. Maudet et al. (2002) found that population differentiation levels among seven ibex (*Capra ibex*) populations affected performance of assignment tests. Further, assignment success between two highly differentiated populations was higher than assignment success between two weakly differentiated populations (Maudet et al. 2002). Additional studies, both empirical and simulated, (Waser and Strobeck 1998; Cornuet et al. 1999; Waits et al. 2000; Berry et al. 2004; Mank and Avise 2004; McLoughlin et al. 2004; Paetkau et al. 2004) have documented a rapid drop in assignment success where F_{st} scores fall below 0.05 (Manel et al. 2005; Hauser et al. 2006). At low levels of genetic differentiation, Bayesian clustering methods are especially susceptible to error (Pearse and Crandall 2004; Manel et al. 2005; Latch et al. 2006). Bayesian methods detect the number of populations present in an unknown sample by determining the number of

populations that minimize Hardy-Weinberg disequilibrium (HWD) and linkage disequilibrium (LD) (Latch et al. 2006). As levels of genetic differentiation among populations decrease, rates of HWD and LD also decrease (Latch et al. 2006).

Others (Paetkau et al. 1995; Smouse and Chevillon 1998; Roques et al. 1999) have suggested that assignment success is expected to increase as a function of population divergence (Bernatchez and Duchesne 2000; Manel et al. 2005). One study even concluded that F_{st} strongly influenced the percentage of individuals correctly assigned by assignment methods (Manel et al. 2002). For example, the percentage of New Zealand's grand skink (*Oligosoma grande*) correctly assigned by both GeneClass and STRUCTURE increased as the populations became more genetically differentiated (Berry et al. 2004). Maximum assignment accuracy was found for the grand skink between F_{st} values of 0.06 – 0.08 (Berry et al. 2004). A slightly wider range of F_{st} values (0.05 – 0.10) has been suggested as necessary for reasonable assignment success (Cornuet and Luikart 1996; Manel et al. 2002; Berry et al. 2004; Hauser et al. 2006).

Low levels of population differentiation among simulated pure strain and hybrid lake trout populations could potentially be counteracted by larger sample sizes or use of additional genetic loci (Cornuet et al. 1999; Hansen et al. 2001; Manel et al. 2005; Hauser et al. 2006). Using larger sample sizes to increase F_{st} values among lake trout populations would not be realistic for future studies for two reasons: (1) sample sizes of simulated populations used to test the algorithms were large, and an empirical study would not likely have the necessary resources to conduct a study of such a magnitude; (2) limited lake trout reproduction on Lake Michigan's mid-lake reef complex (Janssen et al. 2006) limits the number of wild-born offspring that could be collected for sampling.

Therefore, instead of using a larger sample size to increase population differentiation, increasing the number of lake trout specific microsatellite markers available for future studies is more reasonable (Pearse and Crandall 2004).

The number of additional microsatellite markers needed for future studies depends on the desired ability to distinguish advanced hybrid categories (Davies et al. 1999). The ability to distinguish B_x and F_2 hybrid individuals is moot because of limited reproduction and recruitment that is occurring under natural conditions (Holey et al. 1995). For these hybrid categories to occur, a generation of F_1 hybrid individuals would first need to survive to adulthood and then to spawn. After F_1 hybrids are produced determining the accurate hybrid category of every individual collected from the mid-lake reef complex is less important because achievement of a self-sustaining population of lake trout would be well on its way. If assigning wild-born individuals to hybrid categories was desired for purposes of monitoring, 50 - 70 loci would be needed to have enough discriminatory power to discern advanced backcrossed individuals from pure strain individuals (Boecklen and Howard 1997; Davies et al. 1999; Vähä and Primmer 2006; Gunnell et al. 2007).

Of greater importance, and a more pragmatic concern, is the ability to distinguish among pure strain and F_1 hybrid individuals to direct future stocking efforts. First generation hybrid individuals could potentially be identified and discriminated from pure strain individuals with a total of 12 – 24 loci and pairwise F_{st} values of 0.12 – 0.21 (Vähä and Primmer 2006; Barilani et al. 2007). Additionally, using loci with a high number of alleles could increase assignment success even with low F_{st} values (Hedrick 1999;

Nielsen et al. 2001). Loci with between 6 and 10 alleles have been suggested as being sufficient for increased population assignment success (Bernatchez and Duchesne 2000).

Selecting genetic markers with high heterozygosity (H) could also increase assignment success of the algorithms (Blott et al. 1999; Hansen et al. 2001). Use of highly polymorphic loci ($H = 0.6 - 0.8$) has been suggested as a means for increasing assignment success when F_{st} values are higher than 0.1 (Manel et al. 2002; Cegelski et al. 2003). The four pure strain populations considered in this study had observed heterozygosities that ranged from 0.402 to 0.515 (Page et al. 2003). Unfortunately, the gain in assignment success may not be found for lake trout by just using more highly polymorphic loci. Manel et al. (2002) found that using highly polymorphic loci instead of loci with lower polymorphism ($H = 0.3 - 0.5$) had little effect on assignment at low levels of population differentiation, such as those found in this study (Manel et al. 2002).

Future efforts should develop additional lake trout specific genetic markers. Once these markers are developed, performance of algorithms considered in this study should be reassessed with the new suite of microsatellite markers. This assessment would need to include simulated pure strain and hybrid populations unless individuals from pure strains and each potential F_1 hybrid cross could be procured from the hatcheries. If F_{st} values among populations are above 0.03, then Bayesian methods of BAPS and STRUCTURE should be reassessed. If possible, a hierarchical method should be explored further. Methods that are most promising for the first step in a hierarchical method are GMA/ONCOR and WhichRun, both of which had very low error rates but limited assignment success. These algorithms could potentially be used to reduce or eliminate some strains from further consideration in step two of the hierarchical

approach. In step two, NewHybrids should be used to assign individuals to their pure strain or hybrid category. NewHybrids should also be further tested by assessing assignment success for hybrid individuals when one correct strains and one unknown strain are included in the baseline prior file. An added benefit of the hierarchical approach is it bases assignment on results of two assignment tests, and comparing results between two assignments could add additional confidence to individual assignment. A hierarchical approach would also take advantage of using both a maximum likelihood and a Bayesian approach. If a method is found that is both accurate and unbiased, then genetic analysis could be conducted for lake trout eggs and fry collected on Lake Michigan's mid-lake reef complex and the method with the highest performance capabilities could be used to assign wild-born offspring to their strains of origin.

LITERATURE CITED

- Adams, J.R., C. Lucash, L. Schutte, and L.P. Waits. 2007. Locating hybrid individuals in the red wolf (*Canis rufus*) experimental population area using a spatially targeted sampling strategy and faecal DNA genotyping. *Molecular Ecology* 16:1823 – 1834.
- Allendorf, F.W. and S.R. Phelps. 1980. Loss of genetic variation in a hatchery stock of cutthroat trout. *Transactions of the American Fisheries Society* 109:537 – 543.
- Anderson, E.C. and E.A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160:1217 – 1229.
- Balloux, F. and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11: 155 – 168.
- Banks, M.A., and W. Eichert. 2000. WHICHRUN (version 3.2) a computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* 91:87 – 89.
- Barilani, M., A. Sfougaris, A. Giannakopoulos, D. Mucci, C. Tabarrone, and E. Randi. 2007. Detecting introgressive hybridization in rock partridge populations (*Alectoris graeca*) in Greece through Bayesian admixture analyses of multilocus genotypes. *Conservation Genetics* 8:343 – 354.
- Baudouin, L., S. Piry, and J.M. Cornuet. 2004. Analytical Bayesian approach for assigning individuals to populations. *Journal of Heredity* 95:217 – 224.
- Beaumont, M.A. and B. Rannala. 2004. The Bayesian revolution in genetics. *Genetics* 5:251 – 261.
- Bernatchez, L. and P. Duchesne. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Science* 57: 1 – 12.
- Berry, O., S.M. Tocher, S.D. Sarre. 2004. Can assignment tests measure dispersal? *Molecular Ecology* 13:551 – 561.
- Blott, S.C., J.L. Williams, and C.S. Haley. 1999. Discriminating among cattle breeds using genetic markers. *Heredity* 82:613 – 619.
- Boecklen, W.J. and D.J. Howard. 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. *Ecology* 78:2611 – 2616.
- Bruford, M.W. and R.K. Wayne. 1993. Microsatellites and their application to population genetic studies. *Genomes and Evolution* 3:939 – 942.

- Cegelski, C.C., L.P. Waits, and N.J. Anderson. 2003. Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology* 12:2907 – 2918.
- Choisy, M., P. Franck, and J.-M. Cornuet. 2004. Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Molecular Ecology* 13:955 – 968.
- Corander, J. and P. Marttinen. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* 15:2833 – 2843.
- Cornuet, J.-M. and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001 – 2014.
- Cornuet, J.-M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153:1989 – 2000.
- Cowles, M.K. and B.P. Carlin. 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91:883 – 904.
- Davies, N., F.X. Villablanca, and G.K. Roderick. 1999. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution* 14:17 – 21.
- DeKoning, J., K. Keatley, R. Phillips, J. Rhyudderch, J. Janssen, and M. Noakes. 2006. Genetic analysis of wild lake trout embryos recovered from Lake Michigan. *Transactions of the American Fisheries Society* 135:399 – 407.
- Efron, B. 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association* 78:316 – 331.
- Eldridge, M.D.B., J.E. Kinnear, and M.L. Onus. 2001. Source population of dispersing rock-wallabies (*Petrogale lateralis*) identified by assignment tests on multilocus genotypic data. *Molecular Ecology* 10:2867 – 2876.
- Eschmeyer, P.H. 1957. The near extinction of lake trout in Lake Michigan. *Transactions of the American Fisheries Society* 85:102 – 119.
- Excoffier, L., G.Laval, and S. Schneider. 2005. ARLEQUIN (version 3.1): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47 – 50.

- Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567 – 1587.
- Falush, D., M. Stephens, and J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7:574 – 578.
- Fetterolf, C.M. Jr. 1980. Why a Great Lakes Fishery Commission and why a sea lamprey international symposium. *Canadian Journal of Fisheries and Aquatic Sciences* 37:1588 – 1593.
- Goldstein, D.B. and D.D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 88: 335 – 342.
- Guinand, B., K.T. Scribner, K.S. Page, K. Filcek, L. Main, and M.K. Burnham-Curtis. 2006. Effects of coancestry on accuracy of individual assignments to population of origin: examples using Great Lakes lake trout (*Salvelinus namaycush*). *Genetica* 127:329 – 340.
- Guinand, B., K.T. Scribner, A. Topchy, K.S. Page, W. Punch, M.K. Burnham-Curtis. 2004. Sampling issues affecting accuracy of likelihood-based classification using genetical data. *Environmental Biology of Fishes* 69:245 – 259.
- Gunnell, K., M.K. Tada, F.A. Hawthorne, E.R. Keeley, and M.B. Ptacek. 2007. Geographic patterns of introgressive hybridization between native Yellowstone cutthroat trout (*Oncorhynchus clarkia bouvieri*) and introduced rainbow trout (*O. mykiss*) in the South Fork of the Snake River watershed, Idaho. *Conservation Genetics* 9:49 – 64.
- Hansen, M.M., E. Kenchington, and E.E. Nielsen. 2001. Assigning individual fish to populations using microsatellite DNA markers. *Fish and Fisheries* 2:93 – 112.
- Hauser, L., T.R. Seamons, M. Dauer, K.A. Naish, and T.P. Quinn. 2006. An empirical verification of population assignment methods by marking and parentage data: hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. *Molecular Ecology* 15:3157 – 3173.
- Hedrick, P.W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313 – 318.
- Hedrick, P.W. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633 – 1638.

- Holey, M.E., R.W. Rybicki, G.W. Eck, E.H. Brown Jr., J.E. Marsden, D.S.Lavis, M.L. Toneys, T.N. Trudeau, and R.M. Horrall. 1995. Progress toward lake trout restoration in Lake Michigan. *Journal of Great Lakes Research* 21 (Supplement 1):128 – 151.
- Janssen, J., D.J. Jude, T.A. Edsall, R.W. Paddock, N. Wattrus, M. Toneys and P. McKee. Evidence of lake trout reproduction at Lake Michigan's mid-lake reef complex. *Journal of Great Lakes Research* 32:749 – 763.
- Kalinowski, S.T. 2003. Genetic Mixture Analysis 1.0. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from <http://www.montana.edu/kalinowski>.
- Kalinowski, S.T., K.R. Manlove, and M.L. Taper. 2008. ONCOR: a computer program for genetic stock identification. Department of Ecology. Montana State University. Available: <http://www.montana.edu/kalinowski/Software/ONCOR.htm>.
- Latch, E.K., G. Dharmarajan, J.C. Glaubitz, and O.E. Rhodes Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* 7:295 – 302.
- LMLTTC (Lake Michigan Lake Trout Technical Committee). 1985. A draft lakewide plan for lake trout restoration in Lake Michigan. In *Minutes of the Lake Michigan Committee, 1985 annual meeting*, Great Lakes Fisheries Commission, Ann Arbor, Michigan.
- Luikard, G. and P.R. England. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* 14:253 – 256.
- Manel, S., P. Berthier, and G. Luikart. 2002. Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology* 16:650 – 659.
- Manel, S., O.E. Gaggiotti, and R.S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution* 20:136 – 142.
- Mank, J.E. and J.C. Avise. 2004. Individual organisms as units of analysis: Bayesian-clustering alternatives in population genetics. *Genetics Research* 84:135 – 143.
- Manoukis, N.C. 2007. FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. *Molecular Ecology Notes* 7:592 – 593.

- Marsden, J.E., C.C. Krueger, and B. May. 1989. Identification of parental origins of naturally produced lake trout in Lake Ontario: applications of mixed-stock analysis to a second generation. *North American Journal of Fisheries Management* 9:257 – 268.
- Maudet, C., C. Miller, B. Bassano, C. Breitenmoser-Würsten, D. Gauthier, G. Obexer-Ruff, J. Michallet, P. Taberlet, and G. Luikart. 2002. Microsatellite DNA and recent statistical methods in wildlife conservation management: applications in Alpine ibex [*Capra ibex (ibex)*]. *Molecular Ecology* 11:421 – 436.
- McLoughlin, P.D., D. Paetkau, M. Duda, and S. Boutin. 2004. Genetic diversity and relatedness of boreal caribou populations in western Canada. *Biological Conservation* 118:593 – 598.
- Moljonen, M.-L., J.J. Pella, and M. Masuda. 2005. Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries and Aquatic Sciences* 62:2143 – 2158.
- Myung, I.J. 2003. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47:90 – 100.
- Neigel, J.E. 2002. Is F_{st} obsolete? *Conservation Genetics* 3: 167 – 173.
- Nielsen, E.E., L.A. Bach, and P. Kotlicki. 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* 6:971 – 973.
- Nielsen, E.E., M.M. Hansen, and L.A. Bach. 2001. Looking for a needle in a haystack: discovery of indigenous Atlantic salmon (*Salmo salar* L.) in stocked populations. *Conservation Genetics* 2:219 – 232.
- Oliveira, R., R. Godinho, E. Randi, N. Ferrand, and P.C. Alves. 2008. Molecular analysis of hybridization between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conservation Genetics* 9:1 – 11.
- Paetkau, D., W. Calvert, I. Stirling, and C. Strobeck. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4: 347 – 354.
- Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13:55 – 65.
- Page, K.S., K.T. Scribner, K.R. Bennett, L.M. Garzel, and M.K. Burnham-Curtis. 2003. Genetic assessment of strain-specific sources of lake trout recruitment in the Great Lakes. *Transactions of the American Fisheries Society* 132:877 – 894.

- Park, S.D.E. 2001. Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis. University of Dublin
- Pearse, D.E. and K.A. Crandall. 2004. Beyond F_{st} : analysis of population genetic data for conservation. *Conservation Genetics* 5:585 – 602.
- Pella, J. and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fisheries Bulletin* 99:151 – 167.
- Pierpaoli, M., Z.S. Birò, M. Herrmann, K. Hupe, M. Fernandes, B. Ragni, L. Szemethy, and E. Randi. 2003. Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology* 12:2585 – 2598.
- Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95:536 – 539.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945 – 959.
- Ranalla, B. and J.L. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 94:9197 – 9201.
- Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248 – 249.
- Roques, S., P. Duchesne, and L. Bernatchez. 1999. Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. *Molecular Ecology* 8:1703 – 1717.
- Shoemaker, J.S., I.S. Painter, and B.S. Weir. 1999. Bayesian statistics in genetics. *Trends in Genetics* 15:354 – 358.
- Smouse, P.E. and C. Chevillon. 1998. Analytical aspects of population-specific DNA fingerprinting for individuals. *Journal of Heredity* 89: 143 – 150.
- SPSS for Windows. 2005. Release 14.0.1. Chicago:SPSS Inc.
- Vähä, J.-P., and C.R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenario and with different numbers of loci. *Molecular Ecology* 15:63 – 72.

- Vázquez-Domínguez, E., D. Paetkau, N. Tucker, G. Hinten, and C. Moritz. 2001. Resolution of natural groups using iterative assignment tests: an example from two species of Australian native rats (*Rattus*). *Molecular Ecology* 10:2069 – 2078.
- Waits, L., P. Taberlet, J.E. Swenson, F. Sandegren, and R. Franzén. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology* 9:421 – 431.
- Wallace, L.E. 2006. Spatial genetic structure and frequency of interspecific hybridization in *Platanthera aquilonis* and *P. dilatata* (Orchidaceae) occurring in sympatry. *American Journal of Botany* 93:1001 – 1009.
- Wares, J.P., D. Alò, and T.F. Turner. 2004. A genetic perspective on management and recovery of federally endangered trout (*Oncorhynchus gilae*) in the American Southwest. *Canadian Journal of Fisheries and Aquatic Sciences* 61:1890 – 1899.
- Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323 – 354.

Table 1.1. Genetic loci and allele frequencies used in the creation of a theoretical population of pure strain lake trout on the mid-lake reef complex from Page et al. 2003. SMD = Marquette strain, LLW = Lewis Lake strain, GLW = Green Lake strain, SLW = Seneca Lake strain.

Locus	Allele	SMD	LLW	GLW	SLW
Sfo1	108	0.040	0.000	0.008	0.007
	110	0.905	0.974	0.947	0.699
	116	0.056	0.026	0.045	0.294
Sfo112	254	0.048	0.027	0.063	0.037
	256	0.040	0.040	0.071	0.224
	258	0.889	0.920	0.865	0.739
	260	0.016	0.000	0.000	0.000
	262	0.008	0.013	0.000	0.000
Sfo18	171	0.599	0.366	0.465	0.748
	173	0.000	0.000	0.025	0.022
	175	0.041	0.004	0.005	0.204
	179	0.000	0.009	0.000	0.000
	181	0.275	0.451	0.449	0.026
	183	0.005	0.112	0.000	0.000
	185	0.005	0.045	0.000	0.000
	187	0.068	0.013	0.040	0.000
	189	0.009	0.000	0.015	0.000
One μ 10	170	0.000	0.007	0.000	0.000
	174	0.893	0.601	0.827	0.750
	178	0.107	0.392	0.173	0.250
Ogo1a	142	0.000	0.013	0.000	0.000
	144	0.087	0.256	0.182	0.149
	146	0.000	0.019	0.000	0.000
	148	0.000	0.058	0.000	0.000
	150	0.762	0.481	0.667	0.306
	152	0.151	0.173	0.152	0.545
Sco μ 19	157	0.005	0.000	0.000	0.000
	159	0.005	0.000	0.000	0.000
	161	0.122	0.057	0.103	0.256
	165	0.009	0.039	0.000	0.000
	167	0.000	0.022	0.005	0.004
	169	0.005	0.000	0.029	0.004
	171	0.275	0.250	0.279	0.415
	173	0.014	0.000	0.010	0.047
	175	0.437	0.478	0.363	0.231
	177	0.086	0.061	0.108	0.043
Ssa85	126	0.018	0.000	0.005	0.000
	132	0.000	0.000	0.000	0.004
	134	0.694	0.403	0.505	0.470
	136	0.063	0.146	0.040	0.000
	138	0.225	0.447	0.450	0.526
	140	0.000	0.004	0.000	0.000

Table 1.2. Potential pure strain and hybrid crosses on the mid-lake reef complex. SMD = Marquette strain, LLW = Lewis Lake strain, GLW = Green Lake strain, SLW = Seneca Lake strain, F₁ = Pure x Pure, B_x = F₁ x Pure, F₂ = F₁ x F₁.

Pure strain	F ₁	B _x	F ₂
SMD	SMDxLLW	SMD-LLWF1xSMD	SMD-LLWF1xSMD-LLWF1
LLW	SMDxGLW	SMD-LLWF1xLLW	SMD-LLWF1xSMD-GLWF1
GLW	SMDxSLW	SMD-GLWF1xSMD	SMD-LLWF1xSMD-SLWF1
SLW	LLWxGLW	SMD-GLWF1xGLW	SMD-LLWF1xLLW-GLWF1
	LLWxSLW	SMD-SLWF1xSMD	SMD-LLWF1xLLW-SLWF1
	GLWxSLW	SMD-SLWF1xSLW	SMD-LLWF1xGLW-SLWF1
		LLW-GLWF1xLLW	SMD-GLWF1xSMD-GLWF1
		LLW-GLWF1xGLW	SMD-GLWF1xSMD-SLWF1
		LLW-SLWF1xLLW	SMD-GLWF1xLLW-GLWF1
		LLW-SLWF1xSLW	SMD-GLWF1xLLW-SLWF1
		GLW-SLWF1xGLW	SMD-GLWF1xGLW-SLWF1
		GLW-SLWF1xSLW	SMD-SLWF1xSMD-SLWF1
			SMD-SLWF1xLLW-GLWF1
			SMD-SLWF1xLLW-SLWF1
			SMD-SLWF1xGLW-SLWF1
			LLW-GLWF1xLLW-GLWF1
			LLW-GLWF1xLLW-SLWF1
			LLW-GLWF1xGLW-SLWF1
			LLW-SLWF1xLLW-SLWF1
			LLW-SLWF1xGLW-SLWF1
			GLW-SLWF1xGLW-SLWF1

Table 1.3. List of programs tested for accuracy and bias.

Program	Method	Reference
GeneClass2	ML-based assignment testing	Piry et al. 2004
GMA	ML-based assignment testing	Kalinowski 2003
WHICHRUN	ML-based assignment testing	Banks and Eichert 2000
NewHybrids	Admixture using Bayesian clustering	Anderson and Thompson 2002
BAPS	Admixture using Bayesian clustering	Corander et al. 2003
STRUCTURE	Admixture using Bayesian clustering	Pritchard et al. 2000

Table 1.4. Summary data for pairwise F_{st} score among four pure strains and three hybrid categories.

	Minimum	Maximum	Mean	% of populations with $F_{st} \leq 0.03$
Pure/Pure	0.0300	0.129	0.0866	0.0
Pure/ F_1	0.00666	0.119	0.0447	33.3
Pure/ B_x	0.00130	0.125	0.0491	41.7
Pure/ F_2	0.000200	0.113	0.0373	48.8
F_1/F_1	0.00708	0.0554	0.0282	46.6
B_x/B_x	0.00173	0.0880	0.0358	47.0
F_2/F_2	-0.000770	0.0575	0.0147	83.8

Table 1.5. Percentage of individuals in four categories assigned ($p \geq 0.80$ or $\text{LOD} \geq 2$) to an origin strain¹ or to an incorrect strain during assignment in three maximum likelihood algorithms and two Bayesian clustering algorithms.

GeneClass2²				
	Pure	F ₁	B _x	F ₂
% Origin	8.9	2.5	22.3	29.8
% Incorrect	35.1	13.7	13.3	7.7
% Ambiguous	56.0	61.3	64.4	62.6
GeneClass2³				
	Pure	F ₁	B _x	F ₂
% Origin	40.0	27.1	38.5	28.3
% Incorrect	1.8	2.9	28.3	1.5
% Ambiguous	58.2	70.1	33.2	70.2
GMA				
	Pure	F ₁	B _x	F ₂
% Origin	39.9	46.2	6.4	40.9
% Incorrect	1.8	0.8	5.9	0.3
% Ambiguous	58.3	53.1	30.1	58.8
WhichRun				
	Pure	F ₁	B _x	F ₂
% Origin	16.3	4.3	8.1	4.6
% Incorrect	0.0	0.0	0.0	0.0
% Ambiguous	83.7	95.7	91.9	95.4
BAPS				
	Pure	F ₁	B _x	F ₂
% Unique	52.9	23.8	19.9	26.6
% Other	11.2	15.9	28.5	13.4
% Ambiguous	36.2	60.4	51.7	60.0
STRUCTURE⁴				
	Pure	F ₁	B _x	F ₂
% Unique	33.8	6.1	--	--
% Other	1.9	0.9	--	--
% Ambiguous	61.2	93.0	--	--

¹ hybrid individuals assigned as a pure origin rather than as a hybrid

² with MCMC simulation of 100 individuals

³ with no MCMC simulations

⁴ USEPOPINFO flag activated

Table 1.6. Percentage of individuals in four categories assigned ($p \geq 0.80$) in NewHybrids and the breakdown of correct and incorrect assignment based on the number of true origin strains used as a prior.

Pure Strain			
	Origin/Unknown	Origin/Non-origin	Non-origin/Non-origin
% assigned	99.6	81.2	77.7
% correctly assigned	100	100	0.0
% incorrectly assigned	0.0	0.0	100
% ambiguously assigned	0.4	18.8	22.3
F₁			
	Origin/Origin	Origin/Non-origin	Non-origin/Non-origin
% assigned	91.5	95.7	82.6
% correctly assigned	100	0.0	0.0
% incorrectly assigned	0.0	100	100
% ambiguously assigned	8.5	4.3	17.4
% assigned as a pure origin strain	0.0	40.8	0.0
B_x			
	Origin/Origin	Origin/Non-origin	Non-origin/Non-origin
% assigned	94.8	88.2	63.4
% correctly assigned	100	0.0	0.0
% incorrectly assigned	0.0	100	100
% ambiguously assigned	5.2	11.8	36.6
% assigned as a pure origin strain	0.0	84.6	0.0
F₂			
	Origin/Origin	Origin/Non-origin	Non-origin/Non-origin
% assigned	67.5	85.2	90.2
% correctly assigned	28.9	0.0	0.0
% incorrectly assigned	71.1	100	100
% ambiguously assigned	32.5	14.8	9.8
% assigned as a pure origin strain	13.5	82.7	0.0

Table 1.7. Comparison of the number of individuals assigned ($p \geq 0.80$) in GeneClass2 when using 100 or 100,000 simulated individuals. Significant differences are marked by an *.

	Assigned to an origin strain					Assigned to a non-origin strain				
	100 simulated individuals	%	100,000 simulated individuals	%	p-value	100 simulated individuals	%	100,000 simulated individuals	%	p-value
Pure strain	358	8.9	775	19.4	0.083	1405	35.1	915	22.9	0.25
F ₁ hybrid	1499	25.0	1321	22.0	0.63	821	13.7	965	16.1	0.75
B _x hybrid	2675	22.3	2174	18.1	0.39	1600	13.3	2443	20.4	0.10
F ₂ hybrid	6256	29.8	6100	29.0	0.94	1608	7.7	1591	7.6	0.85

Table 1.8. The number of individuals assigned to origin and non-origin strains in WhichRun with a relaxed assignment threshold. Confidence based on LOD ≥ 2 and LOD ≥ 1.5 are presented. Significant differences are marked by an *.

	Assigned to an origin strain					Assigned to a non-origin strain				
	LOD		LOD		p-value	LOD		LOD		p-value
	≥ 2	%	≥ 1.5	%		≥ 2	%	≥ 1.5	%	
Pure strain	651	16.3	960	24.0	0.386	1	0.0	8	0.2	0.321
F ₁ hybrid	260	43.0	576	96.0	0.0370*	1	0.0	11	0.2	0.004*
B _x hybrid	969	81.0	1754	14.6	0.0570	5	0.0	34	0.3	0.003*
F ₂ hybrid	965	46.0	2098	10.0	0.000*	6	0.0	44	0.2	0.000*

Table 1.9. Percentage of pure strain individuals correctly and incorrectly assigned in three maximum likelihood algorithms. Data for $p \geq 0.80$ and $p \geq 0.50$ assignment thresholds (GeneClass2 and GMA), $\text{LOD} \geq 2$ and $\text{LOD} \geq 0.5$ (WhichRun), and pure assignment are presented. P-values represent the significance found in ANOVA/Kruskal-Wallis tests. Significant tests p-values are marked by ⁺. Superscripts indicate which threshold levels are significantly different as found in LSD post hoc tests/Mann-Whitney U-tests.

	Correct Assignment				Incorrect Assignment			
	≥ 0.80 or LOD \geq 2	≥ 0.50 or LOD \geq 0.5	Most likely	p- value	≥ 0.80 or LOD \geq 2	≥ 0.50 or LOD \geq 0.5	Most likely	p- value
GeneClass2 ⁺	40.0	67.0	73.8	0.102	1.8 ²	17.2	26.2 ²	0.018*
GMA	39.9	67.0	73.1	0.108	1.8 ³	17.1	26.9 ³	0.018*
WhichRun	16.3 ¹	48.3	74.8 ¹	0.009*	0.0 ^{4,5}	5.1 ^{4,6}	25.2 ^{5,6}	0.007*

⁺run with no MCMC simulations

Table 1.10. Percentage of F₁ hybrid individuals assigned with pure assignment in three maximum likelihood algorithms based on the most likely and second most likely origin populations.

	Assigned to origin strain(s)		Assigned to non-origin strain(s)	
	% assigned to 1 origin strain	% assigned to 2 origin strains	% assigned to 1 origin and 1 non-origin strain	% assigned to 2 non-origin strains
GeneClass2*	0.0	35.5	58.8	5.7
GMA	11.0	60.6	27.1	1.3
WhichRun	0.0	36.6	58.5	4.9

* without MCMC simulations

Table 1.11. Number of pure strain and F₁ individuals assigned (pure assignment) to an origin strain and incorrectly assigned in three maximum likelihood algorithms when the baseline file contained pure strain and F₁ populations.

	# assigned to an origin strain	%	# assigned to a non-origin strain	%
GeneClass2*				
Pure strain	2216	55.3	1789	44.7
F ₁	1426	23.8	4574	76.2
GMA				
Pure strain	1713	42.8	2291	57.2
F ₁	1816	30.3	4184	69.7
WhichRun				
Pure strain	2215	55.3	1789	44.7
F ₁	1422	23.7	4578	76.3

* without MCMC simulation

Table 1.12. Performance capabilities of two Bayesian clustering algorithms when clustering of pure strain and F₁ individuals was performed with pure strain and F₁ populations in the baseline prior file. Assignment based on a p ≥ 0.50 threshold.

	# assigned	%	Most assigned from one population	Mean # assigned from one population	Most assigned to one cluster	Mean # assigned to one cluster	Mean # from each strain assigned to each cluster
BAPS							
Pure strain	3307	8.3	922	827	895	331	82.7
F ₁	4168	69.5	768	695	827	417	69.5
STRUCTURE*							
Pure strain	541	1.4	103	70	44	28	7.0
F ₁	541	9.0	44	22	53	13	2.2

* USEPOPINFO flag activated for baseline populations

Table 1.13. Assignment success ($p \geq 0.80$) of NewHybrids when the unknown populations were created based on the individual's assignment in GMA.

	# correctly assigned	%	# incorrectly assigned	%	# ambiguously assigned	%
Correctly assigned by GMA	2690	62.6	2	0.0	1602	37.3
Incorrectly assigned by GMA	0	0.0	1642	67.5	790	32.5

Table 1.14. Results of bias tests for each pure strain in five assignment algorithms.

GeneClass2			
	Bias	p - value	Biased to
GLW	No	0.078	--
LLW	No	0.15	--
SLW	No	0.368	--
SMD	No	0.0737	--
GMA			
	Bias	p - value	Biased to
GLW	No	0.0784	--
LLW	No	0.156	--
SLW	No	0.368	--
SMD	No	0.0737	--
WhichRun			
	Bias	p - value	Biased to
GLW	Yes	0.0498	SMD
LLW	No	no error	--
SLW	No	no error	--
SMD	Yes	0.0041	GLW
BAPS			
	Bias	p - value	Biased to
GLW	Yes	< 0.0001	SMD
LLW	Yes	< 0.0001	GLW
SLW	Yes	< 0.0001	SMD
SMD	Yes	< 0.0001	GLW
STRUCTURE			
	Bias	p - value	Biased to
GLW	Yes	0.01	SLW & SMD
LLW	No	0.368	--
SLW	No	0.607	--
SMD	Yes	0.0389	SLW

Table 1.15. Results of bias tests for each F₁ hybrid in five assignment algorithms.

Maximum Likelihood				Bayesian Clustering			
GeneClass2				BAPS			
	Bias	p - value	Biased to		Bias	p - value	Biased to
GLWSLW	Yes	< 0.0001	SLW	GLWSLW	Yes	< 0.0001	SLW
LLWGLW	Yes	< 0.0001	LLW	LLWGLW	Yes	< 0.0001	GLW & LLW
LLWSLW	Yes	< 0.0001	LLW	LLWSLW	Yes	< 0.0001	SMD
SMDGLW	Yes	< 0.0001	SMD	SMDGLW	Yes	< 0.0001	GLW & LLW
SMDLLW	Yes	< 0.0001	LLW	SMDLLW	Yes	< 0.0001	SLW
SMDSLW	Yes	< 0.0001	SLW	SMDSLW	Yes	< 0.0001	SMD
GMA				STRUCTURE			
	Bias	p - value	Biased to		Bias	p - value	Biased to
GLWSLW	Yes	< 0.0001	GLW & SLW	GLWSLW	Yes	< 0.0001	SLW
LLWGLW	Yes	< 0.0001	GLW & LLW	LLWGLW	Yes	< 0.0001	LLW
LLWSLW	Yes	< 0.0001	LLW & SLW	LLWSLW	Yes	< 0.0001	SLW
SMDGLW	Yes	< 0.0001	GLW & SMD	SMDGLW	Yes	< 0.0001	SMD
SMDLLW	Yes	< 0.0001	LLW & SMD	SMDLLW	Yes	< 0.0001	LLW
SMDSLW	Yes	< 0.0001	SLW & SMD	SMDSLW	Yes	< 0.0001	SMD
WhichRun							
	Bias	p - value	Biased to		Bias	p - value	Biased to
GLWSLW	Yes	< 0.0001	SLW				
LLWGLW	Yes	< 0.0001	LLW				
LLWSLW	Yes	< 0.0001	LLW & SLW				
SMDGLW	Yes	0.0117	GLW & SMD				
SMDLLW	Yes	< 0.0001	LLW				
SMDSLW	Yes	< 0.0001	SLW				

Table 1.16. Results of each bias test performed for NewHybrids where the bias occurs.

One origin / one incorrect references					
	Strain 1	Strain 2	Bias	p - value	Biased to
GLWSLW	LLW	GLW	Yes	< 0.0001	F ₁
GLWSLW	LLW	SLW	Yes	< 0.0001	F ₁
GLWSLW	SMD	GLW	Yes	< 0.0001	F ₁
GLWSLW	SMD	SLW	Yes	< 0.0001	F ₁
LLWGLW	GLW	SLW	Yes	< 0.0001	GLW
LLWGLW	LLW	SLW	Yes	< 0.0001	LLW
LLWGLW	SMD	GLW	Yes	< 0.0001	F ₂
LLWGLW	SMD	LLW	Yes	< 0.0001	F ₂
LLWSLW	GLW	SLW	Yes	< 0.0001	F ₁
LLWSLW	LLW	GLW	Yes	< 0.0001	F ₁
LLWSLW	SMD	LLW	Yes	< 0.0001	F ₁
LLWSLW	SMD	SLW	Yes	< 0.0001	F ₁
SMDGLW	GLW	SLW	Yes	< 0.0001	GLW
SMDGLW	LLW	GLW	Yes	< 0.0001	GLW B _x
SMDGLW	SMD	LLW	Yes	< 0.0001	SMD
SMDGLW	SMD	SLW	Yes	< 0.0001	SMD
SMDLLW	LLW	GLW	Yes	< 0.0001	F ₁
SMDLLW	LLW	SLW	Yes	< 0.0001	LLW
SMDLLW	SMD	GLW	Yes	< 0.0001	F ₁
SMDLLW	SMD	SLW	Yes	< 0.0001	SMD
SMDSLW	GLW	SLW	Yes	< 0.0001	F ₁
SMDSLW	LLW	SLW	Yes	< 0.0001	F ₁
SMDSLW	SMD	GLW	Yes	< 0.0001	SMD
SMDSLW	SMD	LLW	Yes	< 0.0001	SMD
Two incorrect references					
	Strain 1	Strain 2	Bias	p - value	Biased to
GLWSLW	SMD	LLW	Yes	< 0.0001	F ₁
LLWGLW	SMD	SLW	Yes	< 0.0001	SMD B _x
LLWSLW	SMD	GLW	Yes	< 0.0001	F ₁
SMDGLW	LLW	SLW	Yes	< 0.0001	SLW B _x
SMDLLW	GLW	SLW	Yes	< 0.0001	F ₂
SMDSLW	LLW	GLW	Yes	< 0.0001	F ₁

Appendix 1.1. Results of t-tests/Mann-Whitney U-tests for comparisons between using 100 and 100,000 simulated individuals for assignment in GeneClass2.

	Assigned to origin strain			Assigned to non-origin strain		
	t / z	d.f.	p-value	t / z	d.f.	p-value
Pure strain	z = -1.732	-	0.083	t = 0.435	6	0.435
F ₁ hybrid	t = 0.445	10	0.666	t = -0.440	10	0.669
B _x hybrid	t = 0.997	22	0.33	t = -1.966	22	0.062
F ₂ hybrid	t = 0.268	40	0.79	t = 0.045	34	0.964

Appendix 1.2. Results of Mann-Whitney U-tests for comparisons between using $\text{LOD} \geq 2$ and $\text{LOD} \geq 1.5$ for assignment in WhichRun.

	Assigned to origin strain		Assigned to non-origin strain	
	z	p-value	z	p-value
Pure strain	-0.866	0.386	-0.992	0.321
F ₁ hybrid	-2.082	0.037	-2.866	0.004
B _x hybrid	-1.906	0.057	-2.943	0.003
F ₂ hybrid	-3.711	0	-3.743	0

Appendix 1.3. Table of ANOVA results comparing the assignment capabilities and error rates of three maximum likelihood algorithms at three assignment thresholds (GeneClass2 and GMA: $p \geq 0.80$, $p \geq 0.50$, and pure assignment, WhichRun: $LOD \geq 2$, $LOD \geq 0.5$, and pure assignment). Significant results are marked with an *.

Correct Assignment					
	d.f.	SS	MS	F	p-value
GeneClass2					
Assignment threshold	2	255693.2	127846.6	2.969	0.102
Residual	9	387571.5	43063.5		
Total	11	643264.7			
GMA					
Assignment threshold	2	250250.2	125125.1	2.88	0.108
Residual	9	391008.8	43445.42		
Total	11	641258.9			
WhichRun					
Assignment threshold	2	666880.2	333440.1	8.283	0.009*
Residual	9	362292.8	40254.75		
Total	11	1029173			
Incorrect Assignment					
	d.f.	SS	MS	F	p-value
GeneClass2					
Assignment threshold	2	122154.5	61077.25	6.524	0.018*
Residual	9	84257.75	9361.972		
Total	11	206412.3			
GMA					
Assignment threshold	2	128791.2	64395.58	6.477	0.018*
Residual	9	89484.5	9942.722		
Total	11	218275.7			

CHAPTER 2

DEVELOPMENT OF NEW MICROSATELLITE LOCI AND MULTIPLEX REACTIONS FOR LAKE TROUT AND THEIR USEFULNESS IN OTHER SALMONID SPECIES

ABSTRACT

Current lake trout (*Salvelinus namaycush*) management plans include stocking multiple hatchery strains into Lake Michigan with the idea that strains best adapted to survival in Lake Michigan will successfully reproduce. Monitoring to identify strains that contribute to natural reproduction and to successful rehabilitation have focused on genetic assignment of individuals to one of four source strains. The possibility of interstrain hybrid lake trout occurring under natural conditions must be considered when assessing reproductive success. At present, origin of interstrain hybrid offspring cannot be determined, in part because of a paucity of genetic markers specifically designed for lake trout. The objective of this research was to develop a suite of lake trout-specific polymorphic microsatellite loci for genetic analysis of lake trout reproduction in Lake Michigan. Eleven lake trout-specific microsatellite loci were developed and grouped into three multiplex reactions. These loci had an average heterozygosity of 0.450 with individual locus heterozygosity ranging from 0.023 to 0.810. The number of alleles per locus ranged from 2 to 22. Reactions were tested for performance with four other salmonid species with varying success. Loci described herein can be used to increase the number of genetic markers for strain identification of lake trout and, possibly, interstrain hybrid detection.

INTRODUCTION

Lake trout (*Salvelinus namaycush*) restoration in Lake Michigan currently focuses on stocking multiple hatchery strains with the idea that contemporary conditions in the lake will select for those strains best adapted to survive (Holey et al. 1995). Because one of the main goals of lake trout management is a self-sustaining lake trout population able to withstand a commercial harvest (Holey et al. 1995), successful strains will be those that reproduce successfully. After reproductively successful strains are determined, stocking can be reassessed and focused towards reproductively successful strains. To determine which strains are reproductively successful, origin of eggs and fry collected from spawning reefs must be determined. If more than one strain successfully reproduces, then interstrain hybrid offspring may be present. Therefore, algorithms used for assigning individuals to strains of origin will need to successfully assign pure strain individuals and hybrids.

Ability to distinguish among hatchery strains and hybrid offspring are limited by a small number of genetic microsatellite markers and low population differentiation. The seven microsatellite markers currently in use (Page et al. 2003) do not allow accurate assignment of hybrid individuals to their strain of origin by any of six algorithms currently used for individual assignment or admixture analysis (Chapter 1). Therefore, genetic origin of lake trout eggs and fry produced under natural conditions cannot presently be determined.

The objective of this study was to develop lake trout specific microsatellite loci and assemble multiplex polymerase chain reactions (PCR) for use in future studies of lake trout and their reproduction. Additionally, loci were assessed for applicability in

studies of four other salmonid species. Increasing the number of genetic markers used in studies of lake trout strain identification could increase performance of assignment methods. Vähä and Primmer (2006) suggested that 12 to 24 microsatellite loci would be necessary for identification of F_1 hybrid individuals when F_{st} scores range between 0.120 and 0.210. Ability to identify the genetic origin of F_1 hybrid individuals will be necessary to focus stocking efforts towards successfully reproducing strains.

METHODS

A lake trout sub-genomic library enriched for multiple microsatellite markers was constructed following the methods of Glenn and Schable (2005). Lake trout DNA was extracted using the Promega Wizard Genomic DNA Purification Protocol (Promega Corp., Madison, WI) from fin clips of a single voucher fish angled from Lake Michigan. Whole genomic DNA was restriction digested using *Rsa* I and *Bst*U I restriction enzymes (two separate digestions) followed by digestion with the *Xmn* I restriction enzyme (Glenn and Schable 2005). Double stranded linkers (Super SNX) were ligated to the DNA fragments to provide a primer-binding site for later PCRs (Glenn and Schable 2005). Linker binding was confirmed through PCR and visualization on a 1% agarose check gel (Glenn and Schable 2005). Amplification was performed on a GeneAmp[®] 9700 (Applied Biosystems, Foster City, CA) with the following temperature regime: 95°C for 2 minutes, 20 cycles (95°C for 20 s, 60°C for 20 s, 72°C for 1.5 min), 72°C for 30 min, and an indefinite hold at 15°C.

Microsatellite enrichment consists of selecting those fragments that possess microsatellites and removing those fragments that do not. Enrichment was performed using complementary microsatellite biotinylated probes and Dynabeads[®] enrichment (Invitrogen, Carlsbad, CA) using oligonucleotide mix 2 or mix 3 of Glenn and Schable (2005). Microsatellite-enriched DNA fragments were recovered through PCR using the previously described thermocycler profile (Glenn and Schable 2005) and cloned using the pGEM[®]-T Easy Vector system (Promega Corp.). Colonies were grown on LB-agar plates in the presence of X-gal and ampicillin according to the manufacturer's protocol (Promega Corp.). Positive colonies were transferred to LB-Amp media and allowed to

grow. After grow-up, 2 μL of each culture was used as template for a PCR reaction with M13/pUC forward and reverse primers to produce templates for DNA sequencing (Glenn and Schable 2005). PCR colonies were purified to remove unincorporated deoxynucleotides and primers using a Multiscreen[®] PCR cleanup kit (Millipore, Billerica, MA). Amplified inserts were cycle sequenced to determine presence or absence of putative microsatellite loci using the BigDye[®] Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and SP6 and T7 promoter primers. The sequencing reaction was purified of unincorporated dideoxynucleotides using a Montage[®] SEQ₉₆ sequencing reaction cleanup kit (Millipore) and sequenced on an ABI PRISM[®] 377XL DNA sequencer (Applied Biosystems).

Geneious v4.0.3 (Drummond et al. 2008) or Primer 3 (Rozen and Skaletsky 2000) was used to develop microsatellite primer pairs from identified microsatellite sequences. Primer pairs were tested for effectiveness in amplification and screened for polymorphism using 44 hatchery lake trout from the Codrington Research Hatchery (Codrington, Ontario, Canada) (Table 2.1). DNA was extracted as previously described, quantified using a Nanodrop[™] ND-1000 spectrophotometer, and normalized to ~20 ng DNA/ μL TLE. PCR reactions (10 μL) consisted of 1x PCR buffer (Fisher Buffer B, Fisher Scientific), 0.6 mM dNTPs, 1.5 mM MgCl_2 , 0.2 μM of each primer, 0.5 units *Taq* (New England Biolabs) and ~60 ng genomic DNA. Initial thermocycler cycling conditions were 94°C for 5 minutes, 35 cycles (94°C for 30 s, 52°C for 30 s, 72°C for 30 s), 72°C for 20 min, and 12°C hold until reactions were either genotyped or stored in the refrigerator overnight. Annealing temperatures, concentrations of MgCl_2 , and primer concentrations were adjusted during further trials to decrease non-specific products

produced during the reaction. PCR products were initially screened on 3% agarose SFR (Amresco, Inc., Solon, OH) gels, at 85V for 90 min, and compared to a 500-bp ladder (Hyberladder VTM, Bionline). Ethidium bromide was incorporated into the agarose gel matrix for visualization of the locus. Criteria for further development was presence of polymorphism (multiple bands or ‘fuzzy’ bands), relatively clean amplification (i.e., no to minor non-specific products), and efficient amplification in all tested individuals.

Eighteen loci matching selection criteria were chosen for further screening. For each locus, the forward primer was 5'-fluorescently labeled, except in one case where the reverse primer was labeled to allow for testing of multiple forward primers to determine the most effective primer pair. Amplicons were genotyped on an ABI PRISM[®] 377XL sequencer (Applied Biosystems) and analyzed using GeneScan v2.1 (Applied Biosystems) with an in-lane standard (GeneFloTM 625 DNA Ladder, Chimerx).

Following determination of polymorphic microsatellite loci, multiplex PCR reactions were developed following the method of Henegariu et al. (1997). A series of amplifications were performed to determine appropriate combinations of loci to simultaneously amplify, optimal primer concentrations for each locus, and dilutions for effective genotyping. Multiplex PCR reactions were run using the finalized reaction conditions specific for each reaction and DNA from 44 lake trout samples. Prior to genotyping, each reaction was diluted (Reaction A: 1 part Sna 79A, Sna 2A, and Sna 13Y PCR: 6 parts TLE, Reaction B: 1 part Sna 19A, Sna 82Y, Sna 63Y, Sna 40V PCR: 3 parts TLE, Reaction C: 1 part Sna 48A, Sna 44Eb, Sna 15E, Sna 64A : 1 part TLE). Genotyping was completed on the same equipment and with the same process as previous screening.

Final genotypes of sampled lake trout were determined by manual scoring of fragment sizes for each locus. Allele frequencies were calculated in GENALEX 6 (Peakall and Smouse 2006) and observed and expected heterozygosity values were calculated in Arlequin v3.01 (Excoffier et al. 2005). Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD), heterozygote deficiency, and heterozygote excess were tested using GENEPOP on the Web (Raymond and Rousset 1995) with Markov chain conditions of 1000 dememorization iterations, 100 batches, and 1,000 iterations per batch. Sequential Bonferroni correction was used to adjust significance levels for multiple pairwise comparisons in the HWE and LD tests (Rice 1989). Microchecker v2.2.3 (Van Oosterhout et al. 2004) was used to check for the presence of null alleles, non-amplifying alleles due to presence of point mutations at a given primer site (Callen et al. 1993; Waits et al. 2000).

Usefulness of loci for studies of other salmonid species was assessed with cross-species amplification tests (Koskinen and Primmer 1999; Williamson et al. 2002; Welsh et al. 2003; Paterson et al. 2004) on brook trout (*Salvelinus fontinalis*);(n = 8), brown trout (*Salmo trutta*);(n = 6), Lahontan cutthroat trout (*Oncorhynchus clarkia henshawi*);(n = 8), and lake whitefish (*Coregonus clupeaformis*);(n = 8). Multiplex reactions used the same reaction conditions and thermocycler profiles as developed for lake trout. For all loci that showed consistent amplification for a given species, observed and expected heterozygosity, number of observed alleles, and size range of the alleles were determined.

RESULTS

Overall, 11 microsatellite loci were developed for lake trout (Table 2.2). Allelic diversity ranged from 2 to 22 alleles per locus (Table 2.3). Global heterozygosity across sampled lake trout was 0.450; individual locus observed heterozygosity ranged from 0.023 to 0.900, and expected heterozygosity ranged from 0.067 to 0.921 (Table 2.2). Five loci did not initially conform to HWE (Sna 15E, Sna 19A, Sna 82Y, Sna 44Eb, and Sna 2A), but following sequential Bonferroni correction only two loci (Sna 15E and Sna 2A) were out of HWE (Table 2.2). Linkage disequilibrium was initially significant for 9 of 55 locus pairs, but following sequential Bonferroni correction, only three locus pairs (Sna 15E/Sna 82Y, Sna 82Y/Sna 44E, and Sna 44Eb/Sna 2A) remained in LD (Table 2.2). Heterozygote deficiency was not significant following sequential Bonferroni correction (Table 2.4). Three loci (Sna 15E, Sna 82Y, and Sna 2A) showed possible evidence of a null allele. Of 11 loci incorporated into three multiplex reactions (Table 2.5), one was a triplex reaction and two were a tetraplex reaction. Final annealing temperatures ranged between 54°C and 56°C, and MgCl₂ concentrations ranged from 1.5mM to 3.0 mM. Primer concentrations ranged from 0.10 µM to 0.38 µM for each primer in the pair.

Success of cross-species amplification varied among the four species tested. For brook trout, seven loci amplified, of which three were polymorphic with 2 - 4 alleles (Table 2.6). Observed heterozygosity ranged from 0.125 to 0.625 and expected heterozygosity ranged from 0.125 to 0.747. Seven loci amplified for brown trout with four polymorphic loci exhibiting 2 - 6 alleles (Table 2.6). Observed heterozygosity for brown trout ranged from 0.500 to 0.800 and expected heterozygosity from 0.409 to 0.844.

Six loci amplified for Lahontan cutthroat trout with three polymorphic loci exhibiting 2 - 3 alleles (Table 2.6). Observed heterozygosity for Lahontan cutthroat trout ranged from 0.429 to 1.0 and expected heterozygosity ranged from 0.233 to 0.692. Only two loci amplified for lake whitefish, of which only one was polymorphic (Table 2.6). The polymorphic locus had three alleles, an observed heterozygosity of 0.500, and an expected heterozygosity of 0.608.

DISCUSSION

I developed 11 lake trout specific microsatellite loci assembled into three multiplex reactions. Addition of these 11 microsatellite loci to the 7 loci currently in use for lake trout studies results in the number of available microsatellite markers for lake trout studies falling within the range suggested necessary for F_1 hybrid detection between genetically divergent strains (Vähä and Primmer 2006): 12 – 24 microsatellite loci with F_{st} scores ranging 0.12 – 0.21. Additional markers will be needed if the ability to distinguish among more advanced hybrid categories (i.e., B_x or F_2 hybrid individuals) is desired. Ability to distinguish among F_1 hybrid categories may be possible with the addition of 12 microsatellite loci recently developed by Rollins et al. (2009).

For studies involving discrimination of B_x hybrid individuals from pure strain individuals, as many as 48 loci with high F_{st} values (i.e., $F_{st} = 0.21$) will be necessary (Vähä and Primmer 2006). This level of discrimination is not currently necessary for studies of lake trout reproduction. If advanced hybrids of lake trout in Lake Michigan exist, F_1 hybrids will first need to hatch, survive to recruitment, and spawn. If this happens, lake trout restoration will be on the way to success, and may not need additional monitoring beyond the F_1 stage. For the immediate future, genetic monitoring, including assignment of advanced hybrids to their origin strain, will be important for evaluating success of lake trout restoration and determining when stocking is no longer necessary. Resources needed to develop more microsatellite or other genetic loci for distinguishing among advanced hybrids may be more appropriately used for other research and management issues, because individuals from advanced hybrid categories do not currently exist. After lake trout reproduce successfully, and current financial issues are

resolved, further loci development may prove fruitful. Using additional genetic markers to discern among advanced lake trout hybrids would be useful for management of lake trout and also enhance studies of occurrence and survival of hybrids or track adaptations to contemporary conditions by the newly formed “Lake Michigan” strain experiences.

The 11 newly-developed markers from my study may also benefit studies of other salmonid species (Koskinen and Primmer 1999; Williamson et al. 2002; Welsh et al. 2003; Paterson et al. 2004). For each species for which these markers were tested, subsets of markers were polymorphic. The observed allele size range for each species suggests that at least some of these markers could be useful for species identification or interspecies hybridization studies. One locus (Sna 48A) also showed an unusual pattern in the Lahonton cutthroat trout. Each Lahonton cutthroat trout genotyped was heterozygous for a 104 and a 110 allele. This could be an artifact of the small number of individuals screened for cross-species amplification with the markers; only 8 Lahontan cutthroat trout were included in screening. Alternatively this could be a duplicate locus (Johnson et al. 1987), a finding that could be tested via inheritance studies (Rodzen and May 2002). Future studies involving Lahontan cutthroat trout will be better able to address the reason for this unusual allele pattern.

Overall, the new lake trout specific microsatellite markers developed in this study have the potential to be useful in a wide variety of studies involving multiple species (Koskinen and Primmer 1999; Williamson et al. 2002; Welsh et al. 2003; Paterson et al. 2004). Most importantly to immediate conservation needs, these markers should help to increase the ability of algorithms to distinguish among pure strain and F₁ hybrids. Future

studies will need to reassess performance of genetic assignment algorithms using these new loci.

LITERATURE CITED

- Callen, D.F., A.D. Thompson, Y. Shen, H.A. Phillips, R.I. Richards, J.C. Mulley, and G.R. Sutherland. 1993. Incidence and origin of “null” alleles in the (AC)_n microsatellite markers. *American Journal of Human Genetics* 52:922 – 927.
- Drummond, A.J., B. Ashton, M. Cheung, J. Heled, M. Kearse, S. Stones-Havas, T. Thierer, and A. Wilson. 2008. Geneious version 4.0.3. Available from <http://www.geneious.com>.
- Excoffier, L., G. Laval, and S. Schneider. 2005. ARLEQUIN, version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47 – 50.
- Glenn, T.C. and N.A. Schable. 2005. Isolating microsatellite DNA loci. *Methods in Enzymology* 395:202 – 222.
- Henegariu, O., N.A. Heerema, S.R. Dlouhy, G.H. Vance, and P.H. Vogt. 1997. Multiplex PCR: Critical parameters and step-by-step protocol. *BioTechniques* 23:504 – 511.
- Holey, M.E., R.W. Rybicki, G.W. Eck, E.H. Brown Jr., J.E. Marsden, D.S. Lavis, M.L. Toney, T.N. Trudeau, and R.M. Horrall. 1995. Progress toward lake trout restoration in Lake Michigan. *Journal of Great Lakes Research* 21 (Supplement 1):128 – 151.
- Johnson, K.R., J.E. Wright Jr., and B. May. 1987. Linkage relationships reflecting ancestral tetraploidy in salmonid fish. *Genetics* 116:579 – 591.
- Koskinen, M.T. and C.R. Primmer. 1999. Cross-species amplification of salmonid microsatellites which reveal polymorphism in European and Arctid grayling, Salmonidae: *Thymallus* spp. *Hereditas* 131:171 – 176.
- Page, K.S., K.T. Scribner, K.R. Bennett, L.M. Garzel, and M.K. Burnham-Curtis. 2003. Genetic assessment of strain-specific sources of lake trout recruitment in the Great Lakes. *Transactions of the American Fisheries Society* 132:877 – 894.
- Patterson, S., S.B. Piertney, D. Knox, J. Gilbey, and E. Verspoor. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes* 4:160 – 162.
- Peakall, R., and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288 – 295.

- Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution* 43:223 – 225.
- Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248 – 249.
- Rollins, M.F., N.V. Vu, I.B. Spies, and S.T. Kalinowski. 2009. Twelve microsatellite loci for lake trout (*Salvelinus namaycush*). *Molecular Ecology Resources* 9:871 – 873.
- Rozen, S. and H.J. Skaletski. 2000. PRIMER 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds S. Krawetz and S. Misener), pp. 365 – 386. Humana Press, Totowa, New Jersey.
- Rodzen, J.A. and B. May. 2002. Inheritance of microsatellite loci in the white sturgeon (*Acipenser transmontanus*). *Genome* 45:1064 – 1076.
- Sloss, B.L., R.P. Franckowiak, and E.L. Murphy. 2008. Development of new microsatellite loci and multiplex reactions for nuskellunge (*Esox masquinongy*). *Molecular Ecology Resources* 8:811 – 813.
- Vähä, J.-P., and C.R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenario and with different numbers of loci. *Molecular Ecology* 15:63 – 72.
- Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software program for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4:535 – 538.
- Waits, L., P. Taberlet, J.E. Swenson, F. Sandegren, and R. Franzén. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology* 9:421 – 431.
- Welsh, A.B., M. Blumberg, and B. May. 2003. Identification of microsatellite loci in lake sturgeon, *Acipenser fulvescens*, and their variability in green sturgeon, *A. medirostris*. *Molecular Ecology Notes* 3:47 – 55.
- Williamson, K.S., J.F. Cordes, and B. May. 2002. Characterization of microsatellite loci in Chinook salmon (*Oncorhynchus tshawytscha*) and cross-species amplification in other salmonids. *Molecular Ecology Notes* 2:17 – 19.

Table 2.1. Strains used for identifying lake trout microsatellite loci.

	Killala Hatchery Strain	Kingscote Lake, Algonquin Park	Lake Manitou Hatchery Strain	Michipicoten Hatchery Strain	Slate Island Hatchery Strain
Source	Island lake	Inland lake	Inland lake	Lake Superior	Lake Superior
N	6	20	13	4	1

Table 2.2. Summary data for 11 microsatellite primer pairs developed for lake trout including the number of alleles at each locus (A), the size range of observed alleles (Range), and observed and expected heterozygosity (H_o and H_e , respectively). Dye labels used for genotyping are located at the beginning of the forward primer in all cases except for one located on the reverse primer.

Locus ID	Primer sequence (5' - 3')	Repeat motif	A	Range	H_o	H_e
Sna 2A ^{++^}	F: <i>6FAM</i> - CTATCGTGCGCCATGAAAC R: GATTCAACCACCGATTCAAC	(AGTGT) ₇	7	138 - 182	0.38636	0.66614
Sna 13Y	F: <i>NED</i> - AAACCCCTTTCAGTTCACC R: CAGTGTGAGAACAAGCAGAG	(CA) ₉	2	156 - 164	0.22727	0.23824
Sna 15E ^{+#^}	F: <i>HEX</i> - TTGGAAATATCTGCTGTAGCC R: AGGAAAGGAAAGTGCTTGTG	(AC) ₁₁ CG(CA) ₂	22	248 - 357	0.80952	0.92083
Sna 19A	F: <i>HEX</i> - GGCCGATGCACTCCTGAC R: TGCTGTAGGCCACCAAATAC	(CT) ₉	2	81 - 85	0.02273	0.06661
Sna 40V	F: <i>6FAM</i> - GTGTCTGCATAAAGCCTTGC R: GAGGCAGAACCGACTCTCTG	(AC) ₈	3	236 - 242	0.06818	0.06714
Sna 44Eb [#]	F: GCAATCACCCCTAACTCAAGC R: <i>HEX</i> - TCCAAGTTGGCTCACTTTAAC	(TC) ₁₂ G(CT) ₈	5	147 - 157	0.65909	0.70272
Sna 48A	F: <i>NED</i> - TGATTTTGATGCGAAGTGGA R: CGGGGAAAGTGCTGGATT	*	5	119 - 155	0.54545	0.51959
Sna 63Y	F: <i>NED</i> - GCACAAGTGTACCGCTTC R: ATCCATCCGTGTTCTCAACC	(GA) ₁₂	2	195 - 199	0.36364	0.50470
Sna 64A	F: <i>6FAM</i> - CACTTCTCCCTTCATCATTTCC R: AGTGGCTGAAACGTCAAACC	(TC) ₈	2	194 - 198	0.15909	0.14812
Sna 79A	F: <i>HEX</i> - AGCTAACTGTCTCTCAAACCTC R: TTTGGTTACTACATGATTCC	(AC) ₁₁	4	113 - 119	0.50000	0.51541
Sna 82Y ^{+#^}	F: <i>6FAM</i> - GAGCGTGTGCGCTTCAGT R: AACACAAATAGTAGGGAGGCAAG	(GA) ₁₆	5	106 - 124	0.52273	0.68391

* Complex microsatellite motif:(GT)₃(GC)₂GATT(GT)₅TA(TG)₅TTA(TG)₉

⁺ Hardy-Weinberg disequilibrium

[#] Linkage disequilibrium

[^] Possible null allele

Table 2.3. Allele frequencies at each of the 11 microsatellite loci developed for lake trout in this study.

Locus ID	Allele Size	Frequency	Locus ID	Allele Size	Frequency
Sna 2A	138	0.011	Sna 40V	236	0.011
	149	0.148		240	0.023
	159	0.023		242	0.966
	164	0.432	Sna 44Eb	147	0.080
	172	0.011		149	0.409
	177	0.364		151	0.011
	182	0.011		153	0.307
Sna 13Y	156	0.864	157	0.193	
	164	0.136	Sna 48A	119	0.023
Sna 15E	248	0.012		132	0.614
	264	0.024		147	0.330
	272	0.036		149	0.023
	276	0.048	155	0.011	
	278	0.012	Sna 63Y	195	0.523
	280	0.071		199	0.477
	284	0.048	Sna 64A	194	0.080
	288	0.060		198	0.920
	295	0.071	Sna 79A	113	0.011
	299	0.024		115	0.568
	303	0.012		117	0.011
	307	0.012		119	0.409
	315	0.024	Sna 82Y	106	0.148
	319	0.060		114	0.466
327	0.202	116		0.011	
331	0.131	118		0.102	
335	0.071	124		0.273	
339	0.012				
343	0.012				
347	0.024				
355	0.024				
357	0.012				
Sna 19A	81	0.034			
	85	0.966			

Table 2.4. Tests for heterozygote deficiency at 11 lake trout microsatellite loci and their significant levels with corresponding sequential Bonferroni-corrected alpha.

Locus ID	p - value	Bonferroni alpha
Sna 64A	1.0000	1.0000
Sna 40V	1.0000	0.5000
Sna 48A	0.7274	0.2425
Sna 13Y	0.5794	0.1449
Sna 79A	0.4767	0.0953
Sna 44E	0.1333	0.0222
Sna 82Y	0.0831	0.0119
Sna 63Y	0.0563	0.0070
Sna 15E	0.0442	0.0049
Sna 19A	0.0345	0.0035
Sna 2A	0.0076	0.0007

Table 2.5. Multiplex reaction conditions for lake trout microsatellite loci developed in this study. Conditions include the concentration of the various PCR cocktail reagents and the multiplex-specific annealing temperature (T_a). All PCRs were 10 μ L total volume with ~60 ng DNA/reaction.

Reaction	Locus	Primer (μM)	Fischer Buffer B	MgCl₂ (mM)	dNTPs (mM)	Taq (U)	T_a (°C)
A	Sna 79A	0.25	1.0x	1.5	0.6	0.5	55
	Sna 2A	0.20					
	Sna 13Y	0.18					
B	Sna 19A	0.10	1.0x	3.0	0.6	0.5	57
	Sna 82Y	0.22					
	Snd 63Y	0.23					
	Sna 40V	0.25					
C	Sna 48A	0.20	1.0x	1.5	0.6	0.5	57
	Sna 44Eb	0.22					
	Sna 64A	0.13					
	Sna 15E	0.38					

Table 2.6. Lake trout microsatellite loci that successfully amplified four other salmonid species' DNA. The number of alleles (A), the size range of observed alleles (Range) and the observed and expected heterozygosities (H_o and H_e , respectively) are presented.

Brook Trout				
Locus ID	A	Range	H_o	H_e
Sna 2A	2	155 - 181	0.125	0.125
Sna 13Y	1	150	0	0
Sna 15E	4	239 - 263	0.28571	0.74725
Sna 19A	1	81	0	0
Sna 48A	4	108 - 124	0.625	0.64167
Sna 63Y	1	192	0	0
Sna 64A	1	196	0	0
Brown Trout				
Locus ID	A	Range	H_o	H_e
Sna 2A	1	189	0	0
Sna 13Y	2	152 - 154	0.5	0.40909
Sna 15E	6	315 - 343	0.8	0.84444
Sna 48A	2	110 - 116	0.6	0.55556
Sna 63Y	2	191 - 205	0.5	0.40909
Sna 64A	1	192	0	0
Sna 82Y	1	119	0	0
Lahontan Cutthroat Trout				
Locus ID	A	Range	H_o	H_e
Sna 2A	3	181 - 213	0.42857	0.69231
Sna 13Y	1	180	0	0
Sna 19A	1	90	0	0
Sna 48A	2	104 - 110	1	0.53333
Sna 63Y	2	187 - 191	0.25	0.23333
Sna 82Y	1	115	0	0
Lake Whitefish				
Locus ID	A	Range	H_o	H_e
Sna 48A	3	99 - 103	0.5	0.60833
Sna 64A	1	184	0	0

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use.

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 4A	F: CAATCAATCAAAGCATACTAATC R: GCCTAACACTCACTCACTTACACTC	(TG) ₈ TCTA(TG) ₈	Monomorphic
Sna 12Aa	F: CAACTATGACTCTGCCATTCTCC R: CAACTGATGGTGCAATAAAGAC	(GT) ₁₁	Monomorphic
Sna 12Ab	F: ATGACTCTGCCATTCTCCTC R: CAACTGATGGTGCAATAAAGAC	(GT) ₁₁	Monomorphic
Sna 31Ac	F: CAGAATCACACTGGACAATAGAAC R: GATTGCACTAGAAGGGAGTG	(CA) ₁₃ CG(CA) ₁₅	Monomorphic
Sna 47Aa	F: TGGGAAAATACTTCTTCACCTC R: AATGACGCTGCTCTCTCTGG	(GA) ₉	Monomorphic
Sna 47Ab	F: GCAATGTCTGTGAAAACAACCTC R: GCAGTAAACCCCTCCCTGTG	(GA) ₉	Monomorphic
Sna 50Aa	F: CTGGCCTGCTCATTAACACC R: CACCCACCCTGTTGTCCTC	(AC) ₈	Monomorphic
Sna 50Ab	F: CATGAAGGCAGGGTAAAGTG R: GTCGCTGTGAGACGGAAAG	(AC) ₈	Monomorphic
Sna 54Aa	F: GGCTGAAGAACAAGGAGCAG R: TGGGTCAAAGCAAGAACAAC	(CA) ₈	Monomorphic
Sna 54Ab	F: AGCAGCAAACCTGAGCACCTAC R: TGGGTCAAAGCAAGAACAAC	(CA) ₈	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 54Ac	F: GAAGCGGTAAGCACTGAAGAC R: TGGGTCAAAGCAAGAACAAC	(CA) ₈	Monomorphic
Sna 57Aa	F: TGTGTTCCCATGCCAATAAAG R: GTTGTTAGTCGTGCGGTTTG	(GA) ₁₀	Monomorphic
Sna 57Ab	F: TGTGTTCCCATGCCAATAAAG R: CCAAATGCAGATGACTTCAGG	(GA) ₁₀	Monomorphic
Sna 65A	F: CAGAGGGTGAATGGGCAAG R: CCATGTCCTGCTGTATTTC	(CA) ₅ TG(CA) ₇	Monomorphic
Sna 75Aa	F: CATGCGATTAAGAGAAAGAGAAAG R: GGCAATACCATTTCAGGGAAC	(GA) ₈	Monomorphic
Sna 75Ab	F: CATGCGATTAAGAGAAAGAGAAAG R: CTTCTGGGCAATACCATTTCAG	(GA) ₈	Monomorphic
Sna 1Ea	F: AAGCTCACTTGTTTAATATGTTGTG R: CCTGATGCAATGAAGAACAC	(TG) ₅₂	Unspecific products
Sna 1Eb	F: AAGCTCACTTGTTTAATATGTTGTG R: TAAAGACGATGAGCCTGATG	(TG) ₅₂	Unspecific products
Sna 1Ec	F: GCTCACTTGTTTAATATGTTGTG R: CCTGATGCAATGAAGAACAC	(TG) ₅₂	Unspecific products
Sna 7Ea	F: GTGTCATTCCAGTCTTTCC R: CCTAGCTAGCAGAATCACAGC	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 7Eb	F: GTGTCATTCCCAGTCTTTCC R: GCCTAGCTAGCAGAATCACAG	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic
Sna 7Ec	F: AGTTCACATTTGACCCCTTG R: CCTAGCTAGCAGAATCACAGC	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic
Sna 7Ed	F: TTGGGTTGTGCTCCCTTTAC R: TAGCTAGCAGAATCACAGC	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic
Sna 7Ee	F: TGTGTCCCCCATAGAGGAAG R: TAGCTAGCAGAATCACAGC	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic
Sna 7Ef	F: TTCCTCTGTGTCCCCCATAG R: TAGCTAGCAGAATCACAGC	(AC) ₁₀ AGA(CA) ₆ G(AC) ₁₃	Monomorphic
Sna 9Ea	F: TCTGTAAGGTCCCTCAATCG R: ACACAGTGAGGTGGTGTCTG	(GA) ₂₄	Monomorphic
Sna 9Eb	F: TCTGTAAGGTCCCTCAATCG R: CTGTCACCTTGTTTGACCAG	(GA) ₂₄	Monomorphic
Sna 9Ec	F: AGAGAGCGGGAGAAAGAGAG R: ACACAGTGAGGTGGTGTCTG	(GA) ₂₄	Monomorphic
Sna 9Ed	F: AGAGAGCGGGAGAAAGAGAG R: TGTTTGACCAGACACTGAGG	(GA) ₂₄	Monomorphic
Sna 9Ee	F: CTATCGCTGGAGGTAAGG R: TGAGGTATCTTGTCTCATGTC	(GA) ₄ N ₉ (AG) ₄ TGG(TA) ₇ N ₈ (GA) ₂₄	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 9Ef	F: CTATCGCTGGAGGTAAGG R: TTTCTTCTAAAGGTAGTGTCTCTG	(GA) ₄ N ₉ (AG) ₄ TGG(TA) ₇ N ₈ (GA) ₂₄	Monomorphic
Sna 23Ea	F: TTTAAAAATGCCCTCAAAGC R: CAGACAGCACATGTAAACTGC	(TG) ₄₂	Stutter
Sna 23Eb	F: TTTAAAAATGCCCTCAAAGC R: AAGAACCATGATCCTCATTACC	(TG) ₄₂	Stutter
Sna 23Ec	F: AGAAAATAGCCCATTTAGCC R: CAGACAGCACATGTAAACTGC	(TG) ₄₂	Stutter
Sna 24Ea	F: AATCTGGATAGGCAGAGACG R: TCAAAGTTCTTTCAACCATCC	(GA) ₈	Monomorphic
Sna 24Eb	F: AATCTGGATAGGCAGAGACG R: CCGAAATCCCAAATATTTACC	(GA) ₈	Monomorphic
Sna 24Ec	F: CGGAGTGAAATCTGGATAGG R: TCAAAGTTCTTTCAACCATCC	(GA) ₈	Monomorphic
Sna 29Ea	F: GTCCGAGTTGCTGTCTTGG R: ATGGCTGTCTAGCTGTCTCC	(TG) ₄₅ (AG) ₃ A ₄ (AG) ₉ N ₁₂ (AG) ₄	Stutter
Sna 29Eb	F: AGAAAAAGATCTGAAAGAGAACG R: ATGGCTGTCTAGCTGTCTCC	(TG) ₄₅ (AG) ₃ A ₄ (AG) ₉ N ₁₂ (AG) ₄	Stutter
Sna 29Ec	F: GTCCGAGTTGCTGTCTTGG R: TGTCTAGCTGTCTCCCATGC	(TG) ₄₅ (AG) ₃ A ₄ (AG) ₉ N ₁₂ (AG) ₄	Stutter

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 29Ed	F: GGTCCGAGTTGCTGTCTTG R: GCTGTCTCCCATGCACTCTC	(TG) ₄₅ (AG) ₃ A ₄ (AG) ₉ N ₁₂ (AG) ₄	Stutter
Sna 29Ee	F: GGTCCGAGTTGCTGTCTTG R: CTAGCTGTCTCCCATGCACTC	(TG) ₄₅ (AG) ₃ A ₄ (AG) ₉ N ₁₂ (AG) ₄	Stutter
Sna 31Ea	F: AGCTAGCAGAATCACACTGG R: GCATGTGTGTGTATGTGTGC	(CA) ₁₃ CG(CA) ₁₅	Stutter
Sna 31Eb	F: AGCTAGCAGAATCACACTGG R: GTGTGTGATTGCACTAGAAGG	(CA) ₁₃ CG(CA) ₁₅	Stutter
Sna 31Ec	F: AGTGATTGTTTAAGGCCTAGC R: GCATGTGTGTGTATGTGTGC	(CA) ₁₃ CG(CA) ₁₅	Stutter
Sna 31Ed	F: AGTGATTGTTTAAGGCCTAGC R: GTGTGTGATTGCACTAGAAGG	(CA) ₁₃ CG(CA) ₁₅	Stutter
Sna 39Ea	F: GCACTGTCAAAAAGTATCC R: CCAGAAAACACACACTCC	(CA) ₂₁	Monomorphic
Sna 39Eb	F: GCACTGTCAAAAAGTATCC R: GCAATTCGTGTTTATTTTCC	(CA) ₂₁	Monomorphic
Sna 39Ec	F: GCACTGTCAAAAAGTATCC R: TCACATCACACAGATGTAGACC	(CA) ₂₁	Monomorphic
Sna 39Ed	F: TCAAAAAGTATCCAGAATAGAC R: GACCAGAAAACACACACTCC	(CA) ₂₁	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 12Va	F: GAGGAGATGGGCACAAACTG R: TCCTAAGCACAAAGTGATAAAC	(CA) ₄ CGTAT(AC) ₁₀	Monomorphic
Sna 12Vb	F: CACACACCCTTGAATGAGTAGG R: TCCTAAGCACAAAGTGATAAAC	(CA) ₄ CGTAT(AC) ₁₀	Monomorphic
Sna 20Va	F: TTTATTCAGAAATGAAGGAGAAGC R: GCATCCCTCTCTGTTTGAGC	(GA) ₁₁	Monomorphic
Sna 20Vb	F: TTTATTCAGAAATGAAGGAGAAGC R: CAGTGTTGGATAGCCATAGCC	(GA) ₁₁	Monomorphic
Sna 27Va	F: TCCTGTATCAGTTGTGCCAAAG R: ACAGGAAAGAACATGACAGG	(AC) ₈	Monomorphic
Sna 27Vb	F: CCATGTCTGAAGTTCCCCTAC R: ACAGGAAAGAACATGACAGG	(AC) ₈	Monomorphic
Sna 32V	F: GTTACACCCTGGCAAAGAGG R: GTTGGTCCCCAAAAGTCC	(CA) ₇	Monomorphic
Sna 48Va	F: TTTGCGCTGACTCTAAGCAC R: GCAGCAGTGTATGTAAAGAGTGTG	(CA) ₇ T(AC) ₁₀ N ₁₄ (CA) ₅	Monomorphic
Sna 48Vb	F: TTCTACCCCCAAGCCATAAG R: GCAGCAGTGTATGTAAAGAGTGTG	(CA) ₇ T(AC) ₁₀ N ₁₄ (CA) ₅	Monomorphic
Sna 59V	F: GCCTCAGAGGTCTCACAACG R: GGAACAGTGGGAACTGTGAAG	(GT) ₉ C(GT) ₂	Unspecific products

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 61Va	F: TGTTGTCACGTATAACACACC R: GTATAGGATGCGGCAAGAGC	(CA) ₁₂	Possible multiple loci
Sna 61Vb	F: TGTTGTCACGTATAACACACC R: TCCATTCTGAACACATCGAGTC	(CA) ₁₂	Possible multiple loci
Sna 7Xa	F: GTGTCTGCACATGCGTCTC R: TCCTGTGTTTGGAGCTTTCAG	(TG) ₇	Monomorphic
Sna 7Xb	F: GCGTCTGTATGTGTGTCTGC R: TCCTGTGTTTGGAGCTTTCAG	(TG) ₇	Monomorphic
Sna 16Xa	F: TGACCAGACCATGACCTTACC R: GACGAGCAACCCTGTGAGA	(CTGT) ₁₂ C(TCTT) ₆	Monomorphic
Sna 16Xb	F: TGACCAGACCATGACCTTACC R: TACACAGGCCAGATGGATG	(CTGT) ₁₂ C(TCTT) ₆	Monomorphic
Sna 32Xa	F: CCAAATTCATGGGTTTGTCC R: AAGCTTTGGTCCTCCTCTTTG	(GT) ₉	Monomorphic
Sna 32Xb	F: AAAGGCTACCGACAAGTTTCC R: AAGCTTTGGTCCTCCTCTTTG	(GT) ₉	Monomorphic
Sna 40Xa	F: CACTTTAGCTTAAGACATGTAGC R: AACGACATGATTTGGCACAG	(CA) ₁₁	Monomorphic
Sna 40Xb	F: CACTTTAGCTTAAGACATGTAGC R: TTCGGTGGTTAGCTCCTAGC	(CA) ₁₁	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 3Ya	F: TTCAGGTTTGAAGGTTAGGTTTG R: GTCTGCACGGGAATGACAG	(CA) ₇ G(AC) ₁₀	Monomorphic
Sna 3Yb	F: TTCAGGTTTGAAGGTTAGGTTTG R: CAGGCTAAAGGAGCATGAAG	(CA) ₇ G(AC) ₁₀	Monomorphic
Sna 37Y	F: TATCCACAGCCAACCTCCTG R: AGGCACCCCTCTCTTTCTTC	(CT) ₃ (CA) ₆	Monomorphic
Sna 57Y	F: CCAGACAGATTGGCAGACC R: TTAAACTCCCATGTAGAGAA	(AC) ₅ N ₂₈ (CA) ₇	Monomorphic
Sna 65Y	F: GGTA AAAACAAAGGGATTG R: CTTCGATGTTCTAATTTC	(TG) ₇	Monomorphic
Sna 68Ya	F: TGGGGCTACTGTGATCTGTG R: TTAAAGATGGGTCTGTGTATC	(CA) ₅ T(AC) ₄ A(AC) ₁₂	Possible ploidy
Sna 68Yb	F: TGATCTGTGTCGCCTCTGTG R: TTAAAGATGGGTCTGTGTATC	(CA) ₅ T(AC) ₄ A(AC) ₁₂	Possible ploidy
Sna 84Y	F: CTGAGCCAAAGGATTTACCTG R: GGAAATGGGTTTTGATAGGAC	(TC) ₂ CT(TC) ₆	Monomorphic
Sna 90Ya	F: GGAGCCATTTGGGACACAG R: CCTTTGGTTCTGCCAGTC	(CT) ₈	Monomorphic
Sna 90Yb	F: TGGCTCTGGTCAAAGTAGTG R: GGTTTGTGTATGTGTGCTTGG	(CT) ₈	Monomorphic

Appendix 2.1. List of lake trout microsatellite loci identified but not finalized for use (continued).

Locus ID	Primer sequence (5' - 3')	Repeat motif	Notes
Sna 62Za	F: AATCGCAACACGTTTATCTCC R: TTGTTCTCTGTTTGGTGTCTCTG	(CA) ₄ AA(CA) ₁₃	Monomorphic
Sna 62Zb	F: TCAGTTGGCTCCGATGAC R: AGGTTTCCTGTGCTGTTGTTTC	(CA) ₄ AA(CA) ₁₃	Monomorphic
