Design of a molecular assay to differentiate ‘white’ from ‘common’ Threespine Stickleback (*Gasterosteus aculeatus*) ecotypes

By
Nathalie L. MacPherson

A Thesis Submitted to
Saint Mary’s University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Biology

March, 2018, Halifax Nova Scotia

© Nathalie MacPherson

Approver: Dr. Anne Dalziel

____
Supervisor

Approved: Dr. Genlou Sun

____
Thesis Examiner

Date: March 30th 2018

Design of a molecular assay to differentiate 'white' from 'common' Threespine Stickleback (*Gasterosteus aculeatus*) ecotypes

Nathalie MacPherson

ABSTRACT

The 'White' Threespine Stickleback is a form of stickleback endemic to Nova Scotia, Canada, which exists sympatrically with the 'common' marine Threespine Stickleback. These fish differ in both morphology and behaviour. White stickleback change colour to an iridescent white during the breeding season rather than blue like the commons. Common males also care for eggs while they are in their nests whereas white males remove eggs from their nests and disperse them throughout the surrounding algae. Aside from male breeding colouration there are no known morphological traits that clearly differentiate white from common ecotypes. Therefore, an effective identification method is necessary to classify females, juvenile males, and mature males outside of the breeding season to study the mechanisms underlying adaptive divergence in colouration and parental care. White and common stickleback do form genetically distinct groups and in this thesis I attempted to develop a molecular assay to identify the fish by using previously identified regions of the stickleback genome with high differentiation between the two ecotypes. I designed primer sets to amplify microsatellite markers from these 'outlier' regions and analyzed allele frequencies of three loci with a discriminant analysis of principal components. I found that the use of only three markers was insufficient to differentiate the ecotypes, so the addition of other markers will be needed to design a successful assay.

March 2018

Table of Contents

ABSTRACT 2
ACKNOWLEGDEMENTS 4
1. INTRODUCTION 5
1.1 Local Adaptation and Ecological Speciation 5
1.2 Threespine Stickleback as a Model Organism for Local Adaptation and Ecological Speciation 6
1.3 The White Stickleback: an Understudied Threespine Stickleback Ecotype 6
1.4 Molecular Methods for Species Identification 8
1.5 Thesis Goals 10
2. MATERIALS AND METHODS 11
2.1 Collection of Threespine Stickleback (*Gasterosteus aculeatus*) Tissue Samples ... 11
2.2 DNA Extractions and Quantification 12
2.3 Sex Identification 12
2.4 Selection of Genomic Regions and Primer Design 13
Table 1 14
2.5 Microsatellite Amplification and Analysis 14
2.6 Statistical Analysis 15
3. RESULTS 16
3.1 Genetic Confirmation of Individual Sex Identification from Field Observation 16
Figure 1 16
3.2 Testing Primers for Variation in Selected Microsatellites on Agarose Gels 16
Figure 2 17
3.3 Multiplex Fluorescent Microsatellite Product Analysis 18
3.4 Scoring Alleles 18
Figure 3. 19
3.5 Statistical Analysis 19
Figure 4 20
4. DISCUSSION 21
4.1 Major Findings 21
4.2 Locus Variation 21
4.3 Future Directions 23
5. LITERATURE CITED 25

ACKNOWLEGDEMENTS

I would like to thank my supervisor, Dr. Anne Dalziel, for giving me the opportunity to work with her and experience the process of scientific research and writing. I will always be grateful to her for teaching me valuable laboratory skills, as well as for her support, direction and feedback, and encouragement throughout my thesis. Her constant positive attitude and enthusiasm made working with her a delightful experience. I also appreciate the help, guidance and good humour of Dr. Laura Weir who provided me with valued assessments and comments on my work and occasional snacks to power through the day. I thank Dr. Tim Frasier for access to his laboratory and equipment as well as teaching me how to use them and answering any of my questions. I express my gratitude to Dr. David Dansereau and Dr. Tim Frasier again as well as my fellow Honours classmates 2017/2018 for their help and evaluations of my work. The continuous help and support from the Dalziel/Weir fish lab was invaluable to me throughout my project. Thank you to SMUWorks for funding. To my friends and family, thanks for believing in me and listening to my whining and complaining when I was stressed. And finally to my fishy friends, thanks you for your generous tissue donations as well as helping me relax when I took mental breaks to care for you all.

1. INTRODUCTION

1.1 Local Adaptation and Ecological Speciation

One of the major objectives of evolutionary biology is to understand the mechanisms leading to the formation of new species (Schluter 2001; Butlin *et al.* 2012; Lowry 2012; Ravinet *et al.* 2017). There is still controversy in the academic community as to how to best define a species (Coyne & Orr 1998; Lowry 2012), but most biologists follow Ernst Mayr's biological species concept. The biological species concept defines a species as a group or population of interbreeding individuals that are reproductively isolated from other taxa and for which the reproductive isolation is genetically based (Coyne & Orr 1998). Therefore, understanding and identifying the causes of reproductive isolation are key to the study of speciation. Speciation can result from arbitrary processes such as genetic drift, or as a result of divergent natural selection (Schluter 2001). The latter process, termed ecological speciation, results from the evolution of reproductive isolation between two groups over time due to local adaptation to distinct environments (Coyne & Orr 1998; Schluter 2001). It has been determined that speciation is often associated with divergent natural selection, and this has led to renewed interest in the ecological conditions promoting this process and its underlying mechanisms (Nosil *et al.* 2009; Butlin *et al.* 2012; Schafer & Wolf 2013). While speciation is often a slow process taking thousands or even millions of years, it can occur very rapidly (Schluter 2001; McKinnon & Rundle 2002). 'Young species' are generally beneficial model systems for the study of ecological speciation and local adaptation and their underlying mechanisms. These populations, currently in the process of adaptation and speciation, are advantageous to study as they are more likely to lead to the discovery of the original mechanisms

causing adaptive divergence which could otherwise be masked by subsequent mutations over time (Samuk 2016).

1.2 Threespine Stickleback as a Model Organism for Local Adaptation and Ecological Speciation

The Threespine Stickleback (G*asterosteus aculeatus)* has often been used as a model organism for the study of local adaptation and ecological speciation (Blouw & Hagen 1990; McKinnon & Rundle 2002; Samuk 2016). Stickleback are small ray finned fish which can be found throughout the northern hemisphere in both marine and freshwater (Blow & Hagen 1990; Bell & Foster 1994; McKinnon & Rundle 2002; Samuk 2016). There are three main lineages of Threespine Stickleback found in the Japan Sea, Pacific, and Atlantic oceans (Toli *et al.* 2016). In each of these lineages, populations have repeatedly, and independently, adapted to new environments, including freshwater (McKinnon & Rundle 2002). These various lineages and populations are collectively referred to as the Threespine Stickleback complex (Blow & Hagen 1990; Bell & Foster 1994; McKinnon & Rundle 2002; Jones *et al.* 2012; Samuk 2016).

1.3 The White Stickleback: an Understudied Threespine Stickleback Ecotype

The 'white' stickleback ecotype is endemic to Nova Scotia and exists in sympatry or parapatry with the 'common' marine Three-spine Stickleback ecotype (Blouw & Hagen 1990; Samuk 2016). While these ecotypes may be found in the same locations they might not always be present at the same time (A. Dalziel, L.Weir, A. Haley, & N. MacPherson unpublished observations). The most prominent distinguishing characteristics between the two forms are the breeding colouration and parental care behaviour of males (Blouw & Hagen 1990; Samuk 2016). During the breeding season, ranging from late April to July, stickleback males will change colour (Blouw & Hagen

1990). Males attract females with their bright colouration and vigorous dancing displays. The males then lead the females back to their nests where the females then decide whether or not to mate with the male (Blow & Hagen 1990). The common Three-spine Stickleback develops a blue-green dorsum, red throat, and blue iris (Blouw & Hagen 1990). Alternatively the white stickleback will become highly conspicuous with a bright white dorsum, red throat, and blue iris (Blow & Hagen 1990). Outside of the breeding season there is no colour difference between white and common stickleback males. Morphologically, the white sticklebacks have moderately smaller body size and also possess slightly shorter pelvic spines than the common Threespine Stickleback (Blouw & Hagen 1990; Samuk 2016). There are a number of behavioural differences between the white and common stickleback which make the white ecotype interesting to study. Nest sites vary between the two forms; as white males will nest only in filamentous algae above the substrate, whereas common males nest on the substrate (Blow & Hagen 1990; Jamieson *et al.* 1992). The white ecotype also exhibits a loss of parental care. Unlike the common males that care for their eggs by fanning fresh water over them, white males remove eggs from their nests and disperse them in the surrounding algae (Blow & Hagen 1990). The white stickleback also appear to be more tolerant to environmental changes than the common stickleback, particularly temperature and salinity variation. White males breed in warmer waters and frequently cross the halocline to reach their nesting sites above the substrate (A. Haley, personal observations).

Research has suggested that the white stickleback has recently diverged from the common marine form as there is very low genetic differentiation between the fishes and ongoing gene flow (McKinnon & Rundle 2002; Samuk 2016). This divergence is thought to have been brought on by both natural and sexual selection (McKinnon & Rundle 2002;

Samuk 2016). In spite of their sympatric existence and low genetic differentiation, the white and common stickleback display reproductive isolation (Blow & Hagen 1990; McKinnon & Rundle 2002; Samuk 2016). Hybrids have been produced in laboratory settings however they have not been documented to occur in nature (Blow & Hagen 1990; Blow 1996). When presented with a choice, females of both ecotypes responded to the white males, but only the white females mated with the white males, as common females eventually left to find a common male (Blow & Hagen 1990; Jamieson *et al.* 1992).

This recent divergence, with the opportunity for hybridization, makes the white and common Nova Scotian ecotypes ideal for the study of speciation and local adaptation (Blow 1996; Samuk 2016). However, as there are no differences in male breeding colouration outside of the breeding season, and juveniles and females of both forms cannot be visually differentiated, morphologically distinguishing the two ecotypes can be very difficult (Blow & Hagen 1990). The inability to identify individual fish is a major obstacle for researchers attempting to compare the ecotypes. Another hindrance to studying these fish is how quickly the breeding colours of the males fade when the fish are caught, stressed, or otherwise disturbed, as the males of both ecotypes are very similar without their breeding colours (Blow & Hagen 1990). This reaction to stress can easily cause doubt as to the identity of the male fish unless released and observed until breeding colouration returns (A. Haley & L. Weir, personal communication). It would therefore be beneficial to implement a reliable molecular identification method to study and work with these ecotypes.

1.4 Molecular Methods for Species Identification

Many species have yet to be identified and categorized, and this can cause problems in conservation, management, education, and research for which it is imperative

to have proper species identification (Brickford *et al*. 2006). Even a single inaccurately classified individual could potentially discredit a study and severely impact a population (Brickford *et al*. 2006; Toli *et al*. 2016). Molecular sex identification can be used to identify individuals in sexually monomorphic species (Toli *et al.* 2016). Similarly, molecular identification can also be implemented for use in differentiating species. This method of identification is particularly helpful in studying cryptic species and is becoming more common because, speciation is not always accompanied by easily discernible morphological differences (Brickford *et al.* 2006). Cryptic species are now commonly differentiated using molecular identification methods, and as DNA sequencing continues to become easier to access, research on cryptic species is becoming more prevalent (Brickford *et al*. 2006).

As previously mentioned, outside of the breeding season the male white sticklebacks are virtually indistinguishable from the common Threespine Stickleback males, and juveniles and females are cryptic year round. The only potentially discernible differences between the white and common stickleback is slightly smaller body size and shorter pelvic spines in the white populations (Blouw & Hagen 1990; Samuk 2016). However, because there is overlap in size distribution between ecotypes (Blouw & Hagen 1990; Samuk 2016), this is not a reliable identification method. Genetic studies of stickleback have greatly facilitated the identifications of genes and mutations that can cause adaptive divergence and speciation and have supplied substantial genetic information about this species (Jones *et al.* 2012). Furthermore, there are genetic differences between the white and common ecotypes (Samuk 2016), and heritable differences in breeding colouration and behaviour (Blow & Hagen 1990). Consequently white and common stickleback can be differentiated into genetically distinct groups

(Samuk 2016; Samuk *et al.* 2017), and we predict that we can design an assay to molecularly differentiate the white stickleback from the marine Three-spine Stickleback.

1.5 Thesis Goals

The common Three-spine Stickleback and the endemic white stickleback are cryptic species outside of the breeding season. This severely limits research efforts as proper identification currently can only be achieved through visual observation of male individuals during the breeding season. The White Threespine Stickleback is in the process of speciation, and genetic differentiation, that is dispersed throughout the genome, is also present in females (Samuk 2016). Samuk (2016) was able to genetically distinguish white and common Threespine Stickleback, however to accomplish this over 10 000 genetic markers were used. The goal of this study is to leverage the data from Samuk (2016) to develop a time and cost effective genotyping assay that could be used for molecular differentiation between these two forms of stickleback. By developing a technique to molecularly identify the white and common stickleback we will facilitate the study of local adaptation and speciation of these ecotypes.

2. MATERIALS AND METHODS

2.1 Collection of Threespine Stickleback (*Gasterosteus aculeatus*) Tissue Samples

A permit was obtained from the Department of Fisheries and Oceans (Fishing Licence #343930) for fish collection. Mature stickleback were caught via dip nets as per SMU Animal Care protocol (17-18A “Collection of sticklebacks and killifish to study the evolution of fish physiology”) so that the ecotype of males (white or common) could be determined by the observation of breeding colours and identification of females could be attempted by observation of behaviour and interactions with males. Fin clips, ranging in size from 2-5 mm^2, were collected from the fish and preserved in 95% ethanol for later use in DNA extractions (Section 2.2).

Samples were acquired from the Nova-Scotia mainland [Canal Lake (GPS coordinates 44.497627° N, 63.900449° W, whites and commons), Lawrencetown (44.645293° N, 63.325352° W, whites), Jeddore Oyster Ponds (44.779638° N, 63.007668° W, whites), and Rainbow Haven (44.654799° N, 63.421140° W, commons)] as well as Cape Breton [Baddeck (46.101757° N, 60.745549° W, whites and commons), Gillis Cove (45.914407° N, 61.054600° W, whites and commons), Blues Cove (45.899065° N, 61.086559° W, whites and commons), and Little Narrows (45.993401° N, 60.979736° W, commons)]. Mainland common stickleback samples were also obtained from Stan King and Dr. Paul Bentzen from the Dalhousie Marine Gene Probe Lab [Lawrencetown Lake Estuary (44.645394° N, 63.347138° W), Mahone Bay (44.449347° N, 64.341259° W), and L’Hebert (43.799539° N, 65.009603° W)]. Of these samples, 20 mainland white males (Canal Lake and Lawrencetown) and 19 mainland common males (Canal Lake, Lawrencetown, and Mahone Bay) were selected for further analysis.

2.2 DNA Extractions and Quantification

Total genomic DNA was extracted from fin clips of 319 fish using the Sigma-Aldrich GenElute^{TM} Mammalian Genomic DNA Miniprep Kit (G1N350, 350 preps) following the instructions for mammalian tissue preparation. The optional 5 min incubation before eluting the DNA and second elution were performed for all samples. Following extractions, genomic DNA samples were stored at -20^{0}C. DNA was then quantified using the Nanodrop^{TM} 1000 spectrophotometer to determine the range of DNA (ng/µL) obtained from the extractions to calculate quantities needed for polymerase chain reactions (PCRs).

2.3 Sex Identification

The sexually dimorphic 3' UTR of isocitrate dehydrogenase (Idh) locus was amplified to confirm the sex of each fish and to test DNA quality (Toli *et al.* 2016). The sex determining gene of Threespine Stickleback is located on a neo X and Y chromosomes which possess different allele sizes and this gene is in linkage disequilibrium with the Idh locus (Toli *et al.* 2016). PCRs were performed, using Promega GoTaq® PCR Core System I, in 25 µL reactions. Final concentrations of the reactions consisted of 1X PCR buffer (Green GoTaq® Flexi Buffer), 2 mM of magnesium chloride solution ($MgCl_2$), 10 µM of each primer (forward and reverse), 0.4 mM of deoxynucleotide triphosphates (dNTPs), and 0.025 U/ µL of GoTaq® DNA Polymerase (Taq). 2 µL of genomic DNA (approximately 20 ng) was added to each reaction and nuclease-free deionized distilled water (ddH_2O) was added to obtain a final volume of 25 µL. PCR amplification conditions were as follows; initial denaturation of 5 min at 95 ^{o}C proceeded by 38 cycles of 15 sec at 95 ^{o}C, 30 sec at 51 ^{o}C, and 30 sec at 72 ^{o}C, followed

by a final extension of 7 min at 72 ^{o}C. The products were then run on a 2% ethidium bromide stained agarose gel to be separated by size and visualized.

2.4 Selection of Genomic Regions and Primer Design

Samuk (2016) was able to genetically differentiate white from common mainland Threespine Stickleback populations by using approximately 20000 single nucleotide polymorphisms (SNPs) identified through genotyping by sequencing (GBS). To locate outlier regions, Samuk (2016) divided the entire Threespine Stickleback genome into 75 000 base pair (bp) regions, many of which were found to differ among mainland white and common stickleback. A subset of these regions with the highest fixation indexes (F_{st}), a measure of population differentiation, were selected to be scanned for microsatellites to design new genomic 'markers' for the most divergent regions (Samuk 2016). Genomic windows (75 000 bp) were downloaded from Ensembl's BROADS Stickleback assembly 1 as txt files and converted to FASTA files to be run in msatcommander.

Msatcommander (Faircloth 2008) version 1.0.6 was used to find microsatellites and design primers in the selected outlier genomic regions from Samuk (2016) with the following parameters: i) perfect repeats of 10-15 dinucleotides and trinucleotides of 8+ repeats. ii) PCR product sizes of 70-80, 81-90, 91-100, and 101-120 bp including the microsatellite oligonucleotides. iii) Primer sizes of 20, 24, or 28 bp, with iv) melting temperatures (T_m) of 58^{0}C, 64^{0}C, or 70^{0}C (Recommendations from Ian Patterson from the Dalhousie Marine Gene Probe Lab, Personal Communication).

Primer pairs from the most variable genomic windows ($F_{st} \geq 0.217$) were selected for testing (12 primer pairs). The primers were tested on both common and white stickleback under regular PCR conditions (25 µL reactions, same volumes and conditions as IDH sex identification PCR) and run on an agarose gel to check that they amplified

loci within range of the expected product size, and to check for variation among individuals and ecotypes. Loci which appeared the most clearly on the gel and displayed the most variation were then chosen (6 primer pairs) to be labelled with fluorophores to be used in multiplex PCRs.

Table 1. Selected primers fluorescently labelled with 5’ modification

Primer Name	Chromosome	Fst	Sequence	Expected Product Size	Fluorophore	Multiplex PCR combination
M1	1	0.258	5'-CGGCAGGAAGAAATGATGCAAGTG-3'	115	FAM	1
			3'-CAGGGAAAGAGGCCAGACATCCAC-5'		-	
M4	1	0.252	5'-ATCTTAGAAAGCTGCATCGCTCTC-3'	100	TAMN	1
			3'-ACTCAAGGCTGTGAAATGACAAGG-5'		-	
M16	6	0.226	5'-GGTCTAAACACCGTTGCTCTGCG-3'	115	-	2
			3'-TGCTGTAATTCCCTACTGTGCATCG-5'		HEX	
M30	11	0.717	5'-ATCAGTGCATTTCTTCGGTGTGTC-3'	80	FAM	2
			3'-CGACTAACAACGAGTGAAACAAC-5'		-	
M33	12	0.257	5'-TCACAGGACTTTCAGAGATTGGAG-3'	120	HEX	1
			3'-AAACTGACGATGGGAGAGATCAGC-5'		-	
M56	21	0.217	5'-CGTTGTTTCCTCCTGGCTTGC-3'	90	HEX	2
			3'-CGCTCTGCTCTGAAGATCATCCAG-5'		-	

2.5 Microsatellite Amplification and Analysis

Multiplex PCRs were carried out using combinations of the fluorescently labelled primers (combo 1; M1, M4, & M33, & combo 2; M30, M16, & M56) and Promega’s GoTaq® PCR Core System I. The PCRs had final concentrations of 1X PCR buffer, 3 mM of $MgCl_2$,

0.2 µM (primers M1, M4, M16, & M30) and 0.125 µM (primers M33 & M56) of each primer sets, 0.4 mM of dNTPs, and 0.05 U/µL of DNA Polymerase. The volume of genomic DNA added was varied to achieve a concentration of 20 ng/25 µL, and the amount of ddH_2O also varied to obtain a final volume of 25 µL. Amplification conditions were recommended by Dr. Tim Frasier (Personal Communication) which consisted of an

initial denaturation of 5 mim at 95 ^{o}C followed by 38 cycles of 30 sec at 95 ^{o}C, 1 min at 58 ^{o}C, and 1 min at 72 ^{o}C, then a final extension of 45 min at 65 ^{o}C.

The multiplex PCR products of the first combination of primers (combo 1; M1, M4, & M33) were diluted by 1/15 and the products from the second combination (combo 2; M16, M30, & M56) were diluted by 1/5 to improve fluorescent signal output. Samples were then prepared to be run on the ABI 3500xL Genetic Analyser in solutions of 0.25 µL of GeneScan^{TM} 600 Liz^{TM} dye Size Standard v2.0 (Life Technologies), 2 µL of diluted samples, and 10 µL of formamide that were then denatured at 95 ^{o}C for 5 min to allow the DNA to remain denatured at room temperature. Products were then size separated by the ABI 3500xL capillary analyser.

Fragment analysis was used to genotype the samples and amplified products were visualized and analyzed using the program GeneMarker® version 2.6.4. Taking into account stutter and the tendency of Taq polymerase to add an A to the end of DNA fragments during PCRs, microsatellite alleles were scored.

2.6 Statistical Analysis

Allele frequencies for each locus were calculated from the scored microsatellites of each region for both white and common populations. Discriminant analysis of principal components (DAPC) was performed on R version 3.4.4 (R Development Core Team 2018) using the ‘adergenet’ version 2.1.1 package (Jombart & Ahmed 2011). This multivariate analysis uses genetic data to identify and describe population structure as well as pinpoint loci that are best suited for discriminating groups (Jombart *et al.* 2010).

3. RESULTS

3.1 Genetic Confirmation of Individual Sex Identification from Field Observation

The sexually dimorphic IDH locus amplifies two alleles in males and one in females, and has been shown to have a zero percent error rate for other Atlantic stickleback populations (Toli *et al.* 2016). Of our fish, all IDH genotype sex identifications corresponded perfectly with individual sex identification from the field (section 2.1 Collection of Threespine Stickleback (*Gasterosteus aculeatus*) tissue samples). Only males were selected for further analysis of genetic divergence between white and common stickleback as their ecotype could be more definitively determined than that of females.

Figure 1. Visualization of IDH sexing locus amplifications. PCR products run out on 2% ethidium bromide stained agarose gel with a 100bp ladder for size reference.

3.2 Testing Primers for Variation in Selected Microsatellites on Agarose Gels

All primers amplified products of the expected size and M1, M4, M16, M30, M33, and M56 also displayed some level of variation among individuals. These were chosen to be fluorescently labeled for use in multiplex PCRs (Table 1). PCRs products were visualized on agarose gels to compare to expected allele sizes and obtain estimates of allele sizes (Figure 1).

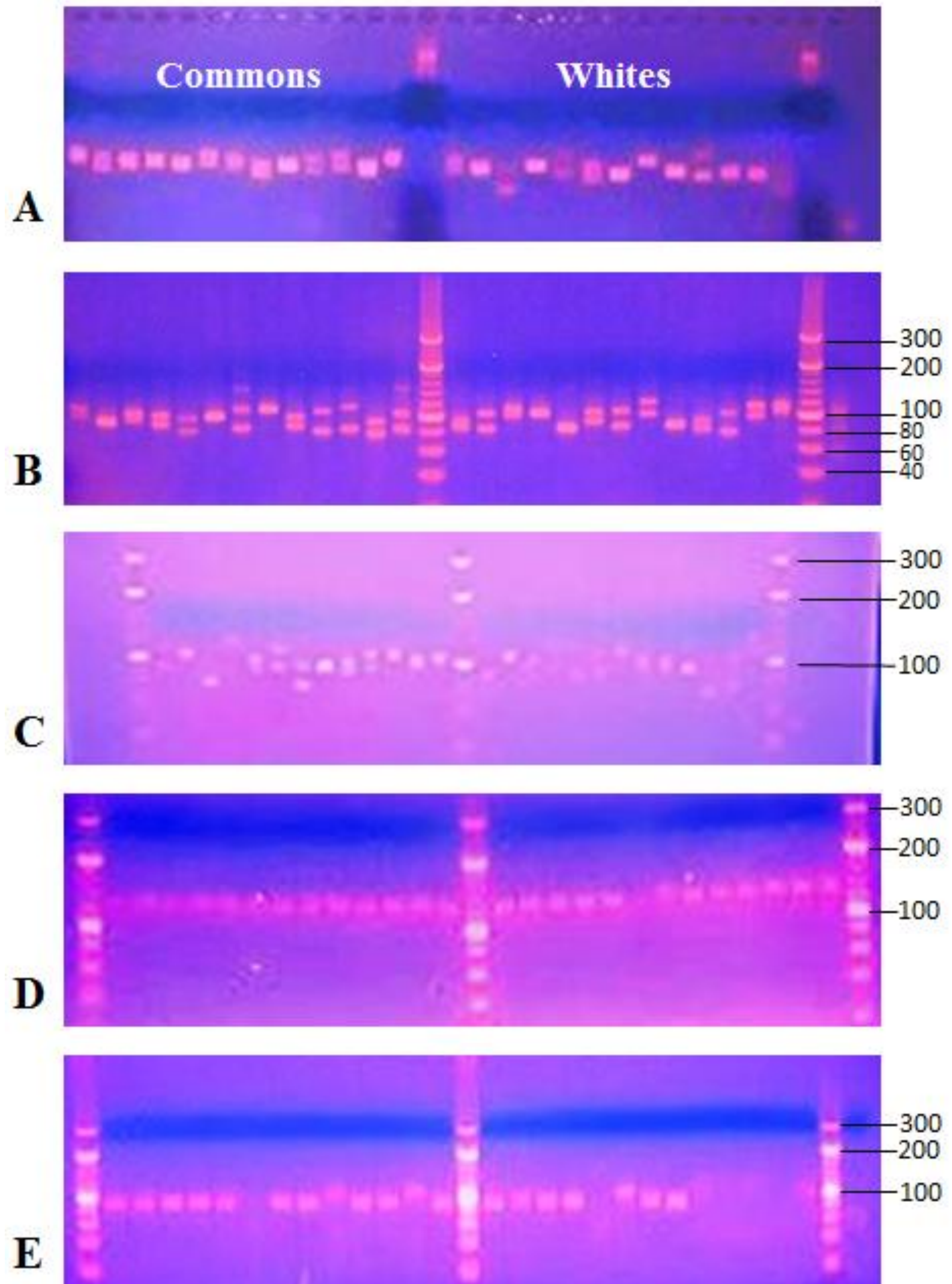

Figure 2. Visualization of microsatellite amplifications. PCR products of the M1 primer set (A), M4 primer set (B), M30 primer set (C), M33 primer set (D), and M56 primer set (E) from white (right half of gel) and common (left half of gel) stickleback run out on 4% ethidium bromide stained agarose gel. A 100bp ladder was run in (A), but did not effectively separate. Thus, a 20bp ladder was used on subsequent gels (B, C, D, & E).

The Thermo Scientific™ O'RangeRuler 20bp DNA Ladder, ready-to-use 20-300bp ladder was found to run around 60bp slower than the PCR products, as well as all other DNA ladders tested in the lab, consequently our microsatellites are approximately 60bp smaller than the size indicated by the ladder in Figure 1.

3.3 Multiplex Fluorescent Microsatellite Product Analysis

The five microsatellite loci presented in section 3.2 were amplified with the fluorescently labelled primers in the combinations summarized in Table 1 and run through the ABI 3500xL. Scoring loci from the multiplex PCRs was not possible for any loci except one (M4). Other loci were run in single reactions using the same PCR conditions, and scored, as there was not sufficient time to further optimize multiplex reactions. However, the M16 primer set was not used as the amplified region interfered with the M56 locus which appeared to have clearer microsatellite peaks.

3.4 Scoring Alleles

The scoring of alleles using the GeneMarker® version 2.6.4 program revealed 19 alleles at the M4 locus, 4 alleles at the M33 locus, and 5 alleles at the M56 locus. Allele frequencies were calculated for these three regions and with the exception of a few alleles that were found in a small number of individuals all alleles appeared in both white and common stickleback (Figure 2).

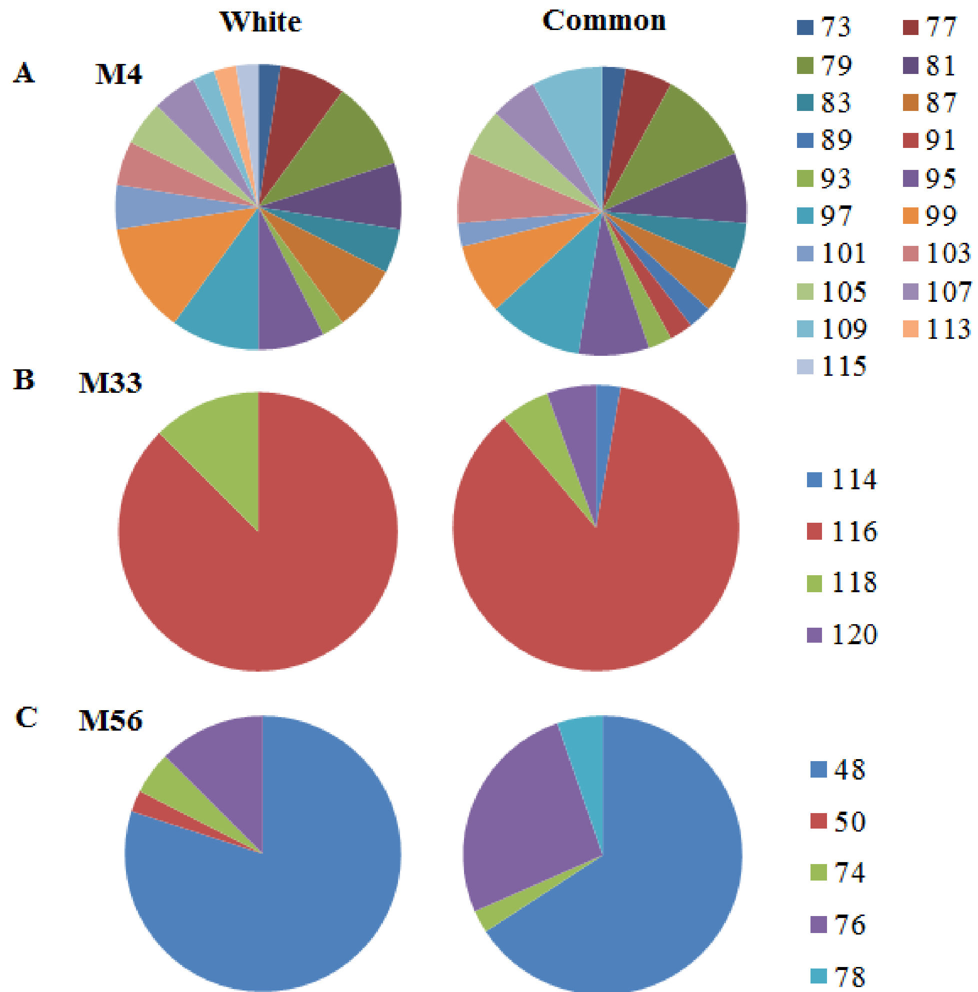

Figure 3. Microsatellite allele frequencies at the M4 (A), M33 (B), and M56 (C) loci of both white and common stickleback. Colours represent different sized PCR products (alleles) and correspond to the colours on the right.

3.5 Statistical Analysis

Discriminant analysis of principal components (DAPC) was unable to clearly differentiate the white and common ecotypes from all locations into distinct groups (Figure 3A). However, there was some degree of separation between Canal Lake white and common stickleback (Figure 3A). This differentiation is largely due to variation at the M4 locus (Chromosome 1) and the M33 locus (Chromosome 12), with little contribution from the M56 locus (Chromosome 21) (Figure 3B).

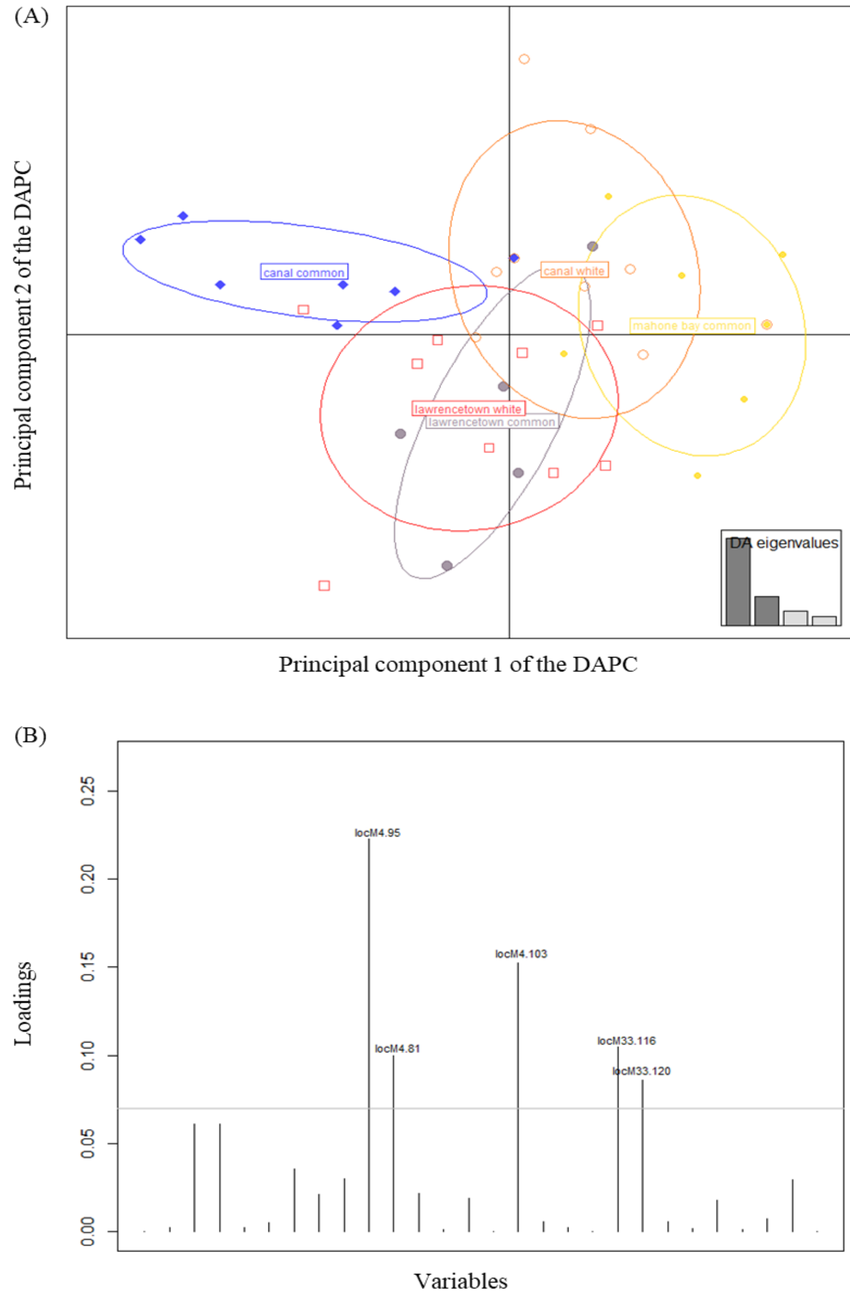

Figure 4. (A) Discriminant analysis of principal components (DAPC) derived from variation in three stickleback microsatellite loci (M4, M30, M56) in ‘white’ and ‘common’ stickleback ecotypes from different locations. The inset on the bottom right displays the eigenvalues for the two axes. Filled markers represent common individuals and empty markers represent white individuals. Locations are displayed with different colours and symbols (canal = Canal Lake; lawrencetown = Lawrencetown Beach; mahone bay = Mahone bay). (B) The factor loadings for each loci (M4, M30 or M56), and the alleles found at each loci, for the first principal component of the DAPC.

4. DISCUSSION

4.1 Major Findings

The use of only three different loci on three chromosomes was insufficient to differentiate white from common Threespine Stickleback ecotypes from mainland Nova Scotia (Figure 4A). Each of the chosen regions displayed considerable variation, as predicted, however this variation was present both between and within populations. It is therefore going to be necessary to select more regions to effectively differentiate these fish ecotypes. Promising candidate loci include the three other regions initially selected for this study (on chromosomes 1, 6, and 11 which could be amplified with primers M1, M16, & M30, respectfully; Table 1); they are all within the genomic regions (75 000 bp) found by Samuk (2016) to have high Fst values (Fst > 0.22), amplified products of expected size, and displayed variation. However, additional loci may need to be selected within closer proximity to, and in linkage disequilibrium with, the SNP regions used by Samuk (2016) to find loci possessing higher divergence between ecotypes that can effectively differentiate the ecotypes.

The addition of more loci from the regions Samuk (2016) determined to be differentiated between ecotypes is expected to be successful in distinguishing white from common stickleback, but the number of additional loci that will be required is unclear. If a successful genetic assay can be developed it would allow identification of breeding fish in a more timely and cost effective manner than would be required by a complete genome scan (Samuk 2016).

4.2 Locus Variation

A wide range of variation was observed at the chosen loci both among individuals and within populations. These differences were not surprizing as marine organisms with

extensive ranges, such as Threespine Sticklebacks, usually present with low genetic differentiation among populations and high variation within populations (DeFavieri *et al.* 2013). This is due to features that cause high variation within populations such as large effective population sizes with a high degree of gene flow, qualities shared by both stickleback ecotypes examined in this study (Nielsen *et al.* 2009; DeFavieri *et al.*2013).

The chosen regions (75 000 bp windows) were from sections of the genome displaying high population differentiation (Samuk 2016; Fst's > 0.2), however the differences in these regions is not fixed. The given genetic evidence suggests these fish are only recently diverged and that there is still a considerable amount of gene flow between populations (Samuk 2016). It is also worth noting that there is potentially recombination occurring between our markers and the ones used by Samuk (2016), in a way that they are not in linkage disequilibrium. It may therefore be necessary to select regions nearer to Samuk's (2016) outlier single nucleotide polymorphisms between white and common Threespine Stickleback populations. The use of several different genomic regions in conjunction would thus be a promising method to identify the ecotypes since there will most likely not be enough differentiation (i.e. no fixed differences) at any one given locus to separate the fish into distinct groups.

As well as designing primers for selected loci, msatcommander (Faircloth 2008) provided expected product sizes. When products were visualized on gels the sizes were slightly larger than those provided by the program. When the sizes observed on the gels were compared to the sizes provided by the ABI, allele sizes were consistently roughly 60 bps larger than expected indicating a problem either with the 20bp ladder used for the gels or possibly improper calibration of the ABI. It was later determined by comparisons to other DNA ladders by other lab members that the 60bp differences observed were due

to faster migration of the 20bp ladder through gels (Y. Gbotsyo, Personal Communication). Regardless, this would not have negatively affected results as the difference in sizes was consistent across all loci and samples.

4.3 Future Directions

The next step in the development of a time and cost effective molecular identification method for these stickleback ecotypes would be to genotype additional microsatellites in the regions detected by Samuk (2016) showing high population differentiation. When msatcommander was used to find microsatellites and design primers, hundreds of primer sets were designed. Of these 80 had been chosen as potential loci. These were narrowed down over the course of the study to the three that were examined in detail. The additional primer sets designed by msatcommander will be used in large multiplex reactions and then sequenced using next generation sequencing technology (Illumina MiSeq) in collaboration with the Dalhousie Marine Gene probe lab. Then, MEGASAT (Zhan *et al.* 2017) will be used to separate and score alleles. MEGASAT (Zhan *et al.* 2017) is a program designed to analyze microsatellites and score alleles, taking into account stutter and non-target amplifications, of large multiplex reactions. Using MEGASAT to analyze many more loci will increase the chances of finding successful combinations of loci that are capable of differentiating white and common stickleback.

Once successful, the identification of either ecotype of stickleback will be possible for all males, females, and juveniles, outside of the breeding season. When we are able to accurately identify females it will be possible select specific individuals of known ecotypes to breed these fish for quantitative genetic studies to map quantitative traits and uncover the genes contributing to the variations in behaviour and morphology (reviewed

by Peichel & Marques 2017). Breeding crosses of these fish is necessary to begin studies of the genetic basis of parental care in these fish as well as the evolution of white breeding colouration of the males. Presently there are no studies on the genes contributing to variation in these traits and the lack of an effective identification method is the main limiting factor in designing crosses for quantitative genetic studies. It will also be possible to survey fish from different areas, and compare mainland vs. Bras d'Or populations. Common sticklebacks from the Nova Scotia mainland and the Bras d'Or lakes are genetically different, so whites from the two regions may be the result of convergent evolution (Samuk 2016). These studies would contribute to the understanding of the underlying mechanisms leading to the divergence of these ecotypes.

5. LITERATURE CITED

Bell M.A., Foster M.A. (1994) Introduction to the evolutionary biology of the threespine stickleback. In *Introduction to the evolutionary biology of the threespine stickleback* (Bell M.A. & Foster S.A., eds) pp. 1-26, Oxford University Press.

Blow D.M., Hagen D.W. (1990) Breeding ecology and evidences of reproductive isolation of a waidesperad stickleback fish (*Gasterosteidae*) in Nova Scotia, Canada. *Biological Journal of the Linnean Society* 39, 195-217

Blow D.M. (1996) Evolution of offspring desertion in a stickleback fish. *Ecoscience* 3, 18-24

Brickford D., Lohman D.J., Sodhi N.S., Ng P.K., Meier R., Winker K., Ingram K.K., Das I. (2006) Cryptic species as a windown on diversity and conservation. *Trends in Ecology & Evolution* 22(3), 148-155

Butlin R., Debelle A., Kerth C., Snook R.R., Beukeboom L.W., Cajas R.F.C., Diao W., Maan M.E., Paolucci S., Weissing F.J., *et al.* (2012) What do we know about speciation? *Trends in Ecology & Evolution* 27, 27-39

Coyne J.A., Orr H.A. (1998) The evolutionary genetics of specation. *Phil. Trans. R. Soc. Lond.* 353, 287-305

DeFaveri J., Shikano T., Shimada Y., Meril J. (2013) High degree of genetic differentiation in marine three-spined sticklebacks (*Gasterosteus aculeatus*). *Molecular Ecology* 22(18), 4811-4828

Faircloth B.C. (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8, 92-99

Jamieson I.G., Blow D.M., Colgan P.W. (1992) Field observations on the reproductive biology of a newly discovered stickleback (*Gasterosteus*). *Canadian Journal of Zoology* 70, 1057-1063

Jombart T., Devillard S., Ballou F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured population. *BMC Genetics* 11, 94-109

Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wi de SNP data. Bioinformatics. doi:10.1093/bioinformatics/btr521

Jones F.C., Grabherr M.G., Chan Y.F., Russell P., Mauceli E., Johnson J., Swofford R., Pirun M., Zody M.C., White S., *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55-61

Lowry D.B. (2012). Ecotypes and the controversy over stages in the formation of new species. *Biological Jounal of the Linnean Society* 106(2), 241-257

McKinnon J.S., Rundle H.D. (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution* 17(10), 480-488

Nielsen E.E., Hemmer-Hansen J., Larsen P.F., Bekkevold D. (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* 18(15), 3128-3150

Nosil P., Harmon L.J., Seehausen O. (2009) Ecological explinations for (incomplete) speciation. *trends in Ecology & Evolution* 24(3), 145-156

Peichel C.L. & Marques D.A. (2017) The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372: 20150486

Ravinet M., Faria R., Butlin R.K., Galindo J., Bierne N., Rafajlovic M., Noor M.A.F., Mehlig B., Westram A.M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* 30(8), 1450-1477

Samuk K.M. (2016). The evolutionary genomics of adaptation and speciation in the threespine stickleback [Thesis]. The University of British Columbia

Samuk K., Owens G.L., Delmore K.E., Miller S.E., Rennison D.J., Schluter D. (2017) Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molectular Ecology* 26, 4378-4390

Schluter D. (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* 16(7), 372-380

Shafer A.B., Wolf J.B. (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters* 16(7), 940-950

Toli E.A., Calboli F.C.F., Shikano T., Merila J. (2016) A universal and reliable assay for molecular sex identification of three-spined sticklebacks (*Gasterosteus aculeatus*). *Molecular Ecology Resources* 16, 1389-1400

Zhan L., Paterson I.G., Fraser B.A., Watson B., Bradbury I.R., Ravindran P.N., Reznick D., Beiko R.G., Bentzen P. (2017) MEGAST: automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources* 17, 247-256