POPULATION GENETICS OF CREEK CHUB (*SEMOTILUS ATROMACULATUS*) IN A POSTGLACIAL, AGRICULTURAL LANDSCAPE

by

Abigail Schnelker

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Biology

Fort Wayne, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mark Jordan, Chair
Department of Biology
Dr. Jordan Marshall
Department of Biology
Dr. Robert Gillespie
Department of Biology

Approved by:

Dr. Jordan M. Marshall
Head of the Graduate Program

For Glenn, Cecilia and Stuart

ACKNOWLEDGMENTS

Thank you to my advisor Dr. Mark Jordan who took a chance on me. I appreciate every moment you took to teach me techniques in the lab, explain concepts to me and make me laugh with you dry sense of humor. Thank you to my committee members, Dr. Robert Gillespie and Dr. Jordan Marshall. Dr. Marshall thank you for joining my committee in the eleventh hour and for helping me when I would get grumpy with my formatting issues. Thank you Dr, Gillespie for being a constant point of positivity throughout my educational career. Thank you to Dar for always being so supportive and positive all these years.

A huge thank you to the students and volunteers who collected the fish used in this study. In both Indiana, and Michigan and the Ohio sampling sites. Thank you to Dr. Michael Blum who provided the Little Miami River genetic and GIS data used in this study and to the EPA of Ohio for maintaining the samples loaned for this study and for keeping the data archived for this study.

I would like to thank my siblings for supporting me through my graduate degree in the special way of teasing me constantly that only siblings can do. Thank you for giving me nieces and a nephew that I love more than I thought possible. I am so proud of you and all you success in life, mom and dad would be proud of you too.

From the bottom of my heart, thank you Kenneth Christmon it is because of your love and guidance that I will graduate without debt and have found the passion that I do for higher education and student advocacy. Thank you to James Velez and Dr. Kristina Creager, the two of you have been invaluable in my education as not only an advocate for students, but for sherpaing me through adulthood. Stephany Alarcon, you brought me out of the shell I built after the loss of mother and have been the friend I needed to knock me upside the head on more than a few occasions. Lynne Koepper, I am so grateful to have reconnected with you as a graduate student. You are my travel companion, my confidant, a fellow geek and my favorite beer maven. Tyler Shuman, you are my Taye Diggs, you are my constant. Thank you to the many friends who have supported me but have not been named, I have 9 bridesmaid dresses that each represent a unique and often hilarious story.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Schnelker, Abigail, R. MS
Institution: Purdue University
Degree Received: May 2019
Title: Population Genetics of Creek Chub (*Semotilus atromaculatus*) in a Postglacial, Agricultural Landscape
Committee Chair: Mark Jordan

The population genetics of species occupying formerly glaciated regions are not only impacted by glacial retreat but also agricultural land use that is typical of such regions. Areas which have experienced glaciation often display a lowered amount of genetic variability and minimal population structure, and these effects become more predominant with increasing distance from a potential refugial population. Meanwhile, agricultural land use over the recent past has also been demonstrated to disrupt population structure distribution through disturbance regimes. The purpose of this study was to assess potential post-glacial and agricultural effects on populations of creek chub (*Semotilus atromaculatus*) in two agricultural watersheds that differ in the glacial history. The Saint Joseph River (SJR) watershed, Indiana and Michigan, USA was entirely glaciated during the last glacial maxima, while the Little Miami River (LMR) watershed in Ohio, USA, is situated on the boundary of the glacier. The degree of agricultural land use also varies between and within the two watersheds. Using eight microsatellite loci, 312 individuals were genotyped from 13 sites in SJR and 2,318 individuals from 29 sites in LMR. Measures of genetic differentiation showed that there was strong differentiation between watersheds. Analyses within watersheds recovered additional but weaker differentiation that was mostly associated with the geography of sub-watersheds and isolation by distance. Proximity to the glacial boundary appeared to play a minimal role in genetic differentiation and genetic variation. Differentiation among localities was not directly associated with the glacial boundary within LMR, and localities in this watershed had lower allelic richness and heterozygosity than those in the fully glaciated SJR. After accounting for the positive correlation of stream distance in LMR using partial Mantel test, both glacial history and agricultural land use were positively correlated with genetic differentiation. However, these predictor variables were also strongly correlated with one another which prevented disentangling the two potential effects. Within SJR, no

relationship of genetic differentiation with agricultural land use was recovered. My study shows that there is not a simple relationship between glacial history, contemporary land use, and genetic differentiation in creek chub. Rather, it appears that the patterns of genetic variation observed may be more closely linked to the dispersal behavior of creek chub within and among watersheds, and the history of effective population size within watersheds.

INTRODUCTION

Glaciation and its effects on genetic variation

Past glaciation events impact extant populations through their continued influence on morphology, geographic distribution, and genetics. The Quaternary Period began 2.6 million years ago (Ma) and is subdivided in to two epochs, the Pleistocene and Holocene, the latter of which constitutes the modern era, beginning 11,700 years before present (Gibbard & Pillans, 2012). The Croll-Milankovitch theory states that the cyclical growth and recession of glaciers coincides with the earth's axial tilt cycle of 41-kyr. Beginning 2.4 Ma, glaciers grew and receded on a 41-kyr cycle until about 900 ka. After this time the cycles expanded to 100 kyr and became more dynamic, with more drastic differences in temperatures between "ice ages" and interglacial periods (Hewitt, 2000). During the last glacial maximum (LGM) approximately 30,000-12,000 years ago, ice encapsulated a large expanse of land throughout most of the North American continent, Eurasia and Antarctica. This colossal expanse of ice and rock slowly carved the landscape, shaping the current topography of much of the Midwestern United States (Ehlers, 2016). The glaciers brought with them not only ice, but rock, soil and in many cases, plant and animals species that were caught along for the ride. Many areas experienced extirpation due to lack of suitable habitat caused by glaciation, or individuals were displaced into fringe populations by these layers of ice (Mee & Moore, 2014).

The late Wisconsinan glacial maxima of the Pleistocene epoch, which receded from what is now the Midwestern United States approximately 12 ka, can still be detected in the genetics of contemporary population of both terrestrial and aquatic species endemic to the area. Genetic studies on both plant and animal species show that during the last ice age many populations were reduced in size due to the lack of suitable habitat in which to live and reproduce (Hewitt, 2000). Recolonization of areas following glacial recession has been shown to decrease genetic diversity of populations, however, many North American species were conserved during these glacial episodes in smaller, refuge populations to the south and in smaller oasis populations (Mee & Moore, 2014). Both relic and colonizing populations play a significant role in the genetic identity of contemporary populations.

Hewitt (2000) has suggested that refugial populations would have accumulated a large amount of genetic diversity as they grew, with new emigrants bringing unique alleles to the populations. Macrorefugia were areas of hospitable habitat south of the glacial terminus in the northern hemispheres. These represent areas that were very large and had sizeable species diversity (Mee & Moore, 2014). However, once glaciers began to recede, the genetic diversity of a population decreases as distance from the refugial site increases. Range expansion following glacial retreat has far reaching impact on terrestrial species. Studies of potential refugia populations *Lilium ceruum* in South Korea, found that the genetic variation of the suspected refugial population was significantly higher than that of range expanded populations that likely lost genetic variation during founding events in northeast China (Vu et al., 2018). Additionally Jordan et al. found that loss of genetic variation preceded deforestation events of the past 200 years and was attributable to an ancient event that coincided with deglaciation (Jordan, Morris, & Gibson, 2009).

In relation to the loss of genetic variation, species in formerly glaciated areas can vary in their response to environmental change, leading to specific population structure. European amphibian populations located in formerly glaciated areas had higher vulnerability to anthropomorphic stressors, such as pollutants from agricultural land use (Dufresnes & Perrin, 2014). Cold adapted lemmings have experienced a drop in genetic diversity since the last glacial maxima suggesting an important role of climate change in shaping population structure (Prost et al., 2010).

Aquatic species are also affected by glacial dynamics. Contemporary fish populations demonstrate the ancient influence of glacial cycles, with fish from glaciated regions exhibiting lowered genetic diversity and larger geographic range sizes than species which have not been directly exposed to glaciation events (Wang, Tsai, Yu, & Lee, 1999). Recolonization after glacial retreat is a gradual process and is believed to have occurred sometime shortly after ice sheet retreat (Ross, 2013). During glaciation, ice dams changed watershed distributions and formed lakes which acted as areas of refuge for plant and animal species.

Glaciation events produced low topographical relief in comparison to unglaciated regions which has resulted in increased agricultural land use. Agricultural land use in the previously Wisconsinan glacial region of Indiana is 89%, compared to 55% in the unglaciated regions (Jacquemin & Pyron, 2011b). Taxonomical and functional differences exist in the species

makeup of streams dependent on the stream's glaciation history, with pre-Wisconsinan streams and unglaciated regions having the largest difference (Jacquemin & Pyron, 2011a).

Contemporary effects of land use on genetic variation in aquatic systems

Colonizing events after a major disturbance are often manifested by a loss of genetic diversity (Hewitt, 2000). As populations move further and further away from a source population, the genetic diversity decreases as the distance from the source population grows. Habitat fragmentation will also increase genetic drift by reducing gene flow among populations. Populations that have experienced a greater amount of disturbance, such as agricultural dredging or a history of glaciation, are expected to have lower genetic diversity than undisturbed populations. Measures of allelic richness, and observed and expected levels of heterozygosity, have been the most commonly used indicators of the genetic diversity of a population (Bernatchez & Wilson, 1998).

Recent ecological disturbances on varying spatial and temporal scales often have indirect effects on the genetic variation of populations. The act of dredging drainage ditches in agricultural areas acts as a fine-scale disturbance on resident aquatic communities. Such fine-scale events have been demonstrated to cause loss of genetic diversity and heterozygosity within a population (Banks et al., 2013). Conversely, Jordan et al. (2013) found a higher amount of allelic richness following agricultural disturbance, which may be due to individuals recolonizing agricultural ditches after dredging from different source populations. Agricultural land use has also caused very large scale disturbance in aquatic habitats, through direct effects such as removal of riparian vegetation and indirect effects such as chemical run off from farming. Anthropomorphic modifications to natural streams, specifically ditching, change the water flow in an area by removing meanders and increasing discharge (Gorman & Karr, 1978). The removal of bank vegetation changes the suitability of the environment in multiple ways, silt levels increase decreasing water clarity and lack of shade increases solar heating and risk of algal blooms (Gorman & Karr, 1978).

Allelic richness is a useful indicator of decreases in population size or past bottleneck events (Foulley & Ollivier, 2006). Heterozygosity is influenced by allele frequencies rather than allele counts, therefore allelic richness takes on an even greater role when making conservation

decisions because the number of alleles is a strong indicator of evolutionary potential of a population (Greenbaum, Templton, Zarmi, & Bar-David, 2014).

Creek Chub biology and genetics

North American fish fauna consists of predominately lentic species and the neararctic zone is the most speciose of the temperate zoological realms. In North America, there are 297 species of Cyprinidae, accounting for 28% of the freshwater fish species on the continent (Ross, 2013). The distribution and genetic diversity of fish populations, is largely influenced by the island-like nature of their habitat (Bernatchez & Wilson, 1998). Creek chub (*Semotilus atromaculatus*) is a small, olive-brown minnow that occurs in small streams and rivers of the midwest to eastern North America, with populations extending from the northeast to southwestern United States (Schemske, 1974). *Semotilus atromaculatus* is so named because of their dark spotting and dorsal fin shape. Opportunistic by nature, creek chub can be found in a variety of habitats, with a preference for smaller, warm, headwater streams, consisting of patches of lotic habitat where they school close to the shoreline and seek shelter in weeded areas for protection. Previous studies have shown that creek chub will use larger streams and reservoirs as a means of migration but will not predominately reside in these habitats (Belica & Rahel, 2008). Similarly they will avoid non-moving water ways and choose to stay near the safety of the shore line. Creek chub distribution is characterized by native habitat, glacial recolonization and introduction through game fisherman (Boizard, Magnan, & Angers, 2009).

Creek chub live in both pristine and disturbed habitat, making them a useful species for the study of anthropomorphic habitat fragmentation as well as historical demography (Skalski, Landis, Grose, & Hudman, 2008). Common species are often used to draw conclusions about fish communities as a whole, especially when studying historical events' impacts on community structure (Whiteley, Spruell, & Allendorf, 2006). Creek chub demonstrate population level responses in life history traits, have a wide distribution, show site fidelity, and an affinity for small streams, allowing chub to be studied in a broad range of conditions (Filgueira, Chapman, Suski, & Cooke, 2016). For example, an examination of the stress hormone cortisol levels in the species found a significant increase in populations from predominantly agricultural areas. These findings suggested that there may be some natural resistance in place, allowing them to persist in degraded environments (Nagrodski, Suski, & Cooke, 2013).

Ice dams and the dynamic channelization created by glaciers parallel conditions created by man-made structures such as dams. The effects man-made dams and reservoirs have on creek chub populations have shown that geographical distance correlates with genetic distance between populations (Skalski et al. , 2008). Genetic variation is often effected by changes to a species' environment or a disturbance, however this can take some time to manifest in the gene pool. Nature is dynamic and there can be a delay between an ecological event and the genetic effect which will manifest. This is known as a time lag, time lags can make distinguishing between contemporary and ancient influences difficult, it is important to match an ecological questions with the correct genetic marker and analysis (Epps & Keyghobadi, 2015).

When conducting analyses of genetic variation, it is important to select a marker that has a level of polymorphism that matches the temporal scale of study. Microsatellite loci are well suited to population level studies at temporal scales spanning from recent to post-glacial time periods. Through slippage during transcription, microsatellites accumulate mutations faster than other markers. Microsatellites, or SSR's, consist of repeated motif units of DNA nucleotides ranging anywhere from 1 to 6 nucleotide repeats that are tandemly arrayed (Beaumont & Hoare, 2003). Microsatellites are also bi-parentally inherited which allows for the identification of heterozygotes and homozygotes. SSR's predominantly occur in areas of noncoding DNA and are therefore assumed to not be subjected to selective pressures (Avise, 1994).

Objectives

The main objective of this study is to examine if there has been a loss in genetic diversity of creek chub populations due to past glaciation events. Watersheds that were entirely glaciated will be compared to those which were partially glaciated. It is expected that the glaciated watershed will have lower genetic diversity and less genetic differentiation among sample localities when compared to the partially glaciated watershed. It will be possible to determine how genetically distinct sampling populations are from one another within watersheds (Boizard et al., 2009) and relate this variation to contemporary land use, considering that forested and agricultural land cover is found in both watersheds with both glacial histories in the study region.

MATERIALS AND METHODS

Study Area and Fish Sampling

Saint Joseph River Watershed Sites

Streams and agricultural ditches were studied in two sub-watersheds of the Saint Joseph River in northeastern Indiana and south-central Michigan (Figure 1) (Jordan, Patel, Sanders, & Gillespie, 2013). Land use in this area is predominantly agricultural, consisting of soybean and corn crop lands (United States Department of Agriculture's Statistic Service 2014). Streams of the Upper Cedar Creek watershed are often channelized, which are modified natural first order streams to increase water flow in an effort to prevent flood damage. This area also has high amounts of pesticide and fertilizer run-off, high amounts of stream terrestrial vegetation and very little traditional riparian habitat (Gorman & Karr, 1978). Seven locations were sampled across three streams within this watershed. The second sub-watershed, East Branch, is located in Michigan and has a combined land cover of deciduous forest and agricultural land use. Three locations were sampled on each of two channelized streams: one located in a forested game preserve and the other with agricultural land cover similar to that of the Upper Cedar Creek sites.

Little Miami River Watershed Sites

The Little Miami River watershed (Figure 2.) is located in southwestern Ohio, USA. and is transected by the Wisconsinan glacial terminus. Collections of creek chubs were made at 34 sites over a four year period from 1999-2003 following the procedures and collection methods described in a study of central stonerollers (Blum et al., 2012). The northernmost sampling sites experienced glaciation and are characterized by low, rolling streams and a flatter terrain than the area south of the terminus. Sampling sites south of the terminus have a more rugged terrain. Land cover in areas which experienced glaciation have a rich sediment deposit and are often used for row crop agricultural. Areas which did not experience glaciation also have agricultural land use, however, there are often more deciduous forests in areas surrounding the streams (Blum et al., 2012). Sampling sites were also chosen with respect to proximity to confluence points between the tributary and mainstream river and are representative "pour" points.

Sample Collection

Fish collected in the SJR were sampled using a backpack electro fisher (100-150-V, 60-Hz, DC Current) over a 125 meter reach at each locality. Sampling of this area initially occurred as a part of a broader study on fish communities in agricultural streams. Creek chub were euthanized with buffered triacane methane sulphonate (MS-222, 250mg/L) and flash frozen on dry ice. Sampling occurred between in 2006-2007 in the Cedar Creek sub-watershed and in 2007-2009 in the East Branch sub-watershed of the Saint Joseph River (Jordan et al., 2013).

LMR fish were sampled in five consecutive summers (1999-2003), using a back pack electro fisher across a 150 meter reach over a thirty minute period at each sampling site. Fish were collected, species identified, and euthanized. The complete methods of fish sampling within the LMR are as described in Blum et al. (2012).

Microsatellite Genotyping

Microsatellites are a useful genetic marker for conservation and population genetics studies due to their high distribution throughout the nuclear genome and their polymorphic nature (Putman & Carbone, 2014). For creek chubs collected from SJR, approximately 20 mg of muscle was dissected from each fish and digested in a solution containing proteinase K at 55 °C. Genomic DNA was extracted using columns lined with silica membrane (DNeasy, Blood & Tissue Kit QIAGEN Sciences, Germantown,MD, U.S.A). These samples were initially genotyped using nine fluorescently labeled primers, in a previous study conducted by Jordan et al. (2013), however, only one of these nine loci was congruent with unpublished data (described below). To make the LMR dataset comparable to the samples collected in SJR, seven microsatellite primers were amplified using fluorescently labelled primers [6-FAM (Blue), HEX (Green), TAMRA (Yellow)] using the polymerase chain reaction (PCR) (Table 1, APPENDIX A).

Touchdown PCR was used to improve amplification specificity. Touchdown PCR uses a cycling method in which the initial annealing temperature is set 10-15°C above the estimated melting temperature (Tm) of the primer being used. By starting the initial annealing process above the estimated Tm, the PCR reaction gets a head start on amplifying the target sequence and eliminates the need for optimization of the reaction. The annealing temperature is lowered

every cycle until Tm is reached, and then annealing continues at this temperature until the cycling is complete. The touchdown method takes advantage of the exponential nature of PCR product, by lowering the annealing temperature every cycle non-specific product is essentially eliminated or subjugated (Korbie & Mattick, 2008).

The PCR reaction conditions (APPENDIX A) are as follows for one 12 µl reaction for loci Seat212, Seat416, Seat204, Seat402, Seat205:PCRs containing 2ng DNA, 1× PCR buffer, 1.5 mM $MgCl_2$ 0.25 µM CAG tag, 0.25 µM forward primer, 0.025 µM tagged primer, 0.05 µM dNTPs, 0.5 U Taq, and 25-50 ng of DNA (Skalski & Grose, 2006). PCR conditions were chosen to mimic previous lab work conducted by Blum et al. (2012). Seat409 samples were ran using the methods described in Jordan et al. (2013).

Amplification problems arose for Seat209 and Seat403 necessitating the modification of PCR conditions and thermocycling conditions (APPENDIX A,). In order to correct for Seat403 (Tables 13, 14), the MgCl concentration was increased from 1.5mM to 2.5 mM, all other reagent concentrations remained the same. Seat 209 reaction conditions were altered to include a higher concentration of Taq to improve amplification. (APPENDIX A) All other reagent concentrations remained the same. Touchdown 65ºC buffer 1 µM $MgCl_2$ 1.5 µM, CAG tag 0.25 µM, forward primer 0.25 µM, reverse primer 0.025 µM, dNTPs 0.05 µM, 0.75 µM Taq 5u/uL, and 25-50 ng of DNA (Garrick T. Skalski & Grose, 2006).

For LMR samples, QIAGEN DNeasy kits for animal tissue (QIAGEN, Valencia California, USA) were used to extract genomic DNA from 2,338 fish, which were genotyped at eight different microsatellite loci. The PCR reactions conditions were as described by Skalski & Grose (2006), PCRs containing 10 ng of DNA, and 1× PCR buffer 1.5 mM $MgCl_2$. 0.25 µM CAG primer, 0.25 µM unlabelled primer, 0.025 µM tagged primer, 50 µM dNTP's, 0.5 U Taq. All reactions were run on an MJ Research Dyad with fluorescently labeled CAG primers (APPENDIX A).

SJR samples were prepared with a 50:1 dilution of formamide, and multiplexed with either two or three loci per plate using three different florescent labels. The groupings in which loci were multiplexed together were also designed to segregate by expected fragment size (Table 1). Samples were electrophoresed at the Yale University DNA Analysis Facility in New Haven, Connecticut and genotyped using GENEMAPPER 3.7. For LMR samples, labeled PCR

amplicons were characterized using a MJ Research Basestation Genetic analyzer and Cartographer® software.

To make the allele calls of the datasets comparable it was necessary to genotype a sample of LMR fish for each locus using the conditions described for SJR. I chose to convert the SJR samples and not the LMR sample allele calls because SJR is a smaller sample and meant less samples would be manipulated.

Landscape Characterization

Landscape characterization was done using L-Thia software (Engel & Theller, 2016). The coordinates of each sampling site were used to provide proportional estimates of land cover types in the watershed upstream of the point. Estimates of land cover were done using the online water delineation (OWL) tool found with-in the L-Thia web application (Engel & Theller, 2016). OWL provides six categories of land cover (water, commercial, agriculture, residential, grass/pasture and forest). These six categories were collapsed down to three, being Agriculture (agriculture, grass/pasture), Forest (forest) and Other (Water, commercial, residential). Landscape characterization was therefore somewhat confounded by the fact that any sampling site downstream of another is affected by its' land coverage and use. For example, a predominantly wooded site downstream of a heavily agricultural area will still manifest the effects of agricultural land use without its self being farmed. Absolute value of the difference in percentage of land cover between localities was used to perform correlation evaluations in PASSaGE (Rosenberg & Anderson, 2011).

Statistical Analysis

Hardy-Weinberg equilibrium (HWE) was examined using GENEPOP for the Web v. 1.2 (Peakall, 2012; Rousset, 2008). GENEPOP software uses exact tests for HWE with the null hypothesis of random union of gametes (Rousset, 2008). Bonferroni correction was applied sample-wide to adjust for Type I error in multiple comparisons [alpha = 0.05/336 tests (42 localities at 8 loci) = $1.5 * 10^{-4}$]. GENEPOP was also used to test for linkage disequilibrium, the nonrandom association between alleles at two loci. Here the null hypothesis is that genotypes at one locus are independent from genotypes at the other locus. The test statistic is the log likelihood ratio statistic (G-test), and Bonferroni correction was applied as described above.

Population structure was examined using four different methods: *F*st, Jost's *D*, AMOVA, and STRUCTURE. F-statistics are used to characterize population genetic structure. Jost's *D* is used as an analogous test of population differentiation, alongside *F*st. AMOVA allows the hierarchical partitioning of genetic variation among populations and predetermined regions. STRUCTURE is used to remove the *a priori* assignment of individuals to sample localities to assess the level of clustering across the dataset

F-statistics are a measure of the deficit of heterozygotes relative to the expected amount under Hardy-Weinberg proportions (Wright, 1943). *F* is the amount that heterozygosity has been reduced, relative to random mating. *Fst* is an integer between zero and one, an *Fst* value of 1 would indicate that alleles are fixed, throughout subpopulations, while an *Fst* value of 0 indicates an equal allele frequency throughout sub populations. *F*-statistics were calculated in GenAlex for correlation tests, and measures of genetic differentiation.

Jost's *D* is a measure of allelic differentiation, rather than a measure of nearness to fixation. *D* has a value of 1 when demes share no alleles and a value of 0 when the same allele is fixed across all demes. Fixation indices and differentiation measures are both important in conservation and population genetics because they provide different insights in to the genetic breakdown of populations (Jost et al., 2018).

Analysis of Molecular Variance (AMOVA) allows for tests of differentiation at different hierarchical levels. I used it to test for: a difference between formerly glaciated and unglaciated sites in the LMR, between glaciated and unglaciated sites, and between SJR and LMR. Each AMOVA was ran using 999 permutations. AMOVA provides several different F statistics as estimates of differentiation, *Frt* is the estimated variance among regions, *Fsr* is the estimate variance among populations within regions, *Fst* is the inbreeding coefficient within subpopulations, relative to the total. *Fst* provides a measure of the genetic differentiation among populations, *Fis* is an estimated variance among Individuals, Fit is a measure of departure from HWE in the entire base population. Bonferroni correction was applied sample-wide to adjust for Type I error in multiple comparisons [alpha = 0.05/3 tests (3 AMOVA comparisons) = $1.7 * 10^{-2}$].

Population structure was further examined in Bayesian cluster analysis using STRUCTURE v.2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). The number of potential populations in the overall sample (K) was accessed using varying

K values between 1 and 20 for the full dataset. Due to the large geographic distance and likely lack of gene flow between SJR and LMR, I also conducted the analysis within each watershed. For LMR, K was assessed for values between 1 and 10 while for the SJR, K was varied from 1 to 13. Previous work the fish populations in SJR found a K value of three, however because different loci were used for this examination the full range of K was investigated (Mark A. Jordan et al., 2013). A preliminary run on the LMR data was conducted for K values 1-20 and found that the K value was low. The data were rerun in STRUCTURE with more permutations at a smaller interval of K between 1 and 10. For each grouping of sample localities, the models were run ten times using a burn in of 5×10^4 followed by 5×10^5 steps (Jordan et al.(2013). Structure Harvester was used to summarize the results, find the highest log likelihood of K [L(K)], and its rate of change between the hypothesized values (ΔK) to estimate the number of clusters in the dataset (Earl & von Holdt, 2011). Once K was identified, the posterior probabilities of assignment of each individual to each cluster (Q) were consolidated from 10 independent runs using CLUMPP in CLUMPAK (Rosenberg, 2004)**.**

Measures of genetic diversity within and among sample localities were made with a range of packages. I used FSTAT to estimate allelic richness. FSTAT utilizes both Nei and Weir-Cockerham estimations of gene diversities and F statistics using randomization methods. FSTAT was developed to eliminate potentially erroneous results of previous genetic differentiation programs by properly handling missing data, small sample sizes and using random number generators to permute the data (Goudet, 1995). The data in this particular study was permuted 999 times. FSTAT does not provide standard error for tests and this was calculated in Excel for allelic richness. Allelic richness is often proportionate to sample area and sample size, therefore it is expected that the Little Miami watershed will have a larger allelic richness and a larger number of overserved alleles than the Saint Joseph River watershed. GenAlex application was used to calculated Observed (Ho) and unbiased expected heterozygosity (He), along with their standard error (Peakall, 2012). The null hypothesis of which is that there is no significant difference. Allelic richness and observed levels of heterozygosity were organized by cluster for the LMR and SJR watersheds, as a means of demonstrating population hierarchy. Analysis of Variance (ANOVA) was performed in excel to determine if the two watersheds were genetically different. ANOVA was also used to determine if glaciation history had any significant genetic

difference between populations. STRUCTURE clusters were also examined with ANOVA to determine if they were statistically different as well.

Possible correlates of population structure were investigated using pairwise matrices of genetic distance (*Fst*) calculated using GenAlex software in excel, geographic distance, glaciation, and percent agricultural land cover. Geographic distance, for both SJR and LMR sampling sites were curvilinear stream distances in km and not Euclidean distances. Correlates were examined using Pattern analysis, spatial Statistics and Geographical Exegesis (PASSaGE) using Mantel and partial Mantel tests at 999 permutations (Rosenberg & Anderson, 2011). The glaciation cover matrix consisted of a binary score with 0 for being on the same side of the glacial boundary, and 1 for different sides. Agricultural land cover was the absolute value of the difference in percentage of land cover between localities based on Lthia land coverage data.

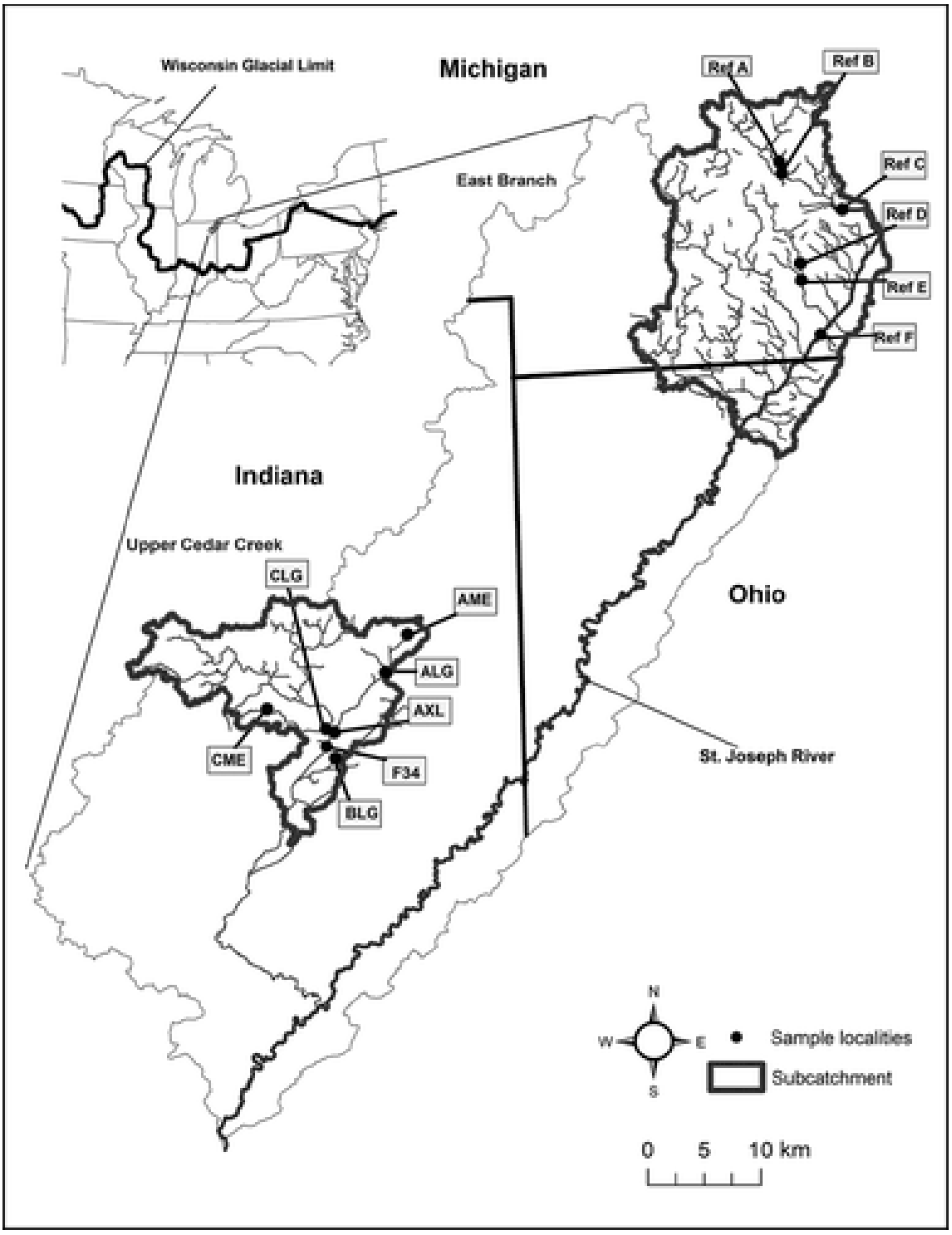


Figure 1 The St. Joseph River watershed and two focal sub-watersheds in Indiana, Michigan, and Ohio, USA. Dots are the locations of the 13 sampling sites. The solid line in the inset map shows the position of the Wisconsinan glacial maximum. (Jordan et al., 2013)

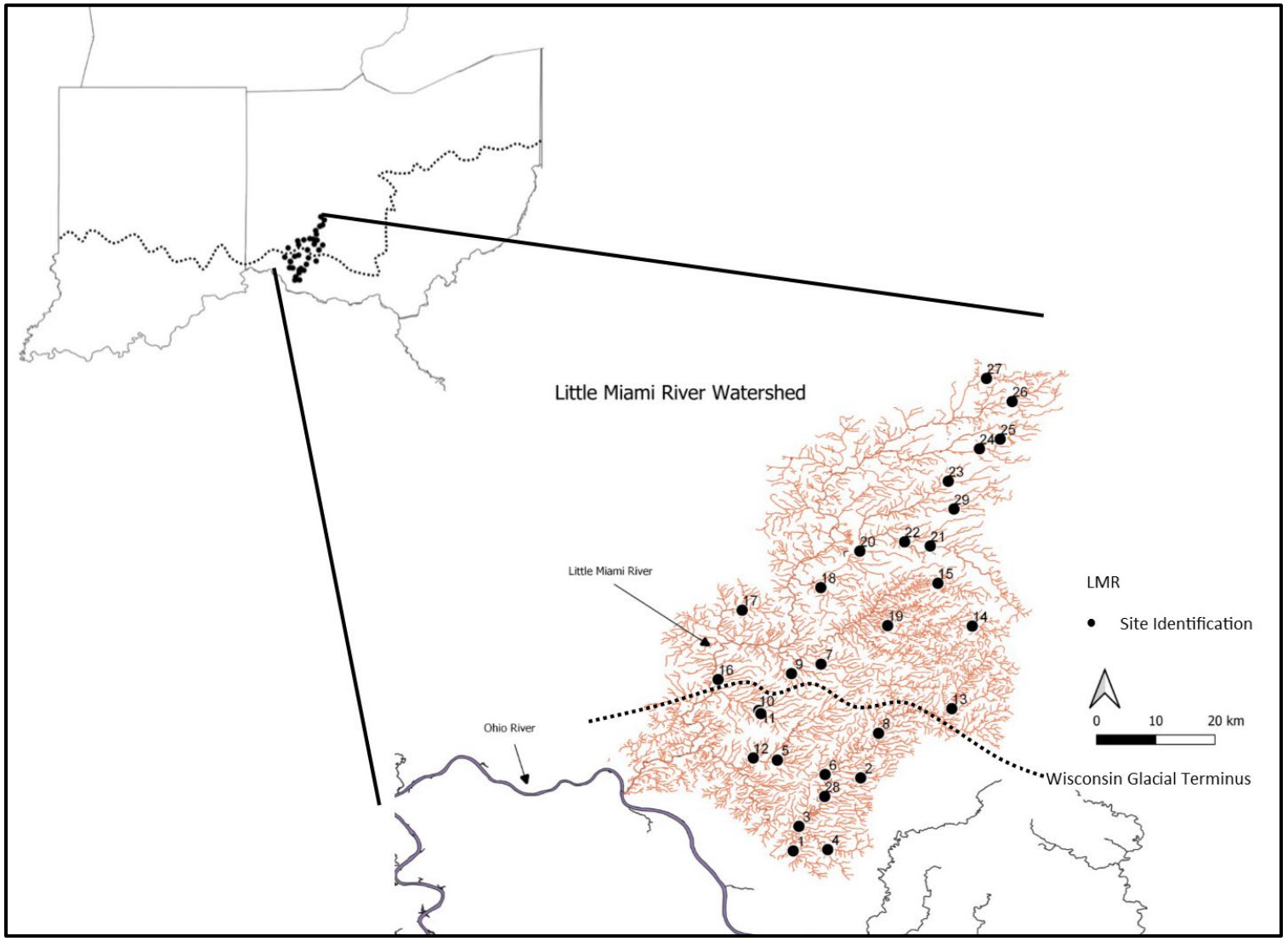


Figure 2 Little Miami River basin In Ohio, USA Dots are locations of 29 sampling sites. The dotted line shows the position of the Wisconsin glacial terminus

Table 1. Microsatellite loci used in the study. Repeat motif and size range were originally described in Skalski et al (2006). The fluorescent label was altered in some loci but the majority were in agreement with that which is stated in the table. Size differential refers to the difference in allele sizes characterized by the Blum lab and those obtained in this study. This information was used to homogenize allele calls between the SJR and LRM datasets. Repeat motif for Seat212 was provided through personal correspondence with Dr. S. P. Hudman, at Truman State University, March 5, 2019.

Locus	Fluorescent Label	Repeat motif	Size Range	Size Differential Between Labs
Seat 212	HEX	$(GT)_{28}$	144-166	14
Seat 403	6-FAM	$(GTAT)_7AC(GTAT)_2$	194-202	12
Seat 409	TAMRA	$(TGAT)_{16}$	219-247	18
Seat 205	HEX	$(AG)_{16}$	196-210	14
Seat 209	6-FAM	$(GA)_{16}$	226-240	17
Seat 402	TAMRA	$(TGTT)_8$	173-185	15
Seat 204	HEX	$(AG)_{15}$	199-214	16
Seat 416	6-FAM	$(GTTT)_7$	164-168	16

RESULTS

Tests of Equilibrium and Summary Statistics

Overall 2,638 creek chub were sampled from 47 locations across the two separate watersheds. Five sample localities from LMR were not included in the following statistical analyses due to low sample size (<5 individuals, Blum et al. 2012). Excluding these sites, there were 312 individuals genotyped from 13 sites in SJR and 2,318 individuals from 29 sites in LMR. The number of alleles per locus in SJR ranged from 2 to 28 while in the LMR it was from 1 to 31 (APPENDIX B). Deviations from Hardy-Weinberg equilibrium were found at several individual loci within sample locality (Table 1). However, there were no consistent deviations for individual loci or sites and therefore no populations or loci were excluded from further statistical analysis. Linkage disequilibrium was found in four pairs of loci across 1175 possible combinations (Seat209 and Seat204 in sites 2 and 5, Seat209 and Seat205 in site 16 and Seat205 and Seat403 in site 42).

Population Structure

Global *F*st across the full dataset was 0.243± 0.052 (p=0.001). Within watersheds, the values were lower with SJR showing slightly more differentiation than LMR (SJR: *F*st = 0.081 ± 0.016 [p = 0.001]; LMR: *F*st = 0.059 ± 0.005 [p = 0.001]). The average overall *D* for LMR was 0.056± 0.017 (p = 0.001).

AMOVA was used to test hierarchical hypotheses related to glacial history and geography. When combined data were classified as two separate watersheds, 34% of the variation was attributed to difference between regions (*F*rt = 0.335, Table 5) with relatively little differentiation among sample localities within each watershed (3%, *F*sr = 0.049). Given the geographic separation of the two watersheds, it is not unexpected that the predominant cause of their genetic differentiation would be attributable to region. Classifying localities by glacial history, regardless of watershed, resulted in only 3% of genetic variation accounted for at the highest hierarchical level (*F*rt = 0.026). The hypothesized effect of glacial history was even weaker when localities within LMR alone were used, with <1% of the genetic variation attributed to this level.

Bayesian Cluster Analysis

STRUCTURE supported the presence of two clusters within the full dataset that corresponded well to the geographical separation of the watersheds (Figure 3) suggested by the *F*st and AMOVA analyses.

Due to the strong differentiation at the regional level, I ran the analyses on each watershed alone to evaluate possible patterns of finer-scale differentiation. STRUCTURE identified three distinct clusters within the LMR watershed (Figure 4) but with high degree of admixture within localities (Figure 5, Table 6). The geographic position of samples localities was correlated with the cluster partitioning. Cluster 1 corresponds with the East Fork of the LMR branch. Sites 5 and 12 also fall with in this cluster. Although this these localities are not within the East Fork branch with the other Cluster 1 sites, the streams are connected to a tributary downstream of a reservoir. Cluster 2 corresponds with the medial sampling sites, located along Caesar's Creek and downstream towards the Little Miami main branch. The final group (Cluster 3) was found in the northernmost sampling sites along and near Massies Creek.

Three clusters within the SJR watershed were found to be the most likely outcome in STRUCTURE (Figures 6 & 7, Table 7). Similar to the outcome in LMR, there was substantial admixture within each cluster suggesting weak differentiation. Cluster 1 was associated with Upper Cedar Creek sample localities while the other two clusters are within the East Branch subwatershed.

Isolation by Distance, Land Cover, and Glaciation

Possible correlations between genetic distance, stream distance, and agricultural land cover and glaciation history were analyzed using Mantel and partial Mantel tests in PASSaGE (Table 8). Analyses were run within the two watersheds. A positive correlation between stream distance and genetic differentiation was found in LMR ($r = 0.14$, $p = 0.027$). When stream distance was held constant in a partial Mantel test, there was a positive correlation between genetic distance and glacial history ($r = 0.31$, $p = 0.001$) but also between genetic distance and the absolute difference in agricultural land cover ($r = 0.32$, $p = 0.002$). The relationship between agricultural land cover and glaciation history was strongly correlated, ($r = 0.99$, $p = 0.001$), making it difficult to separate the two potential effects.

In SJR, there was a positive correlation between stream distance and genetic distance in *(r* = 0.20, p = 0.002; Table 8). When holding stream distance constant, no relationship between absolute difference in percent agricultural land cover and genetic distance was recovered. (*r* = -0.003, p = 0.28).

Genetic Variation

The mean allelic richness of individual loci within LMR ranged from 1.48 to 4.64, with a mean over all loci of 2.59 (standard error ±0.03; Table 8). Allelic richness in SJR ranged from 1.60 to 5.57 and had a mean of 3.19 ± 0.09 over all loci. When tested with ANOVA, the higher value in SJR was statistically supported ($F_{1,40}$ =193.99, $p \leq 0.001$). A similar difference was found using observed heterozygosity ($F_{1,40}$ = 22.35, $p \leq 0.001$).

The expectation that glaciation reduces genetic variation was not supported by the data (Table 9). When LMR was divided by glaciation classification, there was no difference in allelic richness ($F_{1,27}$ = 0.45, $p = 0.50$) or observed heterozygosity ($F_{1,27}$ = 0.82, $p = 0.37$). When SJR localities are added to the glaciated group, allelic richness ($F_{1,40}$ = 9.57, $p = 0.004$) and observed heterozygosity was found to be higher in glaciated sites ($F_{1,40}$ = 7.41, $p = 0.01$).

Genetic variation among STRUCTURE clusters within watersheds (Table 9). Within LMR, three clusters were not found to be different in allelic richness ($F_{2,26}$ = 2.7, $p = 0.086$) or observed heterozygosity ($F_{2,26}$ = 0.13, $p = 0.88$) In SJR there was also no difference found in either measure (A: $F_{2,10}$ = 0.57, $p = 0.58$; Ho: $F_{2,10}$ = 0.40, $p = 0.68$).

Table 2 Sites with deviations from Hardy-Weinberg equilibrium for each locus following sample-wide Bonferroni correction ($p <= 0.0001$)

Locus	Sites
Seat212	38
Seat209	30, 33, 34
Seat402	37,38
Seat204	33
Seat409	-
Seat205	32
Seat416	-
Seat403	30, 32, 42

Table 3 Pairwise *Fst* values for SJR sampling sites. *Dest* could not be calculated due to missing data.

	Pop 30	Pop 31	Pop 32	Pop 33	Pop 34	Pop 35	Pop 36	Pop 37	Pop 38	Pop 39	Pop 40	Pop 41	Pop 42
Pop 30	***												
Pop 31	0.020	***											
Pop 32	0.009	0.014	***										
Pop 33	0.014	0.026	0.010	***									
Pop 34	0.009	0.022	0.008	0.010	***								
Pop 35	0.014	0.037	0.024	0.026	0.021	***							
Pop 36	0.021	0.034	0.024	0.024	0.021	0.026	***						
Pop 37	0.011	0.024	0.017	0.020	0.014	0.025	0.025	***					
Pop 38	0.058	0.088	0.068	0.072	0.061	0.069	0.068	0.061	***				
Pop 39	0.064	0.086	0.070	0.080	0.069	0.081	0.079	0.071	0.028	***			
Pop 40	0.019	0.030	0.023	0.031	0.025	0.027	0.028	0.027	0.053	0.058	***		
Pop 41	0.019	0.037	0.024	0.027	0.023	0.026	0.031	0.030	0.055	0.063	0.019	***	
Pop 42	0.079	0.100	0.082	0.079	0.079	0.074	0.079	0.091	0.112	0.125	0.091	0.087	***

Table 4 Pairwise *Fst* and *Dest* values for Little Miami River sites. Pairwise *Fst* values are below the diagonal and pairwise *Dest* above.

	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10	Pop11	Pop12	Pop13	Pop14	Pop15	Pop16	Pop17	Pop18	Pop19	Pop20	Pop21	Pop22	Pop23	Pop24	Pop25	Pop26	Pop27	Pop28	Pop29
Pop1	***	0.026	0.020	0.019	0.029	0.035	0.070	0.023	0.108	0.070	0.050	0.002	0.019	0.043	0.062	0.056	0.028	0.110	0.062	0.060	0.040	0.031	0.053	0.146	0.113	0.074	0.077	0.017	0.047
Pop2	0.034 ***		0.018	0.035	0.040	0.012	0.057	0.007	0.093	0.059	0.057	0.027	0.011	0.054	0.055	0.074	0.036	0.086	0.059	0.051	0.035	0.044	0.059	0.161	0.137	0.062	0.086	0.006	0.050
Pop3	0.032	0.014 ***		0.019	0.032	0.041	0.067	0.029	0.109	0.050	0.046	0.015	0.030	0.055	0.066	0.084	0.039	0.105	0.075	0.059	0.041	0.049	0.074	0.154	0.118	0.073	0.092	0.022	0.068
Pop4	0.030	0.021	0.014 ***		0.037	0.028	0.072	0.025	0.097	0.047	0.036	0.017	0.023	0.049	0.070	0.066	0.038	0.091	0.064	0.054	0.047	0.047	0.069	0.144	0.118	0.064	0.076	0.024	0.062
Pop5	0.033	0.022	0.018	0.020 ***		0.029	0.073	0.035	0.103	0.072	0.065	0.009	0.033	0.048	0.079	0.072	0.033	0.105	0.071	0.058	0.048	0.051	0.074	0.122	0.101	0.058	0.079	0.018	0.067
Pop6	0.040	0.014	0.027	0.019	0.019 ***		0.044	-0.003	0.074	0.051	0.048	0.024	0.000	0.039	0.057	0.047	0.037	0.065	0.042	0.035	0.036	0.040	0.039	0.124	0.123	0.041	0.066	0.002	0.041
Pop7	0.052	0.031	0.038	0.037	0.035	0.026 ***		0.065	0.032	0.049	0.048	0.053	0.077	0.033	0.028	0.017	0.040	0.066	0.020	0.029	0.025	0.032	0.043	0.103	0.102	0.054	0.086	0.051	0.060
Pop8	0.031	0.007	0.017	0.014	0.017	0.006	0.031 ***		0.083	0.061	0.056	0.025	0.003	0.047	0.063	0.067	0.039	0.087	0.061	0.049	0.042	0.046	0.048	0.152	0.137	0.054	0.071	0.005	0.039
Pop9	0.066	0.046	0.057	0.047	0.046	0.037	0.017	0.037 ***		0.074	0.069	0.084	0.108	0.040	0.047	0.038	0.070	0.077	0.044	0.042	0.054	0.061	0.066	0.096	0.098	0.062	0.090	0.082	0.070
Pop10	0.053	0.032	0.030	0.026	0.035	0.029	0.026	0.030	0.037 ***		0.007	0.041	0.068	0.050	0.074	0.053	0.046	0.085	0.062	0.019	0.043	0.043	0.048	0.088	0.075	0.036	0.049	0.052	0.062
Pop11	0.046	0.034	0.030	0.022	0.035	0.030	0.028	0.030	0.038	0.007 ***		0.034	0.061	0.049	0.071	0.055	0.040	0.084	0.058	0.033	0.043	0.047	0.060	0.102	0.083	0.039	0.051	0.047	0.076
Pop12	0.022	0.017	0.011	0.011	0.007	0.017	0.027	0.014	0.039	0.022	0.020 ***		0.026	0.033	0.059	0.054	0.023	0.085	0.053	0.038	0.030	0.030	0.045	0.112	0.090	0.042	0.060	0.015	0.047
Pop13	0.029	0.009	0.018	0.014	0.017	0.007	0.038	0.004	0.050	0.035	0.034	0.014 ***		0.060	0.081	0.075	0.046	0.102	0.073	0.055	0.048	0.050	0.059	0.161	0.144	0.059	0.072	0.003	0.055
Pop14	0.039	0.028	0.030	0.025	0.023	0.023	0.018	0.022	0.020	0.026	0.027	0.017	0.028 ***		0.024	0.023	0.025	0.045	0.021	0.016	0.019	0.016	0.038	0.064	0.051	0.046	0.066	0.034	0.035
Pop15	0.049	0.031	0.038	0.037	0.038	0.032	0.016	0.031	0.025	0.039	0.041	0.031	0.040	0.014 ***		0.028	0.026	0.029	0.005	0.049	0.030	0.032	0.057	0.131	0.112	0.076	0.097	0.054	0.048
Pop16	0.045	0.036	0.043	0.032	0.032	0.026	0.011	0.030	0.019	0.027	0.030	0.026	0.034	0.013	0.016 ***		0.027	0.061	0.012	0.011	0.019	0.014	0.032	0.079	0.078	0.040	0.050	0.048	0.040
Pop17	0.037	0.023	0.026	0.024	0.020	0.026	0.024	0.022	0.037	0.027	0.027	0.016	0.026	0.016	0.018	0.018 ***		0.054	0.023	0.022	0.011	0.012	0.038	0.103	0.082	0.034	0.044	0.017	0.032
Pop18	0.071	0.046	0.059	0.048	0.051	0.036	0.035	0.042	0.040	0.046	0.049	0.043	0.051	0.024	0.018	0.031	0.032 ***		0.031	0.077	0.067	0.070	0.098	0.154	0.142	0.082	0.112	0.079	0.083
Pop19	0.048	0.031	0.040	0.033	0.033	0.024	0.012	0.029	0.022	0.032	0.032	0.026	0.035	0.012	0.005	0.008	0.016	0.018 ***		0.033	0.021	0.024	0.043	0.105	0.098	0.055	0.076	0.047	0.046
Pop20	0.051	0.031	0.036	0.032	0.031	0.026	0.020	0.027	0.025	0.016	0.023	0.024	0.031	0.014	0.030	0.012	0.020	0.043	0.021 ***		0.002	0.000	0.011	0.033	0.029	0.014	0.024	0.029	0.024
Pop21	0.039	0.021	0.025	0.026	0.024	0.023	0.015	0.022	0.028	0.024	0.026	0.017	0.025	0.012	0.018	0.012	0.011	0.037	0.013	0.008 ***		0.000	0.023	0.069	0.056	0.032	0.047	0.022	0.028
Pop22	0.035	0.024	0.027	0.024	0.024	0.023	0.017	0.022	0.029	0.023	0.026	0.016	0.025	0.010	0.018	0.009	0.010	0.036	0.013	0.007	0.003 ***		0.016	0.070	0.057	0.035	0.042	0.026	0.017
Pop23	0.042	0.028	0.036	0.031	0.031	0.021	0.020	0.021	0.029	0.023	0.030	0.021	0.026	0.018	0.027	0.015	0.021	0.044	0.020	0.012	0.013	0.009 ***		0.079	0.085	0.042	0.048	0.047	0.012
Pop24	0.080	0.073	0.075	0.065	0.052	0.056	0.048	0.063	0.043	0.042	0.052	0.050	0.069	0.030	0.060	0.035	0.050	0.072	0.047	0.021	0.034	0.033	0.033 ***		0.005	0.077	0.090	0.123	0.101
Pop25	0.073	0.070	0.065	0.060	0.049	0.061	0.053	0.063	0.049	0.041	0.049	0.046	0.069	0.028	0.058	0.039	0.046	0.074	0.050	0.022	0.032	0.030	0.039	0.007 ***		0.071	0.078	0.103	0.096
Pop26	0.053	0.033	0.041	0.033	0.028	0.025	0.028	0.026	0.031	0.020	0.023	0.022	0.030	0.023	0.039	0.021	0.021	0.043	0.028	0.013	0.018	0.019	0.020	0.037	0.038 ***		0.006	0.038	0.046
Pop27	0.055	0.045	0.051	0.039	0.038	0.035	0.043	0.034	0.044	0.027	0.030	0.030	0.036	0.033	0.049	0.025	0.026	0.058	0.038	0.018	0.026	0.022	0.023	0.043	0.042	0.006 ***		0.055	0.052
Pop28	0.030	0.008	0.016	0.015	0.012	0.009	0.027	0.006	0.040	0.028	0.028	0.011	0.005	0.019	0.029	0.025	0.014	0.042	0.025	0.021	0.014	0.015	0.022	0.056	0.053	0.021	0.029 ***		0.042
Pop29	0.039	0.024	0.033	0.028	0.028	0.022	0.027	0.017	0.030	0.028	0.036	0.021	0.024	0.016	0.023	0.018	0.018	0.037	0.021	0.016	0.015	0.010	0.007	0.040	0.043	0.021	0.024	0.020 ***	

Table 5. Analysis of Molecular Variance (AMOVA) results for the three analyses that partition localities by geography, glacial history overall and within the Little Miami River watershed. Estimated variance among regions (*F*rt), the estimate variance among populations within regions (*F*sr), overall fixation index (*F*st), and estimated variance among individuals (*F*is). *F*it is a measure of departure from HWE in the entire base population.

Comparison	Sources of Variation	df	% Variation	*F*-statistic	p-value
Two separate watersheds	Among Regions	1	34%	*F*rt= 0.335	<0.001
	Among Populations	40	3%	*F*sr= 0.049	<0.001
	Among Individuals	2588	6%	*F*st= 0.368	<0.001
	Within Individuals	2630	57%	*F*is= 0.101	<0.001
	Total	5259	100%	*F*it= 0.432	<0.001
Glaciated and unglaciated (both watersheds)	Among Regions	1	3%	*F*rt= 0.026	<0.001
	Among Populations	40	13%	*F*sr= 0.131	<0.001
	Among Individuals	2588	8%	*F*st= 0.154	<0.001
	Within Individuals	2630	76%	*F*is= 0.101	<0.001
	Total	5259	100%	*F*it= 0.239	<0.001
Glaciated and unglaciated (LMR alone)	Among Regions	1	1%	*F*rt= 0.009	<0.001
	Among Populations	27	4%	*F*sr= 0.044	<0.001
	Among Individuals	2289	6%	*F*st= 0.053	<0.001
	Within Individuals	2318	89%	*F*is= 0.059	<0.001
	Total	4635	100%	*F*it= 0.108	<0.001

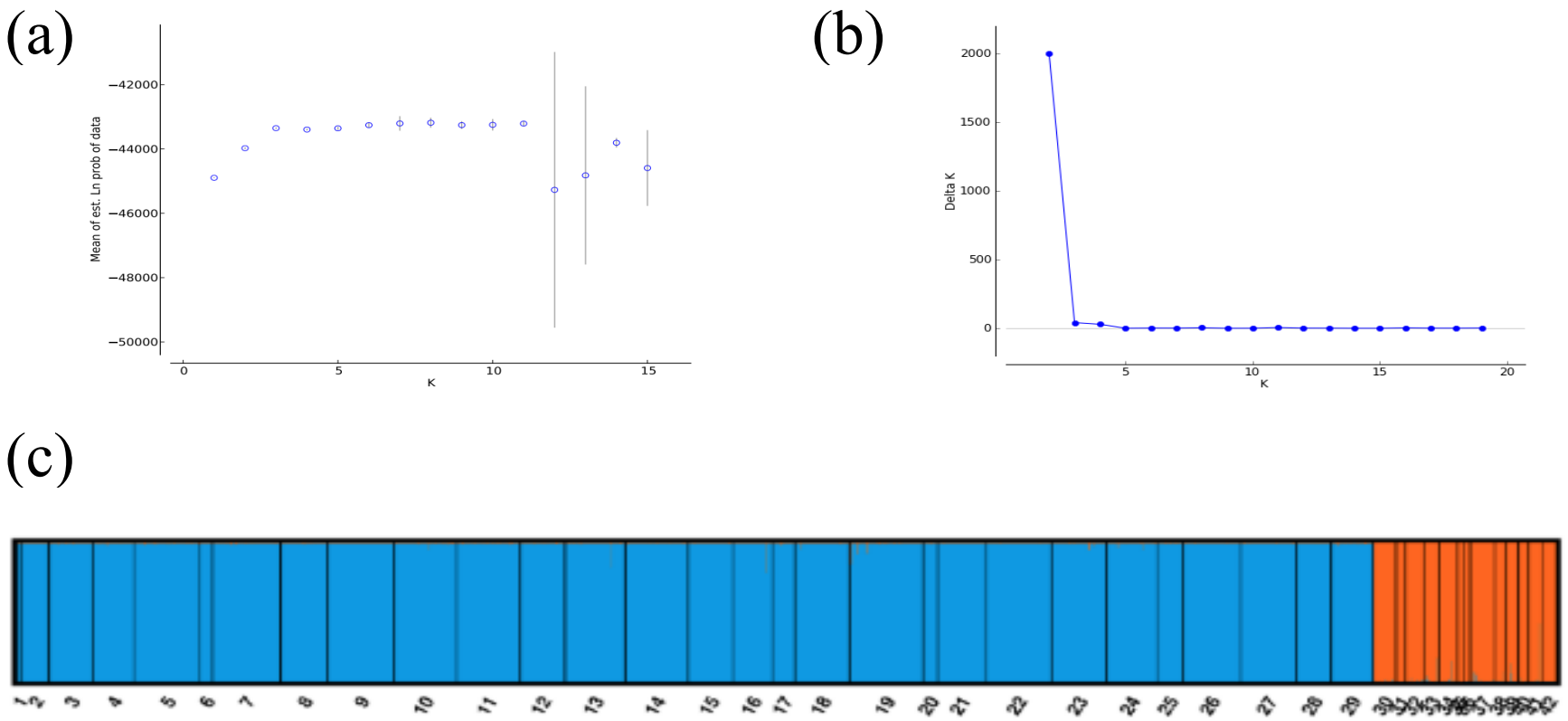
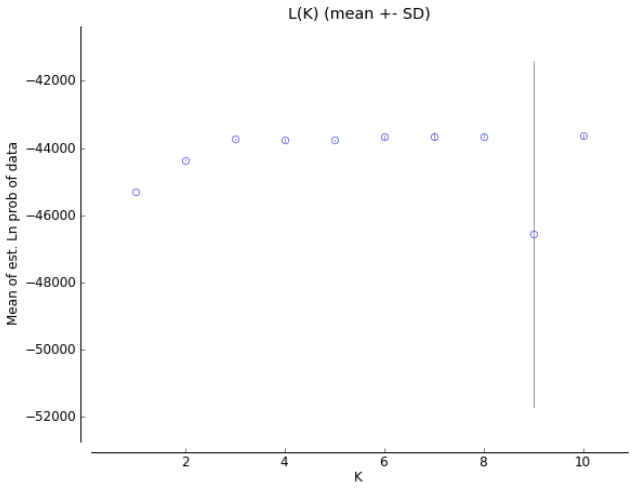
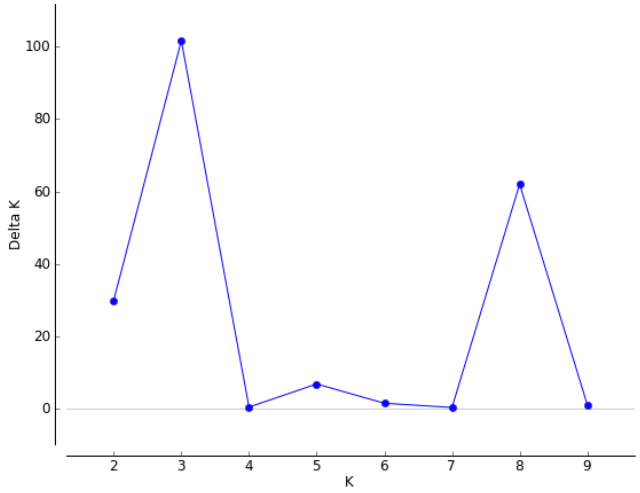


Figure 3 The outcome of STRUCTURE analysis of the combined Little Miami River and Saint Joseph River samples. The number of putative populations (K) is related to: (a) the natural log likelihood of the data as a function of K[L(K)] and (b) the rate of change of the likelihood function (ΔK) (Evanno, Regnaut, & Goudet, 2005) and (c) the proportional membership of individuals (Q) in the three clusters identified from the dataset (Orange = Cluster 1, Blue = Cluster 2).

(a)



(b)



(c)

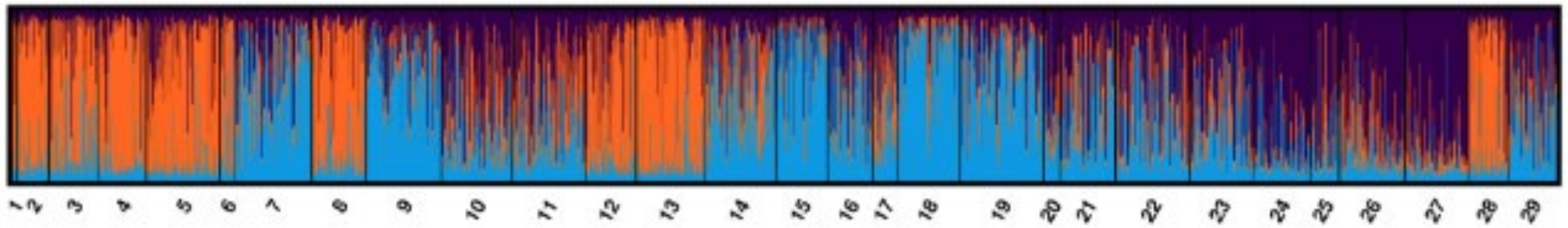


Figure 4 The outcome of STRUCTURE analysis of the Littke Miami River samples. The number of putative populations (K) is related to: (a) the natural log likelihood of the data as a function of K[L(K)], (b) the rate of change of the likelihood function (ΔK) (Evanno et al., 2005), and (c) the proportional membership of individuals (Q) in the three clusters identified from the data set (Orange = Cluster 1, Blue = Cluster 2, Purple= Cluster 3)

Table 6 Mean posterior probability of individual assignment to a cluster (Q) identified by STRUCTURE within the Little Miami River watershed. The highest probability is in boldface.

Site ID	Site #	Cluster 1	Cluster 2	Cluster 3
Poplar Creek	1	**0.760**	0.162	0.078
Five Mile Creek	2	**0.775**	0.106	0.119
Barnes Run	3	**0.672**	0.138	0.190
NB Clover Lick	4	**0.733**	0.145	0.122
Brushy Fork	5	**0.717**	0.189	0.094
Pleasant Run	6	**0.667**	0.138	0.196
Lick Run	7	0.138	0.23	**0.632**
Soloman Run	8	**0.726**	0.144	0.130
First Creek	9	0.109	0.239	**0.652**
O'Bannon Creek	10	0.255	**0.462**	0.283
O'Bannon Creek	11	0.313	**0.402**	0.285
Lick Fork	12	**0.614**	0.267	0.119
Turtle Creek East	13	**0.790**	0.111	0.100
Cowan Creek	14	0.268	0.258	**0.474**
Todd Fork	15	0.115	0.166	**0.719**
Salt Run	16	0.140	0.360	**0.499**
Turtle Creek	17	0.346	**0.358**	0.296
Flat Fork	18	0.109	0.100	**0.792**
Dutch Creek	19	0.134	0.216	**0.649**
Buck Run	20	0.174	**0.520**	0.305
Grog Run	21	0.206	0.384	**0.409**
Painter's Run	22	0.215	**0.444**	0.341
NB Caesar's Creek	23	0.167	**0.551**	0.282
SF Massie's Creek	24	0.069	**0.738**	0.193
NF Massie's Creek	25	0.151	**0.636**	0.214
Lisbon Fork	26	0.179	**0.666**	0.155
NF Little Miami	27	0.116	**0.781**	0.103
Crane Run	28	**0.731**	0.156	0.113
SB Caesar's Creek	29	0.190	**0.456**	0.354

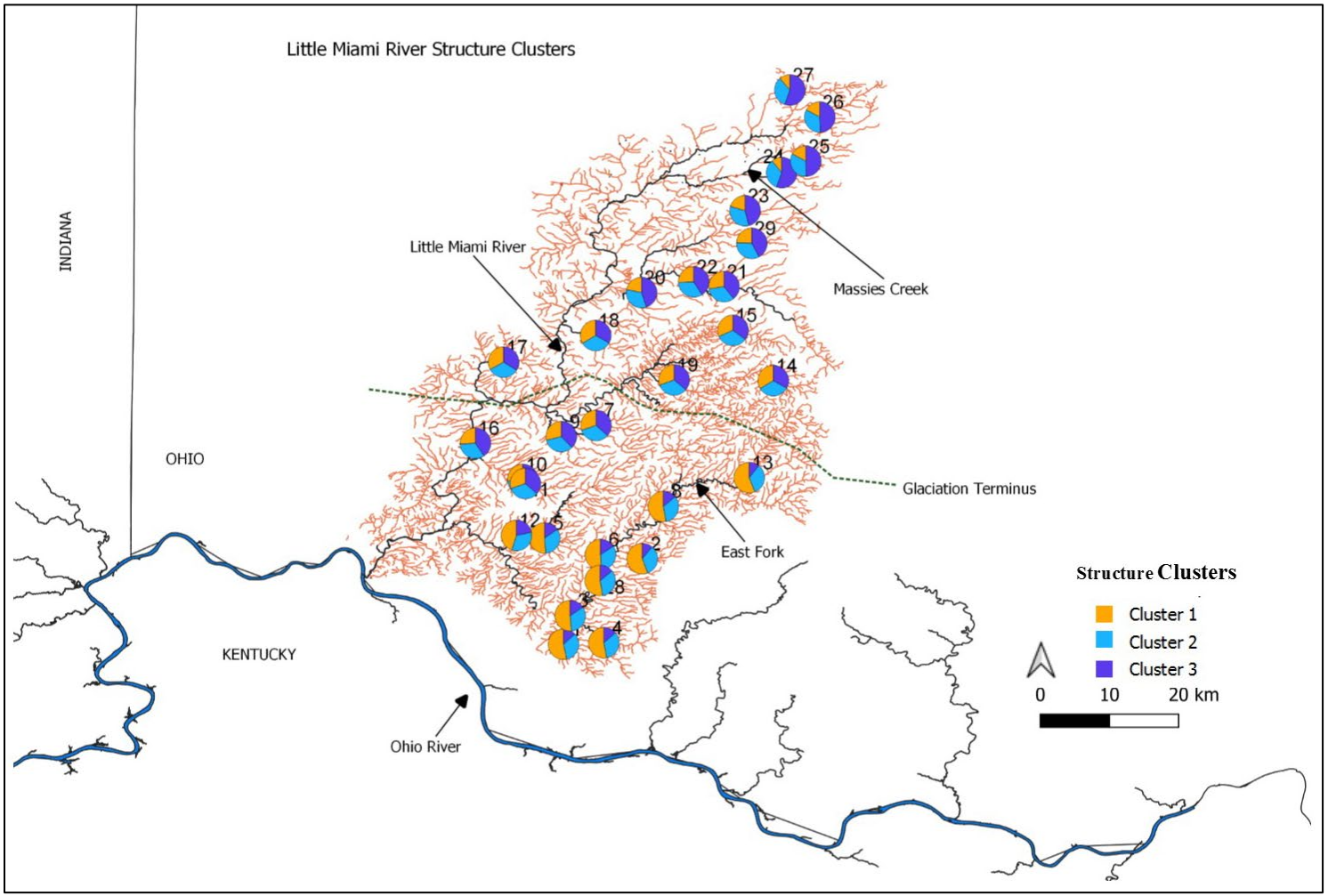


Figure 5 Little Miami River watershed with the average assignment of individuals to one of three clusters identified by STRUCTURE for each sample locality (Orange = Cluster 1, Blue = Cluster 2, and Purple = Cluster 3).

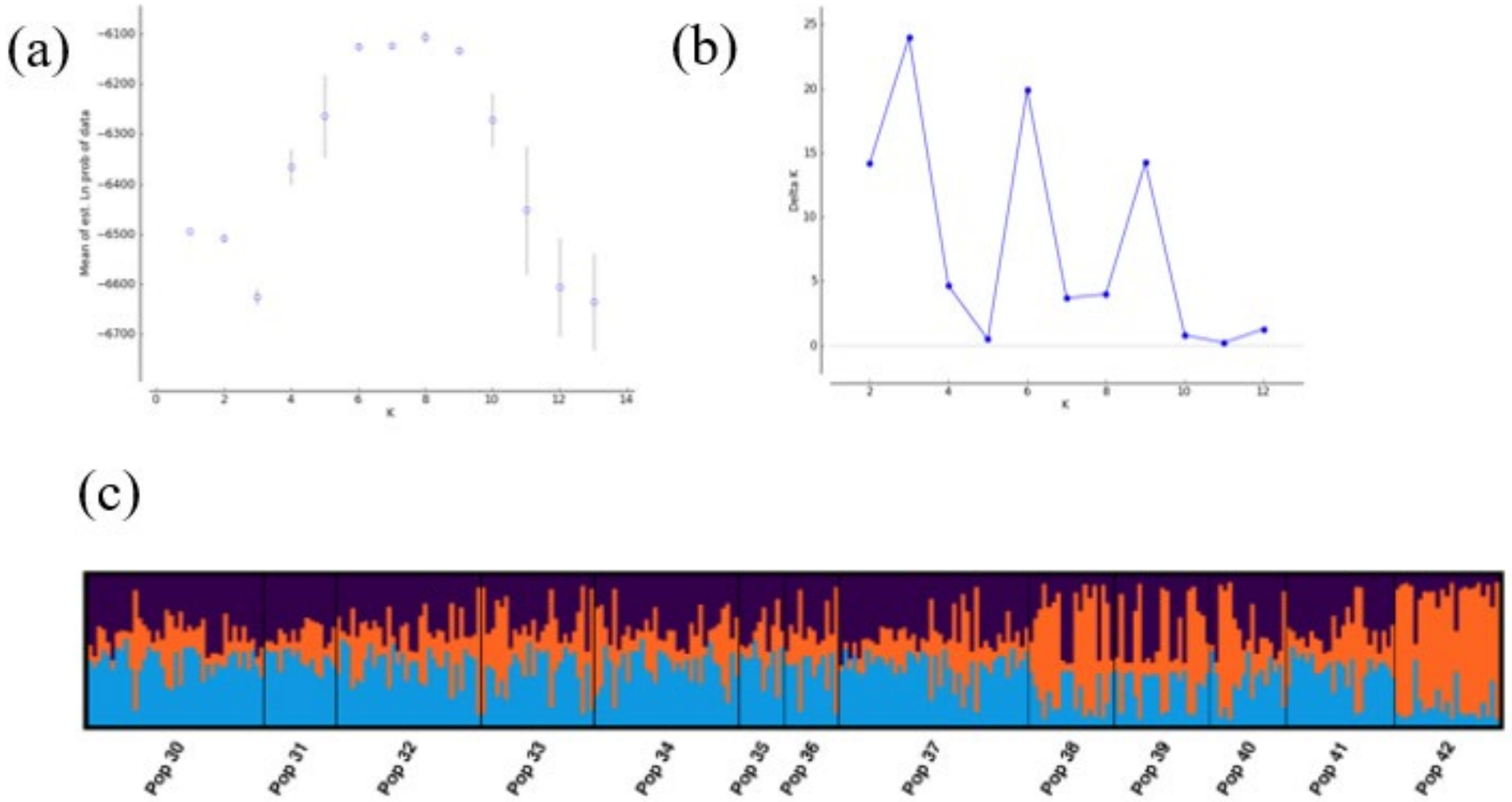


Figure 6 The outcome of STRUCTURE analysis of the Saint Joseph River samples. The number of putative populations (K) is related to: (a) the natural log likelihood of the data as a function of K[L(K)], (b) the rate of change of the likelihood function (ΔK) (Evanno et al., 2005), and (c) the proportional membership of individuals (Q) in the three clusters identified from the data set (Orange = Cluster 1, Blue = Cluster 2, Purple= Cluster 3)

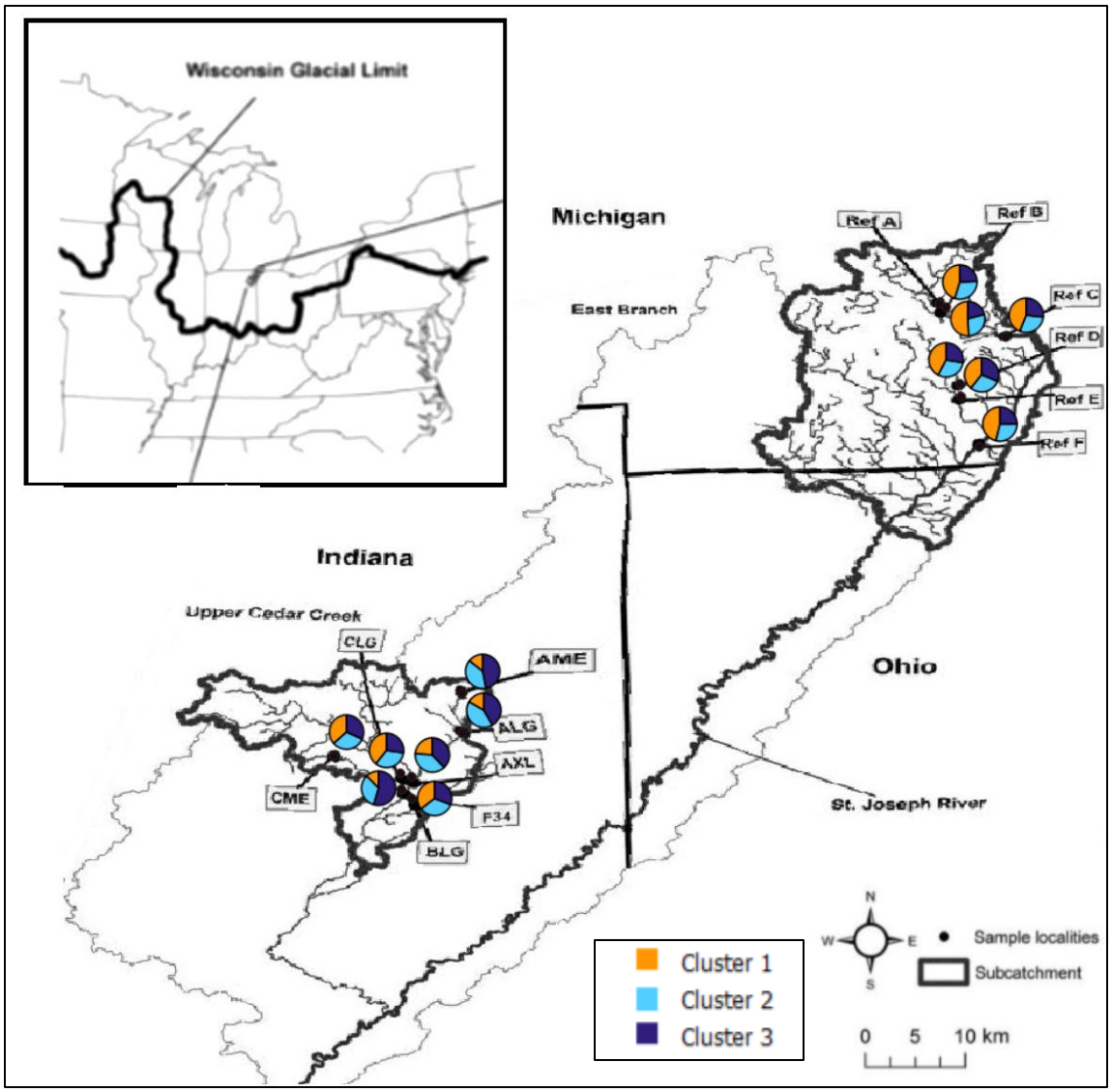


Figure 7 Saint Joseph River watershed with the average assignment of individuals to one of three clusters identified by STRUCTURE for each sample locality (Orange = Cluster 1, Blue = Cluster 2, Purple= Cluster 3)

Table 7. Mean posterior probability of individual assignment to a cluster (Q) identified by STRUCTURE within the Saint Joseph River watershed. The highest probability is in boldface.

Site ID	Site #	Cluster 1	Cluster 2	Cluster3
AME	30	**0.420**	0.384	0.196
ALG	31	**0.442**	0.383	0.175
AXL	32	**0.393**	0.344	0.263
BLG	33	**0.355**	0.304	0.341
CLG	34	**0.388**	0.347	0.265
CME	35	**0.422**	0.363	0.214
F-34	36	**0.352**	0.328	0.320
REF A	37	0.393	**0.395**	0.212
REF B	38	0.203	0.273	**0.523**
REF C	39	0.254	0.361	**0.385**
REF D	40	0.288	0.348	**0.364**
REF E	41	0.364	**0.366**	0.271
REF F	42	0.130	0.123	**0.747**

Table 8. Correlation tests of pairwise differences among sample localities in PASSaGE. Mantel tests were performed between genetic distance (*F*st) and curvilinear stream distance. Partial Mantel test were performed holding curvilinear stream distance constant while correlating *F*st with % agricultural landcover and glaciation history. Tests were run within watersheds.

Test	Watershed	Correlation	*r*	p-value
Mantel	LMR	*F*st & Curvilinear stream distance	0.140	0.027
	SJR		0.200	0.002
Partial Mantel	LMR	*F*st & |% agricultural landcover|	0.320	0.002
	LMR	*F*st & glaciation history	0.310	0.001
	SJR	*F*st & |% agricultural landcover|	-0.003	0.280

Table 9. Mean (± standard error) of allelic richness (A) and observed heterozygosity (Ho) of hypothesized groups within and among LMR and SJR. Number of sample sites (N)

Comparison	Group	N	A	Ho
Watershed	LMR	29	2.59 ±0.07	0.51 ±0.01
	SJR	13	3.19 ±0.09	0.57 ±0.02
Glaciation history	LMR glaciated	13	2.57 ±0.09	0.51 ±0.04
	LMR unglaciated	16	2.60 ±0.09	0.50 ±0.01
	LMR + SJR glaciated	26	2.88 ±0.07	0.54 ±0.01
STRUCTURE clusters	LMR Cluster 1	10	2.67 ±0.12	0.51 ±0.01
	LMR Cluster 2	11	2.56 ±0.10	0.51 ±0.01
	LMR Cluster 3	8	2.49 ±0.14	0.50 ±0.01
	SJR Cluster 1	7	3.21 ±0.13	0.57 ±0.02
	SJR Cluster 2	2	3.10 ±0.24	0.60 ±0.06
	SJR Cluster 3	4	3.20 ±0.16	0.56 ±0.03

DISCUSSION

Ancient and contemporary events continue to impact extant populations. However, the close link between glaciation history and modern agricultural land use can challenge attempts to understand the influences on the genetic makeup of populations. The main objective of this study was to determine if there has been a loss in genetic diversity within and between creek chub populations due to past glaciation events. Microsatellites were chosen as the genetic marker of interest in this study because of their highly polymorphic nature, making SSR's incredibly useful for studying recent evolutionary events among populations (Putman & Carbone, 2014). The differing glaciation histories of the study sites allowed for the unique opportunity to compare the genetic makeup of populations from glaciated and unglaciated streams.

Population structure

Population structure was examined using several different methods. I found a consistently large regional distinction between the two watersheds and a clear lack of population structure attributable to glaciation history. The overall *Fst* of the two watersheds combined was 0.243, indicating a possibility of shared alleles between the two watersheds, however, given the regional distinction of the two watersheds, this is more likely an indication of historic colonization patterns and not recent isolation between watersheds. Creek chub are not a highly mobile species. Work conducted by Belica (2008), demonstrated that creek chub moved an average of only 50.4 m over a 12 week period. *Fst* has been shown to be difficult to interpret when the data set in question is highly polymorphic (Banks et al., 2013). *Fst* increases with the rate of population turnover, where population recovery precedes colonization by a small founder population, however, *Fst* decreases when repopulation is done by a constant influx of immigrants (Banks et al., 2013). Analysis of Molecular Variance (Table 5) between the two watersheds gave an indication that the two watersheds were distinct based on regional indicators, the differentiation of the two watersheds is attributable to geographic distance and not glaciation history. Bayesian cluster analysis in STRUCTURE revealed that the two watersheds constituted two separate populations (Figure 3) the prediction that the two watersheds would be two distinct

populations was supported. Molecular data support the conclusion that these two watersheds are isolated from one another.

Using AMOVA, LMR was examined as two regions based on glaciation history, the regional distinction was responsible for only 1.0% of the variation in the populations. This indicates that individual variation has a higher influence on genetic variation in the LMR than a regional distinction based on glaciation. STRUCTURE analysis also did not cluster LMR sampling sites by glaciation history, rather, STRUCTURE identified three distinct clusters in the LMR (Figure 4), but with high degree of admixture among localities. The limited exchange model states that dispersal is restricted to habitats within a tributary's main stream reach, which would result in a pattern of genetically distinct subpopulations, consistent with isolation by distance (Thornbugh & Gido, 2010). The outlier sites, 5 and 12, within cluster 1 of LMR demonstrated that, although lotic species are often deterred from migration through lentic habitat, they are downstream of the main stream cluster, making it possible the populations would migrate through an undesirable area of confluence, such as a reservoir (Hudman & Gido, 2013). Located in the East Fork State Park, in Clermont county Ohio, USA, the William H. Harsha Lake is located within the Little Miami Watershed (Hedeen, 2008). This lake is located along the East Fork and separates the main cluster 1 sites from the two outlier sites, 5 and 12. Also known as the East Fork Lake, the site is a naturally occurring lake and is the site of two abandoned gold mines. Located in an area of recreation, there is a possibility of introduction by game fisherman, but it is possible that the areas have not been separated long enough to express a barrier effect. Creek chub examined in a partially impounded river system in Kansas, USA displayed a similar spatial structuring pattern (Hudman & Gido, 2013).

SJR samples examined in STRUCTURE (Figures 6 &7, Table 7) were clustered into three groupings. These results partially support previous work by Jordan et al (2013), in which three distinct clusters based on catchment location and land cover. In conjunction with correlation tests, this work suggested that land cover was a larger influencing factor in population structure than geographic distance. The STRUCTURE grouping in this study were not as strong as those in prior work. There were fewer alleles in this study than in the 2013 study on the same sampling sites which may have made the groupings less robust. Population structure of creek chub in channelized agricultural ditches in previous studies was most influenced by watershed characteristics and geomorphology, impacting the length of the individuals, and the

abundance and the biomass of the creek chub. Instream habitat and water chemistry were not notable influences on population structure (Smiley, King, & Fausey, 2017). Smiley's study was unique, as it was the first to relate the importance of channel size and shape to creek chub population makeup. The biomass of creek chub also decreased with increased agricultural land use (Smiley et al., 2017). The STRUCTURE groupings within the SJR were not significantly different from one another in either allelic richness or observed heterozygosity, further supporting the weak groupings after Bayesian analysis.

Stone lapping minnows in the Upper Nan river basin in Thailand have similar Bayesian cluster results to the two watersheds in this study. Clusters were associated with river topology, however, the use of multiple programs to determine population structure demonstrated there was a high degree of admixture and interbreeding between closely located populations (Jaisuk & Senanan, 2018), similar to the weak clustering values of both the LMR and SJR watersheds.

The population structure detected in both watersheds is representative of both historical and recent habitat influences. Fragmentation through both natural anthropomorphic means can lead to genetic isolation and place these smaller demes at risk of extirpation. Within aquatic habitats unobstructed corridors of dispersal are important to maintain genetic diversity. Disruption of these corridors, through natural or anthropomorphic means, can have a negative impact on species and gene dispersal and diversity (Blakney, Loxterman, & Keeley, 2014). Although an anadromous species, sockeye salmon populations have felt the genetic effects of postglacial colonization, similar to that experienced by creek chub in the study areas. Both species were restricted in their dispersal due to glacial barriers and both have a marked pattern of recolonization, particularly in geologically young habitats. Sockeye salmon populations have a high degree of genetic divergence between populations, suggestive of a bottleneck event due to colonization (Ramstad, Woody, Sage, & Allendorf, 2004).

Correlation Tests

Partial Mantel tests of correlation (Table 8) in the LMR determined that agricultural land cover and glaciation history had nearly identical correlation coefficients in regards to population structure. The American Midwest hosts some of the best farm land in the world, the soil is mineral rich, and is found in areas which have experienced glacial cyclical melting and deposition of nutrients from organic life forms (von Engeln, 1914). The concept of comparing

glaciated areas with unglaciated and their agricultural benefit has been a continued interest for ecologist and geologist alike. In 1914 von Engeln wrote a comparative paper exploring the influence of glacial history on land use. Studies in countries like France and Germany were skewed by modern favorable climate, however, when the United States was examined based on farm land value, previously glaciated areas were valued three times higher than historically unglaciated areas (1914). The close linkage between agricultural land use and glaciation has made it difficult to assess the impacts these forces have on extant populations. Previous work on the LMR using Central Stonerollers demonstrated a clear division by glaciation history in terms of genetic and species diversity (Blum et al., 2012). However, this area also has a developed agricultural system which makes it difficult to assess modern effects from historical events. This demonstrates how different species within the same environment can express responses to stressors, both contemporary and historical, in markedly different ways. The population structure identified in this study was influenced more by stream location and confluence point s than glaciation history.

There was a weak correlation between stream distance and genetic distance in both the LMR (R=0.14) and SJR (R=0.20) (Table 8). Partial Mantel tests performed on the LMR data exhibited positive correlations between genetic distance and both absolute agricultural land cover (R=0.32) and glaciation history (R=0.31), while holding geographic distance constant. Mantel and partial Mantel tests have been found to have a high degree of type 1 error, this makes it difficult to distinguish between true Isolation by Distance and true landscape effects (Balkenhol, Waits, & Dezzani, 2009). Substrate makeup and percent agricultural land cover are significant predictors of allelic richness in Central Stonerollers in the same watershed in Ohio (Blum et al., 2012). The relatively weak isolation by distance values support the notion that habitat fragmentation may predate recent anthropogenic habitat alterations. A significant pattern of isolation by distance can be obscured over time by the effect of random genetic drift (Blakney et al., 2014).

Isolation by distance was expected, and, as the results indicate, less genetic variation was present in the loci examined in this study than was demonstrated in previous work on creek chub in the SJR (Jordan et al., 2013). Absence of evidence for isolation by distance where it is expected can be attributed to discontinuity in the colonization process, with founder events occurring at the time of establishment (Leblois, Rousset, Tikel, Moritz, & Estoup, 2000).

Correlation tests performed in PASSaGE indicated that there was no correlation between agricultural land cover and genetic distance in the SJR. These results do not support previous work by Jordan in this same watershed area (2013). Leblois et al. (2000) studied an invasive species of frog in Australia, the results were congruent with my own in that, there was a lack of evidence for isolation by distance (R=0.00072), it is theorized that this will occur in newly established populations (< 10 years old). The areas in this study are a part of the native range for creek chub and are historically habituated by the species, however, with dredging and other disturbances occurring regularly it is possible that the populations in question are exhibiting this phenomena as well. Streams in highly agricultural areas tend to have a lower habitat quality. Changes in hydrology due to agriculture land use can vary by stream, but often manifest in losses of species diversity within streams, and in some cases, extirpation (Allan, 2004).

Genetic Variation

The prediction that glaciated sites would have a lower genetic diversity than unglaciated sites was not supported. Other studies which attempted to quantify the genetic effects of glacial recolonization and contemporary land use consistently draw the conclusion that a loss of genetic variation is often attributed to founder events and bottlenecks associated with dispersal from source populations (Costello, Down, Pollard, Pacas, & Taylor, 2003). Allelic richness is directly influenced by sample size and is often skewed by small sample sizes (Kalinowski, 2005). The Little Miami River samples sizes were larger, the sample sites were more numerous and the watershed area as a whole was larger than the Saint Joseph River sampling sites. I predicted that the Little Miami River would show a higher degree of genetic diversity than the Saint Joseph River watershed due to its proximity to the glacial boundary. This prediction was not supported. SJR was statistically more diverse than the LMR, in measures of both allelic richness and observed heterozygosity after ANOVA tests. In ANOVA tests on glaciation history where watershed data was combined, the glaciated sites were statistically more diverse than the unglaciated sites. This is likely attributable to the higher allelic richness and observed heterozygosity of the SJR skewing the measures, more so than the actual glaciation history of the sites. There was a not significant difference between the glaciated and unglaciated sites within the LMR in respect to allelic richness or observed heterozygosity. ANOVA tests of the three

STRUCTURE clusters, did not indicate a significant difference in either measures of allelic richness or observed heterozygosity, giving further indication of weak groupings.

The LMR was predicted to have a higher degree of allelic richness due to the larger watershed area, larger sample sizes and the watershed's location in the Ohio River basin (APPENDIX B). Being that it is located further south, the LMR is theoretically closer to a historical source population. A high degree of admixture in the gene pool caused by an influx of migrants following a founding event can increase the allelic richness of an area. Gene flow and genetic drift can often work against one another following founding events, gene flow causes an increase in a population's genetic diversity and drift causes a decrease (Greenbaum et al., 2014).

Studies on suspected refugium populations provide evidence of a reduced gene pool with increased distance from a putative population and a reduced gene pool when descended from microrefugium (Mee & Moore, 2014). Examples of this phenomena are Jollytail fish in Australia, which provided evidence for a reduced gene pool, due to small refugial populations and the subsequent expansion of these populations following glacial melt, similar to the reduced gene pools in the LMR and SJR. This large palaeolake allowed previously separated populations to temporarily intermix, however, with already exhausted gene pools, these populations suffered further genetic decline when the palaeolakes receded leaving what is predominately the modern landscape (Zattara & Premoli, 2005). Studies of white fish in Alaska also demonstrated a decrease in a genetic variation as distance from a putative refugia increased (Harris & Taylor, 2009). The LMR spans the glacial boundary, and as a result, expressed a reduced genetic variability.

A similar study of bull trout populations supported that intrapopulation diversity, A and H, was largely influenced by historical factors. Interpopulation diversity, although influenced by ancient recolonization patterns, was more influenced by the degree of connectivity between sites and contemporary factors influencing dispersal. Barriers, whether naturally occurring or manmade, greatly influence molecular variation and demonstrates the importance of habitat fragmentation on dispersal, which directly impacts habitat structure (Costello et al., 2003). The clustering patterns found in these watersheds do indicate a degree of interbreeding in sub catchments, such as the Upper Cedar Creek sub catchment in the SJR. Colonization patterns from refugial populations often do indicate a higher amount of interpolation differentiation as intrapopulation variation decreases with distance form a refuge population.

CONCLUSION

The main objective of this study was to examine if there has been a loss in genetic diversity of creek chub populations due to past glaciation events. Multiple tests indicated that glaciation history is not a key influencer on population structure. Population structure appears to be more attributable to the dispersal patterns of creek chub rather that glaciation history. Furthermore, the expectation that glaciation reduces genetic variation was not supported by the data. Allelic richness did not support the prediction that there would be a lowered amount of genetic diversity in the Saint Joseph watershed. Historically glaciated sites were found to have higher amounts of genetic variation when watershed data was combined, however, this was attributable to the Saint Joseph watershed's higher genetic variation overall. It was predicted that the glaciated watersheds would have lower genetic diversity and less genetic differentiation among sample localities when compared to the partially glaciated watershed. This prediction was not supported. Genetic variation indicators, such as allelic richness, did not support the prediction that there would be a lowered amount of genetic diversity in the Saint Joseph watershed compared to the Little Miami River watershed. STRUCTURE placed the two watersheds in to two distinct groups, however, when the structure of the populations in question were examined with glaciation as a quantifying measure the support was not present and the differentiation between the two watersheds in question was more attributable to individuals rather than regions. Agricultural land use was found to be correlated with population structure in the Little Miami watershed, but the support for this prediction was not found in the Saint Joseph River sampling sites.

REFERENCES

Allan, J. D. (2004). Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology Evolution and Systematics, 35*, 257-284. doi:10.1146/annurev.ecolsys.35.120202.110122

Avise, J. C. (1994). *Molecular Markers, Natural History, and Evolution*. New York: Chapman & Hall New York (511 pp.).

Balkenhol, N., Waits, L. P., & Dezzani, R. J. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography, 32*(5), 818-830. doi:DOI 10.1111/j.1600-0587.2009.05807.x

Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., . . . Peakall, R. (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology and Evolution, 28*(11), 670-679. doi:10.1016/j.tree.2013.08.005

Beaumont, A. R., & Hoare, K. (2003). *Biotechnology and Genetics in Fisheries and Aquaculture* (1 ed. Vol. 1). School of Ocean Sciences University of Wales, Bangor, UK: Blackwell Science Ltd.

Belica, L. A. T., & Rahel, F. J. (2008). Movements of creek chubs, Semotilus atromaculatus, among habitat patches in a plains stream. *Ecology of Freshwater Fish, 17*(2), 258-272. doi:10.1111/j.1600-0633.2007.00277.x

Bernatchez, L., & Wilson, C. C. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology, 7*, 431-452.

Blakney, J., Loxterman, J., & Keeley, E. (2014). Range-wide comparisons of northern leatherside chub populations reveal historical and contemporary patterns of genetic variation. *Conservation Genetics, 15*(4), 757.

Blum, M. J., Bagley, M. J., Walters, D. M., Jackson, S. A., Daniel, F. B., Chaloud, D. J., & Cade, B. S. (2012). Genetic diversity and species diversity of stream fishes covary across a land-use gradient. *Oecologia, 168*(1), 83-95. doi:DOI 10.1007/s00442-011-2078-x

Boizard, J., Magnan, P., & Angers, B. (2009). Effects of dynamic landscape elements on fish dispersal: the example of creek chub (Semotilus atromaculatus). *Molecular Ecology, 18*(3), 430-441. doi:10.1111/j.1365-294X.2008.04044.x

Costello, A. B., Down, T. E., Pollard, S. M., Pacas, C. J., & Taylor, E. B. (2003). The influence of history and contemporary stream hydrology on the evolution of genetic diversity within species: An examination of microsatellite DNA variation in bull trout, *Salvelinus confluentus* (Pisces: Salmonidae). *Evolution, 57*(2), 328-344.

Dufresnes, C., & Perrin, N. (2014). Effect of biogeographic history on population vulnerability in european amphibians. *Conservation biology, 29*, 1235-1241.

Earl, D. A., & Von Holdt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*, 359-361. doi:10.1007/s12686-011-9548-7

Ehlers, J. R. (2016). *The Ice Age*: Chichester, England : Wiley-Blackwell.

Engel, B., & Theller, L. (2016). Long Term Hydrological Impact Analysis (L-THIA). http://lthia.agriculture.purdue.edu/

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: distangling historical and contemporary influences and inferring change. *Molecular Ecology, 24*, 6021-6040.

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

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data:Linked Loci and Correlated Allele Frequencies. *Genetics Society of America*(164), 1567-1587.

Filgueira, R., Chapman, J. M., Suski, C. D., & Cooke, S. J. (2016). The influence of watershed land use cover on stream fish diversity and size-at-age of a generalist fish. *Ecological Indicators, 60*, 248-257. doi:10.1016/j.ecolind.2015.06.006

Foulley, J.-L., & Ollivier, L. (2006). Estimating allelic richness and its' diversity. *Livestock Science, 101*, 150-156.

Gibbard, P., & Pillans, B. (2012). The Quaternary Period. In *The geological Timescale* (pp. 980-1004): Elsevier B. V.

Gorman, O. T., & Karr, J. R. (1978). Habitat structure and stream fish communities. *Ecology, 59*(3), 507-515.

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity, 86*(6), 485-486.

Greenbaum, G., Templton, A. R., Zarmi, Y., & Bar-David, S. (2014). Allelic Richness following Population Founding Events-A Stochastic Modeling Framework Incorporating Gene Flow and genetic Drift. *PLOS ONE, 9*(12).

Harding, J. S., Benfield, E. F., Bolstad, P. V., Helfman, G. S., & Jones, E. B. D. (1998). Stream Biodiversity: The Ghost of Land Use Past*(25),* 14843. Retrieved from http://ezproxy.library.ipfw.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.46631&site=eds-live&scope=site

Harris, L. N., & Taylor, E. B. (2009). Pleistocene Glaciations and the contemporary genetic diversity in a Beringian fish, the broad white fish, Coregonus nasus (Pallas): Inferences from microsatellite DNA variation. *Journal Of Evolutionary Biology, 23*, 72-86.

Hedeen, S. (2008). *The Little Miami Wild and Scenic River Ecology and History*

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 405*(6789), 907.

Hudman, S. P., & Gido, K. B. (2013). Multi-scale effects of impoundments on genetic structure of creek chub (Semotilus atromaculatus) in the Kansas River basin. *Freshwater Biology, 58*(2), 441-453. doi:Doi 10.1111/Fwb.12079

Jacquemin, S. J., & Pyron, M. (2011a). Fishes of Indiana streams: current and historic assemblage structure. *Hydrobiologia, 665*(1), 39-50. doi:10.1007/s10750-011-0602-y

Jacquemin, S. J., & Pyron, M. (2011b). Impacts of past glaciation events on contemporary fish assemblages of the Ohio River basin. *Journal of Biogeography, 38*(5), 982-991. doi:10.1111/j.1365-2699.2010.02457.x

Jaisuk, C., & Senanan, W. (2018). Effects of landscape features on population genetic variation of a tropical stream fish, Stone lapping minnow, *Garra cambodgiensis,* in the upper Nan River drainage basin, northern Thailand. *PeerJ*. doi:10.7717/peerj.4487

Jordan, M. A., Morris, D. A., & Gibson, S. E. (2009). The influence of historical landscape change on genetic variation and population structure of a terrestrial salamander (Plethodon cinereus). *Conservation Genetics, 10*(6), 1647-1658. doi:10.1007/s10592-008-9741-8

Jordan, M. A., Patel, D., Sanders, K. E., & Gillespie, R. B. (2013). The relative roles of contemporary and ancient processes in shaping genetic variation of a generalist fish in a

catchment dominated by agriculture. *Freshwater Biology, 58*(8), 1660-1671. doi:10.1111/fwb.12157

Jost, L., Archer, F., Flanagan, S., Gaggiotti, O., Hoban, S., & Latch, E. (2018). Differentiation measures for conservation genetics. *Wiley: Evolutionary Applications*(11), 1139-1148.

Kalinowski, S. T. (2005). HP-Rare 1.0: A computer program for performing rarefacation on measures of allelic richness. *Molecular ecology, 5*, 187-189.

Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols, 3*, 1452. doi:10.1038/nprot.2008.133

Leblois, R., Rousset, F., Tikel, D., Moritz, C., & Estoup, A. (2000). Absence of evidence for isolation by distance in an expanding cane toad (Bufo marinus) population: an individual-based analysis of microsatellite genotypes. *Molecular Ecology, 9*(11), 1905.

Mee, J. A., & Moore, J.S. (2014, 05//). The ecological and evolutionary implications of microrefugia, Editorial. *Journal of Biogeography,* pp. 837-841. Retrieved from http://ezproxy.library.ipfw.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=95516120&site=eds-live&scope=site

Nagrodski, A., Suski, C., & Cooke, S. (2013). Health, condition, and survival of creek chub (Semotilus atromaculatus) across a gradient of stream habitat quality following an experimental cortisol challenge. *Hydrobiologia, 702*(1), 283-296. doi:10.1007/s10750-012-1331-6

Peakall, R. A. (2012). *GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.* : *Bioinformatics*.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics Society of America*(155), 945-959.

Prost, S., Smirnov, N., Fedorov, V. B., Sommer, R. S., Stiller, M., Nagel, D., . . . Hofreiter, M. (2010). Influence of Climate Warming on Arctic Mammals? New Insights from Ancient DNA Studies of the Collared Lemming Dicrostonyx torquatus (Lemming Population Dynamics). *PLoS ONE, 5*(5), e10447. doi:10.1371/journal.pone.0010447

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution, 4*(22), 4399-4428.

Ramstad, K., Woody, C., Sage, G., & Allendorf, F. (2004). Founding events influence genetic population structure of sockeye salmon (Oncorhynchus nerka) in Lake Clark, Alaska. *Molecular Ecology, 13*(2), 277-290. doi:10.1046/j.1365-294X.2003.2062.x

Rosenberg, M. S., & Anderson, C. D. (2011). PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis.Version 2. *Methods in Ecology & Evolution, 2*(3), 229-232.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes, 4*(1), 137-138. doi:DOI 10.1046/j.1471-8286.2003.00566.x

Ross, S. T. (2013). *Ecology of North American Freshwater Fishes* (1 ed.).

Rousset, F. (2008). GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources, 8*(1), 103-106. doi:DOI 10.1111/j.1471-8286.2007.01931.x

Schemske, D. W. (1974). Age, Length and Fecundity of the Creek Chub, Semotilus atromaculatus(Mitchill), in Central Illinois. *American Midland Naturalist, 92*(2), 505.

Skalski, G. T., & Grose, M. J. (2006). Characterization of microsatellite loci in the creek chub (Semotilus atromaculatus). *Molecular Ecology Notes, 6*(4), 1240-1242. doi:10.1111/j.1471-8286.2006.01502.x

Skalski, G. T., Landis, J. B., Grose, M. J., & Hudman, S. P. (2008). Genetic structure of creek chub, a headwater minnow, in an impounded river system. *Transactions of the American Fisheries Society, 137*(4), 962-975. doi:Doi 10.1577/T07-060.1

Smiley, P. C., King, K. W., & Fausey, N. R. (2017). Influence of Different Habitat Factors on Creek Chub (Semotilus atromaculatus) within Channelized Agricultural Headwater Streams. *24*(sp8), 18-44. doi:10.1656/045.024.0sp804

Thornbugh, D. J., & Gido, K. B. (2010). Influence of spatial positioning within stream networks on fish assemblage structure in the Kansas River basin, USA. *Canadian Journal of Fish and Aquatic Science, 67*, 143-156.

von Engeln, O. D. (1914). Effects of Continental Glaciation on Agriculture. Part I. *Bulletin of the American Geographical Society, 4*(46), 241-264. doi:10.2307/201310.

Vu, S., López-Pujol, J., Herrando-Moraira, S., Son, S., Suh, G., & Chung, M. (2018). Comparison of genetic variation between northern and southern populations of Lilium

cernuum (Liliaceae): Implications for Pleistocene refugia. *PLoS One, 13*(1), e0190520. doi:10.1371/journal.pone.0190520

Wang, H.-Y., Tsai, M.-P., Yu, M.-J., & Lee, S.-C. (1999). Influence of glaciation on divergence patterns of the endemic minnow, Zacco pachycephalus, in Taiwan. *Molecular Ecology, 8*(11), 1879-1888. doi:10.1046/j.1365-294X.1999.00787.x

Whiteley, A. R., Spruell, P., & Allendorf, F. W. (2006). Can common species provide valuable information for conservation? *Molecular Ecology, 15*(10), 2767-2786. doi:10.1111/j.1365-294X.2006.02972.x

Wright, S. (1943). Isolation by Distance. *Genetics, 28*, 114-138.

Zattara, E. E., & Premoli, A. C. (2005). Genetic structuring in Andean landlocked populations of Galaxias maculatus : effects of biogeographic history. *Journal of Biogeography, 32*(1), 5-14. doi:10.1111/j.1365-2699.2004.01164.x

APPENDIX A

Reaction Conditions for Polymerase chain reactions

Touchdown 65°C Conditions followed those steps outlined by Skalski and Grose (2006).

Temp	Time	Cycles
94 °C	2 minutes	1
94 °C,	20 Seconds	21
65 °C decreasing 0.5 °C every cycle	20 Seconds	
72 °C	30 Seconds	10
94 °C	20 Seconds	
55°C	20 Seconds	
72°C	30 Seconds	
72°C	10 minutes	1

LMR PCR Reaction Conditions. PCR Conditions as described by Blum (2012).

Temp	Time	Cycles
95 °C	1 minute	1
95 °C,	30 Seconds	11
72 °C decreasing 0.8 °C every cycle	90 Seconds	
95 °C	30 Seconds	22
54 °C	30 Seconds	
72°C	90 Seconds	
72°C	15 Minutes	1

Touch down 60°C. Conditions Followed that of Jordan (2013).

Temp	Time	Cycles
94 °C	2 minutes	1
94 °C,	20 Seconds	21
60 °C decreasing 0.5 °C every cycle	20 Seconds	
94 °C	20 Seconds	10
50 °C	20 Seconds	
72°C	30 Seconds	
72°C	10 Minutes	1

APPENDIX B

Observed number of alleles at each locus of each sample locality in the Little Miami River (LMR) watershed and Saint Joeseph River (SJR) watershed. North Branch (NB), South Fork (SF), North Fork (NF).

Watershed	Site ID	Site Name	Locus							
			Seat 212	Seat 209	Seat 402	Seat 204	Seat 409	Seat 205	Seat 416	Seat 403
LMR	1	Popular Creek	1	9	1	4	3	7	4	3
	2	Fivemile Creek	3	17	2	7	5	7	5	3
	3	Barnes Run	3	23	2	7	5	9	4	3
	4	NB Cloverlick Creek	3	16	3	7	7	10	5	4
	5	Brushy Fork	5	27	3	6	5	10	5	3
	6	Pleasant Run	3	14	2	6	5	8	4	3
	7	Lick Run	3	24	3	4	3	6	5	3
	8	Soloman Run	4	23	3	6	5	11	7	4
	9	First Creek	3	16	3	5	4	6	3	3
	10	O'Bannon Creek	5	28	4	5	4	12	5	3
	11	O'Bannon Creek	4	23	3	6	4	10	5	3
	12	Lick Fork	4	33	2	6	5	13	5	3
	13	Turtle Creek East	5	22	2	6	4	11	5	3
	14	Cowan Creek	3	21	3	6	4	7	4	2
	15	Todd Fork	3	27	3	6	2	6	3	4
	16	Salt Run	4	26	3	7	3	11	3	3
	17	Turtle Creek	5	18	3	5	4	8	4	3
	18	Flat Fork	4	15	3	3	4	9	3	2
	19	Dutch Creek	7	31	4	5	4	11	5	4
	20	Buck Run	3	12	2	5	5	8	4	2
	21	Grog Run	4	20	3	6	4	9	4	4
	22	Painters Run	4	28	3	6	5	9	7	3
	23	NB Caesars Creek	3	25	4	5	5	7	6	3
	24	SF Massies Creek	4	22	3	6	3	9	5	5
	25	NF Massies Creek	3	16	3	6	3	7	3	4
	26	Lisbon Fork	3	29	2	6	3	9	5	3
	27	NF Little Miami	3	22	3	5	3	10	5	2
	28	Crane Run	3	18	2	6	4	9	5	3
	29	SB Caesers Creek	5	18	3	7	4	7	5	3
SJR	30	ALG	6	28	5	6	4	7	7	5
	31	AME	6	14	5	3	2	7	4	4
	32	AXL	9	18	3	4	5	10	7	7
	33	BLG	7	20	3	4	6	8	5	7
	34	CLG	6	17	5	4	6	8	7	5
	35	CME	4	11	4	3	3	5	3	5
	36	F34	3	12	4	3	3	5	7	5
	37	Ref A	7	17	10	4	2	8	5	3
	38	Ref B	7	4	7	4	4	4	6	3
	39	Ref C	7	10	5	5	3	5	4	4
	40	Ref D	6	16	7	5	4	4	7	4
	41	Ref E	5	8	3	5	4	6	4	7
	42	Ref F	6	21	3	8	3	6	2	9

APPENDIX C

Site identification number, site name, glaciation history, sample size (N), watershed area (ha) and measures of genetic diversity [mean ± standard error: observed heterozygosity (Ho) unbiased expected heterozygosity (He), and allelic richness (A)] for the Little Miami River watershed.

Site Identification Number	Site Name	Glaciation history	Sample size (N)	Watershed area (ha)	Measures of genetic diversity [mean± standard error]		
Little Miami Watershed					Ho	uHe	A
1	Popular Creek	Unglaciated	7	29.6	0.46 ±0.12	0.54 ± 0.13	2.85 ± 0.52
2	Fivemile Creek	Unglaciated	48	22.3	0.52 ± 0.1	0.51 ± 0.1	2.59 ± 0.36
3	Barnes Run	Unglaciated	75	21.4	0.46 ±0.1	0.46 ± 0.1	2.44 ± 0.39
4	NB Cloverlick Creek	Unglaciated	71	34.7	0.52 ± 0.10	0.5 ± 0.1	2.58 ± 0.37
5	Brushy Fork	Unglaciated	111	30.3	0.51 ± 0.08	0.54 ± 0.09	2.73 ± 0.39
6	Pleasant Run	Unglaciated	22	20.1	0.56 ± 0.09	0.56 ± 0.09	2.80 ± 0.41
7	Lick Run	Unglaciated	116	29.7	0.49 ± 0.08	0.5 ± 0.08	2.50 ± 0.37
8	Soloman Run	Unglaciated	81	25.36	0.55 ± 0.09	0.55 ± 0.09	2.75 ± 0.39
9	First Creek	Unglaciated	113	33.7	0.53 ± 0.08	0.52 ± 0.08	2.53 ± 0.33
10	O'Bannon Creek	Unglaciated	107	22.4	0.49 ± 0.08	0.49 ± 0.08	2.48 ± 0.34
11	O'Bannon Creek	Unglaciated	109	33.1	0.45 ± 0.1	0.45 ± 0.099	2.40 ± 0.38
12	Lick Fork	Unglaciated	77	17.5	0.52 ± 0.11	0.53 ± 0.10	2.78 ± 0.46
13	Turtle Creek East	Unglaciated	103	38.3	0.54 ± 0.11	0.52 ± 0.10	2.61 ± 0.4
14	Cowan Creek	Glaciated	107	17.3	0.53 ± 0.09	0.53 ± 0.08	2.59 ± 0.36
15	Todd Fork	Glaciated	77	47.9	0.51 ± 0.08	0.5 ± 0.08	2.47 ± 0.39
16	Salt Run	Unglaciated	68	39.2	0.49 ± 0.08	0.55 ± 0.08	2.73 ± 0.4
17	Turtle Creek	Glaciated	38	59.5	0.46 ± 0.1	0.5 ± 0.09	2.55 ± 0.4
18	Flat Fork	Unglaciated	93	32.1	0.48 ± 0.1	0.48 ± 0.09	2.33 ± 0.33
19	Dutch Creek	Glaciated	126	40.4	0.51 ± 0.08	0.52 ± 0.08	2.59 ± 0.39
20	Buck Run	Glaciated	24	37.3	0.52 ± 0.07	0.53 ± 0.08	2.59 ± 0.36
21	Grog Run	Glaciated	83	22.3	0.5 ± 0.08	0.5 ± 0.09	2.52 ± 0.39
22	Painters Run	Glaciated	112	22.8	0.49 ± 0.08	0.53 ± 0.08	2.59 ± 0.38
23	NB Caesers Creek	Glaciated	94	34.7	0.56 ± 0.07	0.59 ± 0.06	2.74 ± 0.31
24	SF Massies Creek	Glaciated	86	43.7	0.51 ± 0.07	0.54 ± 0.06	2.56 ± 0.25
25	NF Massies Creek	Glaciated	44	32.6	0.47 ± 0.09	0.49 ± 0.09	2.43 ± 0.33
26	Lisbon Fork	Glaciated	99	37.1	0.5 ± 0.08	0.51 ± 0.08	2.52 ± 0.35
27	NF Little Miami	Glaciated	95	31.9	0.52 ± 0.08	0.5 ± 0.08	2.48 ± 0.36
28	Crane Run	Unglaciated	60	35.1	0.47 ± 0.09	0.52 ± 0.1	2.58 ± 0.38
29	SB Caesers Creek	Glaciated	72	31.4	0.6 ± 0.07	0.6 ± 0.07	2.79 ± 0.34

Site identification number, site name, glaciation history, sample size (N), watershed area (ha) and measures of genetic diversity [mean ± standard error: observed heterozygosity (Ho) unbiased expected heterozygosity (He), and allelic richness (A)] for the Saint Joseph River watershed.

Site Identification Number	Site ame	Glaciation history	Sample size (N)	Watershed area (ha)	Measures of genetic diversity [mean± standard error]					
Saint Joseph					Ho		uHe		A	
30	ALG	Glaciated	39	25.25	0.55	± 0.05	0.66	± 0.06	3.16	± 0.36
31	AME	Glaciated	16	2.99	0.52	± 0.07	0.62	± 0.07	2.97	± 0.37
32	AXL	Glaciated	32	69.36	0.52	± 0.07	0.67	± 0.07	3.26	± 0.40
33	BLG	Glaciated	25	17.24	0.54	± 0.07	0.68	± 0.06	3.32	± 0.37
34	CLG	Glaciated	32	17.48	0.61	± 0.06	0.7	± 0.05	3.32	± 0.3
35	CME	Glaciated	10	3.72	0.6	± 0.05	0.68	± 0.05	3.13	± 0.33
36	F34	Glaciated	12	294.28	0.67	± 0.08	0.7	± 0.06	3.34	± 0.37
37	Ref A	Glaciated	42	36.42	0.54	± 0.06	0.64	± 0.06	3.06	± 0.35
38	Ref B	Glaciated	19	42.02	0.6	± 0.05	0.73	± 0.04	3.29	± 0.20
39	Ref C	Glaciated	21	76.72	0.51	± 0.05	0.65	± 0.06	3.04	± 0.34
40	Ref D	Glaciated	17	17.87	0.61	± 0.05	0.7	± 0.05	3.31	± 0.35
41	Ref E	Glaciated	24	25.55	0.67	± 0.07	0.69	± 0.05	3.21	± 0.34
42	Ref F	Glaciated	23	181.29	0.53	± 0.09	0.63	± 0.09	3.20	± 0.45