Environment and Natural Resources Trust Fund (ENRTF)
M.L. 2013 Minnesota Aquatic Invasive Species Research Center
Sub-Project Work Plan

Date of Report: July 31, 2017

Date of Next Status Update Report: January 31, 2018

Date of Work Plan Approval: February 22, 2017

Sub-Project Completion Date: December 31, 2018

Project Completion Date: June 30, 2019

Does this submission include an amendment request? Yes

SUB-PROJECT TITLE: MAISRC Sub-Project #9: Population genomics of zebra mussel spread pathways, genome sequencing and analysis to select target genes and strategies for genetic biocontrol.

Sub-Project Manager: Michael A. McCartney, Ph.D.

Organization: University of Minnesota – Minnesota Aquatic Invasive Species Research Center (MAISRC)

Mailing Address: Dept of Fisheries, Wildlife and Conservation Biology, 135 Skok Hall, 2003 Upper Buford Circle

City/State/Zip Code: Saint Paul, MN55108

Telephone Number: (612) 301-7703

Email Address: mmccartn@umn.edu

Web Address: http://www.maisrc.umn.edu/

Location: Statewide throughout Minnesota, with limited collection of specimens from Mississippi River tributaries and Great Lakes in Illinois, Michigan, Ohio, and Wisconsin

Total ENRTF Sub-Project Budget:	**Sub-Project Budget:**	**$427,950**
	Amount Spent:	**$58,546**
	Balance:	**$369,404**

Legal Citation: M.L. 2013, Chp. 52, Sec. 2, Subd. 06a

Appropriation Language:
$4,350,000 the first year and $4,350,000 the second year are from the trust fund to the Board of Regents of the University of Minnesota to develop and support an aquatic invasive species (AIS) research center at the University of Minnesota that will develop new techniques to control aquatic invasive species including Asian carp, zebra mussels, and plant species. This appropriation is available until June 30, 2019, by which time the project must be completed and final products delivered.

I. SUB-PROJECT TITLE: Population genomics of zebra mussel spread pathways, genome sequencing and analysis to select target genes and strategies for genetic biocontrol.

II. SUB-PROJECT STATEMENT:

Today, zebra mussel invasions of inland lakes in MN are clearly on the increase, so our research on prevention focuses on identifying invasion sources and pathways for spread to MN inland lakes. A zebra mussel "invasion pathway" connects a "source" water body (Great Lake, inland lake, or river reach) that is infested with zebra mussels to a "destination" water body (inland lake or river) that is infested through transport of zebra mussels from the source or sources. Preventing further spread critically depends upon identifying invasion sources and pathways, so that pathways can be interrupted—for example by inspecting and decontaminating watercraft and equipment. We hear the term "pinch points" to refer to locations where concentrated prevention can be placed to stop the movement of zebra mussels and other aquatic invasive species (AIS)—locating pinch points in MN is an ultimate goal of the research in this proposal.

Continuing from the foundations we created in the "Phase I" subproject in 2014-2016, **Activity 1** in the present ""Phase II" sub-project takes our **genetic research on identifying sources and pathways for zebra mussel invasions** in MN much further. Our research is independent from yet complements other research (including a MAISRC project) that maps and analyzes boat traffic in MN, providing *crucial yet indirect* evidence for sources and pathways. Our genetic studies are one-of-a-kind because they provide *direct* evidence. We use highly variable population genetic markers called microsatellite DNAs, and variable DNA positions in the zebra mussel genome—Single Nucleotide Polymorphisms, or SNPs—to genetically type zebra mussel populations, and assign these populations to the source waters from which they were carried to infest new waters.

Well-informed programs of inspection and decontamination of boats, docks and lifts should be our first focus to reduce zebra mussel spread and impacts. However, prevention cannot stop all new invasions, particularly in MN, with >11,000 lakes and > 4,650 boat ramps (includes DNR + local + private). Increasingly, MN and other US states are researching options for treating infested lakes (so far, mainly with mollusk pesticides) to reduce zebra mussel populations and their ecological and economic impacts, or to attempt to eradicate them from water bodies. Chemical pesticides may be effective at the earliest stages of infestation, when populations are small and clusters of mussels can be enclosed within barriers to contain the pesticide. Unfortunately, lakes with large well-established infestations are not candidates for these chemical controls. Recently, interest has grown in management—even of well-established zebra mussel populations—with biological controls. Parasites might be one option, because the parasite can replicate, its population growing and spreading throughout the lake. But so far no effective parasites have been isolated. Genetic biocontrol agents may also be candidates, because some recently developed agents can replicate and spread via transmission to offspring; others delivered (e.g. to early life stages) in ways that conceivably could achieve control of large populations. Interest is therefore rapidly growing in more conventional (e.g. *RNA-interference* or *RNAi*) and in emerging technologies (e.g. *CRISPR/Cas9* "gene-editing," coupled to "gene drives"—details on pg. 6) for zebra and quagga mussel biocontrol.

The first step in a research plan to develop genetic biocontrol technology for any invasive species is to find weak points in life processes, and target genes that control these. **Activity 2** in this work plan **is the sequencing and analysis of the zebra mussel genome and transcriptomes for the purpose of finding these target genes. Activity 2 will produce the first complete sequence of the zebra mussel genome—a resource of critical importance to finding weaknesses in the genome that can be targeted for control.** The genome sequence displays all of the DNA base pairs that make up the entire hereditary material for the zebra mussel, patched together and lined up in sequence (from the left end of chromosome 1 through the right end of chromosome 32) using "bioinformatic" analyses on supercomputers. Next, the sequence of DNA bases is analyzed to find and name the genes that are encoded (most of it does not encode genes), and finally, a select group of genes to target for biocontrol is searched for. In Activity 2, we *search* the zebra mussel genome for target genes, and in Activity 3, we *choose which genes* to search for. The timing of these activities will be in sequence such that the results from Activity 3 inform Activity 2.

The zebra mussel genome will be an invaluable resource for zebra mussel biology, but **to complete its analysis and to develop strategies for its application to genetic biocontrol, we will create a formal research network in Activity 3**. The outcomes are the following. 1) We form the Dreissenid Mussel Genome Collaborative (DMGC): a team of zebra mussel biologists, invertebrate biologists with specialized expertise, genome scientists, and experts in genetic biocontrol. We (McCartney lab) along with the DMGC perform literature research and write a white paper that provides the rationale for sequencing the zebra mussel genome, and chooses "target genes." (2) We convene a workshop to review and analyze results from the genome and the search for gene targets. In the workshop, we also plan research on genetic modification, on techniques for delivery to zebra mussels, and on strategies for releasing modified animals into lakes for population control. At least one paper will be generated from the workshop and subsequent funding sources identified.

III. SUB-PROJECT STATUS UPDATES:

Sub-Project Status as of July 31, 2017

For Activity 1 (genetics of spread), we expanded our analysis of Minnesota water bodies and added samples of zebra mussels from the Great Lakes, to produce a more comprehensive study of spread. All Great Lakes samples in our collection from summer 2016 were genotyped with 9 microsatellite markers, and these samples collected through 2016 from MN were analyzed by genetic clustering, assignment and ABC invasion model testing. We also launched the 2017 sampling season, visiting 31 new waterbodies in MN and collecting from 23 of these (out of 75 new MN sites listed in the research addendum, Appendix 2; 8 water bodies either had too few mussels to collect, or had access issues that we will solve). For Great Lakes samples, it was necessary to develop a test to quickly and reliably distinguish between zebra and quagga mussels, because Great Lakes collections contain both species, and most are dominated by quagga mussels. Our SNP test is refined for zebra mussels, and to avoid the expense of submitting samples of the wrong species, we tested a quick molecular assay (modified from the literature) validated it on sequenced DNA from zebra and quagga mussels, and now use it routinely on these collections.

We processed all samples that were genotyped in Phase I (with microsatellites), as well as the newly extracted Great Lakes samples, and submitted them for genomic SNP analysis [using the University of Minnesota Genomic Center's (UMGC) assay refined for zebra mussels that was completed in Phase I (December 2016)]. We performed initial analyses of SNP data (examining effects of filtering parameters, filtering SNP data, scoring SNP markers, initial clustering analysis…) and found that with conservative parameters 3320 SNPs could be scored for each of 439 mussels, with no missing data); 10 times or more can be scored with less filtering. This important step shows that the SNP analysis generates a very large number of scorable markers (approximately the number expected), and shows the route we can take to increase the number of markers to study relationships between important source water bodies (e.g. Lake Minnetonka, St. Croix and Mississippi Rivers, Great Lakes).

On Activity 2, we completed the bulk of our lab's work, planned for May-July, that was required to launch the sequencing of the genome. For this, we needed new zebra mussel tissue from animals of known gender. This information is critical (e.g. there are male and female specific genes of interest to us) but was lacking from the genome we sequenced in Phase I—that genome was generated simply to help isolate and score SNP markers. We collected large mature zebra mussels from Pelican Brook (Crow Wing Co.), sexed them by microscopy of gonads in our lab, then extracted very high molecular weight DNA using a specialized DNA protocol that we have developed, which generates DNA of an average length of 60,000 bases—ideal for the long-read sequencing being done this summer and fall. Also for Activity 2, we collected specimens and preserved material for the following transcriptomes: [larval (D-stage, umbonal stage, pediveliger), and female adult and male adult (gonad, mantle) from high calcium environments (transcriptomes 3-8 in research addendum)].

For Activity 3, we selected target genes, contacted and have held continued discussions with developmental biologists, CRISPR/Cas9 and RNAi biotechnology experts, a population genomics expert, and with bivalve biologists who are candidates for the genome collaborative.

For dissemination and outreach, we made 7 presentations to public audiences and to MN DNR. Two papers are in press from Phase I work, and we completed revisions and resubmitted a manuscript for *Biological Invasions*: analysis of spread on the entire microsatellite data set from MN.

Amendment request: July 31, 2017 We request that the completion date for the formation of the genome collaborative and the draft of the white paper be moved to November 2017, to allow sufficient time for contacting new collaborators for the work on shell formation genes and byssal thread synthesis. Both of these are very large research communities and we need more time to complete this work. Our work from February through July has been dominated by lab and field work required to advance the goals of Activities 1 and 2. Meanwhile we have made the necessary contacts who have helped us select target genes—the most pressing need for now -- prior to building transcriptomes. Transcriptome building in our lab is ongoing through September/October.

Sub-Project Status as of January 31, 2018
Sub-Project Status as of July 31, 2018
Final Report Summary: February 28, 2017

Overall Sub-Project Outcomes and Results:

IV. SUB-PROJECT ACTIVITIES AND OUTCOMES:

ACTIVITY 1: Identifying spread pathways through genetic and genomic analysis

Description: Phase II of our research that analyzes markers and variations in the genome in zebra mussels collected from dozens of MN invaded waters, to identify the lakes and rivers where invasions originated, and the geographic pathways over which they were carried from source to destination water body (by boats, docks, and boat lifts, as well as water flow). These pathways of past and ongoing spread, reconstructed by evaluating changes in genomes of zebra mussels as they colonized the state's waters, are used to directly "source" invasions—ultimately, to guide where prevention efforts could be located on the map to best intercept spread.

Mussels will be collected from 54 lakes and river reaches in MN, and from 18 sites across the Great Lakes and Mississippi River basins (24 other sites completed 2014-2016). These include important source and destination waters in each MN region of spread, and high-likelihood sources out of state. Each mussel will be genetically typed (genotyped) with 9 microsatellite markers, and about 5,000 genomic SNP markers. SNP markers will be genotyped in collaboration with the UMN Genomics Center (UMGC) by a method (Sequence Based Genotyping) used to type crop plants at UMN, and by mapping to the zebra mussel genome. Mussels will be assigned to source populations using genetic clustering methods, and their pathways of spread determined by contrasting alternative invasion scenarios. Favored scenarios are selected based on statistical criteria, so we have an estimate of the strength of evidence for a particular pathway.

For example, in 2016 we tested how often Mille Lacs Lake was assigned as the source population for other MN zebra mussel lakes. Mille Lacs is a 128,000-acre, extremely popular fishing lake, from which departing boaters often trailer their boats and launch in other lakes. Since Mille Lacs was infested in 2005 (the earliest large MN lake), it is widely viewed as a "super-spreader," and so receives great attention from MN DNR watercraft inspection. We have tested 38 lakes that were zebra mussel infested after 2005 (many, after the 2009 zebra mussel population explosion in Mille Lacs). These tests each contrasted a scenario A in which Mille Lacs was the source for the other inland lake, to a scenario B, in which it was not (i.e. infestation of Mille Lacs and the other

lake followed independent pathways). Scenario B was favored in every one of these tests, with high statistical confidence. This provides strong evidence that, contrary to expectations, Mille Lacs has not yet been a major source. An implication for management is that watercraft inspection of boats departing Mille Lacs must be effective and should be continued—otherwise we would expect many more infestations of lakes we know (from watercraft inspection data) to be highly connected to Mille Lacs by boat traffic.

What instead have been the sources for inland lakes in MN? Many remain unresolved—and resolving these is a major goal of research in Activity 1. Our increased ability to identify sources and pathways comes from the greater power of 5,000 SNP markers compared to the power of 9 microsatellite markers. Mussels from Lakes Minnetonka and Gull are genetically closely related and when we analyzed them with 9 microsatellite markers, they all fell together into one large undifferentiated group. But the same mussels analyzed with 5,000 SNPs fell out into 2 groups by home lake—Minnetonka or Gull. This greater power will allow us to resolve the pattern of the MN invasion to a much greater extent, and will provide much clearer information to management.

Summary Budget Information for Activity 1:

ENRTF Budget: $149,773
Amount Spent: $ 24,475
Balance: $125,148

Activity Completion Date:

Outcome	Completion Date
1. *30 mussels from each of 72 waterbodies genotyped and analyzed using microsatellite DNA markers*	August 31, 2018
2. *10-15 mussels from each of 72 waterbodies genotyped and analyzed using SNP markers*	December 31, 2018
3. *Findings summarized and management recommendations made to DNR; results published*	December 31, 2018

Activity 1 Status as of July 31, 2017:

Sample collection and microsatellite analysis. We launched the 2017 sample collection season, and traveled to 31 water bodies to date (75 total is the target). We successfully collected from 23 of these water bodies. Several lakes in this list were not identified in the original research scope (see research addendum, appendix 2) but were collected opportunistically, and may be analyzed as alternates because (1) they are located in regions that will be informative to understand spread in MN, and (2) they were recently infested and are destination lakes with possible hub-lake origins.

We sampled the following lakes originally included in project scope: *No mussels found*: Upper Hay (Crow Wing), Turtle (Douglas). *Collected*: Virginia (Carver), Waconia (Carver), Lower Whitefish (Crow Wing), Pine River (Crow Wing), Latoka (Douglas), Green (Kandiyohi), Minnewaska (Pope), Pleasant (Ramsey), Sucker (Ramsey), Vadnais (Ramsey), White Bear (Washington). *Visited, no access obtained yet:* Ogechie (Mille Lacs), East Sylvia (Wright).

We sampled the following additional lakes alternates for analysis: *No mussels found*: Round (Aitkin), Pocket (Douglas), Calhoun (Kandiyohi), Florida (Kandiyohi), Washington (Meeker). *Collected*: Andrew (Douglas), Lobster (Douglas), Louise (Douglas), Mill (Douglas), Stella (Meeker), Emily (Pope), Sylvia (Stearns), Big Birch (Todd), Forest (Washington), John (Wright), Clearwater (Wright and Stearns).

We genotyped samples from Lakes Superior, Erie (western: 2 sites), St Clair, Huron (2 sites) and Michigan (2 sites) at 9 microsatellite markers. We have analyzed the new microsatellite dataset (all the samples collected from Minnesota up through 2016 + the Great Lakes samples) by genetic clustering to identify source population, assignment and ABC invasion model testing to identify pathways of spread. These results continue to support the surprising lack of contribution of Mille Lacs and Prior Lakes to more recent infestations, and they futher confirm

genetic patterns in clustered invasions in all Detroit Lakes, Alexandria, and Brainerd Lakes regions. Also surprisingly, we have found little evidence for genetic structure between and within Great Lakes, based on microsatellites, to date—contrary to results from the literature. These samples are being analyzed at SNP makers now. We completed revisions of the manuscript for *Biological Invasions*: analysis of spread on the entire microsatellite data set from MN [all data except for samples analyzed from the Great Lakes, and from some sites analyzed after Phase I funding ended in 2017.

SNP (population genomic) genotyping and analysis. We tested and used a molecular assay (adapted from the literature) to ID zebra and quagga mussels—this test was needed prior to SNP genotyping Great Lakes collections (which are dominated by quagga mussels), so as to avoid expensive genotyping of quagga mussels in this project. Quagga mussels have not yet spread widely in the state and genotyping them requires method development. We processed for SNP genotyping all samples genotyped in Phase I with microsatellites in order to validate previous results generated from a different technique. We also processed for SNP genotyping the newly-extracted Great Lakes samples and submitted all of the samples for genomic SNP analysis [using UMGC's assay refined for zebra mussels in Phase I (December 2016)]. To date we have received data from 439 mussels, and over 500 more are pending at UMGC.

We performed initial analyses of SNP data to examine the effects of data-filtering parameters. We (the McCartney lab) then filtered the SNP data, scored SNP genotypes, and performed initial clustering analysis. While the results of the latter are too preliminary to describe in great detail, we can report that there is a striking concordance between SNP and microsatellite data for the major patterns of structure in the data (e.g. the uniqueness of Mille Lacs and regional clusters), and that we have uncovered genetic patterns with finer resolution in the SNP data that may be useful in studying spread. Using conservative parameters, we scored 3320 SNPs for 439 mussels. More than ten times as many can be scored by relaxing parameters, at the cost of more missing data per mussel. But this process is flexible and can be "tuned" to obtain the optimal resolution to decipher spread patterns.

Activity 1 Status as of January 31, 2018
Activity 1 Status as of July 31, 2018

Final Report Summary:

ACTIVITY 2: Genome sequencing and analysis to find target genes for biocontrol.

Description: Here we search the zebra mussel genome for genes that control crucial life processes, so that in future research, these genes can be modified, inserted back into zebra mussels, and ultimately propagated through populations for their biological control. Emerging molecular technologies are in reach to "edit" these genes to make precise mutations, or to otherwise turn genes off or down and disturb their function, but the first step is to identify **gene targets**. To do so and for the first time we sequence the entire zebra mussel genome, use "bioinformatic" analyses on supercomputers to identify genes and search them to find target genes, then confirm target gene identity using "transcriptomes"—these are libraries of all the genes expressed in the networks or pathways that control crucial life process (like surviving periods of high water temperature).

First the genome is sequenced—i.e. the DNA bases making up the entire sequence of DNA that forms the hereditary material for the zebra mussel, along all of its chromosomes, is determined. In the case of zebra mussels, we generate this genome sequence from a combination of sequences of billions of short DNA sequence fragments ("short reads": completed in 2016) and long sequence fragments or long reads (a more complex and time consuming effort to be completed in this work plan). *Second, short and long DNA fragments are patched together or "assembled*," using computer programs, into contiguous stretches of DNA. For example, all of the DNA bases along each of the 32 chromosomes is assembled into 32 groupings (and other subgroupings of genes that are passed to on to offspring in linked sets). *Third, the assembled genome is "annotated" (genes are located, named and their functions identified)*. Annotation requires additional data that allows genes to be identified, named and their function determined— we will obtain these data from "transcriptomes." Each transcriptome contains sequences of all of the hereditary messages (encoded in "messenger RNAs") that are expressed by a life stage of zebra mussel under a set of conditions (e.g. all of the RNAs expressed by a batch of adults mussels exposed to their highest tolerated temperature will be used to make one of the heat-stress

transcriptomes). RNA provides the intermediate genetic code to turn the mussel DNA into proteins and other molecules required to grow tissues, develop organs, shells and other structures, and perform biological functions. We sequence transcriptomes using RNA-Seq technology, and then computer scientists (bioinformatic analysts) use these transcriptome sequences to help identify genes and annotate the genome. And, finally, *the fourth step is to find target genes by homology searching.* Homology searching looks through the genome for DNA sequences that are similar to DNA sequences in other (well-studied) species, due to descent from common ancestors shared between zebra mussels and these other species. The key is to select *a priori* a good set of candidate genes on which to focus these searches. These candidate genes are known in other species to control processes that are good bets to be targets or weaknesses to exploit in biocontrol—we offer a few examples below under Activity 3, where we describe how we choose which target genes to search for.

Summary Budget Information for Activity 2:

ENRTF Budget: $ 198,827

Amount Spent: $ 21,002

Balance: $ 177,825

Activity Completion Date:

Outcome	Completion Date
1. *Long read genome sequencing*	December 2017
2. *RNA-Seq of transcriptomes (genes expressed in different life stages under various environmental conditions)*	December 2017
3. *Bioinformatics: genome assembly and annotation*	March 2018
4. *Bioinformatics: search for 3 high-potential sets of target genes*	March 2018

Activity 2 Status as of July 31, 2017: Our first task was to extract high quality, long fragments of DNA from mussels of known gender. The purpose is to construct the full genome from a single male mussel, and to use long-read sequencing to ensure that we can assemble this challenging genome sequence. We do not know whether zebra mussels have sex chromosomes. Earlier genetic studies have not identified sex chromosomes that are recognizable (like the human X and Y) in zebra mussel chromosome preparations examined under the microscope, but chromosome "morphological" differences are absent from many organisms that possess chromosomal sex determination. Based on a survey of animals, males have the best chance of having two different sex chromosomes. Starting in May and continuing through June 2017, we collected large mature zebra mussels from Pelican Brook in Crow Wing County. Then we sexed them under the microscope by viewing live sperm or developing ova in gonad tissue. We dissected several tissues from over 3 dozen large animals sexed in this way. Next we used highly specialized DNA extraction techniques (the QIAGEN Genomic Tip technique and a protocol that McCartney and Mallez developed) and gentle handling methods and have isolated DNA from a subset of these animals that is on average 20-60,000 bases in length—more than 4 times longer than our previous DNA isolations. UMGC is building test libraries for long read sequencing from these DNA extracts.

We also collected specimens from Pelican Brook and from Lake Minnetonka and preserved material for the following transcriptomes: [larval (D-stage, umbonal stage, pediveliger), female adults and male adults (gonad, mantle) from high-calcium environments (transcriptomes 3-8 in research addendum, Appendix 3)].

Activity 2 Status as of January 31, 2018

Activity 2 Status as of July 31, 2018

Final Report Summary:

ACTIVITY 3: Choosing target genes, biocontrol technologies and strategies

Description: The zebra mussel genome will be an invaluable resource for zebra mussel biology, but to enable its complete analysis and application to genetic biocontrol, we will create a formal research network. This Dreissenid Mussel Genome Collaborative (DMGC) will bring together zebra mussel biologists, invertebrate biologists with specialized expertise, genome scientists, and experts in genetic biocontrol. Activity 3 outcomes and products are: **Outcome 1**: **Forming the DMGC and laying out plans in a "white paper,"** and **Outcome 2: Workshop: harnessing the zebra mussel genome for biocontrol.**

In the DMGC, we have 4 teams and 4 themes—each a topic for the white paper (**Outcome 1**), and the workshop on genome analysis and applications to control (**Outcome 2**). Here are the teams with some example applications; the first 3 concern target genes: *(1) Developmental biology*—for example, we would search for genes that control formation of the shell of the larva. This is different from the shell of the adult mussel, and we know from other bivalves that a major gene that controls development of the larval shell is the gene called *engrailed* (isolated from fruit flies, where it serves a very different function—forming the body segments). Silencing or editing *engrailed* might generate unshelled larvae that would die before settling out on lake bottoms. *(2) Byssal threads*—these are the strong fibers that marine mussels use to attach to rocks on the coast, and zebra mussels use to attach to lake bottoms, vegetation, boat hulls and lifts. Here we might search for genes expressed in the mussel foot that control fiber growth rate, which we could modify to yield a mussel that attaches less firmly to boats and vegetation, and is spread from lake to lake less frequently. The "non-GMO" version of this is the quagga mussel, which forms many fewer, poorer attachment-strength fibers; quagga mussels seldom invade inland lakes and one explanation is that they are poor hitchhikers on watercraft. *(3) Thermal tolerance*—this includes a legion of candidate genes that respond to heat stress by protecting tissues or scavenging their heat-damaged byproducts, or genes that enhance immune response in a heat-stressed mussel exposed to disease. In this case, perhaps we would silence an immune system gene whose low expression leads to higher mortality during summer die-off events, such as was described in the Pacific oyster.

Team/theme *(4) Biocontrol technologies and strategies*: methods for turning off or turning down the expression of genes*,* using a technology called *RNA interference (RNAi)*. This team also covers *gene-editing methods* that allow for precise mutations to be created within target genes. Gene editing is coupled to *gene-drives*, which are used to drive these mutations to high frequencies in populations—*CRISPR/Cas9* is the name for one of these agents. This team would also investigate technologies for delivery of the modifying agent to zebra mussels; e.g. they would address challenges in delivering *CRISP/Cas9* constructs*.* Microinjection of these into embryos is impractical for the numbers of modified offspring needed for a zebra mussel control program. *Sperm mediated gene transfer* might instead be used to generate sperm into which mutant genes have been transferred. These "transgenic sperm" could be released onto mussel beds to fertilize eggs released by female mussels *in lakes.* After that, the gene drive mechanism takes over, generating offspring with two copies of the mutation, one on the modified chromosome and one the "gene-driven" chromosome, and a chain reaction follows, driving the mutation to high frequencies in the lake.

In February 2017, we start with in-depth literature research, and contact potential contributors and reviewers for a white paper. This paper provides the background and the rationale for sequencing the zebra mussel genome, and for genetic modification for biocontrol. The paper also focuses on the choice of the target genes; although depending on our literature research and discussions, we may replace one or more of these targets with others, if we choose another target as more feasible than the three candidates chosen so far (teams/themes 1-3 above). We seek 4-6 people to help write, and another 4-6 people to review the white paper. It is important that this effort is completed before the 2017 field season, during which the McCartney lab will build transcriptome libraries. If for example teams 1-3 were to decide to eliminate one of these target processes and substitute another, we would have time to build the replacement transcriptome.

Twelve invitees (3 from each of 4 teams) + 1 UMGC genome scientist + 1 MSI informatics analyst + 4 graduate student support staff will then convene the workshop in Minnesota in Spring 2018. Day 1 begins with UMN personnel, who introduce the genome project. Next, major features of the draft genome are presented, with attendees provided access to viewing its features on workstations throughout the workshop (summaries will be provided in advance). Afternoon of day 1 through morning of day 2: presentations on the 4 theme topics from invitees, with roundtable discussions of the "next steps"—in genome analysis, in research on technologies for genetic modification and delivery to zebra mussels, and in research to investigate strategies for hypothetical release programs. The final 1.5 days will be spent drafting manuscript and proposal sections.

Summary Budget Information for Activity 3:

ENRTF Budget: $ 79,500
Amount Spent: $ 13,070
Balance: $ 66,430

Activity Completion Date:

Outcome	Completion Date
1. *Collaborative formed, "White paper," choosing target genes for transcriptome sequencing, draft*	~~May~~ November 2017
2. *Workshop held: harnessing the zebra mussel genome for biocontrol (technologies and strategies)*	April 2018
3. *Manuscript(s) completed*	December 2018

Activity 3 Status as of July 31, 2017: We have begun to contact potential members of the collaborative. For developmental genes, we have contacted Professor Gary Wessel from Brown University, an expert on embryonic development, particularly of the germ line (cells giving rise to gonads, eggs and sperm) and genes that control developmental origins of reproductive tissue in invertebrates. Wessel's laboratory also has introduced CRISPR/Cas9 gene edits into invertebrates (sea urchins and other species) and is experienced with larval culuture. Our colleague Eric Hendrickson at UMN will collaborate on CRISPR/Cas9, and our colleagues Chris Merkes and Jon Amberg at USGS are interested in collaborating on using our genome to perform RNAi experiments to turn off and turn down the expressionof target genes (different ones than we will target with gene editing).

We have added plans to include shell formation genes among (now) 4 classes of candidate target genes for biocontrol. We plan to sequence, identify and characterize expression of genes responsible for shell formation, using the Pacific oyster genome as a resource. Thirty-nine genes controlling shell formation in adults and larvae were identified with high confidence from the Pacific oyster genome, and many more are present. The oyster genome answered many questions about genetic control of shell building in bivalves. Study of zebra mussel shell formation and calcification will benefit greatly from the oyster genome and from communication with bivalve genome scientists that worked on that project. Zebra mussel shell formation genomics holds great promise for providing target genes for biocontrol. Oysters build shells in calcium rich environments. In contrast, some zebra mussel populations live in geographic regions in which calcium poor lakes (below about 12 mg/l concentration) exist. In these waters, zebra mussels are incapable of growing shell rapidly, and worldwide the species is generally absent from waters below this calcium "threshold." Calcium concentration is the single most important water chemistry factor that can predict which lakes are habitable by zebra mussels, and is among the most glaring "weak points" that may be exploited for biocontrol.

Populations from relatively calcium-poor environments [Gilbert Mine Pit (Ore-be-gone) in St. Louis County, and sites we have already collected from Lake Superior (Apostle Islands National Lakeshore, WI)] will be used to build mantle transcriptomes (mantle is the major shell building organ), and these will be compared to mantle transcriptomes from animals from calcium rich environments, and to the oyster genome and other bivalve shell formation genes in this comparative genomic study. This plan does not require new work, as adult mantle was

included in the research addendum (transcriptomes 7-8 in research addendum, Appendix 3) but it signals a change in emphasis. McCartney has several past connections with the "biomineralization" research community while working with colleagues who studied molting and hardening of the blue crab shell, post-molt. This research community is large and studies the biochemistry, cell and molecular biology of mineral deposition into structural tissue by animals, plants and algae. McCartney will continue to search this community and members of the Pacific oyster genome project for collaborators/advisors.

Hence our present set of target genes for transcriptome sequencing includes genes controlling (1) embryonic development, (2) byssal thread formation, (3) thermal tolerance, and (4) shell formation and biomineralization, and these 4 processes will be reviewed in the white paper.

Activity 3 Status as of January 31, 2018
Activity 3 Status as of July 31, 2018

Final Report Summary:

V. DISSEMINATION:
Description: Results will be delivered to professional managers and the scientific community through presentations at professional meetings, reports and journal articles. **At the local scale**, we also will continue to communicate our work in presentations and meetings with lake associations, watershed districts, MN County AIS managers, and other Local Government Units. Results will be regularly communicated to the Center Advisory Board and Technical Committee of MAISRC. We will also continue to communicate activities and outcomes **to the general public** via the MAISRC Website, Facebook page, Newsletter and through mass media interactions that continue to occur due to frequent interest in zebra mussel research.

At the scale of the state of MN, regular updates **to MN DNR** to provide advice for management decisions is an important function of the zebra mussel research program. In Phase I, we accomplished this by regular discussions and presentations at DNR Statewide AIS Advisory Committee (SAISAC) meetings, since McCartney was a SAISAC member from 2014-2016. MAISRC Associate Director Becca Nash continues in that capacity and will ensure that we regularly update and give presentations to SAISAC—there is considerable interest at DNR in our pathways research, and in the genome project. McCartney is a member of 2 other committees in the MN DNR Invasive Species Program (Pilot Project Permit Review and Zebra Mussel Control and Management Committee) and will regularly advise these committees on progress on biocontrol options. To better facilitate DNR communication and access to our findings by DNR managers, we are exploring new ways to use the MAISRC and DNR web resources to deposit research updates.

At the state to regional to national scales and beyond our plan starts at the two largest professional meetings in the surrounding region for state, federal, and university scientists who work on zebra and quagga mussels. For reaching MN professionals, the Upper Midwest Invasive Species Conference has become the best event. In 2016, it was held in LaCrosse and typically rotates between WI and MN. The next meeting may be held in MN or shifted to another state in the Upper Midwest. Attendance has grown annually, and our presentations in 2016 reached more professionals than ever before—*and* members of the lay public (e.g. watershed district personnel) who are now attending. The 2018 conference will be a perfect venue to communicate the work from the entire project. We also request funding to attend the International Association of Great Lakes Research conference, held annually in the Great Lakes Region (in Detroit for 2017). This scientific meeting (associated with the Journal of Great Lakes Research) is the premier conference for reaching the greatest number of scientists worldwide that work on zebra and quagga mussels. McCartney will attend and present, and use this venue to recruit professionals for the Dreissenid Mussel Genome Collaborative (DMGC). Our plan is to connect the work of the DMGC to the Invasive Mussel Collaborative (IMC: https://invasivemusselcollaborative.net/about/). The IMC consists of a network of federal scientists (U.S. Geological Survey, Great Lakes Commission, National Oceanic

and Atmospheric Administration, Great Lakes Fishery Commission), as well as a large number of professionals in state and local agencies concerned with research and management of zebra and quagga mussels. IMC is a management, research and communication network that is "…being established to advance scientifically sound technology for invasive mussel control to produce measurable ecological and economic benefits."

Status as of July 31, 2017:

All presentations are co-authored by McCartney and Mallez and are listed by presenter:

Local scale, presentations by M.A. McCartney

"Perspectives on the Lake Washington zebra mussel infestation," Lake Washington Improvement Association Meeting, Dassel MN, May 6, 2017.

"Zebra mussel pathways of spread, and strategies for biocontrol: an update on population genetic and genomic research," L'Homme Dieu Lake Association Meeting, Alexandria, MN; July 14, 2017.

Local scale, presentations by S. Mallez

"A better future for MN waters: Studying the pattern of spread of zebra mussels," Zebra Mussel Prevention Workshop, St. Croix River Association Meeting, Taylors Falls MN, April 24, 2017.

"Studying the spread of zebra mussels: Population genetic and genomic research," Pelican Property Owners Association - Annual Meeting, Pelican Lake MN, July 22, 2017.

State scale, presentations by M.A. McCartney

"An update on genetic research on zebra mussels at the MN Aquatic Invasive Species Research Center," MN Coalition of Lake Associations Annual Meeting, St. Cloud, June 7, 2017.

"Zebra mussel pathways of spread, and strategies for biocontrol: an update on population genetic and genomic research," MN DNR Fisheries Research Meeting, UMN St. Paul, June 26, 2017.

"Zebra mussel pathways of spread, and strategies for biocontrol: an update on population genetic and genomic research" MN DNR Invasive Species Program, Ecological and Water Resources, July 25, 2017.

Status as of January 31, 2018:

Status as of July 31, 2018:

Final Report Summary:

VI. SUB-PROJECT BUDGET SUMMARY:

A. Preliminary ENRTF Budget Overview:

*This section represents an overview of the preliminary budget at the start of the project. It will be reconciled with actual expenditures at the time of the final report. See the Sub-Project Budget document for an up-to-date project budget, including any changes resulting from amendments.

Budget Category	$ Amount	Explanation
Personnel:	$273,530	Salary for PI (McCartney), FTE = 0.83 (10 months) for year 1, 0.66 FTE (8 months) for year 2; fringe rate = 33.7%. Postdoctoral Fellow (Sophie Mallez), FTE = 1.0 for 1.5 years; 20.75% fringe rate; undergraduate assistant, 6 month per each year, 15 hours per week.
Professional/Technical Services and Contracts:	$138,950	Fees to the University of MN Genomics Center for genome sequencing ($47,000) and genotyping ($35,251); fees to the MN Supercomputing Institute for bioinformatic analysis ($37,000); repair of laboratory

		instruments ($4200), express mail and mailing fees to ship specimens ($500)
Equipment/Tools/Supplies:	$8,790	Laboratory: reagents and plasticware, supplies for molecular biology, laptop computer; Field: supplies for collecting zebra mussel samples; Office: supplies, software, software updates
Capital Expenditures over $5,000:	$0	
Travel:	$6,680	**MN**: Vehicle rental, fuel, lodging, meals for trips to collect specimens within the state ($1,866) **Domestic**: Vehicle rental, fuel, lodging, meals to collect specimens from Lakes Erie, St. Clair, and Michigan + Illinois and Kankakee Rivers to understand source of spread to MN (total = $1,934). Domestic travel also covers professional conferences (total = $2,880). The first is the Upper Midwest Invasive Species Conference (McCartney and Mallez). This is held every 2 years (in MN, WI…) and is the most appropriate conference for reaching the MN audience of professionals in state agencies and academia. We also budgeted for McCartney to attend the International Association of Great Lakes Research conference (IAGLR) held each year throughout the Great Lakes Region. This is the premier conference for federal, state, and university scientists in the US and Canada who study zebra and quagga mussels. McCartney will attend to make contacts to build the Dreissenid Mussel Genome Collaborative.
Other:	$0	
TOTAL ENRTF BUDGET:	**$427,950**	

Explanation of Use of Classified Staff: none

Explanation of Capital Expenditures Greater Than $5,000: none

Number of Full-time Equivalents (FTE) Directly Funded with this ENRTF Appropriation: 2.0

Number of Full-time Equivalents (FTE) Estimated to Be Funded through Contracts with this ENRTF Appropriation: none

B. Other Funds:

Source of Funds	$ Amount Proposed	$ Amount Spent	Use of Other Funds
State: Updraft Award, University of Minnesota Informatics Institute			
	$5,000	$0	
TOTAL OTHER FUNDS:	**$5,000**	**$**	

VII. SUB-PROJECT STRATEGY:

A. Sub-Project Team/Partners:

Partners funded on this project

- **UMN Genomics Center**: genome sequencing, sequencing strategy design, development of SNP genotyping methods, SNP genotyping ($82,250)
- **MN Supercomputing Institute**: bioinformatic analysis of genome and transcriptomes ($37,000)

Project partners not receiving funding

- **MN DNR** for continued assistance with specimen collections and general field assistance. MN DNR Invasive Species program biologists collect from recently infested lakes.
- **National Park Service**, **Northland College, NOAA Great Lakes Environmental Research Laboratory, Illinois Natural History Survey:** assistance with collections from Great Lakes, Illinois River and other sites out of state

B. Sub-Project Impact and Long-term Strategy: MN DNR, MN Counties, watershed districts and other lake managers will utilize our results on spread pathways. The results will help plan geographic pinch-points for prevention, and we expect that the results will continue to break from expectations. We expect the data to show a large role for regional stepping-stone invasion after an area receives its first infested water body—the resulting clustered invasions of lakes have driven a lot of the invasion rate increase in MN. We expect to find evidence for invasions from sources that may implicate vectors that need closer attention. This includes vectors that transport attached mussels (not veliger larvae in water) such as "non-transient" boats (e.g. those stored in marinas for extended periods) and water-related equipment (e.g. docks and lifts). We also expect a larger role for rivers as sources than is currently recognized, and we may produce some evidence for import from outside MN. Users of the zebra mussel genome will be many—at first, the dreissenid mussel research community worldwide will make use of the publicly available database and we will promote this with the collaborative. The opportunities for this to promote progress towards management solutions are considerable. A sequenced genome is a resource that should be available for every invasive species.

Techniques and products generated (the white paper, workshop, the draft genome and the results from high value target searches) will provide the foundations for future research on genetic biocontrol of zebra mussels. In Phase III, we will complete the analysis and annotation of the genome in collaboration with the Genome Collaborative, and publish this product in a highly visible journal. Many forces, some beyond the science and technology, will direct choices on future research and implementation of biocontrol. We will coordinate at the start with efforts to analyze risk and assess ethical challenges. The 2016 MAISRC Research Needs Assessment invites work to "… identify and address the ethical and regulatory challenges for the release of genetically modified organisms for invasive species control." While our workshop will focus on the *effectiveness* of technologies for control of AIS and not on regulatory or ethical issues, there will be discussion about safety—e.g., containment and reversibility of gene drives, and we intentionally take on RNAi technology, in part due to its perceived lower risk. We will also assist with the planning of the risk/bioethics symposium. That meeting clearly needs to happen before any funds are spent on developing genetic biocontrol technologies. Technologically, gene drives are clearly the most effective currently available way to spread genetic modifications throughout a lake population, but the risks may be perceived to be too great, and we want to work towards implementation within a decade. Future directions for the invasions pathways work would include an integrated analysis of the genetic data and boater behavior data, especially to examine discordance between boater and genetic pathways; e.g. the case of Mille Lacs Lake. Also, how would we (based on the genetic data) plan geographic pinch-points to better interfere with invasion pathways? And, patterns of chronic regional short-distance spread are revealed by genetic data—what are the causes? Above, we implicate high risk vectors that transport live zebra mussels (docks, lifts, non-transient boats)—how can we collect data to study these risks, and can we provide any information on areas in the state that merit increased attention to block transport by these vectors?

C. Spending History:

Funding Source	M.L. 2012 or FY13	M.L. 2013 or FY14	M.L. 2014 or FY15	M.L. 2015 or FY16
MN Clean Water Fund	$532,111			

VIII. ACQUISITION/RESTORATION LIST: N/A

IX. VISUAL ELEMENT or MAP(S): N/A

X. ACQUISITION/RESTORATION REQUIREMENTS WORKSHEET: N/A

XI. RESEARCH PROPOSAL: Attached as research addendum

XII. REPORTING REQUIREMENTS:
Periodic work plan status update reports will be submitted no later than [July 31, 2017**], [**January 31, 2018**], and [**July 31, 2018**]. A final report and associated products will be submitted within two months of the anticipated sub-project completion of [**December 31, 2018**].**