Title: The zebra mussel genome project: developing a new resource for invasion biology and biocontrol research

Running header: Zebra mussel genome project

Authors: Michael A. McCartney[1], Sophie Mallez[2], Daryl Gohl[3], Kenneth Beckman[3]

Contact Information:

[1]mmccartn@umn.edu

[1,2]University of Minnesota, Minnesota Aquatic Invasive Species Research Center and Dept. of Fisheries, Wildlife and Conservation Biology, 2003 Upper Buford Circle, St. Paul, MN, 55108 USA

[3]University of Minnesota Genomics Center, 2231 6th Street SE, Minneapolis, MN 55455

ABSTRACT

Rapidly falling costs and advances in sequencing and informatics have made genome sequencing projects far more accessible to researchers in all of the life sciences, including invasion biology. A complete genome is now the most efficient way to identify and characterize genes controlling traits that contribute to invasiveness. At the genomic level, moreover, tremendous power is available to investigate fundamental questions in invasion science (e.g. the relative roles of pre-adaptations vs. post-colonization adaptive evolution in invasion success), and genomic analysis provides new options for development of control technologies. Yet relatively few invasive species genomes have been sequenced, and even fewer of these genomes have been put to use to study invasiveness. In this perspective, we describe an ongoing effort to sequence the genome of the zebra mussel and how this resource might aid in the development of future biocontrol strategies. We invite dreissenid biologists and others to join us in annotating and analyzing this genome, so that its full potential in understanding and controlling this highly destructive animal can be realized.

Introduction

A sequenced genome will soon become a routine part of any research program in biology. Costs drop almost monthly, and data collection and analysis technology development is moving so fast that projects often take advantage of new inventions while underway. One spectacular example (early 2018) is the Mexican axolotyl, an unprecedented long-read sequencing effort that required the creation of a new algorithm just to assemble its 32 gigabase genome, which is 10 times the length of *Homo sapiens* (Nowoshilow et al. 2018). Most of the life sciences can now benefit from the power of genomics and the new questions it can help to ask and answer. Invasion biology is no different. In this report, we briefly review contributions of genomics to the discipline to date, and describe our ongoing effort to sequence the zebra mussel genome.

Native to a small region of southern Russia and the Ukraine (Stepien et al. 2014), zebra mussels (*Dreissena polymorpha* Pallas 1771) have spread throughout European (Karatayev et al. 1997; Karatayev et al. 2003) and North American (Benson 2014) fresh waters to become one of the world's most prevalent and damaging aquatic invasive species (Karatayev et al. 2007). Fouling of water intake pipes costs the power generation industry over $3 billion USD from 1993-1999 in the Laurentian Great Lakes region alone (O'Neill 2008), where dreissenids are a large and complex economic burden to hydropower, recreation and tourism industries and lakefront property owners (Bossenbroek et al. 2009; Limburg et al. 2010). Ecological damage, the extent of which is just beginning to be understood, arises from the tendency for dense infestations to smother and outcompete native benthic species, and remove huge volumes of planktonic organisms from lakes and rivers. Among the noteworthy impacts are widespread population declines and local extinctions of native freshwater mussels and other invertebrates,

damage to fish populations in some cases (Karatayev et al. 1997; Lucy et al. 2014; McNickle et al. 2006; Raikow 2004; Strayer et al. 2004; Ward and Ricciardi 2014), and dramatic restructuring of aquatic food webs (Bootsma and Liao 2014; Higgins and Vander Zanden 2010; Mayer et al. 2014). The congener *D. rostriformis "bugensis"* (or *D. bugensis*: the quagga mussel), while still nowhere near as widespread as the zebra mussel in North American lakes, has ecologically replaced zebra mussels in much of the Laurentian Great Lakes proper and in parts of Europe, and may lead to even greater ecological damage in those systems (Karatayev et al. 2011; Matthews et al. 2014; Nalepa and Schloesser 2014).

The ongoing European and North American invasions spurred an explosion in research effort on *Dreissena*—particularly focused on physiology, autecology, and ecosystem impacts (see (Schloesser and Schmuckal 2012) for a bibliography from 1989 - 2011). Aside from molecular systematic and population genetic studies (Brown and Stepien 2010; Gelembiuk et al. 2006; Mallez and McCartney 2018; May et al. 2006; Stepien et al. 2014), comparatively little genetic work has been accomplished, with transcriptomes from a few tissues (Soroka et al. 2018; Xu and Faisal 2010) the only genomic resources. With the sequence of the zebra mussel genome, we will provide a powerful new resource. We hope to bring together dreissenid mussel researchers and others who can analyze it in appropriate detail, and apply it to better understand and cope with this fascinating but highly destructive animal. Therefore, the first goal of this paper is to advertise the project and invite collaboration.

Our other goal is to consider more broadly the potential contributions of genomics to invasion biology. Three years ago, Rius et al. (2015) reviewed applications of Next-Generation Sequencing technologies to the study of biological invasions, and it is our intent to update their

valuable review. In just 3 years, a sequenced reference genome has become an accessible goal or a resource that is an already available for the study of high-priority invasive species. In this paper, we describe some applications of genome projects to broader questions in invasion biology and towards the development of control technologies; some specific to dreissenids.

Genomics in invasion biology

To illustrate the availability of genomic resources to invasion biologists, we searched for assembled genomes from the 100 "world's worst" alien invasive species according to IUCN (Lowe et al. 2000) on Genbank's Genome resource (https://www.ncbi.nlm.nih.gov/genome/). Twenty-eight of these species have assembled genomes available of varying degrees of quality (Table 1)—a sizable resource for invasion biologists. Of course, there are many reasons for sequencing a genome, and several of these 28 projects were launched because of economic value or use as model species [e.g. *Oncorhynchus mykiss* (rainbow trout), *Sus scrofa* (pig), *Capra hircus* (goat), *Mus musculus* (mouse)]. Moreover, for only 8 of these species did we find that invasiveness was a topic of discussion in publications announcing the genome sequence. This is similar to what Rius et al. (2015) noted—I.e. that invasion biology *per se* has driven interest in genome projects projects in only a minority of cases.

Genomic studies of evolution of invasiveness

So then what does a complete genome provide for invasion biology research? For one, of great ongoing interest is whether and how invasions are facilitated by adaptive evolution (Cristescu

2015; Lee 2002; Sax et al. 2007). Genomic analysis provides unequalled power for identifying "invasiveness" genes and for characterizing their mode of evolution.

One good example is the Southeast Asian fruit fly *Drosophila suzukii*, which is rapidly expanding in Europe and North America since arriving about 2008 on these continents (Asplen et al. 2015). Unlike other (genetically more well-characterized) *Drosophila, D. suzukii* shows the unusual behaviors of egg laying and larval feeding on ripening rather than fermenting fruit, and as a consequence has become a damaging pest of soft fruits (e.g. blueberries, blackberries, strawberries). As part of research to develop integrated pest management, nuclear genomes, mitogenomes, and transriptomes were recently sequenced and analyzed (Ometto et al. 2013). To examine adaptive molecular changes associated with the ecological shift to ripening fruit, Ramasamy et al. (2016) analyzed the repertoire of 131 genes involved in olfaction throughout the genus—those encoding odorant receptors and other receptor proteins expressed in antennae, and the odorant binding proteins. They found several instances of gene loss, duplication and positive selection within these gene families along the *D. suzukii* lineage—candidate adaptations that facilitated the switch in larval feeding and egg laying behaviors and promoted the success of this host plant shift. This study could not have been accomplished without genomic resources.

The Asian longhorned beetle (*Anoplophora glabripenniss)* causes damage to > 100 tree species worldwide, and belongs to the beetle family containing the most species capable of feeding on woody plants. Its genome sequence (McKenna et al. 2016) included a large repertoire of enzymes that can digest wood, including several acquired through horizontal gene transfer from bacteria and fungi. The medfly *(Ceratitis capitata)* is able to locate and feed on a

diversity of host plants, and its genome (Papanicolaou et al. 2016) shows “expansion” (by gene duplication) of chemosensory and visual genes, and others that encode detoxification of plant secondary compounds and synthetic pesticides. Similarly, expansions of gene families encoding immunity, diapause, and insecticide resistance are among the evolutionary changes within the Tiger mosquito (*Aedes ablopictus*) genome that may have promoted its range expansion throughout the world since the 1960’s (Chen et al. 2015).

In each of these cases, the extent of genomic changes involved (gene family expansions, changes in gene order and the like) suggests they arose prior to invasion. The issue of whether adaptations that favor invasiveness are pre-adaptive or whether they evolve rapidly, during and after establishment is of great academic and applied interest (Lee 2002; Ricciardi et al. 2017). Consider invasive plants, in which genomes of weeds have been shown to be smaller than genomes of non-weedy plants (Kuester et al. 2014). Shorter generation times, smaller seeds, and higher growth rates are associated with weediness and smaller genome size, but it is not clear whether small genomes promoted the evolution of weedy traits, or whether genome size reduction was selected for, post-invasion (Kuester et al. 2014). In each of the cases described above, comparative genomic analysis will allow future researchers to mine these genomes to learn much more about the rate and mode of the evolution of key invasiveness traits.

Genomics to study the invasiveness of dreissenids

It is clear that changes in transportation networks (e.g. canal building, opening of shipping channels, ballast water discharge) were the events that initiated primary invasions of European

and North American waters (Karatayev et al. 2007; Pagnucco et al. 2015). Several biological characters, however, are responsible for the rate of spread of zebra and quagga mussels across both continents, while other traits have limited the range of suitable habitats. Genomics offers a path to understanding these traits at the genetic level and in the future, this understanding may provide the tools needed to develop control strategies.

The fibers that zebra and quagga mussels use to anchor themselves to hard surfaces are known as byssal threads. These are key innovations (unique in freshwaters) that allow dreissenids to attach to virtually any hard surface underwater (rocks, plants, woody debris, other mussels) and to boat hulls, plants entangled on boats and trailers, docks, boat lifts and other recreational equipment—allowing rapid rates of spread between water bodies (Collas et al. 2018; De Ventura et al. 2016; Johnson et al. 2001).

Byssal threads are complex extracellular fibers secreted by the bivalve foot, and their underwater adhesion properties and role in biofouling have motivated detailed study, with the marine blue mussels *Mytilus* being most well-characterized (Brazee, Carrington 2006; Lee et al. 2011; Peyer et al. 2009). Byssal threads in *Dreissena polymorpha* differ from those of *Mytilus* in fundamental ways, reflecting their deep convergent evolution in two different subclasses (Heterodonta and Pteriomorpha). First, the regions of the byssus—(a) the thread proximal and (b) distal to the foot, and (c) the plaques (structures that cement the thread to surfaces)—differ from each other in protein composition in *Mytilus* but not in zebra mussels, where each region shows a similar modified protein composition (Waite et al. 2005). Second, the rare amino acid 3,4-diphenylhydroxydoamine (DOPA) is an important modifier of proteins in the plaques and cuticle of *Mytilus* fibers, where it confers mechanical and adhesive properties; in zebra mussels

DOPA is present but in much lower quantities (Rzepecki, Waite 1993). Third, and unexpectedly, zebra mussel fibers (given their environment of less hydrodynamic stress), are stiffer and stronger than those of marine species (Brazee, Carrington 2006).

Quagga mussels are now numerically dominant in the Lower Great Lakes and have invaded a number of reservoirs in the Colorado River system in the southwest US (Benson 2014). While they dominate nearby lake bottom in eastern Lake Erie and western Lake Ontario, zebra mussels outnumber them on boats that have remained in the water for extended periods in harbors (Karatayev et al. 2013). This suggest that poorer attachment abilities may help explain why quaggas have invaded so many fewer inland water bodies in North America than have zebra mussels. Notably, quaggas build lower attachment-strength fibers than zebra mussels, and anchor them more slowly in flow (Peyer et al. 2009).

Expression of genes associated with byssogenesis has been studied in zebra mussels (Xu, Faisal 2010) but a majority of mRNAs that are either up or down-regulated during the synthesis of the byssus could not be identified. Comparative analysis of zebra and quagga mussels would provide testable hypothesis about genetic differences between the two species in the control of fiber synthesis and attachment. In both zebra and quagga mussels, we are using RNA sequencing (RNA-Seq) of transcripts from the foot (the byssus-secreting structure) following experimental induction of byssogenesis (Xu, Faisal 2010) to launch these comparative studies. A complete *D. polymorpha* genome and further annotation of genes expressed in the foot would benefit from ongoing *Mytilus* transcriptomic and proteomic analyses, which have discovered byssal thread foot proteins that were not found earlier in the fibers themselves (DeMartini et al. 2017; Qin et al. 2016). Recent work on proteins in the fibers of quagga mussels (Rees et al.

2016) and our transcriptomes will facilitate comparisons to zebra mussels.

Dreissena thermal biology has received some scrutiny by physiologists, and broad thermal tolerance and ability to adjust it to local conditions have clearly played a role in invasion success. Zebra mussels have higher lethal temperature limits, and they spawn at higher water temperatures in North America than in Europe (McMahon 1996, Nichols 1996). Populations in the Mississippi River provide a good illustration of their breadth of temperature tolerance. In the Lower Mississippi River zebra mussels are found south to Louisiana, where, without cooler water refuges within the river, they persist near their lethal limit of 29-30°C for 3 months during the summer, and for 3 months in the winter the river is at 5-10°C (Allen et al. 1999). In contrast, zebra mussels in the Upper Mississippi River encounter water temperatures > 25°C for just 1 month of the year, and < 2°C for about 3 months (data from USGS gauge from St. Paul, MN). Seasonal scheduling of growth and reproductive effort appears to be responsible for at least some of the adaptation/acclimation to conditions in the lower river, as populations in Louisiana shift their shell and tissue growth to the early spring and stop growing in summer (Allen et al. 1999) while more northerly populations grow tissue and spawn in summer months (e. g. Borcherding 1991; Claxton, Mackie 1998).

A properly annotated genome sequence could accelerate research on thermal adaptation in dreissenids. There is a vast literature on heat-inducible (e.g. “heat shock”) genes and proteins; in fact, marine bivalves and other intertidal invertebrates have been favored subjects (reviewed in Feder, Hofmann 1999). More recently, RNA-sequencing of transcriptomes in heat stressed animals has been accomplished in several invertebrate animals, including mollusks (Porcelli et al. 2015). The freshwater mussel *Villosa lienosa* was the subject of a small-

scale study (5 heat-stressed animals, 5 unexposed), using RNA-Seq. The authors identified a diversity of expressed genes associated with heat stress, including each of the major components of a classic "heat shock" response pathway, and the endoplasmic reticulum protein unfolded protein response (UPR^{ER}), including molecular chaperones, antioxidants, immune factors, cytoskeletal elements and mediators of apoptosis (programmed cell death; Wang et al. 2012) Extensive transcriptome sequencing of stress genes in the Pacific oyster genome project (Zhang et al. 2012) revealed most of the same genes and a few others in temperature stress trials. It is possible that survival in high temperatures in natural environments could be related to genes not involved in thermal tolerance *per se*— immune-surveillance genes, for example. Studies of selective summer mortality in Pacific oyster compared gene expression profiles between genotypes that survived and died, and showed that a set of immune response genes was positively associated with summer survival (Fleury and Huvet 2012). To improve the genomic resources available for studying thermal tolerance in zebra mussels, we have generated transcriptomes from gill tissue in animals exposed to periods of low (24°C), medium (27°C) and high (30°C) chronic temperature stress.

Water chemistry plays a large role in limiting spread of zebra mussels and calcium concentration is the most important single water chemistry parameter (e. g. Mellina, Rasmussen 1994; Whittier et al. 2008). There is evidence that biomass of zebra mussels within water bodies is limited by ambient Ca^{2+} concentrations, and evidence for threshold concentrations below which populations cannot persist. In North America, few inland lake populations are found at concentrations below 20 mg/L Ca^{2+} (Cohen, Weinstein 2001). At several sites along the St. Lawrence River, Mellina and Rasmussen (1994) found no zebra

mussel populations below 15 mg/L Ca^{2+}, while Jones and Ricciardi (2005) showed a decline in biomass of zebra and quagga mussels across a concentration range from 25 to 12 mg/L, with quagga mussel populations absent below 12 and zebra mussels absent below 7.5 mg/L. These thresholds are much higher than those for native sphaerid and unionid bivalves, which regularly occur at concentrations below 5 mg/L (McMahon 1996; McMahon 2002; Strayer 1993).

The mechanism(s) underlying poor tolerance of low Ca^{2+} in dreissenids have received relatively little study. Rearing success and percent of normal larvae were found to decline with Ca^{2+} concentration in laboratory studies of larval development (Sprung 1987). Vinogradov et al. (1993) showed that zebra mussel adults were unable to regulate Ca^{2+} concentrations in their circulatory fluid (hemolymph) at ambient concentrations < 12-14 mg/L (i.e. the animals lose Ca^{2+} to the surrounding water), and lower pH values further reduce their ability to regulate. Moreover, survival, reproductive output, somatic growth and shell growth have each been found to decline with calcium levels in experimental trials (Baldwin et al. 2012; Hincks, Mackie 1997).

In dreissenids and other bivalves, the shell is constructed of calcium carbonate of different crystal forms (typically calcite in adult and aragonite in larval shells) that are deposited in an organic matrix, either through an extracellular mechanism or one mediated by cells within the mantle tissue (Mount et al. 2004; Weiner, Traub 1984). Correlations between environmental Ca^{2+}, shell strength and calcification in some species, considered along with the evidence for selection on shell strength for predator defense in freshwater molluscs (Lewis, Magnuson 1999; Russell-Hunter et al. 1981 and references within), suggest that shell calcification may be the process responsible for low calcium sensitivity in dreissenids.

Genome sequences from bivalves have revealed a surprisingly large number of genes involved in shell formation. Searches of the complete Pacific oyster genome for similarity to known shell formation genes in other molluscs identified > 1,800 candidate genes, showed that some major genes are lacking, and revealed diversification of others, including a large variety of variants related to nacrein (Zhang et al. 2012), a component of the iridescent material inside the shell. Nacrein is also used to build pearls, and the pearl oyster (*Pinctada fucata martensii*) genome shows duplications in the nacrein gene family; one of the shell matrix-protein gene families whose diversity has been generated by tandem duplication to form gene clusters at 14 different loci (Takeuchi et al. 2016). Components of the shell-formation genome and proteome in *P. f. martensii* (Du et al. 2017) includes proteins related to collagen and others that are similar to the chondroitin sulfotransferase enzymes found in vertebrate bone.

For *D. polymorpha*, our specific interest would be in identifying genes related to calcification of the shell or "biomineralization" – the process whereby the protein-based shell matrix nucleates crystals of calcium carbonate, and orients their formation into the highly organized layers that compose the shell. Expressed Sequence Tags (ESTs) of messenger RNA's of the shell-building mantle tissue in the tropical pearl oyster *Pinctada margaritifera* (Joubert et al. 2010) were analyzed and a group of putative biomineralization-related genes were identified: 55 genes due to similarity to genes in other *Pinctada* species, 14 due to similarity to genes in more distantly related bivalves, and 13 due to similarity to genes in gastropods (a different class in Phylum Mollusca). For dreissenids, the most closely related bivalves for which sequence information is available at the genomic and/or transcriptomic levels are *Mytilus* (Murgarella et al. 2016), *Modiolus* and *Bathymodiolus* (Sun et al. 2017), all members of Family Mytylidae. We

are currently using RNA-Seq of mantle libraries to identify biomineralization-related genes. Half of these libraries were prepared from mussels collected from calcium-rich (35 mg/L) and half from mussels collected from calcium-poor (13-14 mg/L) water bodies as a way to infer genes that may be up or down-regulated in response to calcium limitation. As sequenced genomes and other genomic resources from molluscs become increasingly available, comparative approaches in evolutionary developmental biology of shell formation and mineralization (Jackson, Degnan 2016) could be employed to investigate mechanisms of sensitivity of dreissenids to low calcium.

The zebra mussel genome sequencing project

Bivalves are a diverse Class of Mollusca with over 10,000 described species in marine and freshwater environments (Appeltans et al. 2012; Bogan 2008). As of this writing (April 2018), complete genomes have been sequenced and analyzed adequately in only 7 species—all of them marine and most of commercial harvest value (Table 2). Yet 21 invasive bivalve species cause damage to aquatic and marine ecosystems worldwide (Sousa et al. 2009), and only one—the golden mussel, *Limnoperna fortunei*— has so far been the subject of a genome sequencing project (Uliano-Silva et al. 2017).

Zebra mussel genome sequencing strategy

Genomes of eukaryotic organisms typically contain millions of DNA segments that do not code for genes and consist of repeated sequence motifs. In fact, over half the genome of humans

and other mammals is comprised of repetitive DNA (de Koning et al. 2011) that arises from transposable elements and other unknown sources. Bivalve genomes are also highly repetitive, which makes assembly of raw data into contiguous sequences (contigs) challenging. The genomes of the two marine mussel species whose genomes have been sequenced – the deep sea *Bathymodiolus platifrons* and the intertidal *Modiolus philippinarum* – are highly repetitive, with 47.9% and 62%, respectively, being composed of repeats and transposable elements (Sun et al. 2017). Repeats are also common in oyster [36% of Pacific oyster *Crassostrea gigas* and 50% of pearl oyster *Pinctada fucata* (Li et al. 2017; Zhang et al. 2012)] and scallop genomes [39% of Yesso scallop and 32% of Chinese scallop (Li et al. 2017; Wang et al. 2017)].

To deal with its likely repetitive nature, we adopted the following approach to sequence the *D. polymorpha* genome. We generated preliminary short read data for a single zebra mussel by sequencing to a depth of approximately 100x on the Illumina HiSeq instrument. An assembly was performed in CLC Workbench (Qiagen Bioinformatics, Redwood City CA) which yielded ~500,000 contigs with an N50 (a measure of assembly contiguity roughly interpretable as a weighted median contig length) of 2.2 kilobases (kb). This is similar to the published assembly of the Mediterranean blue mussel (*Mytilus galloprovincialis*) genome, which was based only on short read data, with ~1.7 million contigs and an N50 of 2.6 kb (Murgarella et al. 2016). To generate a high-quality zebra mussel reference genome, we are obtaining 100x coverage with the Pacific Biosciences (PacBio) Sequel Single Molecule Real Time (SMRT) sequencing platform, which is capable of producing sequencing reads that are tens of kb in length. Such long reads resolve much of the ambiguity in repetitive regions as the reads are long enough to span many repeats and anchor them to unique sequences. A sequencing depth of 100x PacBio combined

with Illumina short read data has been shown to be effective for high-quality assembly of eukaryotic genomes, including the completion of a single 25 Mb contig that spans all of *Drosophila melanogaster* chromosome arm 3L (Berlin et al. 2015).

The technologies for obtaining long-range genomic scaffolding information are rapidly evolving. Additional technologies such as nanopore sequencing (Jain et al. 2018), Hi-C (Burton et al. 2013), optical mapping, and synthetic long read approaches employed by 10x Genomics (Zheng et al. 2016) have been successfully used to improve genome assemblies and for long-range mapping of polymorphisms to parental chromosomes [i.e. haplotype phasing (Moll et al. 2017; Seo et al. 2016)]. We are also planning to incorporate Hi-C to further improve long-range scaffolding of the zebra mussel genome.

With the ability to generate increasingly long sequencing reads, a major challenge is isolating high-quality DNA of sufficient length, in quantities large enough to take full advantage of these technologies. We isolated >100 ug of genomic DNA from an individual zebra mussel from Duluth/Superior Harbor in Lake Superior using a Qiagen Genomic Tip 100/G kit. Pulsed-Field Gel Electrophoresis indicated a broad size distribution from 20-120 kb (not shown). To create a PacBio library, the genomic DNA was needle sheared to an average size of approximately 40 kb, SMRTbell adapters were ligated, and the final library was size selected for molecules >20 kb on the PippinHT (Sage Science). An Agilent TapeStation Genomic DNA assay indicated that the average size of the final sequencing library was >20 kb.

To date, we have generated 168.97 gigabases (Gb) of sequencing data on the PacBio Sequel. This represents an estimated coverage of 77-105x, based on estimates of genome size ranging from 1.6-2.2 G. The N50 for subreads (PacBio terminology for sequence read partitions

that can be used, in our case, for assembly) is 16,524 bp, validating the high quality of our input DNA and PacBio sequencing library. In order to build gene models and to functionally annotate the zebra mussel genome, we have also acquired expression data from 3 different adult tissues (mantle, foot, and gill) using RNA-Seq, and are continuing to collect RNA-Seq data from embryos and larvae spanning a range of developmental stages. In addition to its utility in gene modeling efforts, studying a large proportion of the expressed transcriptome will also provide information about tissue and stage-specific gene expression patterns that may help inform bio-control efforts.

Applications of genomics: Development of biocontrols

In a few cases, invasive species genome projects have been motivated by the goal to discover new biocontrol strategies. For example, vector-directed biocontrol drove the sequencing of the genomes of the invasive mosquito species that carry malaria (*Anopheles gambiae*: Holt et al. 2002) and those that carry yellow fever, dengue and Zika viruses (*Aedes aegypti:* Nene et al. 2007). Sequencing of the genome of the crown-of-thorns sea star (*Acanthaster planci* spp. group) identified the genes for an array of molecules released when animals aggregate to spawn—including a large number of unique ependymin-family proteins active in the central nervous system of many animals and their putative receptors (Hall et al. 2017). This communication system may be a target for biocontrol using synthetic peptides that mimic aggregation cues. With the exception of attempts to identify parasites and other natural enemies (Molloy 1998), no biological control efforts have been attempted against zebra or quagga mussels. Below we describe technologies under development for genetic modification

that could, given our new genomic resources, potentially be applied to dreissenids for control.

Genetic modification biotechnologies

Molecular biologists have invented several techniques with which they can deliver foreign DNA, or make precise edits in the native DNA of organisms. The CRISPR/Cas9 gene editing system has received the greatest recent attention for applications in biological conservation, including control of invasive species—due to low cost, rapid experimental turn-around time, and potential for spreading genetically edited alleles throughout wild populations—even when they lower fitness—through a mechanism known as a "gene drive" (Burt 2003; Gantz, Bier 2015; Gantz et al. 2015). The CRISPR/Cas9 system works by using a Cas9 endonuclease that can be directed, by a gene-specific guide RNA included in the engineered construct, to cleave a 20-basepair-long DNA sequence in virtually any genome (Fig. 1). Flanking the guide RNA is the payload sequence that contains the desired gene edit. Cas9 cleavage of the non-engineered homologous chromosome initiates a DNA repair process that, in the properly engineered construct, will convert the non-engineered into the engineered copy, making the edited gene homozygyous. This allows for super-Mendelian inheritance that has been demonstrated in laboratory studies (Gantz, Bier 2015; Gantz et al. 2015; Hammond et al. 2015), and that can, in theory, rapidly drive the edited gene to high frequencies in natural populations (Champer et al. 2016; Esvelt et al. 2014). Laboratory demonstrations of how this might be used in control, to date, all come from mosquito vectors of disease—including edits that confer host resistance to carrying malarial parasites (Gantz et al. 2015), and others that code for female sterility mutations to lower host fitness (Hammond et al. 2015)—but the possible applications are

virtually limitless.

Nonetheless, there is considerable recent discussion and controversy about the release of CRISPR/Cas9 into the environment, with two issues of concern. The first is biosafety and the regulatory oversight of the technology. Several members of the scientific community, including some who developed the technology, have made pleas to strictly control technology development until the ecological and ethical risks of gene drives can be adequately addressed (Akbari et al. 2015; Bohannon 2015; Caplan et al. 2015; Oye et al. 2014). As a consequence, protocols for ecological risk evaluation by the international system that regulates testing and release of genetically modified live organisms are now being developed more formally (Hayes et al. 2018). With the risks come enormous potential benefits, so the creation of a framework for ecological risk assessment of CRISPR/Cas9 and similar technologies is essential.

The second issue, ironically, is whether CRISPR/Cas9 gene drives will ever impact natural populations enough to create risk (or benefit). Using both mathematical population genetic theory (Deredec et al. 2008; Drury et al. 2017; Noble et al. 2017a; Noble et al. 2017b; Unckless et al. 2017) and direct characterization of mutations (Champer et al. 2017; Drury et al. 2017), several recent studies have examined the evolution of resistance to gene drives. The extent to which resistance will affect prospects for CRISPR/Cas9-based control is not entirely clear. One study predicts that CRISPR/Cas9 gene drives are too efficient for resistance mutations to slow their propagation throughout the range of invasive species—and that unintended transmission (e.g. to native-range populations) remains likely (Noble et al. 2017a). Several other studies, however, suggest that resistance will hamper the spread of a gene drive unless, beforehand**,** constructs are carefully designed (Noble et al. 2017b; Unckless et al. 2017), and focal

populations are screened for Cas9 target sequence polymorphisms (Drury et al. 2017). It may be that resistance evolves so readily that environmental risk has been overestimated, but research on the fate of CRISPR/Cas9 gene drive in natural populations is just beginning.

A still more-recent but promising approach to biocontrol uses components derived from the CRISPR/Cas9 system described above—in this case, to create synthetic barriers to reproduction of invasive species in the wild. It uses a modified protein (dCas9) that, rather than being used to edit genes and initiate a gene drive, allows for control of gene expression (Qi et al. 2013). Maselko et al. (2017) developed a system in which dCas9, paired with a guide RNA molecule, precisely locates a target gene in the genome (as in the CRISPR/Cas9 system above), binds to its promoter sequence and drives the target gene to overexpress its gene product. Target genes, for which overexpression is known to be lethal, can then be chosen to control invasive populations (Fig. 1).

When an engineered strain mates with a wild type, the heterozygous offspring die from overexpression of the gene, off the wild type promoter. The result is synthetic incompatibility, or immediate “post-zygotic” reproductive isolation between engineered and wild type, with the proof of concept demonstrated in yeast (Maselko et al. 2017). Mating between individuals of the engineered strain produce offspring that can survive because the promoter has been mutated to prevent the dCas9/guide RNA construct from binding to it. The use of the system in invasive species could involve releases of engineered individuals, that by mating with wild type, would suppress population mean fitness as in sterile insect biocontrol designs (Maselko et al. 2017). Since dCas9 does not cleave the homologous chromosome, this system does not cause gene conversion leading to a gene drive, thus avoiding any increased environmental risk of that

outcome. But since there is a fitness deficit for the engineered strain (incompatibility) and no gene drive to counter it, the down side would be a need to periodically release engineered individuals. Determining how often and how large these releases would need to be requires population genetic modeling, which remains to be done. This technology is not immune to some forms of resistance (e. g., survival of individuals due to mutation(s) in the promoter sequence that prevents the dCas9/guide RNA construct from binding to it) and this also needs consideration.

Target genes for genetic modification

The first step forward in research on genetic modification requires selection of target genes and biological processes that, when modified, will produce the desired fitness effect (lethality, reduced viability, infertility). Availability of genome sequences is essential for selecting target genes and designing constructs. For example, Drury et al. (2017) generated genomic sequences from 4 global populations of the flour beetle *Tribolium castaneum* to examine population variation in Cas9 sites in target genes. Edits in these genes are expected to produce a range of fitness costs from their effects on eye pigmentation, female and male fertility, and insecticide sensitivity. Maselko et al. (2017) used the yeast genome to search for target genes that when modified would produce lethal overexpression, then searched population genomic data from rice and fruit flies to look for variants in dCas9 target sites within promoter regions.

Among possible targets in *Dreissena* are genes controlling byssal thread synthesis, thermal tolerance, and shell formation and mineralization—all processes with data available to

advise homology searching and with clear biological significance. Genes controlling embryonic development are also prime targets. The Pacific oyster genome project (Zhang et al. 2012) produced data on gene expression across 38 embryonic and larval stages—for example it showed that about 800 genes start their transcription between the last embryonic and 1st larval stage. The project provided functional studies of specific genes expressed across stages, information on genetic regulation of organogenesis, and on male and female-specific genes expressed in gonad. A large number of developmental genes were also identified from the Pearl oyster and Yesso scallop genomes (Wang et al. 2017; Zhang et al. 2012).

Developmental genes or domains can be conserved at the sequence level, sometimes across broad phylogenetic distances (e.g. *Hox* gene homeodomains), which will aid in their identification. A recurring theme is "co-option" for new functions of genes in animal evolution, and this is seen in mollusks. For example, a *nanos* gene copy controls germline differentiation in *Drosophila*, and the *Tis11* gene is not involved in embryogenesis in vertebrate animals from which it was isolated, yet both genes have been recruited to control spiral cleavage divisions in mollusk embryos (Chan, Lambert 2011; Rabinowitz et al. 2008). Studies of spatial pattern of expression also implicated *vasa* and *nanos* gene family members in germ cell development in oysters, snails and other animals (Dill, Seaver 2008; Rabinowitz et al. 2008); knockdown of *vasa* expression by RNA interference was later confirmed to lower oyster fertility by inhibiting gonad development (Fabioux et al. 2009). The arthropod segmentation gene *engrailed* controls embryonic shell (protoconch) formation throughout molluscs, as does *dpp-BMP2/4*, a gene that specifies the dorso-ventral axis in arthropods and vertebrates (Jackson, Degman 2016; Nederbragt et al. 2002; Wanninger, Haszprunar 2001).

The zebra mussel genome: a community resource

It is impossible for us to envision; let alone to take advantage of the range of applications of this genome to research and management. We recognize, moreover, that to properly analyze it we will need assistance from experts from a number of unrelated disciplines—biomineralization, comparative and evolutionary genomics, developmental biology, materials science, physiology and physiological ecology—to name some that come to mind. With this review, we encourage interested individuals to collaborate on a cross-disciplinary effort to annotate and analyze the genome, and to formulate research on applications. Worldwide the dreissenid mussel research community is large and diverse, and we need its help in this important effort.

ACKNOWLEDGEMENTS

We thank Benjamin Auch in the Genomics Center and Kevin Silverstein of the Minnesota Supercomputing Institute at the University of Minnesota for their exceptional contributions to the sequencing and analysis of this genome. The zebra mussel genome project was funded by a grant from the Minnesota Environmental and Natural Resources Trust Fund, by the Minnesota Aquatic Invasive Species Research Center, and by private donations.

REFERENCES

Akbari OS, Bellen HJ, Bier E, Bullock SL et al (2015) Safeguarding gene drive experiments in the laboratory. *Science* 349:927-928

Allen YC, Thompson BA, Ramcharan CW (1999) Growth and mortality rates of the zebra mussel, *Dreissena polymorpha,* in the Lower Mississippi River. *Canadian Journal of Fisheries and Aquatic Sciences* 56:748-759

Appeltans W, Ahyong Shane T, Anderson G et el (2012) The magnitude of global marine species diversity. *Current Biology* 22:2189-2202

Asplen MK, Anfora G, Biondi A et al (2015) Invasion biology of spotted wing Drosophila (*Drosophila suzukii)*: a global perspective and future priorities. *Journal of Pest Science* 88:469-494

Baldwin BS, Carpenter M, Rury K, Woodward E (2012) Low dissolved ions may limit secondary invasion of inland waters by exotic round gobies and dreissenid mussels in North America. *Biological Invasions* 14:1157-1175

Benson AJ (2014) Chronological history of zebra and quagga mussels (Dreissenidae) in North America, 1988-2010. In: Nalepa TF and Schloesser DW (eds) Quagga and Zebra Mussels : Biology, Impacts, and Control (2nd Edition). CRC Press, Boca Raton, FL, pp. 9-32

Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotech* 33:623-630

Bogan AE (2008) Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. *Hydrobiologia* 595:139-147

Bohannon J (2015) Biologists devise invasion plan for mutations. *Science* 347:1300

Borcherding J (1991) The annual reproductive cycle of the freshwater mussel *Dreissena polymorpha* (Pallas) in lakes. *Oecologia* 87:208-218

Bossenbroek JM, Finnoff DC, Shogren JF, Warziniack TW (2009) Advances in ecological and economical analysis of invasive species: dreissenid mussels as a case study. In: Keller RP, Lodge DM, Lewis MA and Shogren JF (eds) Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management. Oxford University Press, Oxford. Oxford University Press, New York, pp. 244-265

Brazee SL, Carrington E (2006) Interspecific comparison of the mechanical properties of mussel byssus. *The Biological Bulletin* 211:263-274

Brown JE, Stepien CA (2010) Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. *Biological Invasions* 12:3687-3710

Burt A (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:921-928

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology* 31:1119-1125

Caplan AL, Parent B, Shen M, Plunkett C (2015) No time to waste—the ethical challenges created by CRISPR. *EMBO reports* 16:1421-1426

Champer J, Buchman A, Akbari OS (2016) Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics* 17:146-159

Champer J, Reeves R, Oh SY et al (2017) Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLOS Genetics* 13:e1006796

Chan XY, Lambert JD (2011) Patterning a spiralian embryo: A segregated RNA for a Tis11 ortholog is required in the 3a and 3b cells of the Ilyanassa embryo. *Developmental Biology* 349:102-112

Chen X-G, Jiang X, Gu J et al (2015) Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proceedings of the National Academy of Sciences* 112:E5907

Claxton WT, Mackie GL (1998) Seasonal and depth variations in gametogenesis and spawning of *Dreissena polymorpha* and *Dreissena bugensis* in eastern Lake Erie. *Canadian Journal of Zoology* 76:2010-2019

Cohen AN, Weinstein A (2001) Zebra mussel's calcium threshold and implications for its potential distribution in North America. Richmond, CA, pp. 1-43

Collas FPL, Karatayev AY, Burlakova LE, Leuven RSEW (2018) Detachment rates of dreissenid mussels after boat hull-mediated overland dispersal. *Hydrobiologia* 810:77-84

Cristescu ME (2015) Genetic reconstructions of invasion history. *Molecular Ecology* 24:2212-2225

de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLOS Genetics* 7:e1002384

De Ventura L, Weissert N, Tobias R, Kopp K, Jokela J (2016) Overland transport of recreational boats as a spreading vector of zebra mussel Dreissena polymorpha. *Biological Invasions* 18:1451-1466

DeMartini DG, Errico JM, Sjoestroem S, Fenster A, Waite JH (2017) A cohort of new adhesive proteins identified from transcriptomic analysis of mussel foot glands. *Journal of The Royal Society Interface* 14:20170151

Deredec A, Burt A, Godfray HCJ (2008) The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* 179:2013-2026

Dill KK, Seaver EC (2008) Vasa and nanos are coexpressed in somatic and germ line tissue from early embryonic cleavage stages through adulthood in the polychaete *Capitella* sp. I. *Development Genes and Evolution* 218:453-463

Drury DW, Dapper AL, Siniard DJ, Zentner GE, Wade MJ (2017) CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Science Advances* 3:e1601910

Du X, Fan G, Jiao Y, Zhang H et al (2017) The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *GigaScience* 6:1-12

Esvelt KM, Smidler AL, Catteruccia F, Church GM (2014) Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* 3:e03401.

Fabioux C, Corporeau C, Quillien V, Favrel P, Huvet A (2009) In vivo RNA interference in oyster – vasa silencing inhibits germ cell development. *FEBS Journal* 276:2566-2573

Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annual Review of Physiology* 61:243-282

Gantz VM, Bier E (2015) The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* 348:442-444

Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, James AA (2015) Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences* 112:E6736

Gelembiuk GW, May GE, Lee CE (2006) Phylogeography and systematics of zebra mussels and related species. *Molecular Ecology* 15:1033-1050

Hall MR, Kocot KM, Baughman KW et al (2017) The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* 544:231-234

Hammond A, Galizi R, Kyrou K et al (2015) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology* 34:78-83

Hayes KR, Hosack GR, Dana GV et al (2018) Identifying and detecting potentially adverse ecological outcomes associated with the release of gene-drive modified organisms. *Journal of Responsible Innovation* 5:S139-S158

Higgins SN, Vander Zanden MJ (2010) What a difference a species makes: a meta–analysis of dreissenid mussel impacts on freshwater ecosystems. *Ecological Monographs* 80:179-196

Hincks SS, Mackie GL (1997) Effects of pH, calcium, alkalinity, hardness, and chlorophyll on the survival, growth, and reproductive success of zebra mussel (*Dreissena polymorpha*) in Ontario lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 54:2049-2057

Holt RA, Subramanian GM, Halpern A et al (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298:129-149

Jackson DJ, Degnan BM (2016) The importance of evo-devo to an integrated understanding of molluscan biomineralisation. *Journal of Structural Biology* 196:67-74

Jain M, Koren S, Miga KH, Quick J et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*:doi:10.1038/nbt.4060

Johnson LE, Ricciardi A, Carlton JT (2001) Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecological Applications* 11:1789-1799

Jones LA, Ricciardi A (2005) Influence of physicochemical factors on the distribution and biomass of invasive mussels (*Dreissena polymorpha* and *Dreissena bugensis*) in the St. Lawrence River. *Canadian Journal of Fisheries and Aquatic Sciences* 62:1953-1962

Joubert C, Piquemal D, Marie B et al (2010) Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. *BMC Genomics* 11:613

Karatayev AY, Burlakova LE, Padilla DK (1997) The effects of *Dreissena polymorpha* (Pallas) invasion on aquatic communities in eastern Europe. *Journal of Shellfish Research* 16:187-203

Karatayev AY, Burlakova LE, Padilla DK, Johnson LE (2003) Patterns of spread of the zebra mussel (*Dreissena polymorpha* (Pallas)): The continuing invasion of Belarussian lakes. *Biological Invasions* 5:213-221

Karatayev AY, Padilla DK, Minchin D, Boltovskoy D, Burlakova LE (2007) Changes in global economies and trade: The potential spread of exotic freshwater bivalves. *Biological Invasions* 9:161-180

Karatayev AY, Burlakova LE, Mastitsky SE, Padilla DK, Mills EL (2011) Contrasting rates of spread of two congeners, *Dreissena polymorpha* and *Dreissena rostriformis bugensis,* at different spatial scales. *Journal of Shellfish Research* 30:923-931

Karatayev VA, Karatayev AY, Burlakova LE, Padilla DK (2013) Lakewide dominance does not predict the potential for spread of dreissenids. *Journal of Great Lakes Research* 39:622-629

Kuester A, Conner Jeffrey K, Culley T, Baucom Regina S (2014) How weeds emerge: a taxonomic and trait-based examination using United States data. *New Phytologist* 202:1055-1068

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* 17:386-391

Lee BP, Messersmith PB, Israelachvili JN, Waite JH (2011) Mussel-Inspired adhesives and coatings. *Annual Review of Materials Research* 41:99-132

Lewis DB, Magnuson JJ (1999) Intraspecific gastropod shell strength variation among north temperate lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 56:1687-1695

Li Y, Sun X, Hu X, Xun X et al (2017) Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nature Communications* 8:1721

Limburg KE, Luzadis VA, Ramsey M, Schulz KL, Mayer CM (2010) The good, the bad, and the algae: perceiving ecosystem services and disservices generated by zebra and quagga mussels. *Journal of Great Lakes Research* 36:86-92

Lowe S, Browne M, Boudjelas S, De Poorter M (2000) 100 of the world's worst invasive alien species: a selection from the global invasive species database. The Invasive Species Specialist Group (ISSG), World Conservation Union (IUCN), Aukland, New Zealand, pp. 12

Lucy F, Burlakova L, Karatayev A, Mastitsky S, Zanatta D (2014) Zebra mussel impacts on unionids: a synthesis of trends in North America and Europe. In: Nalepa TF and Schloesser DW (eds) Quagga and zebra mussels: biology, impact, and control, 2nd edn. CRC Press, Boca Raton, FL, pp. 623-634

Mallez S, McCartney MA (2018) Dispersal mechanisms for zebra mussels: population genetics supports clustered invasions over spread from hub lakes in Minnesota. *Biological Invasions*, https://doi.org/10.1007/s10530-018-1714-3

Maselko M, Heinsch SC, Chacón JM, Harcombe WR, Smanski MJ (2017) Engineering species-like barriers to sexual reproduction. *Nature Communications* 8:883

Matthews J, Van der Velde G, Bij de Vaate A, Collas FPL, Koopman KR, Leuven RSEW (2014) Rapid range expansion of the invasive quagga mussel in relation to zebra mussel presence in The Netherlands and Western Europe. *Biological Invasions* 16:23-42

May GE, Gelembiuk GW, Panov VE, Orlova MI, Lee CE (2006) Molecular ecology of zebra mussel invasions. *Molecular Ecology* 15:1021-1031

Mayer C, Burlakova L, Eklöv P et al (2014) Benthification of freshwater lakes: exotic mussels turning ecosystems upside down. In: Nalepa TF and Schloesser DW (eds) Quagga and Zebra Mussels : Biology, Impacts, and Control, 2 edn. CRC Press, Boca Raton, FL, pp. 575-586

McKenna DD, Scully ED, Pauchet Y et al (2016) Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *BMC Genome Biology* 17:227

McMahon RF (1996) The physiological ecology of the zebra mussel, *Dreissena polymorpha*, in North America and Europe. *American Zoologist* 36:339-363

McMahon RF (2002) Evolutionary and physiological adaptations of aquatic invasive animals: r selection versus resistance. *Canadian Journal of Fisheries and Aquatic Sciences* 59:1235-1244

McNickle GG, Rennie MD, Sprules WG (2006) Changes in benthic invertebrate communities of South Bay, Lake Huron following invasion by zebra mussels (*Dreissena polymorpha*), and potential effects on lake whitefish (*Coregonus clupeaformis*) diet and growth. *Journal of Great Lakes Research* 32:180-193

Mellina E, Rasmussen JB (1994) Patterns in the distribution and abundance of zebra mussel (*Dreissena polymorpha*) in rivers and lakes in relation to substrate and other physicochemical factors. *Canadian Journal of Fisheries and Aquatic Sciences* 51:1024-1036

Moll KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, Sadowsky MJ, Stupar RM, Tiffin P, Miller JR, Young ND, Silverstein KAT, Mudge J (2017) Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, Medicago truncatula. *BMC Genomics* 18:578

Molloy DP (1998) The potential for using biological control technologies in the management of Dreissena spp. *Journal of Shellfish Research* 17:177-183

Mount AS, Wheeler AP, Paradkar RP, Snider D (2004) Hemocyte-mediated shell mineralization in the eastern oyster. *Science* 304:297-300

Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C (2016) A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS ONE* 11:e0151561

Nalepa TF, Schloesser DW (2014) Quagga and Zebra Mussels : Biology, Impacts, and Control (2nd Edition). CRC Press, Boca Raton, FL, pp. 816

Nederbragt AJ, van Loon AE, Dictus WJAG (2002) Expression of Patella vulgata Orthologs of *engrailed* and *dpp-BMP2/4* in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Developmental Biology* 246:341-355

Nene V, Wortman JR, Lawson D et al (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316:1718-1723

Noble C, Adlam B, Church GM, Esvelt KM, Nowak MA (2017a) Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *bioRxiv*

Noble C, Olejarz J, Esvelt KM, Church GM, Nowak MA (2017b) Evolutionary dynamics of CRISPR gene drives. *Science Advances* 3:e1601964

Nowoshilow S, Schloissnig S, Fei J-F et al (2018) The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554:50-55

O'Neill CR, Jr. (2008) The silent invasion: Finding solutions to minimize the impacts of invasive quagga mussels on water rates, water infrastructure and the environment. U.S. House of Representatives Committee on Natural Resources – Subcommittee on Water and Power. Washington, D.C., pp. 1-13

Ometto L, Cestaro A, Ramasamy S et al (2013) Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biology and Evolution* 5:745-757

Oye KA, Esvelt K, Appleton E et al (2014) Regulating gene drives. *Science* 345:626-628

Pagnucco KS, Maynard GA, Fera SA, Yan ND, Nalepa TF, Ricciardi A (2015) The future of species invasions in the Great Lakes-St. Lawrence River basin. *Journal of Great Lakes Research* 41:96-107

Papanicolaou A, Schetelig MF, Arensburger P et al (2016) The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biology* 17:192

Peyer SM, McCarthy AJ, Lee CE (2009) Zebra mussels anchor byssal threads faster and tighter than quagga mussels in flow. *Journal of Experimental Biology* 212:2027-2036

Porcelli D, Butlin RK, Gaston KJ, Joly D, Snook RR (2015) The environmental genomics of metazoan thermal adaptation. *Heredity* 114:502-514

Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173-1183

Qin C-l, Pan Q-d, Qi Q et al (2016) In-depth proteomic analysis of the byssus from marine mussel *Mytilus coruscus*. *Journal of Proteomics* 144:87-98

Rabinowitz JS, Chan XY, Kingsley EP, Duan Y, Lambert JD (2008) Nanos Is Required in Somatic Blast Cell Lineages in the Posterior of a Mollusk Embryo. *Current Biology* 18:331-336

Raikow DF (2004) Food web interactions between larval bluegill (*Lepomis macrochirus)* and exotic zebra mussels (*Dreissena polymorpha*). *Canadian Journal of Fisheries and Aquatic Sciences* 61:497-504

Ramasamy S, Ometto L, Crava CM et al (2016) The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. *Genome Biology and Evolution* 8:2297-2311

Rees DJ, Hanifi A, Manion J, Gantayet A, Sone ED (2016) Spatial distribution of proteins in the quagga mussel adhesive apparatus. *Biofouling* 32:205-213

Rzepecki L, Waite J (1993) The byssus of the zebra mussel, *Dreissena polymorpha*. I: Morphology and *in situ* protein processing during maturation. *Molecular Marine Biology and Biotechnology* 2:255-266

Ricciardi A, Blackburn TM, Carlton JT et al (2017) Invasion science: A horizon scan of emerging challenges and opportunities. *Trends in Ecology & Evolution* 32:464-474

Rius M, Bourne S, Hornsby HG, Chapman MA (2015) Applications of next-generation sequencing to the study of biological invasions. *Current Zoology* 61:488-504

Russell-Hunter W, Burky A, Hunter R (1981) Inter-population variation in calcareous and proteinaceous shell components in the stream limpet, *Ferrissia rivularis*. *Malacologia* 20:255-266

Sax DF, Stachowicz JJ, Brown JH et al (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology & Evolution* 22:465-471

Schloesser DW, Schmuckal C (2012) Bibliography of *Dreissena polymorpha* (zebra mussels) and *Dreissena rostriformis bugensis* (quagga mussels): 1989 to 2011. *Journal of Shellfish Research* 31:1205-1263

Seo J-S, Rhie A, Kim J et al (2016) *De novo* assembly and phasing of a Korean human genome. *Nature* 538:243-247

Soroka M, Rymaszewska A, Sańko T et al (2018) Next-generation sequencing of *Dreissena polymorpha* transcriptome sheds light on its mitochondrial DNA. *Hydrobiologia* 810:255-263

Sousa R, Gutiérrez JL, Aldridge DC (2009) Non-indigenous invasive bivalves as ecosystem engineers. *Biological Invasions* 11:2367-2385

Sprung M (1987) Ecological requirements of developing *Dreissena polymorpha* eggs. *Archiv für Hydrobiologie Supplement* 79:69-86

Stepien CA, Grigorovich IA, Gray MA, Sullivan TJ, Yerga-Woolwine S, Kalayci G (2014) Evolutionary, biogeographic, and population genetic relationships of dreissenid mussels, with revision of component taxa. In: Nalepa TF and Schloesser DW (eds) Quagga and Zebra Mussels: Biology, Impacts and Control, 2 edn. CRC Press, London, GBR, pp. 403-444

Strayer DL (1993) Macrohabitats of freshwater mussels (Bivalvia:Unionacea) in streams of the northern Atlantic Slope. *Journal of the North American Benthological Society* 12:236-246

Strayer DL, Hattala KA, Kahnle AW (2004) Effects of an invasive bivalve (*Dreissena polymorpha*) on fish in the Hudson River estuary. *Canadian Journal of Fisheries and Aquatic Sciences* 61:924-941

Sun J, Zhang Y, Xu T et al (2017) Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution* 1:0121

Takeuchi T, Koyanagi R, Gyoja F et al (2016) Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters* 2:3

Uliano-Silva M, Dondero F, Dan Otto T et al (2017) A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel *Limnoperna fortunei*. *GigaScience*:gix128

Unckless RL, Clark AG, Messer PW (2017) Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* 205:827-841

Vinogradov GA, Smirnova NF, Sokova VA, Bruznitsky AA (1993) Influence of chemical composition of the water on the mollusk *Dreissena polymorpha*. In: Nalepa TF and Schloesser DW (eds) Zebra Mussels: Biology, Impacts, and Control, 2 edn. CRC Press, Boca Raton, FL, pp. 283-293

Waite JH, Andersen NH, Jewhurst S, Sun C (2005) Mussel adhesion: Finding the tricks worth mimicking. *The Journal of Adhesion* 81:297-317

Wang R, Li C, Stoeckel J, Moyer G, Liu Z, Peatman E (2012) Rapid development of molecular resources for a freshwater mussel, *Villosa lienosa* (Bivalvia:Unionidae), using an RNA-seq-based approach. *Freshwater Science* 31:695-708

Wang S, Zhang J, Jiao W, Li J et al (2017) Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution* 1:0120

Wanninger A, Haszprunar G (2001) The expression of an engrailed protein during embryonic shell formation of the tusk-shell, Antalis entalis (Mollusca, Scaphopoda). *Evolution & Development* 3:312-321

Ward J, Ricciardi A (2014) Impacts of *Dreissena* on benthic macroinvertebrate communities—Predictable patterns revealed by invasion history. In: Nalepa TF and Schloesser DW (eds) Quagga and zebra mussels: biology, impacts, and control, 2 edn. CRC Press, Boca Raton, FL, pp. 599-610

Weiner S, Traub W (1984) Macromolecules in mollusc shells and their functions in biomineralization. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 304:425-434

Whittier TR, Ringold PL, Herlihy AT, Pierson SM (2008) A calcium-based invasion risk assessment for zebra and quagga mussels (*Dreissena* spp). *Frontiers in Ecology and the Environment* 6:180-184

Xu W, Faisal M (2010) Gene expression profiling during the byssogenesis of zebra mussel (Dreissena polymorpha). *Molecular Genetics and Genomics* 283:327-339

Zhang G, Fang X, Guo X et al (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49-54

Zheng GXY, Lau BT, Schnall-Levin M et al (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology* 34:303-311

FIGURE LEGEND

Figure 1. Strategies of zebra mussel genetic modification. The strategies on the left involve genomic editing to interupt the biological function of target genes; with or without gene drives to spread the modification throughout populations. The strategy on the right involve insertion of a gene activator to drive over-expression of genes that create post-zygotic barriers to reproduction. This would lower fitness via “gamete wastage” in engineered populations.

Table 1. Sequenced genomes available from 100 of the world's worst alien invasive species. The five columns with bold italic headings provide descriptions of the length and quality of the sequenced genomes**.** Assembly level: contig is a term for an assembled contiguous stretch of DNA sequence; scaffold refers to when a set of contigs is ordered and placed in the correct orientation; chromosome level is when biological chromosomes are assembled is relative completion (some gaps may remain). The number of contigs provides a metric for the assembly quality; in general the smaller the number the larger the contig length. Contig N_{50} is roughly a measure of the shortest contig length in the data encompassing 50% of the genome in basepairs (bp). Genome length is the total length of the assembled genome in gigabase pairs (Gb).

Common name	**Taxon or group**	**Strain/isolate**	**Impacts**	***Assembly level***	***Number of scaffolds***	***Number of contigs***	***Contig N50 (bp)***	***Genome length (Gb)***	**Year submitted**
Rabbit	Mammal		Degrades biodiversity, particularly in introduced areas that lack predators	Chromosome	3,318	84,024	64,648	2.737	2005
Frog chytrid fungus	Fungus	JAM81	Cause of many amphibian declines and extinctions	Scaffold	127	510	318,114	0.024	2011
Comb jelly	Aquatic invertebrate		Invasive carnivore that consumes zooplankton	Scaffold	5,100	24,927	11,914	0.156	2011
Argentine ant	Terrestrial invertebrate		Often displaces native ants	Scaffold	3,030	18,227	35,858	0.220	2011
Red imported fire ant	Terrestrial invertebrate		Highly damaging nuisance species and pest of crop plants, livestock	Scaffold	69,511	90,219	14,677	0.396	2011
Mouse	Mammal	C57BL/6J	Economic pests, carriers of human disease, several negative impacts on invaded ecosystems	Chromosome	262	750	32,273,079	2.794	2012

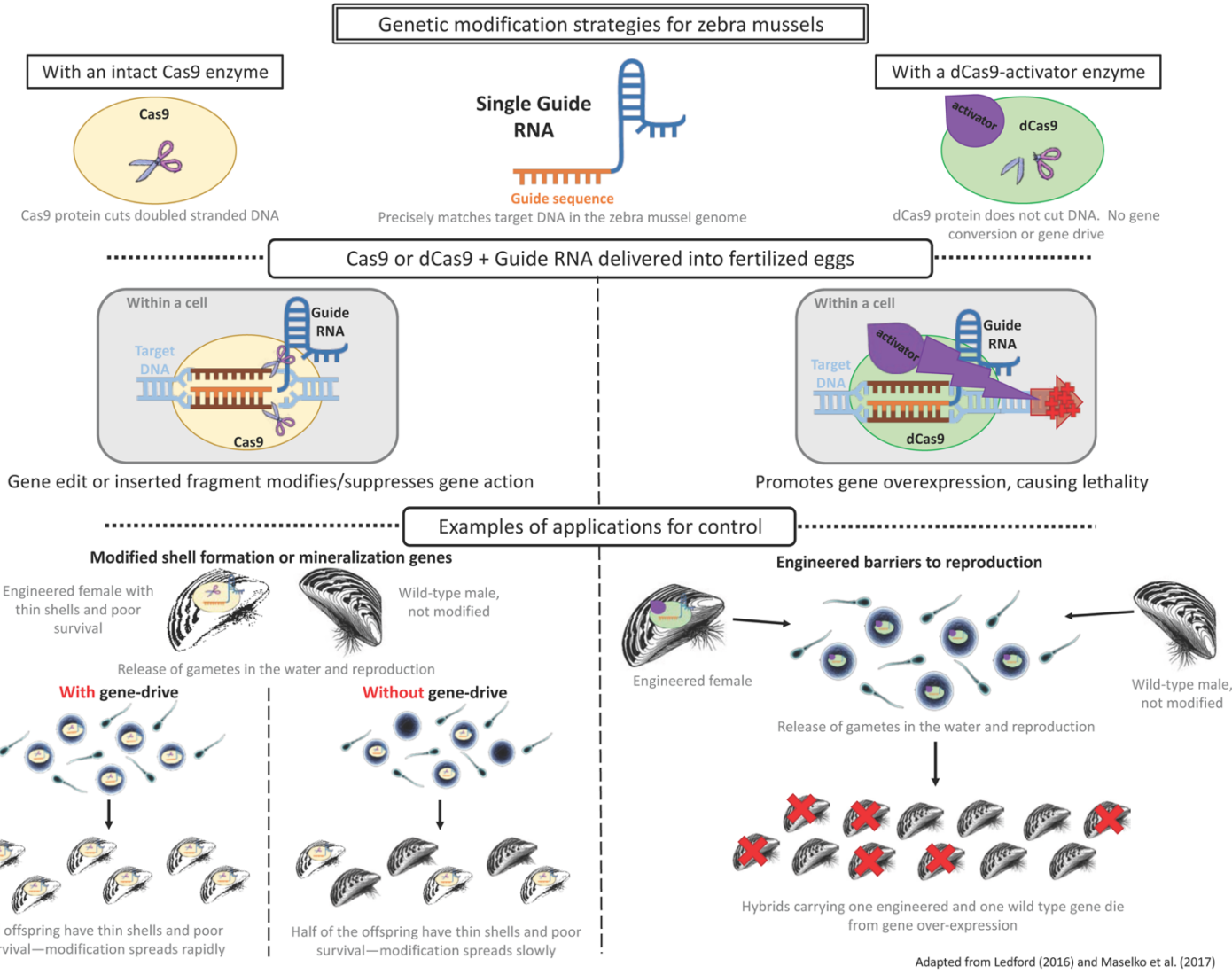
Macaque	Mammal		Lower native bird diversity by eating eggs and chicks, and competing for food	Chromosome	7,625	87,764	86,040	2.947	2013
Crayfish plague	Protist	APO3	Water mold lethal to European crayfish	Scaffold	835	4,659	36,439	0.076	2014
Common carp	Fish		Uproots aquatic vegetation, causing declines in, plants, other fishes and water quality	Chromosome	9,378	53,088	75,080	1.714	2014
Phytophthora root rot	Fungus	MP94-48	Highly damaging with broad host range	Scaffold	5,777	5,831	24,715	0.054	2015
Little fire ant	Terrestrial invertebrate	WASHAW1	Stinging ants that displace native species and harm crop plants	Scaffold	77,788	103,610	37,912	0.324	2015
Starling	Bird	715	Outcompetes native birds for nesting sites and damages fruits and other crops	Scaffold	2,361	22,666	151,865	1.037	2015
Asian tiger mosquito	Terrestrial invertebrate	Foshan	Widespread vector of yellow, dengue and Chikungunya fever viruses	Scaffold	154,782	355,061	18,430	1.923	2015
Avian malaria	Protist	SGS1	Parasites of birds, causing wide-ranging levels of mortality	Chromosome	514	724	583,861	0.023	2016
Sweet potato whitefly	Terrestrial invertebrate	MEAM1	Pest of vegetable crops and ornamentals with vast host range	Scaffold	19,751	31,571	84,501	0.615	2016
Goat	Mammal		Voracious grazers with great impacts on vegetation and	Chromosome	29,907	30,399	26,244,591	2.923	2016

			cascading effects, particularly on islands						
Asian longhorned beetle	Terrestrial invertebrate	ALB-LARVAE	Wood feeding pest of trees in forests and urban settings	Scaffold	9,867	26,749	80,490	0.707	2017
Mediterrenean blue mussel	Aquatic invertebrate		Marine mussel that displaces native species	Scaffold	1,002,334	1,136,100	2,627	1.500	2017
Rainbow trout	Fish	Swanson	Preys upon and outcompetes native fishes, and hybridizes with native trout	Chromosome	139,800	559,855	13,827	2.179	2017
Domestic cat	Mammal	Cinnamon	Voracious predators on native birds, reptiles and mammals responsible for several extinction events	Chromosome	4,525	4,909	41,915,695	2.522	2017
Pig	Mammal	201423004	Feral pigs are pests of crops and property, dig up native vegetation, prey on several native species	Chromosome	14,157	14,818	6,372,407	2.755	2017
Red deer	Mammal	*hippelaphus*	Strong impacts on native forest flora and fauna in invaded range	Chromosome	11,479	406,637	7,944	3.439	2017
Bullfrog	Amphibian	Bruno	Preys upon and outcompetes native amphibians	Scaffold	1,544,635	2,124,505	5,415	6.250	2017
Golden apple snail	Aquatic invertebrate	SZHN2017	Voracious feeder on crops and native vegetation	Chromosome	24	746	1,072,857	0.440	2018
Western mosquito fish	Fish	NE01/NJP1002.9	Causes decline and extinction of other small	Scaffold	2,943	73,682	17,511	0.599	2018

		native fishes through competition						
Leafy spurge	Land plant	Aggressive weed	Scaffold	1,633,094	2,242,201	605	1.125	2018
Cane toad	Amphibian	Toxic skin glands poison predators upon ingestion, endangering native species	Contig	N/A	31,391	167,498	2.552	2018

Table 2. Sequenced genomes from bivalve molluscs

Species	Family	Common name	Commercial interest	Assembly level	Number of scaffolds	Number of contigs	Contig N50 (bp)	Genome length (Mb)	Reference
Bathymodiolus platifrons	Mytilidae	Hydrothermal vent mussel	None	Scaffold	65,662	272,497	12,602	1,658.2	Sun et al. 2017
Chlamys farreri	Pectinidae	Zhikong (Chinese) scallop	Wild harvest and culture	Scaffold	96,024	148,999	21,500	779.9	Li et al. 2017
Crassostrea gigas	Ostreidae	Pacific oyster	Hatchery culture—leads aquatic animals in global harvest	Scaffold	7,659	30,460	31,239	557.7	Zhang et al. 2012
Crassostrea virginica	Ostreidae	Eastern oyster	Wild harvest and hatchery culture	Chromosome	11	669	1,971,208	684.7	Gómez-Chiarri et al. 2015
Mizuhopecten (Patinopectin) yessoensis	Pectinidae	Yesso scalllop	Culture from wild seed	Scaffold	82,659	120,022	65,014	987.6	Wang et al. 2017
Modiolus philippinarum	Mytilidae	Phillipine horse mussel	None	Scaffold	74,573	301,873	18,389	2,629.6	Sun et al. 2017
Pinctada martensii	Pteriidae	Akoya pearl oyster	Cultured pearls	Chromosome	5,039	85,944	21,518	991.0	Unpublished


Genetic modification strategies for zebra mussels
With an intact Cas9 enzyme
Cas9
Cas9 protein cuts doubled stranded DNA
Single Guide RNA
Guide sequence
Precisely matches target DNA in the zebra mussel genome
With a dCas9-activator enzyme
activator
dCas9
dCas9 protein does not cut DNA. No gene conversion or gene drive
Cas9 or dCas9 + Guide RNA delivered into fertilized eggs
Within a cell
Guide RNA
Target DNA
Cas9
Gene edit or inserted fragment modifies/suppresses gene action
Within a cell
Guide RNA
activator
Target DNA
dCas9
Promotes gene overexpression, causing lethality
Examples of applications for control
Modified shell formation or mineralization genes
Engineered female with thin shells and poor survival
Wild-type male, not modified
Release of gametes in the water and reproduction
With gene-drive
Without gene-drive
All offspring have thin shells and poor survival—modification spreads rapidly
Half of the offspring have thin shells and poor survival—modification spreads slowly
Engineered barriers to reproduction
Engineered female
Wild-type male, not modified
Release of gametes in the water and reproduction
Hybrids carrying one engineered and one wild type gene die from gene over-expression
Adapted from Ledford (2016) and Maselko et al. (2017)