

3-24-2018

# Standing genetic diversity and selection at functional gene loci are associated with differential invasion success in two non-native fish species

Kyle Wellband  
*University of Windsor*

Harri Pettitt-Wade  
*University of Windsor*

Aaron T. Fisk  
*University of Windsor*

Daniel D. Heath  
*University of Windsor*

Follow this and additional works at: <https://scholar.uwindsor.ca/biologypub>

 Part of the [Biology Commons](#)

---

## Recommended Citation

Wellband, Kyle; Pettitt-Wade, Harri; Fisk, Aaron T.; and Heath, Daniel D., "Standing genetic diversity and selection at functional gene loci are associated with differential invasion success in two non-native fish species" (2018). *Molecular Ecology*, 27, 7, 1572-1585.  
<https://scholar.uwindsor.ca/biologypub/1229>

This Article is brought to you for free and open access by the Department of Biological Sciences at Scholarship at UWindsor. It has been accepted for inclusion in Biological Sciences Publications by an authorized administrator of Scholarship at UWindsor. For more information, please contact [scholarship@uwindsor.ca](mailto:scholarship@uwindsor.ca).



MR. KYLE W WELLBAND (Orcid ID : 0000-0002-5183-4510)

Article type : Original Article

Title: Standing genetic diversity and selection at functional gene loci are associated with differential invasion success in two non-native fish species

Kyle W. Wellband<sup>1</sup>, Harri Pettitt-Wade<sup>1</sup>, Aaron T. Fisk<sup>1</sup>, Daniel D. Heath<sup>1,2\*</sup>

<sup>1</sup>Great Lakes Institute for Environmental Research, University of Windsor.

<sup>2</sup>Department of Biological Sciences, University of Windsor.

\*Corresponding author: Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Ave., Windsor, Ontario, Canada, N9B 3P4; Tel: 519 253-3000 (ext 3762); Fax: 519 971-3616; email: dheath@uwindsor.ca.

Keywords: biological invasions; non-indigenous species; natural selection; founder effect; protein-coding SNPs

Running title: Drift, selection, and goby invasion success

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.14557

This article is protected by copyright. All rights reserved.

Accepted Article

## Abstract:

Invasive species are expected to experience a unique combination of high genetic drift due to demographic factors while also experiencing strong selective pressures. The paradigm that reduced genetic diversity should limit the evolutionary potential of invasive species and thus their potential for range expansion has received little empirical support, possibly due to the choice of genetic markers. Our goal was to test for effects of genetic drift and selection at functional genetic markers as they relate to the invasion success of two paired invasive goby species, one widespread (successful) and one with limited range expansion (less successful). We genotyped fish using two marker types: single nucleotide polymorphisms (SNPs) in known-function, protein-coding genes and microsatellites to contrast the effects of neutral genetic processes. We identified reduced allelic variation in the invaded range for the less-successful tubenose goby. SNPs putatively under selection were responsible for the observed differences in population structure between marker types for round goby (successful) but not tubenose goby (less successful). A higher proportion of functional loci experienced divergent selection for round goby, suggesting increased evolutionary potential in invaded ranges may be associated with round goby's greater invasion success. Genes involved in thermal tolerance were divergent for round goby populations but not tubenose goby, consistent with the hypothesis that invasion success for fish in temperate regions is influenced by capacity for thermal tolerance. Our results highlight the need to incorporate functional genetic markers in studies to better assess evolutionary potential for the improved conservation and management of species.

## Introduction:

Human activities have altered the global distribution of species. We have transported numerous species from their native ranges and introduced them into novel areas where, without the aid of humans, dispersal would not have been possible (e.g. Ricciardi, 2006). Many non-indigenous species have little impact on the communities to which they were introduced. However, some species can have wide-spread and damaging effects on the ecosystem or to economic activities in the

Accepted Article

introduced range, and hence become defined as “invasive” (Colautti & MacIsaac, 2004). An important component of the management of non-indigenous species is identifying which ones are at high risk of becoming invasive to direct limited management resources where they will have the most impact: by preventing the establishment or mitigating the effects of the worst invaders while it is still feasible (Kolar & Lodge, 2001). Unfortunately, a major failing of existing risk assessment frameworks is they do not consider the evolutionary potential of invaders, nor the role selection may play in the invasion process (Strayer, Eviner, Jeschke, & Pace, 2006; Whitney & Gabler, 2008).

Biological invasions expose species to a unique combination of evolutionary forces. The stochastic demographic processes associated with colonization, founder events, and bottlenecking suggest that genetic diversity of invasive populations should be reduced by strong genetic drift effects (Nei, Maruyama, & Chakraborty, 1975). Organisms are simultaneously exposed to novel environmental conditions that should result in strong natural selection (Sakai et al., 2001). This combination of evolutionary forces has generated predictions that invaders should experience reduced evolutionary potential as a result of the loss of genetic diversity due to drift. However, empirical assessments have demonstrated that reductions in genetic diversity due to drift during invasion are not severe (Dlugosch & Parker, 2008) and putative adaptive evolution of species during invasion is a common occurrence (Whitney & Gabler, 2008). The apparent disconnect between theory and empirical evidence may be a result of limitations in the approach to measuring genetic diversity and the invasive organisms available for study.

The vast majority of studies of genetic diversity in invasive species to date have used classically neutral genetic markers such as microsatellite or mitochondrial DNA sequence data. These markers are only a proxy for genome-wide variation and even substantial changes in neutral microsatellite diversity may not be reflective of meaningful change in diversity at protein-coding loci, depending on their standing genetic frequency in the native populations (Liu, Chen, Wang, Oh, & Zhao, 2005). Indeed, correlations between microsatellite diversity and quantitative trait variation are weak (Reed & Frankham, 2001) and quantitative trait variation does not exhibit the same decrease during invasion that neutral diversity does (Dlugosch & Parker, 2008). In contrast, protein-coding

gene polymorphisms are expected to evolve in response to selection in addition to the stochastic effects of drift. Such genetic markers are relevant to predicting the evolutionary potential of invasive species and the consequences of changes in diversity at these loci likely impact long-term invasion success. Methods to measure genome scale functional variation in nucleotide diversity have become widely accessible in the past decade (Ellegren, 2014). Furthermore, genomic resources for non-model organisms now facilitate the characterization of protein-coding gene variation for invasive species with the goal of identifying the genomic basis of adaptive invasive phenotypes (Chown et al., 2015; Ellegren, 2014; Rius, Bourne, Hornsby, & Chapman, 2015).

An important component in determining the role of genetic diversity in predicting invasion success is the availability of “unsuccessful invasions” for comparison (Zenni & Nuñez, 2013). Failed invasions are difficult to study for the obvious reasons that the organisms simply do not exist, or occur at low enough densities to make replicated collections a serious challenge. We have previously demonstrated the utility of a comparative approach, using taxonomically and geographically paired non-indigenous species that differ in the success of their invasions (extent of range expansion), to investigate the role of neutral genetic diversity (Wellband et al., 2017) and dietary niche breadth (Pettitt-Wade, Wellband, Heath, & Fisk, 2015) in predicting invasion success. We use the same approach here to compare standing genetic variation and population structure at functional genetic loci for two invasive species that differ in their extent of post-establishment range expansion. While the present study only considers one pair of species, and thus our power to make generalizations about the phenomenon we document is limited, our results provide insight into a potential mechanism contributing to the differential performance of these closely related species.

Round goby (*Neogobius melanostomus*) and tubenose goby (*Proterorhinus semilunaris*) are gobiid fish species native to the Black and Caspian Seas of Eastern Europe (Kottelat & Freyhof, 2007). Both species were first discovered in the St. Clair River in the North American Great Lakes basin in 1990 (Jude, Reider, & Smith, 1992). The source populations for invasions of both these species have been genetically traced back to the same northern tributary rivers of the Black Sea (Brown & Stepien, 2009; Stepien & Tumeo, 2006) and ballast water mediated introductions have

Accepted Article

been suggested as the most likely vector causing invasion. Despite originating from the same place and having been introduced at the same time, these species have markedly different extent of range expansion and impact in North America. Round goby have rapidly spread throughout the entire Great Lakes basin and reached high population densities with detrimental effects on other species (Corkum, Sapota, & Skora, 2004), while tubenose goby have remained relatively restricted in distribution and at low population densities where they occur (Grant, Shadle, & Andraso, 2012; Kocovsky et al., 2011). One possible explanation for their differential range expansion is differences in evolutionary potential between these species that have facilitated the more successful species and restricted the less-successful species in adapting to the novel invaded-range environments.

To investigate the role of functional genetic variation in explaining the differential success of these species invasions, we characterized genetic diversity and population structure based on neutral (microsatellite) and functional protein-coding single nucleotide polymorphism (SNP) genetic markers for round and tubenose goby. We tested two hypotheses. The first, that the highly successful invasive species had a higher standing functional genetic diversity and suffered less severe reductions from invasion source population than the less successful invasive species. The second, that selection played a greater role in explaining invaded range population divergence at functional genetic markers (greater divergence than predicted by neutral markers) among invasive populations for the highly successful invasive species. Greater divergence at functional markers for highly successful species may reflect a signature of adaptation that has facilitated increased range expansion relative to the less successful species. We developed species-specific protein-coding SNP markers from a transcriptome previously generated using RNA sequencing data for the two target species. We genotyped individuals from several introduced populations and the putative invasion source populations for both species using these SNP markers as well as microsatellite markers. We compare genetic diversity and patterns of genetic structure produced by functional SNPs and neutral microsatellite markers to identify the evolutionary forces driving the scope of invasion between these two related, but ecologically divergent species. Our results highlight the value in screening functional genetic diversity for increased accuracy in the prediction of invasion success.

Methods:

### *Sample Collection*

To test for the effects of genetic drift, specifically the loss of standing genetic variation, and natural selection as a result of colonization, we collected gobies from multiple sites in North America and one site in Europe. Round goby and tubenose goby were collected in 2012 and 2013 using a combination of seine netting and angling from six sites in North America (Figure 1) and one site from the port city of Kherson, Ukraine in Europe. North American sites were chosen to represent the geographic extent of range expansion in North America and the European site was chosen to represent the putative source population for these invasions based on existing studies of mitochondrial and microsatellite variation (Brown & Stepien, 2009; Neilson & Stepien, 2009). While the original founders may have included individuals from other European sites, the object of this study was to characterize differences in population structure in the invaded range as a result of post-establishment processes. Thus, the single European site was included to provide reference for whether either species of goby suffered a loss of genetic diversity, for any reason, during the invasion. Gobies were euthanized in accordance with the law and a fin clip was removed and preserved in a saturated salt solution (700 g/L Ammonium Sulfate, 25 mM Sodium Citrate, 20 mM Ethylenediaminetetraacetic acid, pH 5.2). DNA was extracted from fin clips using a glass filter-based binding procedure (Elphinstone, Hinten, Anderson, & Nock, 2003).

### *SNP characterization*

Previously generated RNA sequencing (RNAseq) data for round and tubenose goby (Wellband and Heath 2017) were used to develop functional gene locus single nucleotide polymorphism (SNP) markers for those two species. Briefly, nine individuals from each species were collected from the Detroit River site (Figure 1). Individuals were collected approximately three months following the collections for genotyping described above. Total RNA extracted from liver tissue was sequenced on two lanes of an Illumina HiSeq2000 using 100 bp pair-end sequencing and

Accepted Article

TruSeq stranded cDNA library construction that generated approximately 25 million paired-end reads per individual. We performed *de novo* transcriptome assembly for each species separately using Trinity v3.0.3 (Grabherr et al., 2011). For complete bioinformatics details regarding the transcriptome assemblies please refer to Wellband and Heath (2017). We used these RNAseq datasets to characterize variable SNPs in protein-coding genes for both species. We followed the Broad Institute's Genome Analysis Tool Kit (GATK) best practices methodology to characterize nucleotide variation among individuals (DePristo et al., 2011; Van der Auwera et al., 2013). First, we used BWA v0.7.12 (Li & Durbin, 2009) to align sequencing libraries for each individual to the transcriptome. We then removed PCR duplicates using Picard Tools (<http://broadinstitute.github.io/picard>). We performed base quality recalibration, indel realignment and variant discovery using GATK v3.6 (McKenna et al., 2010). We filtered detected variants using standard hard filtering parameters recommended for RNAseq experiments (DePristo et al., 2011; Van der Auwera et al., 2013). The specific parameters we used for each step are available in the supplementary material as a Unix shell script. To compare levels of standing genetic variation between round and tubenose goby we used a Fisher's Exact Test implemented in R v3.3.1 (R Core Team 2016) to test for a difference in the proportion of variable sites detected for each species.

Following variant characterization, we annotated SNPs. We characterized open reading frames of the assembled transcripts using GeneMarkS-T v5.1 (Besemer, Lomsadze, & Borodovsky, 2001) and then used SnpEff v4.2 (Cingolani et al., 2012) to characterize the functional relevance of each SNP (e.g. coding or 5' / 3' UTR, synonymous or nonsynonymous). We then used LEMONS (Levin et al., 2015) to predict exon – exon boundaries in assembled transcripts. We used this information to target SNPs that were far enough from exon – exon boundaries that we could design primers from the available transcript sequence to amplify genomic DNA without interference from introns. Additionally, we used gene function information for these transcripts generated during the annotation of the RNAseq project (Wellband & Heath 2017) to select genes involved in potentially adaptive processes (e.g. oxidative stress, immune system processes, metabolism) that reasonably may



Accepted Article

be expected to have experienced selection. We selected 96 transcripts with SNP variants for each species and designed 'SNP flanking primers' to target a 175 - 225 bp region around the SNP using the default settings with BatchPrimer3 v1.0 (You et al., 2008; accessed online at: <http://probes.pw.usda.gov/batchprimer3/index.html>). Forward-specific and reverse-specific universal adaptors were added to the 5' end of the primers to facilitate addition of sequencing adaptors and individual barcodes. To test the primer sets we used 12.5 uL PCR reactions that contained: 20 mM Tris-HCl pH 8.75, 10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.1% Triton X-100, 0.1 mg/mL BSA, 200 μM each dNTP, 200 nM forward and reverse primers, 0.5 U of taq polymerase (Bio Basic Canada Inc., Markham, ON) and 10-20 ng of gDNA. Thermocycling conditions were 95 °C for 2 minutes, 35 cycles of 95 °C for 15s, 60 °C for 15s and 72 °C for 30s, followed by 72 °C for 5 minutes. Primer sets that failed to amplify or exhibited larger than expected fragments were excluded from library preparation.

### *Genotyping*

#### *SNP genotyping:*

We amplified the SNP markers designed above following the GTseq methodology of Campbell *et al.* (2015). This method uses a nested PCR approach to first amplify the targeted loci in a multiplex PCR reaction and a second round of PCR to add sequencing adaptors and individual barcodes to samples. All primers for each species were combined and diluted to an individual primer concentration of 200 nM. Multiplex PCRs were performed in 7 uL volumes for each individual that contained 3.5 uL of 2X Multiplex Plus MasterMix (Qiagen, Inc., Toronto, Canada), 1.5 uL of multiplex primer pool and 2 uL of genomic DNA. Thermocycling conditions were 95 °C for 15 minutes, 15 cycles of 95 °C for 30s, 60 °C for 30s and 72 °C for 1 min, followed by 72 °C for 2 minutes. PCR products were diluted 20-fold by adding 133 uL of ddH<sub>2</sub>O. We added sequencing adaptors and identifying barcodes to each individual with a second 10 uL PCR reaction that contained 5 uL of 2X Multiplex Plus MasterMix (Qiagen, Inc., Toronto, Canada), 1 uL of 10 uM Ion-A -

Accepted Article

barcoded primer, 1 uL of 10 uM Ion-P1 primer and 3 uL of diluted PCR product from the first reaction. Thermocycling conditions for the second PCR were 95 °C for 15 minutes, 10 cycles of 95 °C for 30s, 60 °C for 30s and 72 °C for 30s, followed by 72 °C for 5 minutes. We combined 5 uL of library from each individual, performed an isopropanol precipitation and a gel extraction of the desired library fragment by cutting the region from 150 bp to 300 bp and recovering the DNA using a commercial gel extraction binding column kit following the manufacturer's instructions (Epoch Life Science, Inc., Sugar Land, TX). The final library was quantified using a DNA High Sensitivity Assay on a Bioanalyzer 2100 (Agilent Technologies Canada Inc., Mississauga, Canada). The library was diluted to 50 pM and prepared for sequencing (emulsion PCR, clean-up and chip loading) using Ion PGM Hi-Q chemistry in an Ion Chef System (Thermo Fisher Scientific, Inc., Streetsville, Canada). The library was sequenced using 850 nucleotide flows (400 bp run) in an Ion 318 Chip v2 on an Ion PGM Sequencer (Thermo Fisher Scientific, Inc., Streetsville, Canada).

The sequencing output was separated by individual barcode using the Torrent Suite Software v5.0.4 (Thermo Fisher Scientific, Inc., Streetsville, Canada) producing a fastq file for each individual. Individual sequences were then trimmed of adaptors using cutadapt v1.11 (Martin, 2011) and mapped to the transcriptomic reference sequences using BWA v0.7.12 (Li & Durbin, 2009). Data processing followed the GATK best practices for base recalibration, indel realignment and variant calling (DePristo et al., 2011; Van der Auwera et al., 2013). We performed joint genotyping on all samples together as recommended and then applied standard hard filtering parameters on the variant set (specific parameters used available in the supplemental material as a Unix shell script). We extracted all SNPs that were variable, and called in at least 80% of individuals. We also excluded individuals that were missing more than 10% of their genotypes. We used PGDSpider v2.0 (Lischer & Excoffier, 2012) to convert the variant call files into other formats for subsequent analysis. We used plink v1.07 (Purcell et al., 2007) to remove loci with a minor allele frequency less than 0.01 and to calculate linkage disequilibrium between pairs of loci in each population. Some of the amplicons we designed contained more than one SNP, so to avoid bias in our data we removed linked SNPs to retain only one SNP per amplicon.

### *Microsatellite genotyping:*

To provide a control for the functional SNP markers and assess divergence and population structure that result from genetic drift alone, we genotyped both goby species at neutral microsatellite markers. For invasive populations of both species we used previously published data for nine microsatellite markers (see Wellband *et al.* 2017). We further genotyped individuals collected from the native range using these microsatellite markers following the procedures described in Wellband *et al.* (2017).

### *Population genetic analyses*

If genetic drift associated with the colonization process has resulted in the loss of genetic diversity in invaded range populations of goby we would expect to see reductions in heterozygosity for both microsatellite and SNP markers and a reduction in allelic richness of microsatellites. To quantify these effects and to assess whether drift has disproportionately affected tubenose goby we calculated observed heterozygosity for all loci (SNP and microsatellite) using the ‘adegenet’ v2.0.1 package (Jombart, 2008) in R v3.3.1 (R Core Team 2016). We also calculated microsatellite allelic richness using the ‘hierfstat’ v0.04-22 package (Goudet & Jombart, 2015). This method uses rarefaction to account for differences in sample size and standardizes the number of alleles to the smallest population by locus combination of individuals. We then used a linear mixed effects model implemented in the ‘lme4’ v1.1-12 package (Bates *et al.*, 2014) in R v3.3.1 (R Core Team 2016) to test for a difference among all populations for observed heterozygosity and allelic richness for each species individually. Models included a fixed effect of population, a random effect for locus and we tested significance of the fixed effect of population using a likelihood ratio test of the full model compared with a reduced model without the population term. For tests where population was a significant predictor of variation in allelic richness or heterozygosity we then used Tukey post-hoc testing, as implemented by the “glht” function in the R package ‘multcomp’ v1.4-6 (Hothorn *et al.*

2008), to investigate all pairwise differences between populations. These comparisons tested whether differences in diversity were due to the initial invasion (Ukraine population vs. invaded population) or secondary range expansion (Core invaded population [DR for round goby and MB for tubenose goby] vs. invaded range edge population). Due to the fact that SNP markers were characterized as variable in the invaded range, a similar test for the SNP markers would be inappropriate as it is unlikely that we captured the true range of nucleotide diversity present for functional genes in these species. Instead we quantified statistically significant shifts in SNP allele frequency from the native population to each invaded population using Fisher Exact tests in R v3.3.1 (R Core Team 2016). We then compared the relative number of significant shifts observed for round and tubenose goby using a  $\chi^2$  test with the expectation that species experiencing a larger effect of genetic drift would exhibit a greater number of markers with dramatic allelic shifts.

If genetic drift was the dominant evolutionary force explaining genetic diversity of biological invaders, we would expect our microsatellite and SNP markers to demonstrate similar patterns of genetic structure among invaded populations. Alternatively, if selection was also important, we would expect our functional SNPs to exhibit different patterns of divergence among groups that reflect the effects of selection due to site-specific environmental conditions. To characterize potential differences in the patterns of population structure between neutral and functional marker types in both species invaded range we used a naïve clustering approach that does not make assumptions about idealized populations, called discriminant analysis of principal components (DAPC, Jombart et al., 2010), to explain allelic variation among individuals. The lack of assumptions about genetic equilibrium made by this approach are an important consideration for invasive populations due to their inherently non-equilibrium nature (Fitzpatrick, Fordyce, Niemiller, & Reynolds, 2011).

Briefly, DAPC uses k-means clustering on the allelic datasets after they have been transformed using principal components analysis. We tested a range of possible genetic groupings (one to ten clusters) for both species while retaining all allelic principal components (PCs). The number of clusters was selected based on the profile of the Bayesian Information Criterion (BIC) for each k-cluster as recommended (Jombart et al., 2010). Discriminant function analysis was then

performed to maximize the difference among groups based on the chosen number of clusters. To avoid over-fitting and ensure reproducibility of the analysis, we used a resampling and cross validation routine to determine the optimal number of allelic PCs to be included in the discriminant function analysis. A stratified random sampling of 90% of the dataset was selected as a training set and DAPC was performed over a range of retained PCs (5 – 40 PCs). The remaining 10% of the dataset was used to validate the analysis. This procedure was replicated 100 times and we chose the number of PCs to retain in the final analysis based on the number of PCs demonstrating the highest mean assignment success and the lowest variability of assignment success as recommended (Jombart et al., 2010). We identified the alleles responsible for discriminating groups on the first discriminant function by examining the loadings of the PCs. To quantitatively compare patterns of population structure characterized by neutral and functional markers we used Fisher's Exact Test to compare the proportion of individuals assigned to a group for each population separately. These comparisons were composed of a 2x2 contingency table for each population where the cells represent the number of individuals with a primary membership (>0.5) to a cluster and the row and columns of the table represent the genetic clusters (K=2, see below) and marker type describing structure respectively.

To test for evidence of selection driving differences in SNP markers among invaded populations of gobies we conducted  $F_{ST}$  outlier tests. Populations of invasive species do not conform to the expectations of migration-drift equilibrium assumed by many models for outlier detection (Lotterhos and Whitlock 2014). To account for the non-island model nature of our sampled populations we used two methods of outlier detection that account for underlying demographic histories: 1) the extended Lewontin-Krakauer method of Bonhomme et al. (2010) and 2) Bayenv2 (Günther & Coop 2013). The method of Bonhomme *et al.* (2010) calculates an  $F_{ST}$  analogue ( $T_{F-LK}$ ) while controlling for co-ancestry using a phylogenetic tree based on Reynold's co-ancestry coefficient (Reynolds, Weir, & Cockerham, 1983). Under the assumption that all of our microsatellite markers are selectively neutral and unlinked to any genes under selection, we calculated Reynold's co-ancestry coefficient among populations with the Ukraine population as the out-group based on the microsatellite datasets for each species using the 'adegenet' v2.0.1 package (Jombart, 2008) in R. We

then used FLK as implemented in R by Bonhomme *et al.* (2010) and available online at <https://qgsp.jouy.inra.fr> (accessed on 19 October 2016) to calculate  $T_{F-LK}$  and associated p-values for the outlier test. Bayenv2 also allows the user to account for non-independence of populations by specifying a variance-covariance matrix. This method performs well when it can be parameterized with many (thousands) known-neutral SNPs (Günther & Coop 2013); however, as we did not have an appropriately large dataset we again used the microsatellite data to provide an estimate of underlying population structure where we used the Reynold's distance calculated above as the co-variance matrix. This usage should be appropriate given that the co-variance and  $F_{ST}$ -based measures are closely related (Günther & Coop 2013). Bayenv2 does not provide an estimate of statistical significance; therefore, to compare outlier tests we performed a spearman rank correlation of the FLK and Bayenv2 test statistics as suggested by Günther and Coop (2013).

Finally, to test whether the inclusion of multiple populations in the outlier analysis for round goby upwardly biased our ability to detect outliers compared to only the two populations available to be compared for tubenose goby, we conducted FLK outlier tests for all pairwise population comparisons of round goby. We compared the average number of outliers detected between pairs of round goby populations to the number detected for tubenose goby using a one-sample T-test.

Results:

#### *SNP characterization*

We characterized 46 092 SNPs in a total of 49.1 million base pairs (Mb) of round goby transcriptome sequence and 28 217 SNPs in total of 50.2 Mb of tubenose goby transcriptome sequence (Table 1). This equated to an average of one variable site every 1065 bp for round goby and 1779 bp for tubenose goby, or an average of 1.8 variable sites per transcript for round goby and 1.2 variable sites per transcript for tubenose goby. The difference in standing genetic variation for functional SNPs between goby species was highly significant (Fisher's Exact Test,  $p < 0.001$ ) and reflected over 60% more variable sites in round goby than tubenose goby. We did not take into

account linkage disequilibrium for these SNPs and round goby has a slight tendency for more SNPs per transcript (Figure S1), thus the effective number of differences may be smaller than we report. However, the number of round goby transcripts that had at least one variable site ( $N = 14284$ ) was nearly twice that of tubenose goby ( $N = 8397$ ) suggesting the differences we observed do reflect higher diversity for round goby. To select candidate SNP markers, we removed variants from the dataset that came from transcripts that did not possess a valid start codon (incomplete or non-coding transcripts) which resulted in 39 408 available SNPs for round goby and 23 591 available SNPs for tubenose goby (Table 1). We selected 96 different transcripts with SNPs from each species to assay. Of the markers we designed, 73 markers for round goby and 80 markers for tubenose goby amplified, were of the expected size and were included in our multiplex assay (Supplementary material Table S1).

#### *Population genetic analyses*

We collected 30 to 50 individuals from each site for all sites except for tubenose goby at Mitchell's Bay where we were only able to collect 25 individuals. We amplified SNPs for 30 individuals from each population; however, we had inconsistent amplification that resulted in approximately half of the samples for the round goby population from Hamilton Harbour and the tubenose goby population from Kherson, Ukraine and approximately one third of the samples for round goby populations from Detroit River and Trent-Severn Waterway sites being represented in the sequencing library (Table 2). These effects are believed to be the result of DNA quality and concentration and the lack of among- individual normalization prior to sequencing. They are unlikely to be the result of any deterministic effects of the markers not amplifying in certain populations. The distribution of reads was not uniform across SNP markers for either species (Supplementary material Table S1). This resulted from the over-representation of a few loci and the loss of several loci for both species. After all quality filters were applied to the data we analyzed data for 48 SNPs for round goby and 34 SNPs for tubenose goby.

The only statistical difference for observed heterozygosity was of SNP markers in tubenose goby (Figure 2; TNG SNP:  $\chi^2 = 16.5$ ,  $df = 2$ ,  $p < 0.001$ ). This difference resulted from approximately 10% higher heterozygosity in the introduced Mitchell's Bay population compared with both the putative invasion source population from Kherson, Ukraine and the other invaded population of Thunder Bay (Figure 2; Tukey Test: MB-KH  $p < 0.001$ , MB-TB  $p = 0.006$ ). In contrast, tubenose goby exhibited a significant loss of microsatellite allelic richness in invaded populations compared with the native range population (Figure 2; Tukey Test: TB-KH  $p < 0.001$ , MB-KH  $p = 0.001$ ). Similarly there were, on average, more significant allele frequency shifts observed for tubenose goby (Table 2, mean of 41.2% of SNPs per population, range 10 – 18 / 34 SNPs) than round goby (Table 2, mean of 11.9% of SNPs per population, range: 0 – 15 / 48 SNPs;  $\chi^2 = 3030.9$ ,  $df = 11$ ,  $p < 0.001$ ).

DAPC identified two genetic clusters in the SNP data for both round and tubenose goby based on BIC (Figure S2). For both species' microsatellite datasets there was support for additional genetic clusters of up to approximately five for both species (Figure S2); however, to simplify the comparison between markers we chose to use  $K = 2$  for the microsatellite markers as well for both species. Adding additional clusters to the microsatellite analysis simply resulted in subdivision of existing groups at  $K = 2$  (Figure S3). For ease of visualization we have restricted the graphs of posterior population assignment probabilities to only those individuals for which we also had SNP data; however, it should be noted that we conducted the analyses on all individuals for which we had data to avoid biases associated with accurately estimating microsatellite allele frequencies with low sample sizes. These issues are lessened for SNP markers where there are only two alleles.

Divergent patterns of population structure were identified between marker types for round goby (Figure 3). These differences were exhibited by general patterns of group membership for sampling sites that switched (e.g. NA grouping with HH and TS based on functional SNPs but DU, CO and DR based on microsatellites; Figure 3) or a substantial proportion of individuals with group membership that switched based on genetic markers (HH and TS; Figure 3). These differences in genetic group membership were statistically significant for three round goby populations (NA, HH, TS; Fisher's Exact Test:  $p < 0.001$ ) but not the others. The pattern was present for analyses that



included all individuals genotyped at microsatellites as well as when the analysis was restricted to only those present in the SNP dataset (results not shown). In contrast tubenose goby did not show dramatic differences in population structure and the significance of changes were non-significant (Figure 4). The SNP principal component loadings that contributed to the discriminant function separating groups indicated allelic differences at SNPs 1, 14, 39, 46, 55, 60, 62, 75, 91, 93, 98 and 111 were responsible for discriminating the two groups for round goby. For tubenose goby, the loadings of the principal components included in the discriminant function indicated allelic differences at SNPs 24, 45, 46, 51, 57, 60, 67 and 80 drove the differences between genetic clusters.

Based on the FLK statistic (Bonhomme *et al.* 2010) we identified 14 SNPs that were  $F_{ST}$  outliers for round goby and three SNPs that were  $F_{ST}$  outliers for tubenose goby (Figure 5). Five of the  $F_{ST}$  outlier SNPs for round goby (SNPs 1, 14, 15, 22 and 109) were significant following false discovery rate (Benjamini & Hochberg, 1995) correction for multiple tests. Four of the SNPs (SNPs 1, 14, 39, 62) were also important for discriminating genetic clusters from the DAPC analysis. None of the three  $F_{ST}$  outliers for tubenose goby (SNPs 45, 60 and 80) were significant following false discovery rate correction; however, all of them were identified as being important for discriminating genetic clusters by the DAPC analysis. The biological functions of the outliers primarily reflected genes involved in the heat-shock and oxidative stress responses for round goby and genes involved in steroid signaling pathways and xenobiotic processing for tubenose goby (Table 3). The outlier results of Bayenv2 were roughly congruent with those of FLK based on Spearman's rank correlation for round goby ( $\rho = 0.26$ ,  $p = 0.05$ ) but not for tubenose goby ( $\rho = -0.03$ ,  $p = 0.88$ ). The differences in these results likely reflect the small number of neutral loci used and the imperfect proxy the Reynold's distance matrix represents for the allelic covariance matrix used by Bayenv2 (Günther & Coop 2013).

The average number of  $F_{ST}$  outliers as determined by FLK between pairs of round goby populations was 4.0 (95% CI: 1.6 – 6.3) and did not statistically differ from the number detected for tubenose goby (one-sample T-test:  $t = 0.9$ ,  $df = 14$ ,  $p = 0.38$ ). However, the distribution of pairwise comparisons was bimodal (Figure S4) with two sets of pairwise population comparisons each

demonstrating 13 outliers, suggesting that it is possible for only two populations to exhibit a similar number of outliers as that observed for all round goby populations overall. This result highlights the population specific nature of adaptive divergence.

#### Discussion:

Biological invasions are expected to suffer from significantly reduced genetic diversity due to genetic drift associated with colonization; however, this has been difficult to demonstrate empirically (Dlugosch & Parker, 2008; Uller & Leimu, 2011). This loss of genetic diversity may result in reduced evolutionary potential; yet, adaptation also appears to be common in biological invasions (Whitney & Gabler, 2008). A critical missing component of the existing studies is a lack of knowledge about less-successful or failed invasions (Zenni & Nuñez, 2013). Here we used a comparative approach to contrast functional and neutral genetic variation for two species of goby with similar invasions histories, but that differ dramatically in the extent of their post-establishment range expansion. We have demonstrated that tubenose goby, the less successful species, exhibited evidence of reduced allelic diversity associated with colonization and also appears to have less evidence of adaptive divergence in its North American invaded range, compared with the considerably more abundant and widespread round goby. While there are many ecological reasons that an invasion may succeed or fail (Sakai *et al.* 2001), genetic variability is the prerequisite for populations to evolve adaptations to the ecological challenges they encounter.

Modest reductions in genetic diversity of approximately 10 – 20% for heterozygosity and allelic richness from native regions to invaded ones are the norm for most invasive species (Dlugosch & Parker, 2008). We observed no obvious reductions in allelic richness for the highly successful round goby but much larger magnitude reductions for the less-successful tubenose goby (approximately 50% for allelic richness). These results are consistent with previous work on round goby that identified no reductions in diversity (Brown & Stepien, 2009), but they stand in contrast to the conclusions of Stepien and Tumeo (2006) who did not observe reductions in diversity for tubenose

Accepted Article

goby. However, their work had limited sample size (11 individuals) and used a single mitochondrial DNA marker (3/4 native range haplotypes present in North America). The selection of our functional SNPs (initially characterized as variable in an invasive population) made it difficult to identify losses of allelic diversity at these markers, but drift effects should still be evident in the form of significant allele shifts. Indeed, we observed those effects and they are much stronger for tubenose goby compared to round goby. It is possible that these effects reflect selection acting during or immediately after transport and establishment in the invaded region; however, the evidence we present of similar reductions in allelic richness at neutral microsatellites suggests the effects of genetic drift alone would have been sufficient to produce these patterns. In characterizing functional SNPs, we observed that round goby had a higher level of nucleotide diversity in coding regions across the whole transcriptome, relative to the tubenose goby. The characterization of SNPs as variable in the invaded range represents ascertainment bias that precludes us from determining whether these differences are a result of drift or historical differences between species in the native range. Regardless, a higher level of nucleotide diversity for functional genes (standing genetic variation) should theoretically provide an increased capacity for adaptive evolution under novel environmental conditions.

An alternative explanation for the differences in allelic diversity between round and tubenose goby is that ecological factors have prevented the success of tubenose goby and led to the decreased diversity that we have observed. If this were the case, we would also expect to see reductions in heterozygosity reflecting a protracted bottleneck (Allendorf 1986). In contrast to the results for allelic diversity, we observed no reductions in heterozygosity for either species and heterozygosity was actually increased in one of the invasive tubenose goby populations relative to the native range. This suggests that there is no ongoing bottleneck for tubenose goby and that reductions in allelic diversity resulted from a founder effect or selection early in the invasion process.

It is widely accepted that the Great Lakes round goby experienced multiple introduction events (Brown & Stepien, 2009; Johansson et al., 2018) and this likely explains the lack of evidence for founder effects or genetic bottlenecks we and others have found. In contrast, much less is known about the likelihood of multiple Great Lakes invasions for tubenose goby. Their discontinuous

geographic distribution in the Great Lakes may suggest that they too experienced multiple introductions; however, we do not believe this to be the case. Allelic richness (microsatellites) was lower and significant allele frequency shifts (SNPs) were more frequent in Thunder Bay (TB) than in the site near the initial North American introduction (MB), suggesting the TB population experienced strong effects of genetic drift. In addition, the North American TNG sites were found to cluster more closely with the other Great Lakes site than with multiple native range populations (data not shown). Furthermore, both Great Lakes sites were found to secondarily cluster with a single native range population, Kherson (KH), consistent with our current understanding of the source of the TNG introduction (Stepien & Tumeo, 2006). We thus believe that the genetic structure among tubenose goby populations and the discontinuous geographic distribution reflect strong genetic drift effects resulting from a secondary founding event in Thunder Bay that was sourced by individuals from the initial site of invasion (MB) coupled with slow natural dispersal this species demonstrates in North America (Kocovsky et al., 2011).

We demonstrated divergent patterns of population structure between neutral and functional gene loci for round goby. The same functional SNP loci that were most important for determining population structure were also identified as being  $F_{ST}$  outliers, implying that divergent selection may be acting to drive differences in allele frequency for certain populations. It is important to note that our characterization of SNP markers differs from reduced-representation genome-wide approaches and other anonymously chosen marker sets in that all of our markers are in the coding regions of known function genes and the majority of them represent non-synonymous variants. Thus, none can truly be considered neutral. While both species of goby showed some evidence of divergent selection at SNP loci, round goby demonstrated an overall larger proportion of SNPs ( $13/48 = 27.1\%$ ) potentially under divergent selection compared with tubenose goby ( $3/34 = 8.8\%$ ). The average pairwise difference between populations of round goby did not differ from that of tubenose goby; however, certain pairs of round goby populations still exhibited large numbers of outliers suggesting that the divergence we observed is population specific. In each of these cases, one of the populations involved was one of those identified as those having a significant switch in group membership

between marker types indicating that these populations that are similar at neutral loci have diverged at functional loci. These results are consistent with the idea that invasive populations locally adapt to different habitat conditions throughout the invaded range. Rapid evolution is known to be a common phenomenon in biological invasions (Whitney & Gabler, 2008) and can lead to higher fitness for local populations of organisms in the invaded range (e.g. Colautti & Barrett, 2013; Kinnison, Unwin, & Quinn, 2008). Furthermore, potential for adaptive evolution is a known factor that can influence the rate of range expansion (García-Ramos & Rodríguez, 2002). Our results thus indicate that the limited expansion of tubenose goby in the Great Lakes basin may be the result of reduced evolutionary potential driven by a lack of genetic diversity. It would be interesting to contrast patterns of functional genetic variation of the North American invasion of tubenose goby with invasions of the tubenose goby throughout Europe where it has been much more successful (e.g. Naseka, Boldyrev, Bogutskaya, & Delitsyn, 2005; Vašek et al., 2011).

$F_{ST}$  outlier approaches have a tendency to generate false positive results when the demographic history of the biological system does not match the assumptions of the model (Lotterhos & Whitlock, 2014). Biological invasions are characterized by complicated demographic histories due to multiple introduction events and subsequent hybridization among these groups (Dlugosch, Anderson, Braasch, Cang, & Gillette, 2015; Dlugosch & Parker, 2008). We used two approaches to detect SNP outliers that account for the recent demographic relationship among populations and explicitly use them as the null model to test for outliers (Bonhomme *et al.* 2010; Günther & Coop 2013), thus controlling for the effects of drift among populations. These types of approaches have been recommended for use with expanding populations (Lotterhos & Whitlock, 2014) and provide support for our results reflecting real divergence among populations due to natural selection. The lack of congruency between outlier approaches for tubenose goby may reflect false positive outliers detected by the FLK analysis but it could simply reflect low power in the Bayenv2 analysis as a result of the small number of loci and the non-parametric comparison we employed (Günther & Coop 2013). While targeting nucleotide variation in protein-coding genes represents a powerful tool to investigate functional variation (De Wit *et al.* 2015), the small number of loci we used (for a sufficient

number of samples required to make population genetic inference) is a limitation of this work. Future work comparing these species should aim to sample a larger number of markers representing both functional and putatively neutral markers that capture the full range of variation present in the native range of these species.

It is also possible that the outlier results we obtained resulted solely from genetic drift associated with range expansion through a process known as gene surfing (Klopfstein, Currat, & Excoffier, 2006). Here, variants can rise to high frequency at the leading edge of the range expansion and can mimic a selective sweep (Currat et al., 2006). This process should lead to the accumulation of deleterious variants in expanding populations and is expected to result in decreased fitness across broad areas of the expanding range as a result of reduced heterozygosity (Peischl & Excoffier, 2015). We did not observe consistent reductions in heterozygosity for SNP markers suggesting that, while gene surfing is a possible explanation for some of SNP divergence it is unlikely to explain all of the outliers we characterized.

Functional genetic markers, like the ones used in this study, reveal important genetic differences among populations that may not be evident based solely on neutral genetic markers. For example, putatively adaptive SNPs have revealed population structure patterns that were not evident for neutral loci in salmonids (Ackerman, Templin, Seeb, & Seeb, 2013; Hand et al., 2016; O'Malley, Jacobson, Kurth, Dill, & Banks, 2013) and invasive invertebrates (Rohfritsch et al., 2013; Tepolt & Palumbi, 2015). More importantly, the functions of such genes can reveal important information about the environmental or ecological forces driving genetic differences among populations. Several of the genes that we identified as having  $F_{ST}$  outliers for round goby are involved in heat shock and oxidative stress responses (Table 3), suggesting that the dramatic range expansion of round goby throughout the Great Lakes basin may, in part, have resulted from adaptive evolution for these traits. These genes are known to be involved in responses to maintain or regain homeostasis in the face of dramatically altered temperature (Richter, Haslbeck, & Buchner, 2010) and are key components of the thermal tolerance of species (Kassahn, Crozier, Pörtner, & Caley, 2009). Thermal tolerance has previously been implicated in the differential range expansion of these two species (O'Neil, 2013;

Xin, 2016; Wellband and Heath 2017), for range expansion of other invasive fish in the Great Lakes (Kolar & Lodge, 2002) and for successful invasion and range expansion of aquatic species in general (Bates et al., 2013). Although our study does not test the same suite of loci in both species, tubenose goby did not show divergence at the heat shock related gene that we assayed. Further work should explicitly test the adaptive divergence of thermal tolerance related markers for these two species and the generality of diversity for genes of these processes in the relative invasion success of a variety of other taxa. This type of functional genetic information may be useful for the improved management of these goby species by identifying populations with greater risk to invade specific un-invaded habitats and for risk assessment of other potential future invaders (Chown et al., 2015) by identifying biomarkers or signatures of diversity that may predict invasion risk.

We previously used a comparative approach to explicitly demonstrate that reduced genetic variation can be associated with less-successful biological invasion (Wellband et al. 2017) and extended that work here to show that, for tubenose goby, the reduction may be due to losses as a result of the demographic process associated with colonization. Furthermore, the effects of drift associated with tubenose goby colonization are observable at functional genetic markers. This may have consequences for the ability of the species to adapt to the novel Great Lakes environment and expand their range following establishment. We used a demographically sensitive approach to detect  $F_{ST}$  outliers and identified functional protein-coding SNPs putatively under selection that explain population structure for these species. Our results suggest that a combination of genetic drift and natural selection are acting to structure invasive populations of gobies and that functional genetic markers are critical for understanding processes influencing the range expansion of invasive species. Our results further implicate adaptive evolution of allele frequencies at genes related to thermal tolerance in the extensive range expansion of round goby, which is consistent with hypotheses regarding the range expansion of invasive fishes in the Great Lakes (Kolar & Lodge, 2002). Our results highlight the need to incorporate functional genetic markers in the assessment of genetic diversity and evolutionary potential of invasive species for improved risk assessment and

management. We advocate the use of genomic approaches to improve the resolution of both demographic and evolutionary processes effecting biological invasions.

Acknowledgements:

We thank Katerina Stojanovic and Sergei Kocodiy for their assistance with sample collection. We also thank the Associate Editor and three anonymous reviewers for constructive comments that improved the quality of this work. Funding from the Canadian Aquatic Invasive Species Network II to DDH and ATF, an Ontario Trillium Scholarship to HPW and an NSERC post-graduate scholarship to KWW supported this work.

Literature cited:

Ackerman, M. W., Templin, W. D., Seeb, J. E., & Seeb, L. W. (2013). Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. *Conservation Genetics*, *14*(2), 483–498. doi:10.1007/s10592-012-0401-7

Bates, A. E., McKelvie, C. M., Sorte, C. J. B., Morley, S. A., Jones, N. A. R., Mondon, J. A., ... Quinn, G. (2013). Geographical range, heat tolerance and invasion success in aquatic species. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20131958. doi:10.1098/rspb.2013.1958

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. doi:10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300.

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for



prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 29(12), 2607–2618. doi:11410670

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & SanCristobal, M. (2010). Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*, 186(1), 241–262. doi:10.1534/genetics.110.117275

Brown, J. E., & Stepien, C. A. (2009). Invasion genetics of the Eurasian round goby in North America: tracing sources and spread patterns. *Molecular Ecology*, 18(1), 64–79. doi:10.1111/j.1365-294X.2008.04014.x

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855–867. doi:10.1111/1755-0998.12357

Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *Evolutionary Applications*, 8(1), 23–46. doi:10.1111/eva.12234

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. doi:10.4161/fly.19695

Colautti, R. I., & Barrett, S. C. H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342(6156), 364–6. doi:10.1126/science.1242121

Colautti, R. I., & MacIsaac, H. J. (2004). A neutral terminology to define “invasive” species. *Diversity and Distributions*, 10(2), 135–141. doi:10.1111/j.1366-9516.2004.00061.x

Corkum, L. D., Sapota, M. R., & Skora, K. E. (2004). The Round Goby, *Neogobius melanostomus*, a Fish Invader on both sides of the Atlantic Ocean. *Biological Invasions*, 6(2), 173–181. doi:10.1023/B:BINV.0000022136.43502.db

- Accepted Article
- Currat, M., Excoffier, L., Maddison, W., Otto, S., Ray, N., Whitlock, M., & Yeaman, S. (2006). Comment on “Ongoing Adaptive Evolution of ASPM, a Brain Size Determinant in Homo sapiens” and “Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans.” *Science*, *313*(5784), 172a. doi:10.1126/science.1122712
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. doi:10.1038/ng.806
- Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, *24*(9), 2095–2111. doi:10.1111/mec.13183
- Dlugosch, K. M., & Parker, I. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17*(1), 431–449. doi:10.1111/j.1365-294X.2007.03538.x
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, *29*(1), 51–63. doi:10.1016/j.tree.2013.09.008
- Elphinstone, M. S., Hinten, G. N., Anderson, M. J., & Nock, C. J. (2003). An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes*, *3*(2), 317–320. doi:10.1046/j.1471-8286.2003.00397.x
- Fitzpatrick, B. M., Fordyce, J. A., Niemiller, M. L., & Reynolds, R. G. (2011). What can DNA tell us about biological invasions? *Biological Invasions*, *14*(2), 245–253. doi:10.1007/s10530-011-0064-1
- García-Ramos, G., & Rodríguez, D. (2002). Evolutionary speed of species invasions. *Evolution*, *56*(4), 661–668. doi:10.1554/0014-3820(2002)056[0661:ESOSI]2.0.CO;2
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A.

(2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome.

*Nature Biotechnology*, 29(7), 644–652. doi:10.1038/nbt.1883

Grant, K. A., Shadle, M. J., & Andraso, G. (2012). First report of tubenose goby (*Proterorhinus semilunaris*) in the eastern basin of Lake Erie. *Journal of Great Lakes Research*, 38(4), 821–824. doi:10.1016/j.jglr.2012.09.019

Hand, B. K., Muhlfeld, C. C., Wade, A. A., Kovach, R. P., Whited, D. C., Narum, S. R., ... Luikart, G. (2016). Climate variables explain neutral and adaptive variation within salmonid metapopulations: the importance of replication in landscape genetics. *Molecular Ecology*, 25(3), 689–705. doi:10.1111/mec.13517

Johansson, M. L., Dufour, B. A., Wellband, K. W., Corkum, L. D., MacIsaac, H. J., & Heath, D. D. (2018). Human-mediated and natural dispersal of an invasive fish in the eastern Great Lakes. *Heredity*. doi:10.1038/s41437-017-0038-x

Jombart, T. (2008). Adegnet: A R package for the multivariate analysis of genetic markers.

*Bioinformatics*, 24(11), 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. doi:10.1186/1471-2156-11-94

Jude, D. J., Reider, R. H., & Smith, G. R. (1992). Establishment of Gobiidae in the Great Lakes Basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 416–421.

Kassahn, K. S., Crozier, R. H., Pörtner, H. O., & Caley, M. J. (2009). Animal performance and stress: responses and tolerance limits at different levels of biological organisation. *Biological Reviews*, 84(2), 277–292. doi:10.1111/j.1469-185X.2008.00073.x

Kinnison, M. T., Unwin, M. J., & Quinn, T. P. (2008). Eco-evolutionary vs. habitat contributions to invasion in salmon: Experimental evaluation in the wild. *Molecular Ecology*, 17(1), 405–414.

doi:10.1111/j.1365-294X.2007.03495.x

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3), 482–490.

doi:10.1093/molbev/msj057

Kocovsky, P. M., Tallman, J. A., Jude, D. J., Murphy, D. M., Brown, J. E., & Stepien, C. A. (2011). Expansion of tubenose gobies *Proterorhinus semilunaris* into western Lake Erie and potential effects on native species. *Biological Invasions*, 13(12), 2775–2784. doi:10.1007/s10530-011-9962-5

Kolar, C. S., & Lodge, D. M. (2001). Progress in invasion biology: Predicting invaders. *Trends in Ecology and Evolution*, 16(4), 199–204. doi:10.1016/S0169-5347(01)02101-2

Kolar, C. S., & Lodge, D. M. (2002). Ecological Predictions and Risk Assessment for Alien Fishes in North America. *Science*, 298(5596), 1233–1236. doi:10.1126/science.1075753

Kottelat, M., & Freyhof, J. (2007). *Handbook of European freshwater fishes*. Kottelat, Cornol, Switzerland and Freyhof, Berlin, Germany.

Levin, L., Bar-Yaacov, D., Bouskila, A., Chorev, M., Carmel, L., & Mishmar, D. (2015). LEMONS – A Tool for the Identification of Splice Junctions in Transcriptomes of Organisms Lacking Reference Genomes. *PLOS ONE*, 10(11), e0143329. doi:10.1371/journal.pone.0143329

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. doi:10.1093/bioinformatics/btp324

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. doi:10.1093/bioinformatics/btr642

Liu, N., Chen, L., Wang, S., Oh, C., & Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6(Suppl

1), S26. doi:10.1186/1471-2156-6-S1-S26

- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23(9), 2178–2192. doi:10.1111/mec.12725
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10. doi:10.14806/ej.17.1.200
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. doi:10.1101/gr.107524.110
- Naseka, A. M., Boldyrev, V. S., Bogutskaya, N. G., & Delitsyn, V. V. (2005). New data on the historical and expanded range of *Proterorhinus marmoratus* (Pallas, 1814) (Teleostei: Gobiidae) in eastern Europe. *Journal of Applied Ichthyology*, 21(4), 300–305. doi:10.1111/j.1439-0426.2005.00685.x
- Nei, M., Maruyama, T., & Chakraborty, R. (1975). The Bottleneck Effect and Genetic Variability in Populations. *Evolution*, 29(1), 1–10. Retrieved from <http://www.jstor.org/stable/10.2307/2407137>
- Neilson, M. E., & Stepien, C. A. (2009). Evolution and phylogeography of the tubenose goby genus *Proterorhinus* (Gobiidae: Teleostei): evidence for new cryptic species. *Biological Journal of the Linnean Society*, 96(3), 664–684. doi:10.1111/j.1095-8312.2008.01135.x
- O'Malley, K. G., Jacobson, D. P., Kurth, R., Dill, A. J., & Banks, M. A. (2013). Adaptive genetic markers discriminate migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*) amid continued gene flow. *Evolutionary Applications*, 6(8), 1184–1194. doi:10.1111/eva.12095
- O'Neil, J. (2013). *Determination of standard and field metabolic rates in two Great Lakes invading*

fish species: round goby (*Neogobius melanostomus*) and tubenose goby (*Proterorhinus semilunaris*). University of Windsor.

- Peischl, S., & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Molecular Ecology*, *24*(9), 2084–2094. doi:10.1111/mec.13154
- Pettitt-Wade, H., Wellband, K. W., Heath, D. D., & Fisk, A. T. (2015). Niche plasticity in invasive fishes in the Great Lakes. *Biological Invasions*, *17*(9), 2565–2580. doi:10.1007/s10530-015-0894-3
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. doi:10.1086/519795
- Reed, D. H., & Frankham, R. (2001). How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, *55*(6), 1095–103. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2001.tb00629.x/full>
- Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, *105*(3), 767–79.
- Ricciardi, A. (2006). Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. *Diversity and Distributions*, *12*(4), 425–433. doi:10.1111/j.1366-9516.2006.00262.x
- Richter, K., Haslbeck, M., & Buchner, J. (2010). The Heat Shock Response: Life on the Verge of Death. *Molecular Cell*, *40*(2), 253–266. doi:10.1016/j.molcel.2010.10.006
- Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015). Applications of next-generation sequencing to the study of biological invasions. *Current Zoology*, *61*(3), 488–504.
- Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F., & Lapègue, S. (2013). Population genomics shed light on the demographic and adaptive histories of European invasion in the

Pacific oyster, *Crassostrea gigas*. *Evolutionary Applications*, 6(7), 1064–1078.

doi:10.1111/eva.12086

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, M., Molofsky, J., With, K. A., ... Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, 32, 305–332.

Stepien, C. A., & Tumeo, M. A. (2006). Invasion Genetics of Ponto-Caspian Gobies in the Great Lakes: A “Cryptic” Species, Absence of Founder Effects, and Comparative Risk Analysis. *Biological Invasions*, 8(1), 61–78. doi:10.1007/s10530-005-0237-x

Strayer, D. L., Eviner, V. T., Jeschke, J. M., & Pace, M. L. (2006). Understanding the long-term effects of species invasions. *Trends in Ecology and Evolution*, 21(11), 645–651. doi:10.1016/j.tree.2006.07.007

Tepolt, C. K., & Palumbi, S. R. (2015). Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. *Molecular Ecology*, 24(16), 4145–4158. doi:10.1111/mec.13294

Uller, T., & Leimu, R. (2011). Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology*, 17(11), 3478–3485. doi:10.1111/j.1365-2486.2011.02509.x

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, 11(1110), 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

Vašek, M., Jůza, T., Čech, M., Kratochvíl, M., Prchalová, M., Frouzová, J., ... Kubečka, J. (2011). The occurrence of non-native tubenose goby *Proterorhinus semilunaris* in the pelagic 0+ year fish assemblage of a central European reservoir. *Journal of Fish Biology*, 78(3), 953–961. doi:10.1111/j.1095-8649.2011.02901.x

Whitney, K. D., & Gabler, C. A. (2008). Rapid evolution in introduced species, “invasive traits” and recipient communities: challenges for predicting invasive potential. *Diversity and Distributions*, 14(4), 569–580. doi:10.1111/j.1472-4642.2008.00473.x

Xin, S. (2016). *Comparison of physiological performance characteristics of two Great Lakes invasive fish species: Round Goby (Neogobius melanostomus) and Tubenose Goby (Proterorhinus semilunaris)*. University of Windsor.

You, F. M., Huo, N., Gu, Y., Luo, M.-C., Ma, Y., Hane, D., ... Anderson, O. D. (2008). BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, 9(1), 253. doi:10.1186/1471-2105-9-253

Zenni, R. D., & Nuñez, M. A. (2013). The elephant in the room: The role of failed invasions in understanding invasion biology. *Oikos*, 122(6), 801–815. doi:10.1111/j.1600-0706.2012.00254.x

#### Data Accessibility:

The raw sequencing data used to characterize functional SNP markers is available on GenBank under the project accession numbers: SRP075124 and SRP075141. The genotype files for both SNP and microsatellite markers and the scripts used to analyze them are available on the Dryad Digital Repository: doi:10.5061/dryad.mb49d60.

#### Author Contributions:

KWW, DDH and ATF conceived of the study, KWW and HPW designed the sampling strategy and collected the samples. KWW conducted the data analysis and wrote the first draft of the manuscript and all authors contributed to revisions of the manuscript.



Table 1: Summary statistics for characterization of single nucleotide polymorphism (SNP) in transcriptome data from nine individuals each of round goby (RG) and tubenose goby (TNG). Open reading frames for transcripts determined by GeneMarkS-T v5.1 (Besemer et al., 2001) and variant types for transcripts with a valid open reading frame identified by SnpEff v4.2 (Cingolani et al. 2012).

Statistic	RG	TNG
Number of transcripts	26 215	23 648
Total bp of transcriptome	49 102 157	50 184 832
Number of SNPs characterized	46 092	28 217
SNP with valid start codon	39 408	23 591
<i>Variant type</i>		
5' UTR variant	2 813	2 180
Missense variant	3 806	2 990
Synonymous variant	10 660	4 963
3' UTR variant	10 435	7 086
Intergenic region	11 660	6 342
other	34	30

Table 2: Sample sizes, summary statistics, and population site codes for round goby and tubenose goby collected throughout their invaded North American range and from the putative source of the invasion in Ukraine. N= sample size, Ar = allelic richness and Ho = observed heterozygosity for microsatellite and single nucleotide polymorphism (SNP) genotyping. AS indicates the proportion of SNP loci with significant allele frequency shifts from the Ukrainian reference populations for each species.

Site	Population Code	Microsatellites			SNPs		
		N	Ar	Ho	N	AS	Ho
<i>Round Goby</i>							
Duluth, MN	DU	32	7.0	0.56	30	0.14	0.34
Collingwood	CO	50	6.7	0.60	28	0.31	0.34
Detroit River	DR	50	6.4	0.46	11	0.00	0.32
Nanticoke	NA	50	7.0	0.60	30	0.22	0.32
Hamilton Harbour	HH	50	6.4	0.53	16	0.02	0.35
Lake Seymour	TS	45	6.8	0.49	11	0.02	0.31
Kherson, Ukraine	KH	35	7.8	0.49	27	-	0.36
<i>Tubenose Goby</i>							
Thunder Bay	TB	47	4.2	0.47	28	0.53	0.33
Mitchell's Bay	MB	25	4.5	0.56	18	0.29	0.43
Kherson, Ukraine	KH	46	8.8	0.59	16	-	0.30

Table 3: Details of single nucleotide polymorphisms (SNPs) showing elevated divergence compared with the demographic patterns of population structure based on microsatellites as detected using the  $T_{F-LK}$  of Bonhomme et al. (2010). Nucl. Variant = nucleotide allele variants, Prot. Variant = amino acid variants following standard IUPAC single letter codes for amino acids. Bold indicates statistical significance following FDR multiple test correction.

	Nucl. Variant	Prot. Variant	Gene Description
<i>Round goby</i>			
<b>SNP_1</b>	<b>T/A</b>	<b>S/T</b>	<b>Growth arrest-specific protein 1</b>
SNP_10	T/G	S/A	DNA repair protein XRCC2
<b>SNP_14</b>	<b>T/C</b>	<b>N/N</b>	<b>Heat shock protein HSP 90-beta</b>
<b>SNP_15</b>	<b>A/G</b>	<b>N/D</b>	<b>Superoxide dismutase Cu-Zn 2</b>
SNP_20	C/T	P/L	Growth hormone receptor
<b>SNP_22</b>	<b>G/A</b>	<b>G/S</b>	<b>Heat shock factor-binding protein 1</b>
SNP_39	T/C	F/F	Glucose-6-phosphatase
SNP_62	G/T	L/L	Heat shock cognate 71 kDa protein
SNP_75	C/T	Y/Y	Actin cytoplasmic 1
SNP_84	G/A	G/E	Group XIIB secretory phospholipase A2-like protein
SNP_105	T/C	I/I	Methyltransferase-like protein 16
SNP_106	C/A	G/K	Caspase-8
<b>SNP_109</b>	<b>T/A</b>	<b>T/T</b>	<b>Serine/threonine-protein phosphatase 4 regulatory subunit 3</b>
<i>Tubenose goby</i>			
SNP_45	G/A	A/T	Deleted in malignant brain tumors 1 protein
SNP_60	C/A	A/A	Membrane-associated progesterone receptor component 2
SNP_80	G/C	R/S	Aryl hydrocarbon receptor

Figure captions:

Figure 1: Map of the North American sampling sites for invasive populations of round goby (RG; grey stars) and tubenose goby (TNG; white stars). See Table 2 for description of sample site codes. The site of initial introduction for both species is the St. Clair River located at the southern tip of Lake Huron indicated by the solid arrow.

Figure 2: Observed heterozygosity (mean  $\pm$  95 CI) for single nucleotide polymorphism (squares) and microsatellite (circles) markers and microsatellite allelic richness in populations of round goby (filled symbols) and tubenose goby (open symbols). Native range populations for both species (KH) are indicated with a shaded box. Confidence interval estimates from linear mixed-effect models implemented in R with a random effect for loci and a fixed effect for population. Statistically significant differences among populations are indicated by lettering and are present for heterozygosity of SNP markers in Mitchell's Bay (MB) tubenose goby population that shows higher diversity than either other population (Tukey Test: MB-KH  $p < 0.001$ , MB-TB  $p = 0.006$ ) and for microsatellite allelic richness of both invaded range populations of tubenose goby relative to the native range (Tukey Test: MB-KH  $p = 0.001$ , TB-KH  $p < 0.001$ ).

Figure 3: Discriminant analysis of principal components barplots of the group membership probabilities for round goby individuals based on microsatellite (A), all single nucleotide polymorphism markers (B) and only the putatively neutral SNPs (C). Results highlight the different patterns of genetic group memberships assigned to individuals by the different classes of markers. Population-level differences between the individuals assigned to each genetic groups based on microsatellites compared with SNPs were statistically significant for three round goby populations (NA, HH, TS; Fisher's Exact Test:  $p < 0.001$ ) but not the others.

Figure 4: Discriminant analysis of principal components barplots of the group membership probabilities for tubenose goby individuals based on microsatellite (A), all single nucleotide polymorphism markers (B) and only the putatively neutral SNPs (C). Results highlight the similar patterns of genetic group memberships assigned to individuals by the different classes of markers in contrast to the results for round goby (Figure 3).

Figure 5:  $F_{ST}$  outlier tests based on the  $T_{F-LK}$  statistic of Bonhomme et al. (2010). Labeled single nucleotide polymorphisms (SNPs) indicate significance based on the  $\chi^2$  distribution for round goby (A) and tubenose goby (B). Stars indicate significance following false discovery rate (Benjamini & Hochberg, 1995) correction for multiple tests. The solid and dashed lines represent smoothed splines of the 99% and 95% intervals of 10 000 permutations of the data simulated under the null model and the dotted line indicates the median.





