Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity

Janna R. Willoughby1,2 | Avril M. Harder1 | Jacob A. Tennessen3 | Kim T. Scribner4,5 | Mark R. Christie1,2

1Department of Biological Sciences, Purdue University, West Lafayette, Indiana
2Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana
3Department of Integrative Biology, Oregon State University, Corvallis, Oregon
4Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan
5Department of Integrative Biology, Michigan State University, East Lansing, Michigan

Correspondence
Janna R. Willoughby, Department of Biological Sciences, Purdue University, West Lafayette, IN.
Email: willoughby@purdue.edu

Abstract
Introduced species often colonize regions that have vastly different ecological and environmental conditions than those found in their native range. As such, species introductions can provide a deeper understanding into the process of adaptive evolution. In the 1880s, steelhead trout (Oncorhynchus mykiss) from California were introduced into Lake Michigan (Laurentian Great Lakes, North America) where they established naturally reproducing populations. In their native range, steelhead hatch in rivers, migrate to the ocean and return to freshwater to spawn. Steelhead in Lake Michigan continue to swim up rivers to spawn, but now treat the freshwater environment of the Great Lakes as a surrogate ocean. To examine the effects of this introduction, we sequenced the genomes of 264 fish. By comparing steelhead from Lake Michigan to steelhead from their ancestral range, we determined that the introduction led to consistent reductions in genetic diversity across all 29 chromosomes. Despite this reduction in genetic diversity, three chromosomal regions were associated with rapid genetic adaptation to the novel environment. The first region contained functional changes to ceramide kinase, which likely altered metabolic and wound-healing rates in Lake Michigan steelhead. The second and third regions encoded carbonic anhydrases and a solute carrier protein, both of which are critical for osmoregulation, and demonstrate how steelhead physiologically adapted to freshwater. Furthermore, the contemporary release of diverse hatchery strains into the lake increased genetic diversity but reduced the signature of genetic adaptation. This study illustrates that species can rapidly adapt to novel environments despite genome-wide reductions in genetic diversity.

KEYWORDS
bottleneck, contemporary evolution, founder effect, hatcheries, rapid evolution, species introductions

1 | INTRODUCTION
Introduced species can successfully colonize habitats where conditions are vastly different from their native range (i.e., novel environments; Lee, 2002; Prentis, Wilson, Dormontt, Richardson & Lowe, 2008) and these introductions present a valuable opportunity to gain a greater understanding of adaptive evolution (Dlugosch & Parker, 2008; Kolbe et al., 2004). By identifying the specific genes and biochemical pathways that respond quickly to selection, we can begin to understand the evolutionary mechanisms that maintain and utilize adaptive genetic variation. Furthermore, introduced species are often descended from only a small handful of individuals, resulting in large
reductions in genetic diversity in the introduced populations (Dlugosch & Parker, 2008). This common finding raises the question: How can introduced species genetically adapt to novel environments despite large reductions in genetic diversity? Answers to that question have broad implications for better understanding how adaptive evolution can affect colonization, range dynamics, speciation and conservation (e.g., can species with depressed population sizes genetically adapt to rapidly changing environments; Carlson, Cunningham & Westley, 2014). To explore the possibility of rapid genetic adaptation in an introduced species, we examined how an ocean-migrating fish was able to adapt to the entirely freshwater environment of the Laurentian Great Lakes (North America).

Which genes, and thus traits, can respond to strong directional selection in a short period of time depends upon the initial genetic variation present in the progenitor population (Vigouroux et al., 2002). Genes that are fixed in the ancestral population are less likely to respond to selection over short timescales because they require specific mutations or chromosomal rearrangements (Orr, 1998; Yeaman, 2013). Thus, soft sweeps, which act on standing genetic variation, are more likely to occur than hard sweeps over ecologically relevant timescales, a result borne out by recent experimental studies (McCoy & Akey, 2017; Wilson, Pennings & Petrov, 2017). Correspondingly, we hypothesized that genes involved in an adaptive response would be found in regions of high genetic diversity in the ancestral population. We further hypothesized that pathways involved in osmoregulation would be under strong directional selection in the entirely freshwater environment (Vellotta et al., 2017).

Steelhead trout (anadromous Oncorhynchus mykiss) is an ecologically, economically and culturally important fish that are native to the northern Pacific and surrounding regions. Both steelhead and rainbow trout, a life history variant that spend their entire lives in freshwater streams before migrating out to the ocean to forage and grow. After spending 1–2 years in the ocean, steelhead return to their natal streams to spawn (Quinn, 2005). By contrast, steelhead from Lake Michigan continue to spawn in streams and rivers, but now use the entire freshwater environment of the Great Lakes as a surrogate ocean (Figure 1). Furthermore, there is no known interbreeding between resident and anadromous life history variants in Lake Michigan (i.e., steelhead and rainbow trout are not, to our knowledge, sympatric; Seelbach & Whelan, 1988), suggesting that steelhead have adapted to the novel freshwater environment. Apart from salinity, there are many additional differences between the introduced and native habitats including water temperature, stream characteristics, and community composition (Mistak, Hayes & Bremigan, 2003; Moyle & Light, 1996; Nobriga, Feyrer, Baxter & Chotkowski, 2005). Although the Lake Michigan populations were self-sustaining, hatchery stocking began anew in the 1950s with the release of small fingerlings that rarely survived to the smolt stage (0.01% survival; Seelbach & Whelan, 1988) and likely never survived to maturity. Beginning in the mid-1980s, larger yearlings of diverse ancestry were released into Lake Michigan (Bartron & Scribner, 2004), permitting survival to the smolt stage of these larger, hatchery-released individuals (90% survival; Seelbach & Whelan, 1988) and possible introgression with the original, California-derived strain.

To identify how steelhead adapted to the Great Lakes, we used genome-wide sequencing to answer two specific questions: (a) How did the introduction event impact the patterns of genome-wide diversity in Lake Michigan steelhead, and (b) how did steelhead genetically adapt to the novel, freshwater environment of Lake Michigan over a nearly 100-year period (approximately 25 generations)? To address these questions, we sampled steelhead from the ancestral range (possibly the original progenitor populations; MacCrimmon & Gots, 1972) see Section 2) and the introduced range both before and after effective hatchery supplementation occurred in Lake Michigan (Figure 1a). We then used a pooled-sequencing approach to identify genome-wide patterns of genetic diversity and genes that responded to selection in the novel environment.

2 | MATERIALS AND METHODS

We sampled steelhead from Lake Michigan (1983, N = 96; and 1998, N = 120; Figure 1) and from the Californian-Pacific coast (California 2005, N = 24; 2014, N = 24; Figure 1), the ancestral range of Lake Michigan steelhead as inferred from written records (MacCrimmon & Gots, 1972). Because the samples collected from California have not been impacted by hatchery introgression (Abadía-Cardoso et al., 2016; Leitwein, Garza & Pearse, 2017), we treated these samples as representative of the ancestral population. Using a pooled-sequencing approach to obtain genome-wide data (Axelsson et al., 2013), we sequenced a total of 264 individuals, in 10 separate libraries, split evenly across 9 lanes of an Illumina HiSeq2500. Our libraries were constructed as follows: (a) 24 individuals from Clear Creek, California, sampled in 2005; (b) 24 individuals from Redwood Creek, California, sampled in 2014; (c,d) 48 individuals from the Bet-sie River sampled in 1983; (e,f) 48 individuals from the Little Manis-tee River sampled in 1983; (g) 24 individuals from the Betsie River in 1998; (h,i) 48 individuals from the Black River in 1998; and (j) 48 individuals from the Platte River sampled in 1998 (Figure 1; Supporting Information Table S1). We removed adapter sequences using Trimmomatic (Bolger, Lohse & Usadel, 2014) and replaced base pairs with a quality score <4 with an “N” prior to further analyses.

To aid in downstream filtering and annotation, we used two reference genomes: the well-annotated Atlantic salmon genome (S. salar GenBank accession GCA_000233375.4; Lien et al., 2016) and the newly updated O. mykiss assembly (GenBank accession GCA_002163495.1). For both genome assemblies, we mapped all reads using BWA-mem (Li & Durbin, 2009), allowing a 4% mismatch for each read, and removed all reads that mapped multiple times.
FIGURE 1 Location of steelhead sampling sites (a) and comparison of environmental conditions between native and introduced populations (b). Numbers in parentheses represent number of individuals sequenced and collection year at each site. Sampling from the native range included only anadromous individuals that are born in freshwater streams but then migrate to the Pacific Ocean as juveniles. Individuals sampled in Michigan are also born in freshwater streams but now use the Great Lakes as a surrogate ocean. Because the Pacific Ocean and Lake Michigan differ in salinity, the sampled steelhead populations were subject to vastly different salt and ion-transport processes in their respective environments (b) [Colour figure can be viewed at wileyonlinelibrary.com]
from initial downstream analyses (Supporting Information Figure S1; but see Supporting Information Figure S2). Next, we combined the mapped reads across sampling sites into three groups: samples collected from California, Lake Michigan 1983, and Lake Michigan 1998. For each group, mapped reads were combined proportionally to the amount of DNA sequenced so that all locations were represented equally in the combined groups. We verified low divergence between sampling locations, to validate our decision to combine reads across sampling sites, using a principal coordinate analysis (average $F_{ST} = 0.016$ across all sites; Supporting Information Figure S3).

Using the combined files, we used POLIMAPS (modified for pool-seq data: https://github.com/jacobtennessen/GOPOPS; Tennessen, Govindaraju, Ashman & Liston, 2014), requiring a minimum depth of $20 \times$ across all populations, to identify SNPs, estimate allele frequencies, and calculate $F_{ST}$.

We first considered the effects of the introduction of steelhead by examining the pattern of genetic diversity across the genome. We compared heterozygosity across chromosomes between the three sampled populations, using allele frequencies estimated from SNPs aligned to the O. mykiss reference genome and nonoverlapping 100 kilobase (kb) windows (Supporting Information Figure S4; Kardos et al., 2015; Rubin et al., 2010). We estimated pooled heterozygosity ($H_p$) as $H_p = 2 \Sigma \Delta_{MAU} \Sigma \Delta_{MIN}(\Sigma \Delta_{MAX}^2 + \Sigma \Delta_{MIN}^2)$, where $\Sigma \Delta_{MAX}$ and $\Sigma \Delta_{MIN}$ are the sums of the major and minor allele counts in the window, respectively (Rubin et al., 2010). Next, we estimated the mean difference in $H_p$ between populations by computing the per cent difference in $H_p$, estimated for each 100 kb window, between populations. We estimated the 95% confidence interval around this estimate using a bootstrapping approach: From the distribution of the per cent difference in $H_p$ in all windows compared between populations, we sampled, with replacement, the total number of windows and calculated the mean difference in $H_p$ 5,000 times. We then estimated the 95% confidence interval from the resulting distribution. These calculations were performed using the R package boot (Canty & Ripley, 2017). We also used a Wilcoxon signed-ranks test to compare the estimates of $H_p$ for each window between populations. Finally, we compared $H_p$, across all chromosome arms among our three sample groups.

We next examined genome-wide patterns of genetic differentiation to identify genes with allele frequency shifts consistent with selection in the novel environment (Lesca et al., 2015). Using both the Atlantic salmon and steelhead reference genomes, we identified genomic regions putatively subjected to strong selection using $F_{ST}$. To limit false outliers, we used $Z$-transformed $F_{ST}$ values and then identified regions with $Z(F_{ST})$ that were greater than 5 standard deviations from the mean (Axelsson et al., 2013; Kardos et al., 2015; Rubin et al., 2010), averaged over 100 kb windows with 50 kb steps (Kardos et al., 2015; Rubin et al., 2010). We examined regions that had high $Z(F_{ST})$ in both California vs. Lake Michigan 1983 and California vs. Lake Michigan 1998 comparisons. We further validated these outlier regions by assessing $F_{ST}$ at these loci with independent library data sets: Lake Michigan 1983 vs. Clear Creek (Pacific population, sampled in California), Lake Michigan 1983 vs. Redwood Creek (Pacific population, sampled in California); Lake Michigan 1998 vs. Clear Creek, and Lake Michigan 1998 vs. Redwood Creek (Figure 1; Supporting Information Figures S5–S7).

Finally, we used the well-annotated Atlantic salmon genome (Samy et al., 2017) to understand the function of our set of outlier loci. Using Blastn (percent identity > 80%, coverage > 80%, e-value cut-off = $10^{-6}$; Camacho et al., 2009), we identified the location of all known Atlantic salmon genomic coding segments (Samy et al., 2017) in the reference O. mykiss genome. We filtered the blast hits to include those that occurred in a single region of the Atlantic salmon genome (i.e., we eliminated two coding sequences that were from disparate regions of the Atlantic salmon genome). To identify functional changes in the candidate genes (i.e., the coding sequences found in regions associated with high $Z(F_{ST})$ in the O. mykiss genome), we examined the allele frequency changes between the California and Lake Michigan population focusing on alleles with large differences (>25%). For these analyses, we considered allele frequencies in reference to the major allele present at each SNP in the Lake Michigan population. To identify the open-reading frames (ORFs) within each outlier region, we extracted the O. mykiss reference genome sequence for all coding sequences with a 100-bp overlap on the 5’ and 3’ ends and identified the open-reading frame using the NCBI Open Reading Frame Finder (Wheeler et al., 2003), requiring all coding regions in each gene to be on the same strand. When multiple fragments with start and stop codons were identified within a single sequence, we selected the longest frame. Finally, we determined whether the SNPs we identified within each open-reading frame resulted in synonymous or nonsynonymous changes by translating the nucleotide sequences and comparing the resulting sequences of amino acids.

3 | RESULTS

We generated a total of 1,935,216,602 paired-end reads (Supporting Information Table S1) across all 10 libraries, which resulted in a mean population read depth of $43 \times$ and a total of 29,521,445 SNPs across all three populations. In all three populations, heterozygosity was lowest at the centromeres and contained a region with extremely high heterozygosity on chromosome 5 (Figure 2a). We subsequently identified this region as an inversion associated with an anadromous life history in steelhead (Pearse, Miller, Abadia-Cardoso & Garza, 2014). Allele frequencies associated with this inversion did not differ among our populations (Figure S8). When making comparisons between the California population and both the 1983 and 1998 Lake Michigan samples, we found three striking patterns (Figure 2a): First, heterozygosity in the ancestral California population was higher than heterozygosity in the Lake Michigan across all chromosomes (84% and 79% of 100 Kb windows had higher heterozygosity in California populations than 1983 or 1998 Lake Michigan populations, respectively). In fact, Lake Michigan 1983 steelhead were characterized by an average 9.49% (95% confidence interval 9.34%–9.64%; $V = 174,720,000$, $p$-value < 0.0001) reduction in genetic diversity across the entire genome and had an average of 150 additional SNPs fixed for a single allele within each 100 kb
window (Figure 2b) compared to the California population. Second, we found that for 68% of the 100 kb windows, average heterozygosity in the 1998 Lake Michigan population was higher than heterozygosity in the 1983 Lake Michigan population, resulting in a mean 4.75% increase (95% confidence interval 4.57%–4.94%; \( V = 143,480,000 \), \( p \)-value < 0.0001) in genetic diversity between 1983 and 1998 (Figure 2c). This pattern likely reflects the successful introgression of hatchery-stocked smolts, which came from diverse sources and were released into Lake Michigan beginning in 1983 (Bartron & Scribner, 2004; Seelbach, 1987). Lastly, we observed that the reduction in heterozygosity in the Lake Michigan populations relative to California depended on position along each chromosome; the difference between California and Lake Michigan 1983 heterozygosity was the smallest near the centromere and largest near the telomeres. However, the difference between California and Lake Michigan 1998 heterozygosity was largest near the centromeres and smallest near the telomeres (Figure 2d). This pattern was not driven by changes in the total number of SNPs (Supporting Information Figure S9), read depth (Supporting Information Figure S10), occurrence of paralogous loci (Supporting Information Figure S11), gene density (Supporting Information Figure S12), or number of repetitive elements (Supporting Information Figure S13). Furthermore, this pattern was found across 48 of 52 chromosome arms and was thus not driven by the accumulation of duplicated genes on specific chromosomes (Kodama, Brieuc, Devlin, Hard & Naish, 2014).

We next examined our data to identify genes with allele frequency shifts consistent with selection in the novel environment using \( Z(F_{ST}) \). Although we identified six total regions with high \( Z(F_{ST}) \), two regions were eliminated from further analyses: (a) a region on chromosome 14, which was characterized by low coverage and few SNPs, had a high degree of variability in \( F_{ST} \) in the region (Supporting Information Figure S14); (b) a region on chromosome 15 that did not contain any
known genes; and (c) a region on chromosome 1 that had high \( F_{ST} \) in California vs. Lake Michigan 1983 but not in the California vs. Lake Michigan 1998 comparison. (The region on chromosome 1 contained a voltage-dependent, calcium channel subunit gene.) We treated the remaining three regions (100–250 kb) as outliers (Figure 3). The three outliers, which were located on chromosomes 4, 8 and 28, were associated with two different functions: metabolism and osmoregulation/acid-base balancing. Inclusion of reads that could have mapped to paralogous loci (i.e., salmonids have many paralogous loci due to a genome duplication event ~125 million years ago (Allendorf & Thorgaard, 1984; Macqueen & Johnston, 2014)) changed the amplitude of the \( F_{ST} \) estimates, but not the location of high \( F_{ST} \) regions (Supporting Information Figure S2; Limborg, Larson, Seeb & Seeb, 2017). Furthermore, we did not find any evidence that the location of transposable elements (Supporting Information Figure S15) or copy number variation (Supporting Information Figure S16) played a role in the adaptation of steelhead to Lake Michigan.

The 150 kb outlier region on \( O. \) mykiss chromosome 4 contained two genes (Figure 4a): (a) GRAM domain containing 4 (GRAMD4) that, when activated by p73, interacts with mitochondria to initiate caspase activation and apoptosis (John, Alla, Meier & Pü, 2010); and (b) ceramide kinase (CERK), an enzyme that catalyses the phosphorylation of ceramide to ceramide-1-phosphate (C-1-P), which facilitates DNA synthesis and cell proliferation (Figure 4b; Gomez-Muñoz et al., 1995; Gomez-Muñoz, Frago, Alvarez & Varela-Nieto, 1997). In GRAMD4, a total of 34 SNPs had large allele frequency differences (Supporting Information Figure S17). By contrast, we found 510 SNPs with large allele frequency changes in CERK, including 5 nonsynonymous SNPs (range of \( F_{ST} \) 0.49–0.83; Figure 4). In general, these 5 nonsynonymous SNPs were characterized by low frequencies in the California population (mean allele frequency = 0.28), approaching fixation in 1983 Lake Michigan (mean allele frequency = 0.74), and a dampening of the...
frequencies in 1998 Lake Michigan (mean allele frequency = 0.59), indicating a soft-sweep occurred in CERK (Figure 4b). We interpret these patterns of allele frequency changes as evidence of positive selection favouring a previously rare allele in this region. The role of CERK on C-1-P induced cell proliferation (Figure 4c) suggests metabolic regulation shifts occurred in the rapidly adapting Lake Michigan population. These shifts presumably resulted from divergent metabolic and/or wound-healing needs in the California and Lake Michigan populations (Rollins, Richardson & Shine, 2015).

The remaining two outlier regions were found on chromosomes 8 and 28 and both included genes associated with osmoregulation and the physiological balancing of acid–base conditions (Figure 5). The outlier region on chromosome 8 contained 2 carbonic anhydrase genes (Figure 5a). In mammals, carbonic anhydrases are expressed in the kidney and are important in acid–base balance (Purkerson & Schwartz, 2007). In teleosts, cytoplasmic carbonic anhydrase isoforms are associated with gill tissue and facilitate CO2 hydration to H+ and HCO3\(^{-}\). These ions are then used as counter-ions for Na\(^{+}\) and Cl\(^{-}\) uptake in freshwater fish, thus contributing to regulation of osmotic pressure (Figure 5c; Georgalis, Perry & Gilmour, 2006). The outlier region on chromosome 28 contained a single gene (Figure 5b), solute carrier family 26 member (SLC26) 6-like. In humans, solute carrier family genes are multifunctional and are often involved in acid–base balancing via movement of monovalent and divalent anions (e.g., Cl, O2). In teleosts specifically, SLC26 is a multifunctional anion exchanger family of genes that act to move a large diversity of anions (Romano, Barca, Storelli & Verri, 2014). We identified, specifically, SLC26 member 6-like gene, which is involved in Cl\(^{-}\) uptake in teleosts and is often detected in base secreting, mitochondria-rich cells in the gill (Figure 5c; Bayaa et al., 2009; Georgalis et al., 2006). One carbonic anhydrase gene (gene 1 Figure 5a) and the solute carrier gene (gene 3 Figure 5b) contained 7 and 1 nonsynonymous SNPs, respectively (Supporting Information Figure S17). Thus, the regions with high \(F_{ST}\) on chromosomes 8 and 28 reflect selection on two different acid–base balancing pathways with specific emphasis on ion uptake associated with freshwater systems.

4 | DISCUSSION

We found evidence of rapid genetic adaptation to a novel environment despite a reduction in genome-wide levels of genetic diversity. Although much of this adaptation likely occurred between the introduction (1880s) and our first sampling point in the new environment (Lake Michigan 1983), much of the variation was maintained through the subsequent sampling period (Lake Michigan 1998). In total, 9.49% of the genetic variation present in the ancestral steelhead populations was lost in the introduced Lake Michigan population resulting in an average of 150 fewer segregating sites per 100 kb window, translating to 2.9 million fewer segregating sites across the entire genome. Additionally, the loss of genetic diversity we documented is likely an underestimate due to the contemporary population declines.

**FIGURE 5** \(F_{ST}\) estimates for all SNPs located within the outlier regions on chromosome 8 (a) and chromosome 28 (b), denoted by the extent of the black bracket, for both California vs. Lake Michigan 1983 and California vs. Lake Michigan 1998. Highlighted regions display the extent of the three genes located within the outlier regions: (1–2) carbonic anhydrases; (3) solute carrier family 26 member 6 (SLC26). Functionally, carbonic anhydrase hydrolyses CO2 resulting in H\(^{+}\) and HCO3\(^{-}\), which are then used to balance somatic pH (c). Similarly, SLC26 facilitates Cl\(^{-}\) uptake from the environment and releases cellular HCO3\(^{-}\) into the environment [Colour figure can be viewed at wileyonlinelibrary.com]
that have occurred in steelhead populations found in northern California (Mills, McEwan & Jennings, 1997). That this reduction in genetic diversity was found across the entire genome is consistent with theoretical predictions of population bottlenecks and founder effects (Frankham, 1997; Leberg, 1992). However, the recovery of genetic diversity after introduction into Lake Michigan differed across chromosomes (i.e., $H_E$ in Lake Michigan 1998 recovered to California levels near the telomeres but not near the centromeres; Figure 2d). We suggest this is a result of hatchery introgression; hatchery-associated alleles may have been preferentially incorporated into the wild population near the telomeres, if selection combined with low recombination near the centromeres inhibited introgression. Thus, the present-day distribution of HP in Lake Michigan steelhead likely reflects the evolutionary interplay among gene flow (here introgression from hatchery strains), recombination, and selection. Furthermore, the genes associated with adaptation in the novel environment were not located near the centromeres (Supporting Information Figure S18) but were rather located in regions with high genetic diversity in the ancestral population (Figure 2b–d). Because genes that are fixed in an ancestral population will require mutations or rearrangements before a response to selection can occur (Orr, 1998; Yeaman, 2013), our results suggest that certain genes, depending upon their location in the genome and thus the variation available to select upon, may be more likely to be involved in adaptation to novel environments.

By examining genes in regions with high genetic differentiation between the California and Lake Michigan population, we found evidence that altered metabolic processes were associated with the genetic adaptation of steelhead to the Great Lakes. The selection on CERK and the associated modulation of C-1-P induced cell proliferation suggests steelhead metabolism played an important role in adaptation to the freshwater environment (Figure 4). This adaptation may have allowed steelhead to take advantage of alternative prey or allocate additional resources to activity in the Lake Michigan habitat (Axelsson et al., 2013; Babbitt, Warner, Pedrigo, Wall & Wray, 2011; Gomez-Muñoz et al., 1997). Intriguingly, C-1-P greatly increases the DNA synthesis of fibroblasts, cells that play a key role in wound-healing pathways (Gomez-Muñoz et al., 1995, 1997; Singer & Clark, 1999). Due to osmotic pressure differences, an open wound in freshwater results in more severe cell lysis compared to a wound in saltwater and these osmotic differences can influence the mechanism by which wounds heal (Enyedi & Niethammer, 2015; Gault, Enyedi & Niethammer, 2014). In the Great Lakes specifically, introduced, parasitic sea lamprey occur at high densities (Hansen et al., 2016), and lamprey attacks result in large wounds resulting in mortality rates of ~40% of parasitized O. mykiss individuals (Swink & Hanson, 1989). Most lamprey species that exist in the native steelhead range are nonparasitic (Potter, Gill, Renaud & Haoucher, 2015) and parasitic Pacific lamprey occur at much lower densities than Lake Michigan sea lamprey and do not rely heavily on salmon as hosts (Beamish, 1980). Thus, we speculate that C-1-P mediated wound healing may be a response to the strong selective pressure imposed by introduced parasitic sea lamprey found at high abundances in Lake Michigan.

We also found evidence of genetic adaptation in osmoregulatory and acid–base balancing pathways on two independent chromosomes. Freshwater fish actively uptake ions from the environment to compensate for salts lost via passive diffusion whereas saltwater fish actively excrete ions to compensate for the passive diffusion of salts into their bodies (Figure 1; Evans, 2010). Anadromous salmonids can actively switch between these two processes (Bystriansky, Richards, Schulte & Ballantyne, 2006; McCormick, 1994), but this flexibility is energetically costly (Morgan & Iwama, 1991) and maladaptive for steelhead residing in an entirely freshwater environment. Correspondingly, we found evidence of a large response to selection in Lake Michigan steelhead for two independent sets of genes, carbonic anhydrases (chromosome 8) and SLC26 (chromosome 28). Previous studies in stickleback (Gasterosteus aculeatus) have also identified SLC26 family genes as being important in adaptation to freshwater systems (Hohenlohe et al., 2010). Both carbonic anhydrases and SLC26 proteins facilitate uptake of ions from the aquatic environment (Figure 5; Georgalis et al., 2006; Purberson & Schwartz, 2007). Additionally, blood pH changes have been observed after salinity alterations (Whiteley, Scott, Breeze & McCann, 2001), a result of a mismatch between carbonic anhydrase activity and ions needed to maintain pH. Thus, the increased ion uptake in Lake Michigan steelhead not only maintains osmotic pressure, but also contributes to effective regulation of body pH in freshwater (Figure 5c). The documentation of large shifts in allele frequencies for genes involved in both metabolic and osmoregulatory pathways provides the first glimpse of how and why introduced Pacific salmonids have been successful in the Great Lakes ecosystem. Furthermore, comparison of orthologous genes in other salmonids introduced into the Great Lakes will provide insight into both the repeatability of evolution and the role of standing genetic variation in rapid adaptation.

It is clear that the stocking efforts that released genetically diverse steelhead strains beginning in 1983 increased genome-wide levels of genetic diversity in Lake Michigan steelhead. Even though the 95% confidence intervals of $H_E$ in Lake Michigan steelhead collected in 1983 overlapped with Lake Michigan steelhead collected in 1998, the point estimates were higher for 1998 steelhead across the genome (68% of 100 kb windows had higher genetic diversity in 1998 steelhead). These rapid, large-scale genome-wide increases in genetic diversity over a 15-year period cannot be explained by mutation or other natural phenomena (Cooper et al., 2005; Sawyer & Hartl, 1992). Furthermore, the increases in genetic diversity did not occur equally across chromosomes; genetic diversity was gained faster towards the telomeres, perhaps reflecting contemporary recombination rates (Roesti, Moser & Berner, 2013; Wong et al., 2010) or decreased purifying selection in telomeric regions (Ellegren et al., 2012; Langley et al., 2012).

While steelhead stocking efforts increased genetic diversity, they have also led to the erosion of the recently derived local adaptation; alleles at particular SNPs that nearly reached fixation in response to selection in the 1983 Lake Michigan population were driven back to more intermediate frequencies in the 1998 Lake Michigan population (Figure 4b). The conservation and management of introduced salmonids in the Great Lakes is complex, but if there is a goal to maintain natural reproduction in these systems, then the release of
nonlocally adapted steelhead and subsequent introgression may result in reduced reproductive success (Fraser, Weir, Bernatchez, Hansen & Taylor, 2011; Hohenlohe et al., 2013; Naish, Seamons, Dauer, Hauser & Quinn, 2013; Rius & Darling, 2014). More broadly, the release of nonlocally adapted individuals for conservation or restoration purposes may serve to increase genome-wide levels of genetic diversity but will likely drive allele frequencies (and thus phenotypes) towards lower fitness at a local or regional scale.

5 | CONCLUSION

By comparing Lake Michigan steelhead to their ancestral population, we were able to uncover genomic patterns of rapid genetic adaptation to the Great Lakes ecosystem despite a reduction in genome-wide genetic diversity. Whether selection occurred immediately upon release or gradually over a number of generations is challenging to disentangle and the mechanisms are not necessarily mutually exclusive (i.e., selection could occur immediately upon introduction and continue afterwards). Given that osmoregulatory homeostasis is critical to survival, we speculate that selection on the carbonic anhydrase and solute carrier genes occurred during or soon after steelhead were introduced into Lake Michigan. Conversely, if the changes to CERK are in fact due to increased wounding by sea lamprey, this response to selection could only have occurred since 1936, when sea lamprey were first recorded in Lake Michigan (Hansen et al., 2016). That genetic adaptation can still occur despite genome-wide reductions in genetic diversity has substantial conservation and management implications; imperiled species with small population sizes may still be able to adapt to changing environmental conditions. Using genomic approaches to better understand which genes, populations and species can rapidly respond to novel and often anthropogenically induced selective forces represents a key component of future conservation research.

ACKNOWLEDGEMENTS

We thank D. Pearse for providing us with the California samples and for constructive feedback on this project and S. Lien for providing us with centromere locations in O. mykiss chromosomes. We also thank the Purdue Genomics Core for their sequencing efforts and M. Blouin, A. Brüniche-Olsen, M. Hale, A. Martinez, S. Mattoo, M. Sparks, M. Sundaram and the Michigan State University Fish Discussion Group for constructive comments and discussion. J.R.W was partially supported by a fellowship through the Purdue Forestry and Natural Resources postdoctoral programme. This research was funded by support to M.R.C. from the Purdue Department of Forestry and Natural Resources and the Department of Biological Sciences.

DATA ACCESSIBILITY

Code and scripts are available at https://github.rcac.purdue.edu/MarkChristieGroup/steelhead-poolseq. Aligned steelhead reads are available via NCBI GenBank, Accession no. SRP126365.

AUTHOR CONTRIBUTIONS

Research was designed by J.R.W. and M.R.C. All authors analysed the data and contributed to the writing and editing of the manuscript.

ORCID

Janna R. Willoughby http://orcid.org/0000-0002-0176-1878
Avril M. Harder http://orcid.org/0000-0001-7371-4002
Mark R. Christie http://orcid.org/0000-0001-7285-5364

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.