

# Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes

Amanda S. Ackiss<sup>1</sup>  | Wesley A. Larson<sup>2</sup> | Wendylee Stott<sup>3</sup>

<sup>1</sup>Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, Wisconsin

<sup>2</sup>U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, Wisconsin

<sup>3</sup>U.S. Geological Survey, Great Lakes Science Center, Ann Arbor, Michigan

## Correspondence

Amanda S. Ackiss, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve Street, Stevens Point, WI 54481, USA.

Email: aackiss@uwsp.edu

## Funding information

Great Lakes Fish and Wildlife Restoration Act; Great Lakes Restoration Initiative; Great Lakes Fishery Trust

## Abstract

Effective resource management depends on our ability to partition diversity into biologically meaningful units. Recent evolutionary divergence, however, can often lead to ambiguity in morphological and genetic differentiation, complicating the delineation of valid conservation units. Such is the case with the “coregonine problem,” where recent postglacial radiations of coregonines into lacustrine habitats resulted in the evolution of numerous species flocks, often with ambiguous taxonomy. The application of genomics methods is beginning to shed light on this problem and the evolutionary mechanisms underlying divergence in these ecologically and economically important fishes. Here, we used restriction site-associated DNA (RAD) sequencing to examine genetic diversity and differentiation among sympatric forms in the *Coregonus artedii* complex in the Apostle Islands of Lake Superior, the largest lake in the Laurentian Great Lakes. Using 29,068 SNPs, we were able to clearly distinguish among the three most common forms for the first time, as well as identify putative hybrids and potentially misidentified specimens. Population assignment rates for these forms using our RAD data were 93%–100% with the only mis-assignments arising from putative hybrids, an improvement from 62% to 77% using microsatellites. Estimates of pairwise differentiation ( $F_{ST}$ : 0.045–0.056) were large given the detection of hybrids, suggesting that reduced fitness of hybrid individuals may be a potential mechanism for the maintenance of differentiation. We also used a newly built *C. artedii* linkage map to look for islands of genetic divergence among forms and found widespread differentiation across the genome, a pattern indicative of long-term drift, suggesting that these forms have been reproductively isolated for a substantial amount of time. The results of this study provide valuable information that can be applied to develop well-informed management strategies and stress the importance of re-evaluating conservation units with genomic tools to ensure they accurately reflect species diversity.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

## KEYWORDS

adaptive divergence, conservation units, coregonines, genomic islands of divergence, hybridization, population genomics, RAD sequencing, species complex

## 1 | INTRODUCTION

Defining conservation units is one of the most fundamental yet challenging aspects of resource management (Coates, Byrne, & Moritz, 2018). Partitioning species into units with substantial reproductive isolation provides managers with the ability to monitor and regulate independently evolving groups that may respond differently to harvest, disease, habitat alteration, or climate change (Allendorf & Luikart, 2007; Ryder, 1986). Over the past several decades, advancements in genetic analysis have provided scientists with powerful tools to estimate the amount of gene flow between species or populations to inform the creation of conservation units (Olsen et al., 2014; Palsbøll, Bérubé, & Allendorf, 2007; Palsbøll, Peery, & Bérubé, 2010; Schwartz, Luikart, & Waples, 2006). With the arrival of the genomics era, the power and accuracy to discern levels of reproductive isolation, inbreeding, and effective population size has vastly improved (Allendorf, Hohenlohe, & Luikart, 2010), and tools such as genome scans have revolutionized our ability to identify and understand adaptive genetic variation (Funk, McKay, Hohenlohe, & Allendorf, 2012; Waples & Lindley, 2018). Despite these advancements, delineating discrete conservation units can still be problematic. For example, taxonomic uncertainty can lead to confusion regarding species boundaries (Bickford et al., 2007; Hey, Waples, Arnold, Butlin, & Harrison, 2003), and observed phenotypic, spatial, temporal, or behavioral differences can be opposed by apparent genetic panmixia (Als et al., 2011; Hoey & Pinsky, 2018; Palm, Dannewitz, Prestegard, & Wickström, 2009).

Perhaps no other group embodies the challenges of defining conservation units better than the coregonines. A subfamily of the Salmonidae, coregonines are comprised of three genera of freshwater and anadromous fishes distributed throughout cold water habitat in North America, Europe, and Asia. The most speciose genus, *Coregonus*, includes the ciscoes and whitefishes, which exhibit an extreme array of phenotype variability that is attributed to recent adaptive radiation into lacustrine habitat following glacial retreat during the Pleistocene epoch (Behnke, 1972; Schluter, 1996; Svärdsön, 1979). Often, distinct phenotypes can be found in sympatry and allopatry, which leads to difficulty in distinguishing a single, monophyletic origin of forms from parallel ecological speciation in individual lakes. Several coregonines exhibit sympatric dwarf and normal forms, including members of the European whitefish *C. lavaretus* species complex, North American whitefish *C. clupeaformis*, cisco *C. artedi*, and least cisco *C. sardinella* (Huitfeldt-Kaas, 1918; Mann & McCart, 1981; Shields, Guise, & Underhill, 1990; Vuorinen, Bodaly, Reist, Bernatchez, & Dodson, 1993). Empirical evidence of hybridization and introgression (Garside & Christie, 1962; Kahilainen et al., 2011; Lu, Basley, & Bernatchez, 2001) raises the question of how to manage forms when reproductive isolation is incomplete and

has led some to suggest that the broad phenotypic variation observed in coregonines is a result of reticulate evolution (Svärdsön, 1998; Turgeon & Bernatchez, 2003). This breadth of taxonomic ambiguity in the coregonines was first termed “the coregonid problem” by Svärdsön (1949) and persists today as the more accurate “coregonine problem” (Eshenroder et al., 2016; Mee, Bernatchez, Reist, Rogers, & Taylor, 2015).

In recent decades, phenotypic data have been combined with genetic analysis in an attempt to untangle the complicated relationships among coregonines. Taxonomic units in coregonine systematics have traditionally relied on morphological characteristics such as head and body shape, morphometrics, and meristics such as gill raker counts (Himberg, 1970; Koelz, 1929; Svärdsön, 1979). Phylogeographic analyses with allozymes, restriction fragment length polymorphisms (RFLPs), mitochondrial DNA (mtDNA), and microsatellites have helped resolve evolutionary relationships among species and forms (Bernatchez & Dodson, 1994; Østbye, Bernatchez, Naesje, Himberg, & Hindar, 2005), and this growing body of research indicates that relying solely on phenotypic traits can be problematic for determining phylogenetic relationships when environmental plasticity occurs (Muir et al., 2013; Todd, 1998; Todd, Smith, & Cable, 1981). Amplified fragment length polymorphisms (AFLPs), mtDNA, and microsatellite analyses all show support for broad colonization of monophyletic lineages followed by the parallel ecological speciation of sympatric species of European whitefish *C. lavaretus* (Hudson, 2011; Hudson, Lundsgaard-Hansen, Lucek, Vonlanthen, & Seehausen, 2017; Præbel et al., 2013). Recently, genomics methods have been applied to the coregonine problem in lake whitefish *C. clupeaformis*, and restriction site-associated DNA (RAD) sequencing, quantitative trait loci (QTL) analysis, and genome scans have provided valuable insight into the evolutionary mechanisms of speciation-with-gene-flow in dwarf and normal forms (Gagnaire, Normandeau, Pavey, & Bernatchez, 2013; Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; Laporte et al., 2015; Rogers & Bernatchez, 2007; Rougeux, Bernatchez, & Gagnaire, 2017).

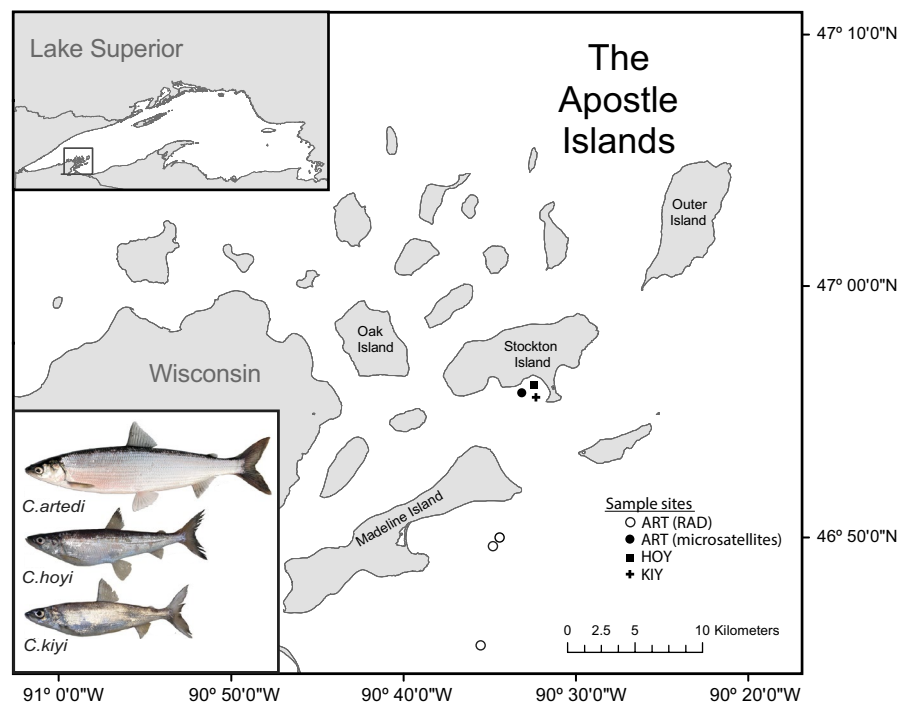
The most extensive regional adaptive radiation within the *Coregonus* genus in North America occurred in the Laurentian Great Lakes, but the detection of genetic differentiation or reproductive isolation among Great Lakes forms has been mostly unsuccessful. Rapid diversification of one or more colonizing lineages (Eshenroder et al., 2016; Smith & Todd, 1984; Turgeon & Bernatchez, 2003) into newly available deepwater habitat following the Wisconsin Glacial Episode resulted in the evolution of at least eight distinct forms within the *C. artedi* species complex (Koelz, 1929; Scott & Crossman, 1998). Morphological differences occur across a variety of traits including body and head shape, lower jaw position, eye size, fin length, and gill raker counts (Koelz, 1929), though subtle variations among forms and between lakes

can often make visually distinguishing them difficult without all possible forms present (Eshenroder et al., 2016; Turgeon et al., 2016). Different forms typically occur in specific depth ranges leading to the hypothesis that depth niche specialization is a primary driver of divergence among forms (Smith & Todd, 1984). Stable isotope analysis further supports niche differentiation, indicating that many of the Great Lakes forms occupy different trophic levels with observed changes in proportion of pelagic and benthic food sources (Schmidt, Harvey, & Vander Zanden, 2011; Schmidt, Vander Zanden, & Kitchell, 2009; Sierszen et al., 2014). All forms likely undergo seasonal spawning migrations, with two of the three most common forms—*C. artedi* and *C. hoyi*—forming near-shore aggregations in November–December and January–March, respectively, and the third—*C. kiyi*—suspected to spawn during November–December (Dryer & Beil, 1968; Koelz, 1929; Yule, Addison, Evrard, Cullis, & Cholwek, 2009; Yule, Stockwell, Evrard, Cholwek, & Cullis, 2006). However, very little is known about behavioral and habitat differences among forms during overlapping periods of spawning, maintaining the possibility that hybridization during spawning events could be preventing genetic divergence. Analyses using RFLPs, mtDNA, and microsatellites have resulted in estimates of low or no genetic differentiation among Great Lakes forms (Bernatchez, Colombani, & Dodson, 1991; Reed, Dorschner, Todd, & Phillips, 1998; Turgeon & Bernatchez, 2003; Turgeon et al., 2016) leaving the question of reproductive isolation—particularly among deepwater forms—unanswered.

Over the past century, anthropogenic impacts have greatly reduced the original diversity of the *C. artedi* species complex in the Great Lakes, underscoring the need for establishing well-informed conservation units in ciscoes. The introduction of invasive forage fish, overfishing, and habitat loss led to large decreases in cisco

abundance and lake-wide extirpation to complete extinction of historically documented deepwater forms (Commission & Christie, 1973; Smith, 1968, 1970; Wells & McClain, 1973). Of the eight accepted forms originally described, only *C. artedi* and three deepwater forms—*C. hoyi*, *C. kiyi*, and *C. zenithicus*—are extant in the Great Lakes (referred to henceforth by specific epithet; Bailey & Smith, 1981; Todd & Smith, 1992). The range of a fifth deepwater form, *C. nigripinnis*, has been reduced to nearby Lake Nipigon (Ontario, Canada), though an extant *nigripinnis*-like form is still periodically caught in Lake Superior (Eshenroder et al., 2016). Lake Superior's peripheral location relative to both large human populations (and associated fishing pressure) and the canal construction that opened the Great Lakes to invasion from non-native species appears to have provided some protection from the impacts that extirpated cisco forms from the other four lakes (Koelz, 1926). Of the four lakes in which members of the *C. artedi* complex remain, Lake Superior is the only lake where all extant Great Lakes forms can still be regularly found (Eshenroder et al., 2016).

Recent evidence in the Great Lakes for declining abundance of invasive fish such as alewife and increasing abundance in cisco (Bronte et al., 2003; Mohr & Ebener, 2005; Schaeffer & Warner, 2007) has led to growing interest in re-establishing lost populations. An understanding of the roles of heritable genetic differences and reproductive isolation in the establishment and persistence of remnant forms is vital for developing both informed conservation units and restoration strategies. The main goal of our study was to employ genomic methods to improve our understanding of genetic variation among these forms. Specifically, we (1) examined genetic differentiation and diversity among putative forms of cisco, (2) compared the performance of SNPs and microsatellites in this system, and (3) leveraged a newly built cisco linkage map (Blumstein et al., 2019) to investigate adaptive divergence



**FIGURE 1** Map of sample sites in the Apostle Islands. INSET (upper right): Location of the Apostle Islands, gray box, within Lake Superior. INSET (lower right): Photos of the three most common cisco forms found in Lake Superior from Eshenroder et al. (2016), used with permission from the author

among forms. In order to remove the potentially confounding factor of distinguishing spatial genetic structure from form-based genetic structure, we focused on a single region in Lake Superior where multiple forms of cisco are found in sympatry, the Apostle Islands.

## 2 | MATERIALS AND METHODS

Tissue samples preserved in >95% ethanol were collected for the three most common cisco forms—*artedi*, *hoyi*, and *kiyi* in the Apostle Islands (Figure 1, Table 1, metadata provided in the supplemental material). Samples of putative *artedi* were collected in November 2017 using top and bottom gillnet surveys off Madeline Island conducted by the Wisconsin Department of Natural Resources. Putative *artedi*, *hoyi*, and *kiyi* were collected in July 2005 off Stockton Island by the U.S. Geological Survey Great Lakes Science Center (GLSC). In addition, we included a small number of available samples from two rare forms of cisco—*zenithicus* and the *nigripinnis*-like—from Minnesota, Michigan, and Ontario waters in Lake Superior collected by the GLSC in summers of 2007, 2012–2015. All samples were identified to putative form using a suite of standardized morphological characteristics (Eshenroder et al., 2016). DNA was isolated from fin tissue samples with Qiagen DNeasy® Blood & Tissue Kits.

### 2.1 | RAD library prep and sequencing

Restriction site-associated DNA (RAD) libraries were prepared following the BestRAD protocol (Ali et al., 2016). Extracted DNA was quantified using a Quant-it™ PicoGreen® dsDNA Assay

(Invitrogen, Waltham, MA) and normalized to a concentration of approximately 50 ng/μl for a 2-μl digestion reaction with the restriction enzyme *SbfI* followed by ligation with barcoded adaptors. Individually barcoded libraries were pooled into master libraries of 96 and fragmented to ~300-500bp with 12–14 30-s cycles in a Q500 sonicator (Qsonica, Newtown, CT). Fragmented DNA containing library adapters was bound to Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen) and washed with buffer to remove nontarget fragments before an incubation step to release DNA from the beads. Following purification with AMPure XP beads (Beckman Coulter, Brea, CA), master libraries were input into the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® at the End Prep step for ligation of master library barcodes, a 250-bp insert size-selection, and a 12-cycle PCR enrichment. Successful size selection and enrichment were confirmed with visualization of products on a 2% agarose E-Gel (Invitrogen). Products underwent a final AMPure XP purification cleanup followed by quantification with a Qubit® 2.0 Fluorometer. All prepared libraries were sent to Novogene (Sacramento, CA) for sequencing on the Illumina NovaseqS4 platform.

### 2.2 | Read processing and SNP filtering

Raw sequences generated from RAD sequencing were processed in the software pipeline Stacks v2.3d (Rochette, Rivera-Colón, & Catchen, 2019). Sequences were demultiplexed by barcode, filtered for presence of the enzyme cut-site and quality, and trimmed in the subprogram *process\_radtags* (parameter flags: `-e SbfI -c -q -r -t 140 --filter_illumina --bestrad`). Filtered reads for each individual were aligned to create matching stacks with *ustacks* following

**TABLE 1** Sample statistics, diversity, and effective population size estimates

Method	Group	Code	N	A	$H_o$	$H_e$	$G_{IS}$	Assignment	$N_e$ (CIs)
RAD	<i>C. artedi</i>	ART	60	98.5%	0.246	0.253	0.028	98.3% <sup>b</sup>	1,834 (1,769–1,906)
	<i>C. hoyi</i>	HOY	21 <sup>a</sup>	94.8%	0.219	0.273	0.196	92.5% <sup>b</sup>	1,701 (1,487–1,985)
	<i>C. kiyi</i>	KIY	25	93.4%	0.217	0.242	0.103	100%	2,126 (1,846–2,506)
	<i>C. nigripinnis</i>	NIG	6	83.1%	0.213	0.261	0.182	–	–
	<i>C. zenithicus</i>	ZEN	4	76.5%	0.186	0.259	0.279	–	–
Microsatellites	<i>C. artedi</i>	ART	30	70.0%	0.534	0.607	0.121	76.7%	91 (42–6,126)
	<i>C. hoyi</i>	HOY	21 <sup>a</sup>	65.0%	0.629	0.659	0.046	61.9%	73 (33-Infinite)
	<i>C. kiyi</i>	KIY	25	67.0%	0.572	0.631	0.093	72.0%	649 (74-Infinite)

Note: N is the number of individuals successfully genotyped, A is the percentage of total sampled alleles found in each group,  $H_o/H_e$  is observed/expected heterozygosity,  $G_{IS}$  is the inbreeding coefficient, Assignment is the percentage of individuals that were correctly assigned to their population of origin in a leave-one-out test, and  $N_e$  is effective population size calculated using the LDNE method and reported with 95% confidence intervals.

<sup>a</sup>Three genotyped *hoyi* samples are suspected to be misidentified *kiyi* from the RAD-based PCA and ADMIXTURE analysis and were removed to prevent bias in estimates of diversity and  $N_e$ .

<sup>b</sup>The one *artedi* that assigned to HOY and the two *hoyi* that assigned to KIY appear to be hybrids. Outside of these putative hybrids, assignment to the ART and HOY groups was 100%.

guidelines suggested from empirical testing to avoid under- or over-merging loci from RAD datasets (Paris, Stevens, Catchen, & Johnston, 2017; parameter flags: `--disable-gapped -m 3 -M 5 -H --max_locus_stacks 4 --model_type bounded --bound_high 0.05`). A catalog of consensus loci built from 51 *artedi* sampled in Lake Huron used for the development of a cisco linkage map (Blumstein et al., 2019) was appended in *cstacks* with an additional 75 individuals in the *C. artedi* species complex from across the Great Lakes, including 19 fish used in the current study: five *hoi* and four *kiyi* from the Apostle Islands and five each of *nigripinnis*-like and *zenithicus* from various locations across Lake Superior (parameter flags: `-n 3 -p 6 --disable_gapped`). Locus stacks for each individual were matched to the catalog using *sstacks* (parameter flag: `--disable_gapped`), data were oriented by locus in *tsv2bam*, and reads were aligned to loci and SNPs were called with *gstacks*. SNPs genotyped in greater than 30% of individuals (parameter flag: `-r 0.3`) were exported with the subprogram *populations* in genepop and variant call format (vcf) files. We isolated SNP calling to single-end reads to ensure equal coverage across sequences, but paired-end assemblies for each locus are available from Blumstein et al. (2019).

Primary SNP filtering was performed with *vcftools* v0.1.15 (Danecek et al., 2011) and included (1) removing loci genotyped in fewer than 70% of individuals, (2) removing individuals missing more than 50% of loci, and (3) removing loci with a minor allele count less than 3. In addition, since all salmonids including members of the *C. artedi* species complex have experienced a recent genome duplication, putatively paralogous loci were identified with the program HDPlot (McKinney, Waples, Seeb, & Seeb, 2017), and any loci with heterozygosity greater than 0.55 or a read ratio deviation greater than 5 and less than -5 were removed. Finally, loci on the same RAD tag may be linked so only the SNP with the highest minor allele frequency on each tag was included in the final dataset. All file format conversions were performed using PGDSpider v2.1.1.5 (Lischer & Excoffier, 2011).

### 2.3 | Microsatellite amplification and genotyping

We used the methods described in Stott et al. (2020) to genotype 12 microsatellites developed for coregonines (Bernatchez, 1996; Patton, Gallaway, Fechhelm, & Cronin, 1997; Rogers, Marchand, & Bernatchez, 2004) and salmonids (Angers, Bernatchez, Angers, & Desgroseillers, 1995; Estoup, Presa, Krieg, Vaiman, & Guyomard, 1993) in *artedi*, *hoi*, and *kiyi* from the Apostle Islands: *Bwf1*, *Bwf2*, *C2-157*, *Cocl23*, *CoclLav6*, *CoclLav27*, *CoclLav32*, *CoclLav72*, *Sfo8*, *Sfo23*, *Str-60*. *Sfo8* consistently amplifies two genomic regions resulting in alleles sorting into upper (U: 215-281bp) and lower (L: 163-193bp) size ranges; therefore, *Sfo8* is considered two loci. Fragment analysis was performed using a Genetic Analyzer 3.0 (Life Technologies), and genotypes were assigned at each locus using GeneMapper 3.7 (Life Technologies). We used Genepop v4 (Rousset, 2008) to conduct exact tests for deviations from Hardy-Weinberg

and linkage equilibrium ( $\alpha = 0.01$ ). Three loci were removed for being out of Hardy-Weinberg Equilibrium in all three forms (*Bwf1*, *Cocl23*, *CoclLav6*), and no loci showed significant linkage disequilibrium.

### 2.4 | Genetic differentiation and diversity

To examine the influence of selection on estimates of differentiation and diversity with our RAD dataset, we employed two statistical tools for the identification of putative outlier loci—BayeScan v2.1 (Foll & Gaggiotti, 2008) and *pcadapt* v4.1.0 (Luu, Bazin, & Blum, 2017). BayeScan uses a Bayesian approach and the multinomial-Dirichlet model to generate  $F_{ST}$ -related measures of population-level differences in allele frequencies and was run on our data with default parameters. *Pcadapt* takes an individual-level approach and uses principal component analysis (PCA) for outlier detection. A scree plot of the first 20 principal components (termed *K* in *pcadapt*) indicated that the optimal *K* was 2 for computing correlations between our RAD loci and *K* principal components. We used Benjamini and Hochberg's (1995) method for correction of the false discovery rate in both BayeScan and *pcadapt* at  $\alpha = 0.05$ . Exploratory analysis comparing the full dataset and a dataset of neutral loci resulted in no substantial differences in estimates of genetic differentiation (see Results), so we proceeded using the full dataset of RAD loci.

To ensure that our putative form designations were appropriate, we used two approaches to assess genetic similarity among individuals. First, we conducted a PCA in the R package "adeget" (Jombart, 2008) for both the RAD and microsatellite datasets. Next, we estimated the number of ancestral populations, *K*, contributing to contemporary genetic clustering for the RAD dataset using the program ADMIXTURE v1.3 (Alexander, Novembre, & Lange, 2009). We tested *K* from 1 to 5 with ADMIXTURE's cross-validation procedure and a *k*-fold of 10 (parameter flag: `--cv = 10`) to examine support for each *K*. ADMIXTURE analysis was only conducted on the RAD dataset because this program is not compatible with microsatellite data.

PCA and ADMIXTURE analysis with RAD genotypes revealed that three individuals originally identified as *hoi* fell within the cluster of *kiyi* samples. Given the discrete clustering of the remaining of *hoi* and *kiyi*, there is strong possibility that these three *hoi* samples were misidentified in the field, so we removed these individuals from both RAD and microsatellite datasets for all further analyses to prevent bias in estimates of diversity, differentiation, and  $N_e$ .

We calculated a variety of summary statistics for the groups comprised of the five forms (*artedi*—ART, *hoi*—HOY, *kiyi*—KIY, *nigripinnis*—NIG, and *zenithicus*—ZEN) including percentage of total observed alleles, observed and expected heterozygosity, and an inbreeding coefficient ( $G_{IS}$ ). Summary statistics for each locus and populations were calculated using both microsatellite and RAD datasets in GenoDive v2.0b23 (Meirmans & Van Tienderen, 2004). Genetic differentiation among all forms was estimated across all loci in the RAD dataset with pairwise  $F_{ST}$  (Weir & Cockerham, 1984) in Genepop and tested using exact tests (Goudet, Raymond, de Meeüs, & Rousset, 1996; Raymond and Rousset, 1995;  $\alpha = 0.01$ ) in Arlequin (Excoffier & Lischer, 2010).

We also calculated locus-specific overall and pairwise- $F_{ST}$  values (Weir & Cockerham, 1984) in Genepop using a dataset that included the three forms with  $n > 10$  (ART, HOY, and KIY). To compare genetic differentiation between RAD and microsatellite datasets, we used GenoDive to estimate standardized pairwise genetic differentiation,  $G'_{ST}$  (Hedrick, 2005), which employs an additional correction for bias from sampling a limited number of populations (Meirmans & Hedrick, 2011).

The rate at which individuals were assigned back to their form with RAD and microsatellite datasets was tested using population assignment in GenoDive for all forms with  $n > 10$ . Assignment was performed by calculating the home likelihood ( $L_h$ ) that an individual genotype is from a specific group given the allele frequencies (Paetkau, Calvert, Stirling, & Strobeck, 1995) using the leave-one-out method to avoid the bias from a target individual's contribution to the allele frequencies of a source population. Zero frequencies were replaced with 0.005, and a significance threshold of  $\alpha = 0.002$  was applied separately to each group over 1,000 replicated datasets (c.f. Perreault-Payette et al., 2017).

Effective population size ( $N_e$ ) was estimated for all forms with  $n > 10$  using RAD and microsatellite datasets with the bias-corrected linkage disequilibrium method (LDNE; Hill, 1981; Waples, 2006; Waples & Do, 2010) in the software package NeEstimator v2.1 (Do et al., 2014). We used a  $p$ -crit of 0.05 for the RAD dataset (Waples, Larson, & Waples, 2016) and 0.02 for the microsatellite dataset (Waples & Do, 2010). For the RAD dataset, only comparisons between sampled loci that were found on different linkage groups (LGs) of the cisco linkage map (Blumstein et al., 2019) were included to correct for physical linkage (Waples et al., 2016).  $N_e$  calculations using the linkage disequilibrium method can be biased slightly downward when individuals from multiple cohorts are included in the sample due to a slight Wahlund effect (7% downward bias on average; Waples, Antao, & Luikart, 2014). However, this small bias should not greatly affect the interpretation of the  $N_e$  results.

Finally, we used EASYPOPv2.0.1 (Balloux, 2001) to compare observed patterns of differentiation and hybridization with simulations encompassing a variety of demographic scenarios. It is important to note that these simulations only modeled neutral processes. It is likely that adaptive divergence played at least some role in the formation of distinct cisco forms; however, the patterns and magnitude of differentiation among forms were similar with our full dataset and with a putatively neutral dataset (see results), indicating that simulations of neutral processes should be useful for interpreting our data. Base parameters for all simulations were informed where possible by cisco life history traits and included random mating, same number of females and males per population (total  $n = 2,000$  per population based on our  $N_e$  estimates for the three main cisco forms), equal proportions of female and male migration, biallelic loci with free recombination, a mutation rate equivalent to that measured in human and fish nuclear genomes ( $\mu = 1.0 \times 10^{-8}$ ; Bernardi & Lape, 2005; Conrad et al., 2011), and the KAM mutation model. First, we modeled the impacts of hybridization on genetic differentiation between simulated populations with similar characteristics to our cisco forms using three scenarios: two populations with a starting level

of differentiation equal to (a) approximately the level we observed among cisco forms with our RAD dataset ( $F_{ST} \approx 0.05$ ), (b) approximately twice the amount observed ( $F_{ST} \approx 0.10$ ), and (c) approximately four times the amount observed ( $F_{ST} \approx 0.20$ ). Preliminary levels of differentiation were achieved with low migration rates ( $m$ ) set over 1,000 generations ( $m = 0.001$ ,  $m = 0.0005$ , and  $m = 0.0001$ , respectively) and followed by 5, 10, or 15 generations of one of three higher migration (i.e., hybridization) rates ( $m = 0.01$ , 0.05, and 0.10) for every initial level of differentiation. Each unique parameter combination was run using a dataset of 1,000 loci and replicated 50 times. Estimates of  $F_{ST}$  for each replicate were generated in the R package "diveRsity" (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013) with a genepop file containing a subset of 100 randomly selected individuals per population (50 females/50 males). Second, we simulated migration between three populations representing our three common cisco forms (ART, HOY, KIY) in order to reconstruct datasets with similar characteristics to our empirical data and compare signatures of hybridization among them. Preliminary differentiation among the three populations (P1–P3) was set at  $F_{ST} \approx 0.05$  using the same method described above of applying a low rate of migration ( $m = 0.001$ ) over an extended period of time (1,000 generations), and we allowed populations to hybridize for 2, 5, or 10 generations with  $m = 0.05$ , which was equal to roughly the proportion of hybrids observed in our wild populations between ART-HOY and HOY-KIY. Since no putative ART-KIY hybrids were observed in our RAD dataset, we chose to implement a one-dimensional stepping-stone model of migration (Kimura & Weiss, 1964) among simulated populations. Each unique parameter set for these simulations was run using the maximum possible number of loci in EASYPOP (8,000), and a genepop file was output containing a subset of 100 randomly selected individuals per population (50 females/50 males) and compared to a reduced dataset of 8,000 neutral loci from our empirical data. The reduced dataset was generated by removing putative outliers from the loci that were placed on a linkage map (see below) and randomly subsampling the remaining markers. ADMIXTURE was used to generate  $Q$ -scores for each individual in both simulated and reduced empirical datasets. Putative hybrids were identified using a  $Q$ -score of less than 70% following similar thresholds applied in the literature (Kapfer, Sloss, Schuurman, Paloski, & Lorch, 2013; Marie, Bernatchez, & Garant, 2011; Weigel et al., 2018), and hybrid combinations were assigned to the two populations representing the largest  $Q$ -scores within the hybrid individual.

## 2.5 | Differentiation across the genome

We examined genetic differentiation across the genome by pairing our data with the *artedi* linkage map constructed by Blumstein et al. (2019). Catalog IDs were identical between the current study and linkage map; therefore, no alignment step was needed to compare loci. To identify putative genomic islands of divergence that were highly differentiated from the rest of the genome, we used a Gaussian kernel smoothing technique (Gagnaire, Pavey, et al., 2013; Hohenlohe et al., 2010;

Larson et al., 2017) that incorporated locus-specific differentiation and genomic position on the linkage map. A window size of 5 cM and a stepwise shift of 1 cM was used for this analysis, and values of genetic differentiation were weighted according to their window position as described by Gagnaire, Pavey, et al. (2013). Highly differentiated windows were identified by randomly sampling  $N$  loci from the genome (where  $N$  was the number of loci in the window) and comparing the average differentiation of those loci to the average differentiation of the loci in the window. This sampling routine was conducted 1,000 times for each window. If a window exceeded the 90th percentile of the sampling distribution, the number of bootstrap replicates was increased to 10,000. Contiguous windows that contained at least two loci and exceed the 99th percentile of the distribution after 10,000 bootstrap replicates were classified as significantly differentiated. We investigated genomic differentiation using four locus-specific metrics: overall  $F_{ST}$  and pairwise  $F_{ST}$  between each of the three putative forms with  $n > 10$  (ART-HOY, ART-KIY, and HOY-KIY). We then plotted overall  $F_{ST}$  across the genome and constructed a bubble plot to visualize the number of significant windows on each LG for each comparison. We found over 50 significant windows for each comparison (see results). Conducting an in-depth investigation of all significant windows was not feasible; therefore, we isolated in-depth analysis to one LG (Cart21) that contained the most significant windows and outlier loci in the dataset.

To investigate this highly differentiated LG, we aligned consensus sequences for all loci from Cart21 to chromosome Ssa05 in Atlantic salmon (*Salmo salar*), which is syntenic to Cart21 in *artedi* (Blumstein et al., 2019). Ssa05 sequences were obtained from genome version ICSASG\_v2 (Lien et al., 2016), and alignments were conducted with BLASTN. The best alignment for each locus was retained, and all alignments had  $e$ -values  $< 1e-58$ . We then visualized the relationship between recombination and physical distance using alignments to Ssa05 and information from the *artedi* linkage map to determine whether this highly differentiated region is characterized by lower recombination. We also obtained annotation information from the Atlantic salmon genome to determine whether genes of interest were co-located with areas of high divergence. Finally, we plotted the allele frequencies of the 10 SNPs on Cart21 with the highest overall  $F_{ST}$  to investigate whether these SNPs show consistent patterns of population structure.

### 3 | RESULTS

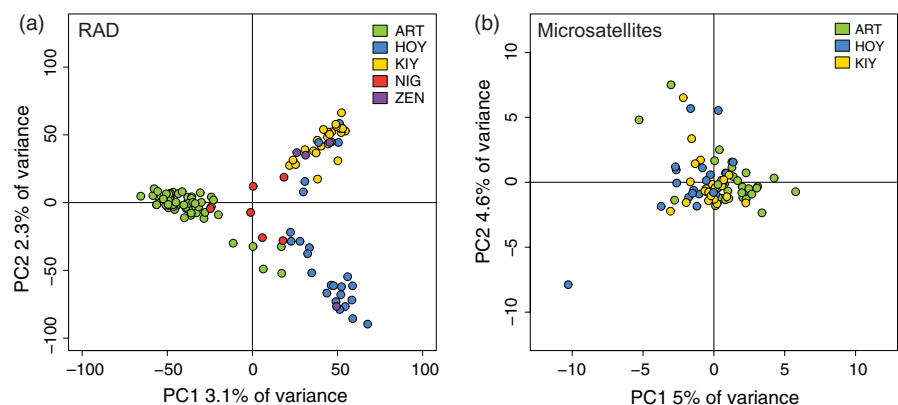
#### 3.1 | Sequencing and genotyping

A total of 137 individuals were RAD sequenced producing more than 455 million reads and an average of 3,346,457 reads per sample. After filtering, 119 individuals with representatives from all five putative cisco forms in Lake Superior were genotyped at 29,068 loci (Table 1). More than half of these loci ( $n = 15,348$  loci) were also placed on the linkage map. Since RAD sequencing and microsatellite amplification were performed on the same *hoi* and *kiy* samples, microsatellite genotypes used in our analyses were restricted to the same individuals that successfully genotyped with our RAD loci. An additional 30 *artedi* from Stockton Island were genotyped at the 9 microsatellite loci.

#### 3.2 | Genetic differentiation and diversity

BayeScan identified 244 outliers, and pcadapt identified 318 outliers, 150 of which were identified by both programs (see Figure S1). All 412 outlier loci were putatively under divergent selection and were removed from our full RAD dataset to generate a neutral dataset of 28,656 loci. We examined both the full and neutral datasets using PCA and ADMIXTURE and found no significant differences between the patterns of differentiation between datasets. Likewise, estimates of pairwise genetic differentiation ( $F_{ST}$ ) decreased by an average of only 0.0039 with the neutral dataset. Results from exploratory analyses with the neutral dataset are included (Figure S2 and Table S1), and all results presented below are for the full RAD dataset unless otherwise stated.

PCA showed a sharp contrast in resolution between marker sets with microsatellites producing one large cluster of overlapped forms across the first two principal components and the RAD dataset producing three major clusters primarily composed of ART, HOY, and KIY (Figure 2). The ART cluster separates from HOY and KIY along the first principal component (PC), and HOY and KIY form discrete clusters along the second PC with three exceptions. Three individuals originally identified as *hoi* fell within the KIY cluster (see methods and ADMIXTURE results below). Of the rare forms, *nigripinnis* (NIG) loosely



**FIGURE 2** Principal components analysis with RAD (a) and microsatellite (b) data. The percentage of variance explained by each principal component (PC) is labeled on the x- and y-axes

grouped in the center of the PCA, with five of the six samples falling out between the ART, HOY, and KIY clusters (Figure 2). The sixth sample fell within the ART cluster, and like the three *hoi* in the KIY cluster, possibly represents a misidentified specimen. Unlike the NIG samples, which suggest the possibility for a distinct cluster with the addition of more specimens, the four ZEN samples closely associated with either the HOY cluster ( $n = 1$ ) or the KIY cluster ( $n = 3$ ). Low representation of both NIG and ZEN in our RAD dataset reduces our ability to draw strong conclusions based on these PCA results, so both groups were unaltered for estimates of diversity, inbreeding, and differentiation and dropped for the remaining analyses.

The most supported number of ancestral populations ( $K$ ) estimated using the cross-validation procedure in ADMIXTURE was two (cross-validation error: 0.498; Figure 3). Examining additional  $K$ s for significant substructuring among forms generated results that corroborated those from the PCA. When  $K = 2$ , the ART cluster splits from HOY, KIY, and ZEN. Individuals in the NIG cluster exhibited mixed ancestry between the two major groups as seen on PC1 of the PCA (Figure 3). When  $K = 3$  (cross-validation error: 0.509), the major genetic ancestries differentiate the ART, HOY, and KIY clusters as seen on PC2. The three putatively misidentified *hoi* first noted in the PCA all had  $Q$  estimates of 100% for the KIY cluster and were removed from further analyses. Additional  $K$ s did not differentiate either the NIG or ZEN cluster but begin to differentiate small subsets of individuals within groups, which was likely statistical noise (see Figure S3).

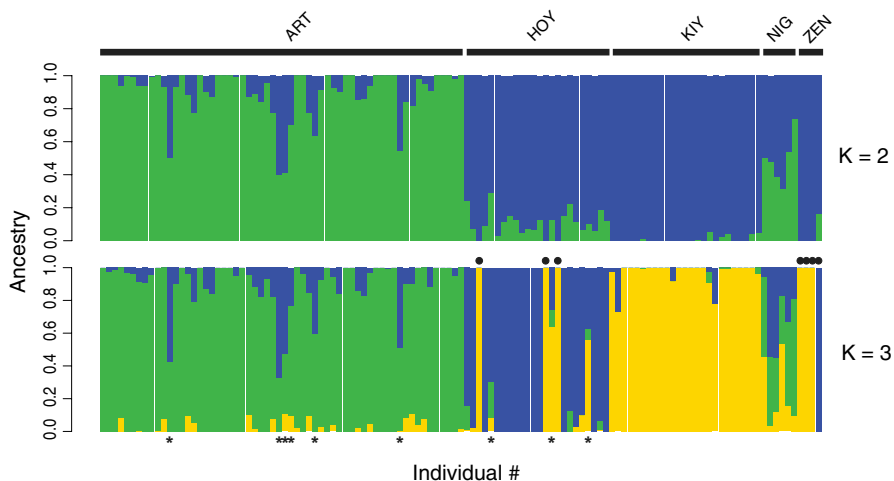
Observed and expected heterozygosity ranged from 0.186–0.246 and 0.242–0.273, respectively, in the RAD dataset and 0.534–0.629 and 0.607–0.659, respectively, in the microsatellite dataset (Table 1). Inbreeding coefficients were not substantially different from zero in both datasets. The largest  $G_{IS}$  was measured in ZEN from only four samples (0.279), and the rest were between  $-0.028$  and  $0.196$ . All estimates of pairwise genetic differentiation among forms with  $n > 10$  (ART, HOY, KIY) with the RAD dataset were significant (Table 2). The magnitude of genetic differentiation followed similar trends observed in the PCA, with ART being slightly more differentiated from deepwater forms ( $F_{ST} = 0.049$ – $0.056$ ). This pattern remained the same with a standardized measure of genetic

differentiation in the RAD and microsatellite datasets (Table 3). All pairwise comparisons in both datasets were significant with higher overall values of  $G'_{ST}$  generated using microsatellites (0.110–0.122) than SNPs (0.060–0.75). See Tables S1 and S2 for locus-specific summary statistics.

Population self-assignment rate using the microsatellite dataset ranged from 61.9% to 76.7% with *artedi* exhibiting the highest likelihood of being assigned back to the ART group (Table 1). Assignment rate with the RAD dataset was 100% for the KIY group and 98.3% and 92.5% for the ART and HOY groups, respectively, a result of one putative *artedi* assigning to HOY and two putative *hoi* individuals assigning to KIY. In the ADMIXTURE analysis, these three individuals exhibited relatively high  $Q$  estimates for ancestry to the populations to which they were assigned (*artedi*,  $Q_{HOY} = 67.4\%$ ; *hoi*,  $Q_{KIY} = 64.2\%$  and  $56.2\%$ ). In the PCA generated from the same data, these individuals were oriented between the main clusters. In the microsatellite dataset with the same *hoi* and *kiyi* samples, neither of the two potentially misclassified *hoi* assigned to the HOY group, with one being assigned to ART and one to KIY. Assignment scores are reported in Tables S3–S5.

Estimates of  $N_e$  with the microsatellite dataset ranged from 73 in HOY to 659 in KIY (Table 1). Only the estimate for ART produced both upper and lower bound confidence intervals ( $N_e$ : 91, CI: 42–6, 126), whereas the estimates for HOY and KIY produced confidence intervals with an “infinite” upper bound. An “infinite” upper bound is typically an indication that the data are not powerful enough to produce an accurate estimate of  $N_e$  given the sample size, population size, and/or marker resolution (Do et al., 2014; Waples & Do, 2010). For the RAD dataset, estimates of  $N_e$  were generated with loci that were placed on the linkage map and ranged from 1,701 to 2,126 with confidence intervals within 10% of these values.

The fact that we observed hybridization coupled with relatively high genetic differentiation was puzzling and prompted us to conduct two types of simulations to investigate how hybridization (i.e., gene flow can influence genetic differentiation). Simulated hybridization over a 15-generation period resulted in declines in genetic differentiation among populations for all tested levels of migration (Figure S4). When the migration rate was set to 5 or 10%, levels of differentiation



**FIGURE 3** Genetic lineages in Apostle Islands ciscoes estimated with ADMIXTURE. Each vertical bar represents a single individual and is colored by the proportion of ancestry ( $Q$ ) assigned to each genetic lineage ( $K$ ). In the  $K = 3$  plot (bottom), filled circles represent putatively misidentified individuals within forms (with  $Q$ -scores of 100% to other forms) and asterisks represent putative hybrids based on a  $Q$ -score threshold of less than 70%



more than halved after only five generations for all initial levels of  $F_{ST}$ . A migration rate of 1% resulted in steadily declining genetic differentiation but only began to approach a halved  $F_{ST}$  after 15 generations.

Simulations of stepping-stone migration between three populations and subsequent ADMIXTURE analysis resulted in an overall pattern similar to that observed in our empirical data (Figure 4, baseline ADMIXTURE plot at generation 0 presented as supplemental Figure S5). One of the major goals of these simulations was to determine whether  $Q$ -scores between 0.1 and 0.2 for alternative forms observed in our empirical dataset could be the result of statistical noise. Results

**TABLE 2** Pairwise differentiation among putative forms calculated with the RAD dataset

Group	ART	HOY	KIY	NIG	ZEN
ART	-				
HOY	<b>0.049</b>	-			
KIY	<b>0.056</b>	<b>0.045</b>	-		
NIG	0.019	<b>0.013</b>	<b>0.034</b>	-	
ZEN	<b>0.061</b>	<b>0.017</b>	0.020	0.014	-

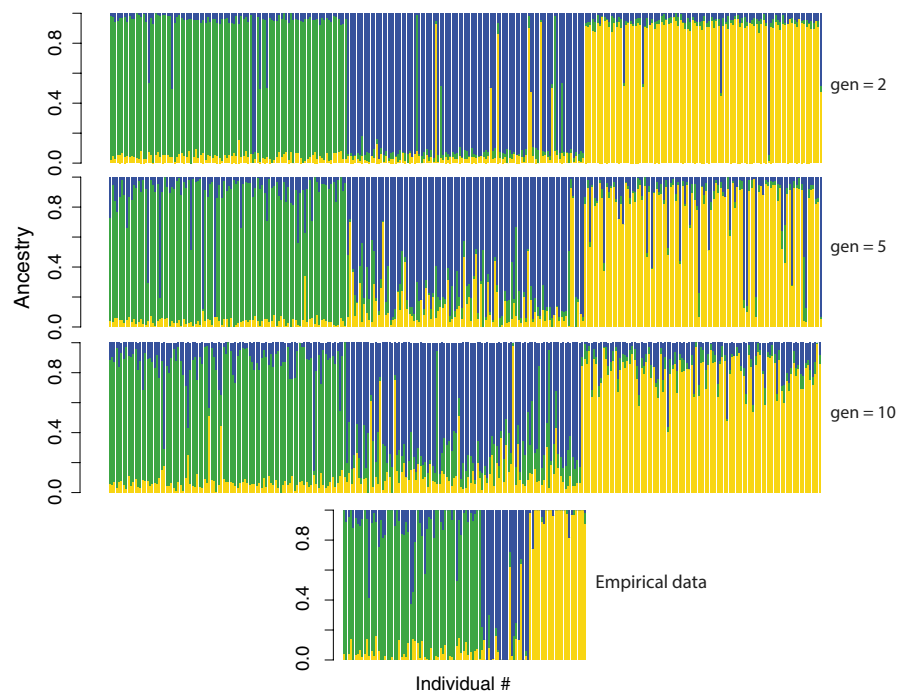
Note:  $F_{ST}$  values on the lower diagonal. Significant values are in bold.

**TABLE 3** Standardized pairwise genetic differentiation,  $G'_{ST}$  (Hedrick, 2005), for SNP and microsatellite datasets

Group	ART	HOY	KIY
ART	-	0.122	0.113
HOY	0.064	-	0.110
KIY	0.075	0.060	-

Note: Values measured using the SNP dataset are in the lower diagonal, and values measured using the microsatellite dataset are in the upper diagonal.

**FIGURE 4** Levels of genetic admixture in three simulated populations after 2, 5, and 10 generations (gen) of stepping-stone migration. Simulations were run with random mating of 1,000 females and 1,000 males in each population using 8,000 biallelic loci, and preliminary conditions that produced a similar level of differentiation observed in our RAD dataset among forms ( $F_{ST} \approx 0.05$ ; 1,000 generations with an island migration rate of 0.001). Each ADMIXTURE plot represents a random subset of 50 males and 50 females from each population. Empirical data were reduced to a dataset of 8,000 randomly selected, neutral loci that could be placed on the linkage map and run in ADMIXTURE for comparison to simulated data



from the simulations with 2 generations of migration (G2), where only F1 migrants are possible, demonstrated that these  $Q$ -scores in the range of 0.1–0.2 can be generated by statistical noise as they were present in this simulation (Figure 4). However, since no forms contained private alleles, we do not have the ability with our empirical data to confidently distinguish statistical noise from true hybrid backcrosses. We documented a relatively high level of hybridization in our empirical dataset, with ART-HOY hybrids comprising 9.9% of the total number of sampled *artedi* and *hoiy*, HOY-KIY hybrids comprising 4.3% of sampled *hoiy* and *kiyi*, and ART-KIY hybrids comprising 0% (Table 4). Simulation results indicated that the proportion of observed ART-HOY hybrids is more similar to that found after 5 generations (G5) of hybridization (G5: 10.6%), whereas the proportion of observed HOY-KIY hybrids is more similar that found after 2 generations of hybridization (G2: 3.5%). Similar to our observed data, ART-KIY hybrids were absent or in low frequency (G5: 1.2%) in all simulations. Coupled with the high levels of genetic differentiation we observed, these results suggest two possible scenarios—that ART-HOY and HOY-KIY hybridization is a relatively recent development or that some pre- or postzygotic mechanism is reducing hybrid fitness.

### 3.3 | Differentiation across the genome

Genetic differentiation across the genome was generally high, with many loci displaying overall  $F_{ST}$  values  $> 0.2$  (Figure 5, Figures S6–S7). This differentiation was also not localized to a few LGs, as every LG had at least one locus with  $F_{ST} > 0.2$ . Kernel smoothing analysis revealed 389 genomic windows that displayed significantly elevated differentiation compared to the rest of the genome (Figure 5, Figures S6–S7). The number of significant windows was highest for the overall  $F_{ST}$  comparison (115),

followed by the ART-HOY comparison (101), ART-KIY comparison (98), and HOY-KIY comparison (75). Significantly differentiated windows were found on all but five LGs (average per LG = 10,  $SD = 10$ , range = 0–36). LG Cart21 displayed the highest number of differentiated windows, prompting us to conduct an in-depth

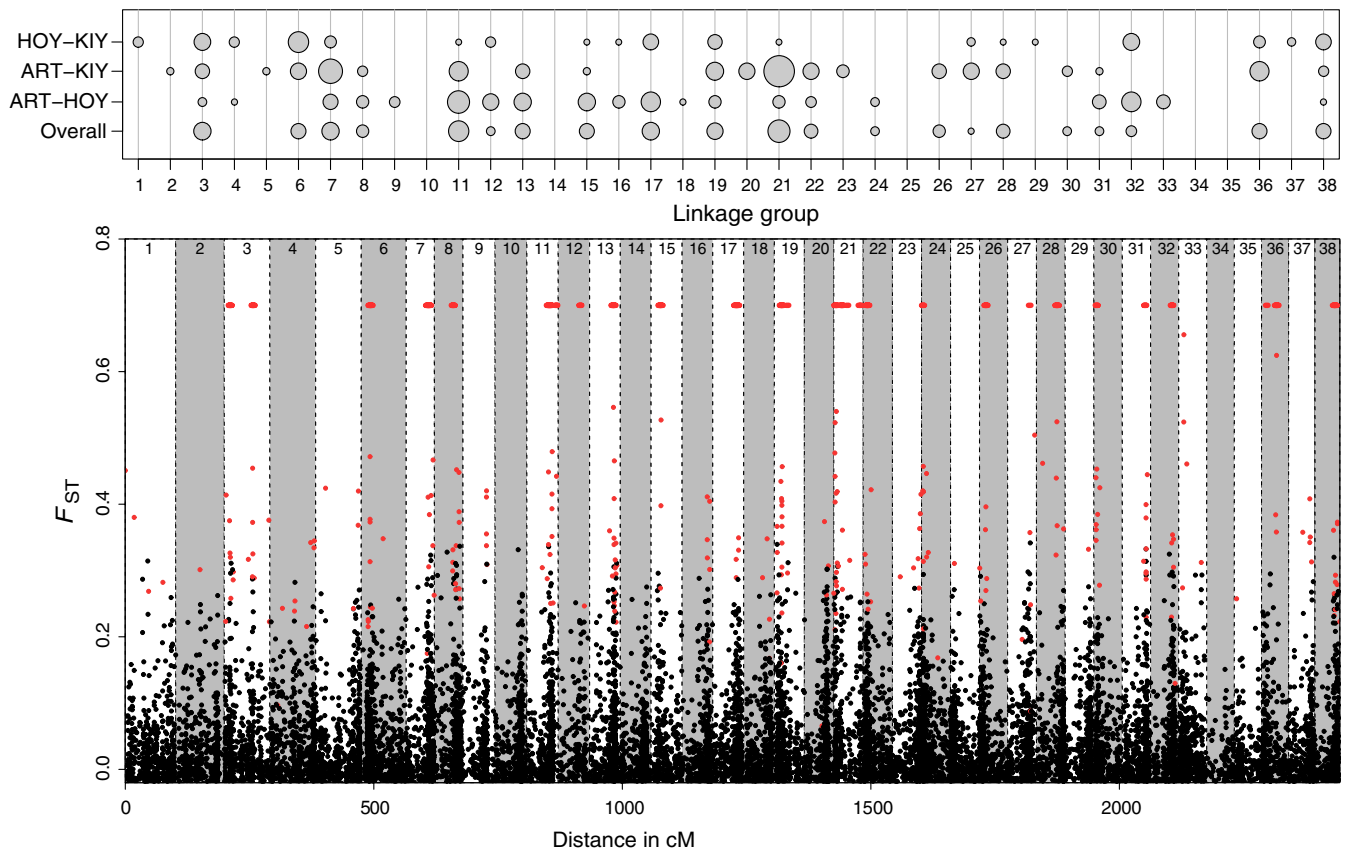
investigation of this LG to investigate patterns and potential drivers of divergence (Figure 6).

Cart21 contained 36 significantly differentiated windows, with the largest number of windows found for the ART-KIY comparison (18), followed by the overall  $F_{ST}$  comparison (13), ART-HOY comparison (4),

**TABLE 4** Type, number, and proportion of hybrids observed in empirical data and simulated populations

Hybrid type	Empirical populations			Simulated populations								
	NA			G2			G5			G10		
	ART-HOY	HOY-KIY	ART-KIY	P1-P2	P2-P3	P1-P3	P1-P2	P2-P3	P1-P3	P1-P2	P2-P3	P1-P3
No. of hybrids	8	2	0	6	7	0	20	27	2	31	38	0
Proportion of hybrids	0.099	0.043	0.000	0.031	0.035	0.000	0.106	0.142	0.012	0.172	0.210	0.000

Note: Data simulated in EASYPOP were comprised of 8,000 biallelic loci from three populations (P1–P3) that experienced stepping-stone migration ( $m = 0.05$ ) for 2, 5, or 10 generations (G2, G5, G10) after 1,000 generations of low migration ( $m = 0.001$ ) to approximate an  $F_{ST}$  similar to that observed between our three major cisco forms (ART, HOY & KIY,  $F_{ST} \approx 0.05$ ). Putative hybrids were identified using a  $Q < 70\%$  and were assigned to the combination of source populations that had the two highest  $Q$ -scores. Empirical data were reduced to a dataset of 8,000 randomly selected, neutral loci and similarly assessed for comparison to simulated data.



**FIGURE 5** Genetic differentiation across the genome visualized with a bubble plot (top) and plot with the overall  $F_{ST}$  of each marker (bottom). The size of each bubble in the bubble plot represents the number of genomic windows that were significantly differentiated from the rest of the genome according to kernel smoothing analysis for each form comparison. The “overall” designation is overall  $F_{ST}$  across the dataset. Each dot in the graph of differentiation across the genome represents a marker, and red lines denote significantly differentiated windows. Red dots are loci that were found to be putatively under divergent selection in either Bayescan or PCAdapt. Linkage groups are separated by dashed lines. Form abbreviations are in Table 1. See Figures S4–S8 for visualizations of genetic differentiation for each chromosome and form comparison

and HOY-KIY comparison (1). We were able to place 351 loci on Cart21, and 12 of these loci displayed overall  $F_{ST}$  values  $> 0.3$  (Figure 6a). The largest cluster of high- $F_{ST}$  loci was found between 0 and 10 cM on the linkage map. This region appears to be characterized by relatively low recombination, as loci found in the first 10 cM of Cart21 span about 25 megabases of the Atlantic salmon genome (Figure 6b). Alignments to the Atlantic salmon genome were possible for 151 loci on Cart21, and these alignments revealed that the highest  $F_{ST}$  loci were found between positions 15 million and 25 million on Ssa05 (Figure 6c). Some of these loci were found in genes with functions that include cell signaling and membrane transport. However, there are over 2,000 genes on Ssa05, making it difficult to reach any robust conclusions about the functional significance of our loci. Allele frequencies at the high- $F_{ST}$  loci were generally the most diverged between ART and the other two forms, KIY in particular (Figure 6d).

## 4 | DISCUSSION

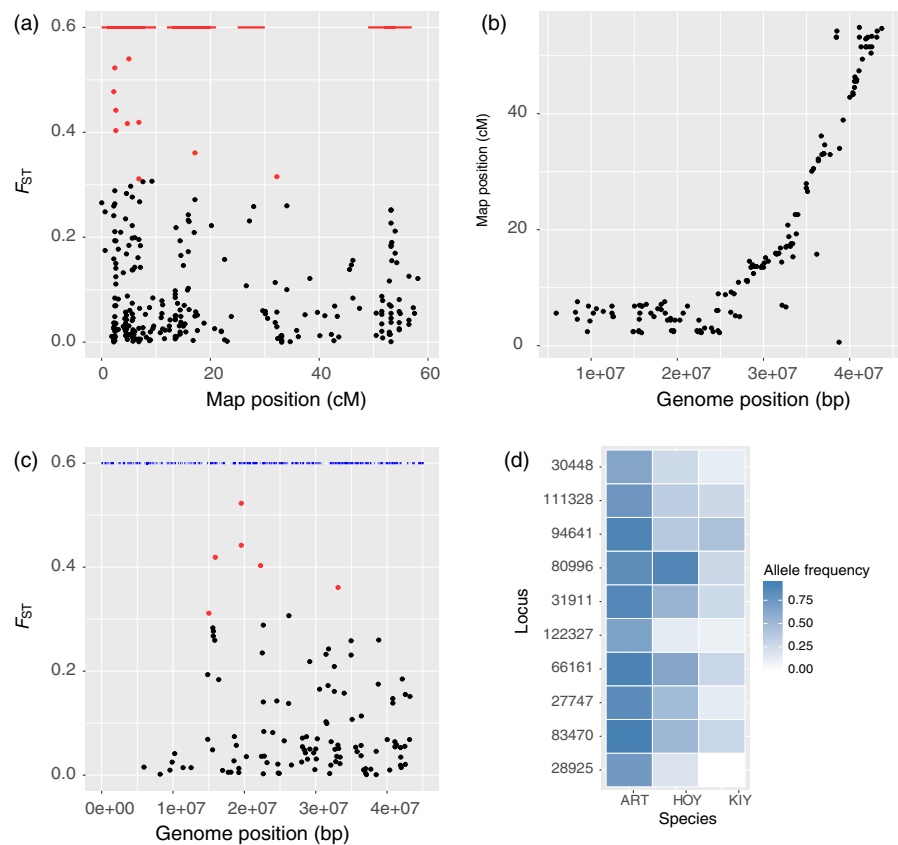
The detection of genetic structure in recently diverged species complexes has proved challenging with traditional genetic methods, prompting the reevaluation with genomics of a taxonomically uncertain species complex in the Laurentian Great Lakes. Using a genome-wide panel of SNPs to examine differentiation and diversity of sympatric coregonines in the Apostle Islands of Lake Superior, we were able to unambiguously assign individuals to the three major forms as well as identify putative hybrids

and misidentified individuals. Despite a century of anthropogenic impacts in the Great Lakes that has seen the extirpation and extinction of historically documented forms in the *C. artedi* species complex, estimates of  $N_e$  and diversity in Apostle Islands populations do not suggest the three major extant forms are experiencing bottleneck effects. Genetic differentiation among forms was notably high despite the presence of hybrids, and simulations suggest either that hybridization at the rates we observed between forms is a relatively recent phenomenon or that the fitness of hybrids is reduced. This second scenario is further supported by the discovery of widespread differentiation between forms across the genome, indicating that much of the divergence observed may be driven by long-term reproductive isolation and drift.

### 4.1 | Hypotheses for high genetic differentiation among forms

Genetic differentiation of the three primary cisco forms in our study (*artedi*, *hoyi*, *kiyi*) was relatively high compared to previous research in cisco using allozymes, mtDNA, microsatellites, AFLPs, and RAD data, which has largely suggested that forms are not diverged within lakes (but see Stott et al., 2020; Turgeon, Estoup, & Bernatchez, 1999; Turgeon et al., 2016). For example, Piette-Lauzière, Bell, Ridgway, and Turgeon (2019) documented neutral  $F_{ST}$  values near or below 0.01 between forms in small lakes within Algonquin Provincial Park (Ontario, Canada) using RAD data, which

**FIGURE 6** Investigation of genetic differentiation on linkage group Cart21, the linkage group with the most significantly differentiated windows. (a) Genetic differentiation (overall  $F_{ST}$ ) at 351 loci that were placed on Cart21 in the cisco linkage map. The top ten loci with the highest  $F_{ST}$  are colored red. Red lines denote significantly differentiated windows. (b) Recombination distance on cisco linkage group Cart21 (y-axis) versus physical distance on Atlantic salmon chromosome Ssa05 (x-axis). (c) Genetic differentiation (overall  $F_{ST}$ ) at 152 loci from Cart21 that successfully aligned to Ssa05, the syntenic chromosome in Atlantic salmon. Six of the top ten loci from panel a aligned to Ssa05 and are colored red. Blue lines indicate the position of genes on Ssa05. (d) Allele frequencies of the loci with the highest  $F_{ST}$  from Cart21. Loci are ordered from highest  $F_{ST}$  (bottom) to lowest. Form abbreviations are found in Table 1



were much lower than the  $F_{ST}$  values of  $\sim 0.05$  observed in our study. Unfortunately, genetic data for cisco in the Great Lakes are relatively sparse; however, a previous study using mtDNA and microsatellites did not find strong evidence of differentiation among forms (Turgeon & Bernatchez, 2003). Results from the microsatellites genotyped in the current study are similar and indicate that these markers are unable to differentiate species in Lake Superior, even though genetic structure was relatively high according to the RAD data. Interestingly, our estimates of genetic divergence among forms more closely mirror two studies in lake whitefish and European whitefish that used RAD sequencing (Feulner & Seehausen, 2019; Gagnaire, Pavey, et al., 2013) than previous studies in cisco that genotyped mtDNA and microsatellites.

The high genetic divergence among forms observed in cisco is not typical of other fishes in the Laurentian Great Lakes. Lake trout (*Salvelinus namaycush*) and brook trout (*Salvelinus fontinalis*) display significant life history polymorphisms in the Great Lakes, with lake trout exhibiting morphologically distinct ecotypes related to depth and brook trout exhibiting both fluvial and adfluvial life histories. Two recent studies used RAD sequencing to investigate life history polymorphism in these species, and neither was able to document strong signals of divergence among forms (Elias, McLaughlin, Mackereth, Wilson, & Nichols, 2018; Perreault-Payette et al., 2017). It is possible that divergence within these forms was reduced through introgression mediated by stocking, as these species were stocked heavily, whereas cisco has not been stocked in consistently large numbers (Baillie, Muir, Scribner, Bentzen, & Krueger, 2016; Wilson et al., 2008). However, it is also possible that reproductive isolation among cisco forms is more complete, reducing the potential for introgression to erode divergence among forms.

Very little is known about the spawning biology of forms outside of *artedi*, although our observation of relatively frequent hybrids between ART-HOY and HOY-KIY suggests that there is at least some overlap in reproductive timing among the three forms. It is also notable that we did not observe any putative hybrids between ART-KIY. These three cisco forms are encountered in different depths in Lake Superior, with *artedi* inhabiting waters  $< 80$  m deep, *hoiy* inhabiting depths between 60 and 160 m, and *kiyi* inhabiting depths from 80 to 200 m (Eshenroder et al., 2016). This depth stratification likely explains our observation that *hoiy* hybridizes with *artedi* and *kiyi*, but *artedi* and *kiyi* do not hybridize with each other.

Our observation that hybrids appear to be relatively common even though genetic differentiation among forms is high may indicate that hybrids possess reduced fitness and potentially genetic incompatibilities. Successful hybridization can homogenize genetic structure within a few generations, as evidenced by our simulations and by a large body of literature in species such as European whitefish and cichlids (reviewed in Seehausen, 2006). For example, Vonlanthen et al. (2012) documented rapid speciation reversal in European whitefish in response to habitat degradation. It is possible that cisco forms in Lake Superior were historically more differentiated, and we are observing the beginning of reverse speciation. However, this is unlikely since Lake Superior has remained the least impacted relative to

the other Great Lakes, and cisco populations in Lake Superior remain robust (see Eshenroder et al., 2016). Alternatively, we hypothesize that lower fitness of hybrids caused by negative interactions between hybrid genomes, such as Dobzhansky-Muller incompatibilities (Dobzhansky, 1936; Muller, 1942), may be at least partially responsible for maintaining high genetic differentiation among forms (Dagilis, Kirkpatrick, & Bolnick, 2019; Ellison & Burton, 2008). Multiple lines of evidence for hybrid incompatibilities have been found in a sister taxon of cisco, lake whitefish (*Coregonus clupeaformis*, reviewed in Bernatchez et al., 2010). For example, Rogers and Bernatchez (2006) found that hybrid backcrosses had  $\sim 6$  times higher mortality than F1 crosses, Renaut, Nolte, and Bernatchez (2009) found disruption of gene expression pattern in backcrosses, and Whiteley, Persaud, Derome, Montgomerie, and Bernatchez (2009) documented reduced sperm performance in backcrosses. Our results combined with those from lake whitefish provide evidence that hybrid backcrosses may experience dramatically reduced fitness in cisco. However, future research is necessary to empirically test this hypothesis and investigate potential mechanisms for genomic incompatibilities in hybrid backcrosses.

## 4.2 | Genetic differentiation across the genome

Comparison of patterns of genomic divergence found in our study with previous research suggests that diversification of cisco forms in the Great Lakes is likely polygenic and that these forms may have been isolated with limited gene flow for a relatively long period of time. We identified over 100 significantly differentiated genomic windows in our study, and genetic differentiation among forms was consistently high across the genome. This result is similar to two other investigations of genomic divergence in coregonines, which also hypothesized that adaptive divergence among forms likely involves a large number of genes (Feulner & Seehausen, 2019; Gagnaire, Pavey, et al., 2013). Specifically, Feulner and Seehausen (2019) investigated genomic divergence in three species of whitefishes (*Coregonus* spp.) from two lakes in Switzerland and found high divergence on all chromosomes, and Gagnaire, Pavey, et al. (2013) assessed divergence between species pairs of lake whitefish across five lakes in the St. John River Basin (Maine, USA, and Québec, Canada) and also found consistently high differentiation across the genome. Additionally, there is a large body of research in lake whitefish demonstrating that traits that differ among species (e.g., body size, growth rate) are controlled by many quantitative trait loci (QTL) (Gagnaire, Normandeau, et al., 2013; Laporte et al., 2015; Rougeux, Gagnaire, Praebel, Seehausen, & Bernatchez, 2019). Most recently, Rougeux et al. (2019) documented numerous genes related to metabolism, immune response, and growth that are differently expressed among forms. Taken together, these studies provide strong evidence that diversification of coregonines typically involves genome-wide selection on a large number of phenotypic traits.

Even though we are relatively confident that divergence among cisco forms is the result of polygenic adaptation, differentiating

adaptively important genomic regions from neutrally evolving ones and disentangling the signatures of historic selection and contemporary drift is extremely challenging (Lotterhos & Whitlock, 2015; Noor & Bennett, 2009; Yeaman, 2015). We hypothesize that cisco forms in Lake Superior diverged in sympatry thousands of years ago, and our genome scans seem to reveal a complex history of both drift and selection consistent with this hypothesis. Sympatric speciation, even when involving many genes of small effect, is predicted to create highly differentiated regions (i.e., genomic islands of divergence), as selectively advantageous loci that are clustered together are protected from between population recombination (Via, 2012; Via & West, 2008; Yeaman & Whitlock, 2011). We believe the first 10–20 cM of LG21 may represent one such island, as this region contained the highest concentration of outlier loci in our dataset and displayed potential evidence of reduced recombination. However, genetic differentiation appeared to be relatively similar across many LGs, with peaks and valleys but few conspicuous regions of differentiation. It is likely that many of these peaks may have been caused by genetic drift resulting from relatively complete and long-term isolation among forms (Via, 2012). Taken together, our results suggest a heterogenous landscape of divergence across the genome that has likely been shaped by both selection and drift. We suggest that future studies attempting to disentangle selection from drift in this system compare cisco forms among multiple lakes and potentially even compare cisco to other coregonine species (e.g., Rougeux et al., 2019).

### 4.3 | Conservation implications

We documented high genetic differentiation among the three major cisco forms in Lake Superior and, based on this information, we suggest that separate conservation units could be constructed for each form. This strategy differs from the current conservation recommendation that the entire *C. artedii* species complex be considered *C. artedii sensu lato* (translation, in the broad sense) and that units of conservation should be designed to preserve environments that have facilitated the evolution of different forms (i.e., lakes) rather than on forms at a larger scale (Turgeon & Bernatchez, 2003; Turgeon et al., 2016). These recommendations were informed by the best available data, which up to this point, have been generated with commonly employed markers for the detection of taxonomic and conservation units—AFLPs, mtDNA, and microsatellites (Reed et al., 1998; Turgeon & Bernatchez, 2003; Turgeon et al., 2016)—none of which have been able to consistently resolve different forms within the Great Lakes (but see Stott et al., 2020 for a single lake example).

Our results suggest that “last generation” markers may be insufficient for capturing differentiation in evolutionarily young species, such as cisco, and highlight the utility of genomic data for designating conservation units in these species. However, it is important to note that we only surveyed animals from one lake, and a larger genotyping effort across the Great Lakes is necessary to accurately inform conservation units for Great Lakes ciscoes. Additionally, our ability to draw conclusions regarding the relationships of the rare forms

*nigripinnis* and *zenithicus* to the three major forms was limited by low sample size; therefore, genotyping additional contemporary and/or historic specimens may help resolve the placement these forms.

The potential of genomic data to revolutionize construction of conservation units has been frequently discussed (reviewed in Allendorf et al., 2010; Funk et al., 2012), and many studies have found that genomic data provide increased resolution for delineating population structure compared to last generation markers, such as microsatellites (Hodel et al., 2017; Vendrami et al., 2017; Wagner et al., 2013). However, conservation units that were constructed with these last generation markers have generally proven to be robust and are usually only updated slightly, if at all, based on genomic data (e.g., Hecht, Matala, Hess, & Narum, 2015; Larson et al., 2014; Moore et al., 2014). Our findings do not follow this pattern and suggest that current conservation unit recommendations for cisco may not accurately reflect the diversity of this species complex. We found that individual-based analyses conducted using genomic data were consistently able to differentiate cisco forms, whereas this was not possible with our dataset of nine microsatellites. We believe that this discrepancy largely stemmed from lack of power with the microsatellite dataset rather than, for example, heterogenous genomic divergence that was captured with genomic data but not with the microsatellites. The discrepancy that we observed between marker types has larger implications for constructing conservation units. Specifically, our data suggest that conservation units constructed based on data from small panels of last generation markers that did not document structure among groups of animals with different phenotypes could potentially be re-evaluated with genomic tools to ensure within-species diversity is being adequately conserved.

### 4.4 | Conclusions and future directions

Our study provides some of the first evidence that cisco forms within the Great Lakes are genetically differentiated. We documented high genetic differentiation among the three major forms in Lake Superior, and highly differentiated markers were distributed across the genome, with islands of divergence found on nearly every linkage group. Additionally, we identified putative hybrids but hypothesize that fitness breakdown of backcrosses may aid in maintaining differentiation among these forms. The results of this study provide the foundation for a new understanding of the ecology and evolution of the *C. artedii* species complex within the Great Lakes. The ability to differentiate forms with genomics provides researchers with a powerful tool for ground truthing morphological phenotypes and identifying cisco species at any life stage. In particular, the ability to identify larval ciscoes will allow researchers to estimate recruitment, which is vital for management and conservation, and will also significantly improve our understanding of early life history characteristics and reproductive dynamics in this species. Finally, our results suggest that some management units created using last generation markers may not adequately reflect species diversity and could be re-evaluated with genomic data.

## ACKNOWLEDGEMENTS

We thank Brad Ray with the Wisconsin Department of Natural Resources and Dan Yule and Mark Vinson with the United States Geological Survey for providing the Lake Superior samples used in this project. We are grateful to Kristen Gruenthal for laboratory support during RAD library preparation, to Garrett McKinney for discussions on analysis, and to Julie Turgeon, John Robinson, and Jared Homola for comments on the manuscript. This research was supported by the Turing High Performance Computing cluster at Old Dominion University and was funded by grants from the Great Lakes Restoration Initiative, the Great Lakes Fishery Trust, and the Great Lakes Fish and Wildlife Restoration Act. Any use of trade, product, or company name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## CONFLICT OF INTEREST

None declared.

## DATA AVAILABILITY STATEMENT

Demultiplexed RAD sequence data used in this research along with corresponding metadata were archived in the NCBI sequence read archive using the publicly accessible Genomic Observatories Metadatabase (GUID: <https://n2t.net/ark:/21547/DPK2>, NCBI: PRJNA600818), and microsatellite genotypes are archived on DRYAD (Ackiss, Larson, & Stott, 2020; <https://doi.org/10.5061/dryad.cjsxksn2m>).

## ORCID

Amanda S. Ackiss  <https://orcid.org/0000-0002-8726-7423>

## REFERENCES

- Ackiss, A., Larson, W., & Stott, W. (2020). Data from: Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes. Dryad Dataset, <https://doi.org/10.5061/dryad.cjsxksn2m>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (rapture): Flexible and efficient sequence-based genotyping. *Genetics*, 202(2), 389–400. <https://doi.org/10.1534/genetics.115.183665>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697. <https://doi.org/10.1038/nrg2844>
- Allendorf, F., & Luikart, G. (2007). *Conservation and the genetics of populations*. Oxford, UK: Blackwell Publishing Ltd.
- Als, T. D., Hansen, M. M., Maes, G. E., Castonguay, M., Riemann, L., Aarestrup, K., ... Bernatchez, L. (2011). All roads lead to home: Panmixia of European eel in the Sargasso Sea. *Molecular Ecology*, 20(7), 1333–1346. <https://doi.org/10.1111/j.1365-294X.2011.05011.x>
- Angers, B., Bernatchez, L., Angers, A., & Desgroseillers, L. (1995). Specific microsatellite loci for brook charr reveal strong population subdivision on a microgeographic scale. *Journal of Fish Biology*, 47(sa), 177–185. <https://doi.org/10.1111/j.1095-8649.1995.tb06054.x>
- Bailey, R. M., & Smith, G. R. (1981). Origin and geography of the fish fauna of the Laurentian Great Lakes basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(12), 1539–1561. <https://doi.org/10.1139/f81-206>
- Baillie, S. M., Muir, A. M., Scribner, K., Bentzen, P., & Krueger, C. C. (2016). Loss of genetic diversity and reduction of genetic distance among lake trout *Salvelinus namaycush* ecomorphs, Lake Superior 1959 to 2013. *Journal of Great Lakes Research*, 42(2), 204–216. <https://doi.org/10.1016/j.jglr.2016.02.001>
- Balloux, F. (2001). EASYPOP (Version 1.7): A computer program for population genetics simulations. *Journal of Heredity*, 92(3), 301–302. <https://doi.org/10.1093/jhered/92.3.301>
- Behnke, R. J. (1972). The systematics of salmonid fishes of recently glaciated lakes. *Journal of the Fisheries Board of Canada*, 29(6), 639–671. <https://doi.org/10.1139/f72-112>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Bernardi, G., & Lape, J. (2005). Tempo and mode of speciation in the Baja California disjunct fish species *Anisotremus davidsonii*. *Molecular Ecology*, 14(13), 4085–4096. <https://doi.org/10.1111/j.1365-294X.2005.02729.x>
- Bernatchez, L. (1996). Réseau de suivi environnemental du complexe la Grande. Caractérisation génétique des formes naines et normales de grand corégone du réservoir Caniapiscou et du lac Séigny à l'aide de marqueurs microsatellites. Retrieved from Rapport présenté par l'université Laval à la viceprésidence Environnement et Collectivités d'Hydro-Québec.
- Bernatchez, L., Colombani, F., & Dodson, J. (1991). Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. *Journal of Fish Biology*, 39, 283–290. <https://doi.org/10.1111/j.1095-8649.1991.tb05091.x>
- Bernatchez, L., & Dodson, J. J. (1994). Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(S1), 240–251.
- Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., ... St-Cyr, J. (2010). On the origin of species: Insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1547), 1783–1800. <https://doi.org/10.1098/rstb.2009.0274>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Blumstein, D. M., Campbell, M. A., Hale, M. C., Sutherland, B. J. G., McKinney, G. J., Stott, W., & Larson, W. A. (2019). Comparative genomic analyses and a novel linkage map for cisco (*Coregonus artedii*) provides insight into chromosomal evolution and rediploidization across Salmonids. *bioRxiv*, 834937. <https://doi.org/10.1101/834937> [Epub ahead of print]
- Bronte, C. R., Ebener, M. P., Schreiner, D. R., DeVault, D. S., Petzold, M. M., Jensen, D. A., ... Lozano, S. J. (2003). Fish community change in Lake Superior, 1970–2000. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(12), 1552–1574. <https://doi.org/10.1139/f03-136>
- Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, 6, 165. <https://doi.org/10.3389/fevo.2018.00165>
- Commission, G. L. F., & Christie, W. (1973). *Review of the Changes in the Fish species composition of Lake Ontario*.
- Conrad, D. F., Keebler, J. E. M., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., ... Genomes, P. (2011). Variation in genome-wide mutation rates within and between human families. *Nature Genetics*, 43(7), 712–714. <https://doi.org/10.1038/ng.862>

- Dagilis, A. J., Kirkpatrick, M., & Bolnick, D. I. (2019). The evolution of hybrid fitness during speciation. *PLoS Genetics*, *15*(5), e1008125. <https://doi.org/10.1371/journal.pgen.1008125>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214.
- Dobzhansky, T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila Pseudoobscura* hybrids. *Genetics*, *21*(2), 113–135.
- Dryer, W. R., & Beil, J. (1968). Growth changes of the bloater (*Coregonus hoyi*) of the Apostle Islands region of Lake Superior. *Transactions of the American Fisheries Society*, *97*(2), 146–158. [https://doi.org/10.1577/1548-8659\(1968\)97\[146:GCOTBC\]2.0.CO;2](https://doi.org/10.1577/1548-8659(1968)97[146:GCOTBC]2.0.CO;2)
- Elias, A., McLaughlin, R., Mackereth, R., Wilson, C., & Nichols, K. M. (2018). Population structure and genomic variation of ecological life history diversity in wild-caught Lake Superior brook trout, *Salvelinus fontinalis*. *Journal of Great Lakes Research*, *44*(6), 1373–1382. <https://doi.org/10.1016/j.jglr.2018.08.006>
- Ellison, C. K., & Burton, R. S. (2008). Interpopulation hybrid breakdown maps to mitochondrial genome. *Evolution*, *62*(3), 631–638. <https://doi.org/10.1111/j.1558-5646.2007.00305.x>
- Eshenroder, R. L., Vecsei, P., Gorman, O. T., Yule, D., Pratt, T. C., Mandrak, N. E., Muir, A. M. (2016). *Ciscoes (Coregonus, subgenus Leucichthys) of the Laurentian Great Lakes and Lake Nipigon*. Retrieved from <http://pubs.er.usgs.gov/publication/70179454>
- Estoup, A., Presa, P., Krieg, F., Vaiman, D., & Guyomard, R. (1993). (CT)n and (GT)n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity*, *71*(5), 488–496. <https://doi.org/10.1038/hdy.1993.167>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology*, *28*(3), 615–629. <https://doi.org/10.1111/mec.14977>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Gagnaire, P.-A., Normandeau, E., Pavey, S. A., & Bernatchez, L. (2013). Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, *22*(11), 3036–3048. <https://doi.org/10.1111/mec.12127>
- Gagnaire, P.-A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*(9), 2483–2497. <https://doi.org/10.1111/evo.12075>
- Garside, E., & Christie, W. (1962). Experimental hybridization among three coregonine fishes. *Transactions of the American Fisheries Society*, *91*(2), 196–200. [https://doi.org/10.1577/1548-8659\(1962\)91\[196:E-HATCF\]2.0.CO;2](https://doi.org/10.1577/1548-8659(1962)91[196:E-HATCF]2.0.CO;2)
- Goudet, J., Raymond, M., de Meeüs, T., & Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics*, *144*(4), 1933–1940.
- Hecht, B. C., Matala, A. P., Hess, J. E., & Narum, S. R. (2015). Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*, *24*(22), 5573–5595. <https://doi.org/10.1111/mec.13409>
- Hedrick, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, *59*(8), 1633–1638. <https://doi.org/10.1111/j.0014-3820.2005.tb01814.x>
- Hey, J., Waples, R. S., Arnold, M. L., Butlin, R. K., & Harrison, R. G. (2003). Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, *18*(11), 597–603. <https://doi.org/10.1016/j.tree.2003.08.014>
- Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, *38*, 7. <https://doi.org/10.1017/s0016672300020553>
- Himberg, K. (1970). A systematic and zoogeographic study of some North European Coregonids. *Biology of Coregonid Fishes*, 219–250.
- Hodel, R. G. J., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: Comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports*, *7*(1), 17598. <https://doi.org/10.1038/s41598-017-16810-7>
- Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evolutionary Applications*, *11*(9), 1732–1747. <https://doi.org/10.1111/eva.12676>
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. *PLoS Genetics*, *6*(2), e1000862. <https://doi.org/10.1371/journal.pgen.1000862>
- Hudson, A. G. (2011). *Population genomics and ecology of parallel adaptive radiations: The Alpine Lake Whitefish*. Dissertation, Bern, Switzerland:Verlag nicht ermittelbar.
- Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P., & Seehausen, O. (2017). Managing cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic gradient in Alpine whitefish (*Coregonus* spp.). *Evolutionary Applications*, *10*(3), 251–266.
- Huitfeldt-Kaas, H. (1918). *Ferskvandfiskenes utbredelse og innvandring i Norge medet tillæg om krebsen*. Oslo: Centraltrykkeriet.
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kahilainen, K. K., Østbye, K., Harrod, C., Shikano, T., Malinen, T., & Merilä, J. (2011). Species introduction promotes hybridization and introgression in *Coregonus*: Is there sign of selection against hybrids? *Molecular Ecology*, *20*(18), 3838–3855. <https://doi.org/10.1111/j.1365-294X.2011.05209.x>
- Kapfer, J. M., Sloss, B. L., Schuurman, G. W., Paloski, R. A., & Lorch, J. M. (2013). Evidence of Hybridization between Common Gartersnakes (*Thamnophis sirtalis*) and Butler's Gartersnakes (*Thamnophis butleri*) in Wisconsin, USA. *Journal of Herpetology*, *47*(3), 400–405. <https://doi.org/10.1670/12-057>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782–788. <https://doi.org/10.1111/2041-210X.12067>
- Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, *49*(4), 561–576.
- Koelz, W. N. (1926). *Fishing industry of the Great Lakes*.
- Koelz, W. N. (1929). Coregonid fishes of the Great Lakes. *Bulletin of the United States Bureau of Fisheries*, *43*(2), 297–643.
- Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A., Dalziel, A. C., ... Bernatchez, L. (2015). RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish

- (*Coregonus clupeaformis*) species pairs. *G3: Genes, Genomes, Genetics*, 5(7), 1481–1491.
- Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E., & Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular Ecology*, 26(2), 554–570. <https://doi.org/10.1111/mec.13933>
- Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, 7(3), 355–369. <https://doi.org/10.1111/eva.12128>
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533, 200. <https://doi.org/10.1038/nature17164>
- Lischer, H. E., & Excoffier, L. (2011). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031–1046. <https://doi.org/10.1111/mec.13100>
- Lu, G., Basley, D., & Bernatchez, L. (2001). Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, 10(4), 965–985. <https://doi.org/10.1046/j.1365-294X.2001.01252.x>
- Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77.
- Mann, G., & McCart, P. J. (1981). Comparison of sympatric dwarf and normal populations of least cisco (*Coregonus sardinella*) inhabiting Trout Lake, Yukon Territory. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(2), 240–244.
- Marie, A., Bernatchez, L., & Garant, D. (2011). Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (*Salvelinus fontinalis*). *Conservation Genetics*, 12, 1215–1227. <https://doi.org/10.1007/s10592-011-0224-y>
- McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, 17(4), 656–669. <https://doi.org/10.1111/1755-0998.12613>
- Mee, J. A., Bernatchez, L., Reist, J. D., Rogers, S. M., & Taylor, E. B. (2015). Identifying designatable units for intraspecific conservation prioritization: A hierarchical approach applied to the lake whitefish species complex (*Coregonus* spp.). *Evolutionary Applications*, 8(5), 423–441.
- Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11(1), 5–18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
- Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- Mohr, L., & Ebener, M. (2005). The coregonine community. Special Publication. *Great Lakes Fishery Commission*, 5, 69–76.
- Moore, J.-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., ... Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: Outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology*, 23(23), 5680–5697. <https://doi.org/10.1111/mec.12972>
- Muir, A. M., Vecsei, P., Pratt, T. C., Krueger, C. C., Power, M., & Reist, J. D. (2013). Ontogenetic shifts in morphology and resource use of cisco *Coregonus artedii*. *Journal of Fish Biology*, 82(2), 600–617. <https://doi.org/10.1111/jfb.12016>
- Muller, H. (1942). *Isolating mechanisms, evolution, and temperature*. Paper presented at the Biol. Symp.
- Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439. <https://doi.org/10.1038/hdy.2009.151>
- Olsen, M. T., Andersen, L. W., Dietz, R., Teilmann, J., Härkönen, T., & Siegmund, H. R. (2014). Integrating genetic data and population viability analyses for the identification of harbour seal (*Phoca vitulina*) populations and management units. *Molecular Ecology*, 23(4), 815–831.
- Østbye, K., Bernatchez, L., Naesje, T., Himberg, K. J., & Hindar, K. (2005). Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology*, 14(14), 4371–4387.
- Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, 4(3), 347–354. <https://doi.org/10.1111/j.1365-294X.1995.tb00227.x>
- Palm, S., Dannewitz, J., Prestegard, T., & Wickström, H. (2009). Panmixia in European eel revisited: No genetic difference between maturing adults from southern and northern Europe. *Heredity*, 103, 82. <https://doi.org/10.1038/hdy.2009.51>
- Palsbøll, P. J., Bérubé, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, 22(1), 11–16. <https://doi.org/10.1016/j.tree.2006.09.003>
- Palsbøll, P. J., Peery, M. Z., & Bérubé, M. (2010). Detecting populations in the 'ambiguous' zone: Kinship-based estimation of population structure at low genetic divergence. *Molecular Ecology Resources*, 10(5), 797–805. <https://doi.org/10.1111/j.1755-0998.2010.02887.x>
- Paris, J. R., Stevens, J. R., Catchen, J. M., & Johnston, S. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, 8(10), 1360–1373. <https://doi.org/10.1111/2041-210x.12775>
- Patton, J. C., Gallaway, B. J., Fechhelm, R. G., & Cronin, M. A. (1997). Genetic variation of microsatellite and mitochondrial DNA markers in broad whitefish (*Coregonus nasus*) in the Colville and Sagavanirktok rivers in northern Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(7), 1548–1556. <https://doi.org/10.1139/f97-062>
- Perreault-Payette, A., Muir, A. M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P., & Bernatchez, L. (2017). Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior. *Molecular Ecology*, 26(6), 1477–1497. <https://doi.org/10.1111/mec.14018>
- Piette-Lauzière, G., Bell, A. H., Ridgway, M. S., & Turgeon, J. (2019). Evolution and diversity of two cisco forms in an outlet of glacial Lake Algonquin. *Ecology and Evolution*, 9(17), 9654–9670. <https://doi.org/10.1002/ece3.5496>
- Praebel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., ... Amundsen, P.-A. (2013). Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution*, 3(15), 4970–4986. <https://doi.org/10.1002/ece3.867>
- Raymond, M., & Rousset, F. (1995). An exact test for population differentiation. *Evolution*, 49(6), 1280–1283.
- Reed, K. M., Dorschner, M. O., Todd, T. N., & Phillips, R. B. (1998). Sequence analysis of the mitochondrial DNA control region of ciscoes (genus *Coregonus*): Taxonomic implications for the Great Lakes species flock. *Molecular Ecology*, 7(9), 1091–1096. <https://doi.org/10.1046/j.1365-294x.1998.00419.x>
- Renaut, S., Nolte, A. W., & Bernatchez, L. (2009). Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (*Coregonus* spp. Salmonidae). *Molecular Biology and Evolution*, 26(4), 925–936. <https://doi.org/10.1093/molbev/msp017>



- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21), 4737–4754. <https://doi.org/10.1111/mec.15253>
- Rogers, S. M., & Bernatchez, L. (2006). The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, 19(6), 1979–1994. <https://doi.org/10.1111/j.1420-9101.2006.01150.x>
- Rogers, S. M., & Bernatchez, L. (2007). The Genetic architecture of ecological speciation and the association with signatures of selection in natural Lake Whitefish (*Coregonus* sp. Salmonidae) species pairs. *Molecular Biology and Evolution*, 24(6), 1423–1438. <https://doi.org/10.1093/molbev/msm066>
- Rogers, S. M., Marchand, M. H., & Bernatchez, L. (2004). Isolation, characterization and cross-salmonid amplification of 31 microsatellite loci in the lake whitefish (*Coregonus clupeaformis*, Mitchell). *Molecular Ecology Notes*, 4(1), 89–92. <https://doi.org/10.1046/j.1471-8286.2003.00578.x>
- Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of Lake Whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8), 2057–2074. <https://doi.org/10.1093/gbe/evx150>
- Rougeux, C., Gagnaire, P.-A., Praebel, K., Seehausen, O., & Bernatchez, L. (2019). Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Molecular Ecology*, 28(19), 4388–4403. <https://doi.org/10.1111/mec.15226>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Ryder, O. A. (1986). Species conservation and systematics: The dilemma of subspecies. *Trends in Ecology & Evolution*, 1(1), 9–10. [https://doi.org/10.1016/0169-5347\(86\)90059-5](https://doi.org/10.1016/0169-5347(86)90059-5)
- Schaeffer, J. S., & Warner, D. M. (2007). *Status and trends of pelagic prey fish in Lake Huron, 2007*. Ann Arbor, MI: US Geological Survey, Great Lakes Science Center Reports.
- Schluter, D. (1996). Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1341), 807–814.
- Schmidt, S. N., Harvey, C. J., & Vander Zanden, M. J. (2011). Historical and contemporary trophic niche partitioning among Laurentian Great Lakes coregonines. *Ecological Applications*, 21(3), 888–896. <https://doi.org/10.1890/09-0906.1>
- Schmidt, S. N., Vander Zanden, M. J., & Kitchell, J. F. (2009). Long-term food web change in Lake Superior. *Canadian Journal of Fisheries and Aquatic Sciences*, 66(12), 2118–2129. <https://doi.org/10.1139/F09-151>
- Schwartz, M. K., Luikart, G., & Waples, R. S. (2006). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22(1), 25–33. <https://doi.org/10.1016/j.tree.2006.08.009>
- Scott, W., & Crossman, E. (1998). *Freshwater fishes of Canada Oakville*. Cambridge, ON: Galt House Publishing.
- Seehausen, O. (2006). African Cichlid fish: A model system in adaptive radiation research. *Proceedings: Biological Sciences*, 273(1597), 1987–1998.
- Shields, B. A., Guise, K. S., & Underhill, J. C. (1990). Chromosomal and mitochondrial DNA characterization of a population of dwarf cisco (*Coregonus artedii*) in Minnesota. *Canadian Journal of Fisheries and Aquatic Sciences*, 47(8), 1562–1569.
- Sierszen, M. E., Hrabik, T. R., Stockwell, J. D., Cotter, A. M., Hoffman, J. C., & Yule, D. L. (2014). Depth gradients in food-web processes linking habitats in large lakes: Lake Superior as an exemplar ecosystem. *Freshwater Biology*, 59(10), 2122–2136. <https://doi.org/10.1111/fwb.12415>
- Smith, G. R., & Todd, T. N. (1984). Evolution of species flocks of fishes in north temperate lakes. In A. A. Echelle, & I. Kornfield (Eds.), *Evolution of species flocks* (pp. 45–68). Orono, ME: University of Maine Press.
- Smith, S. H. (1968). Species succession and fishery exploitation in the Great Lakes. *Journal of the Fisheries Board of Canada*, 25(4), 667–693. <https://doi.org/10.1139/f68-063>
- Smith, S. H. (1970). Species interactions of the alewife in the Great Lakes. *Transactions of the American Fisheries Society*, 99(4), 754–765. [https://doi.org/10.1577/1548-8659\(1970\)99<754:SIOTAI>2.0.CO;2](https://doi.org/10.1577/1548-8659(1970)99<754:SIOTAI>2.0.CO;2)
- Stott, W., MacDougall, T., Roseman, E. F., Lenart, S., Chiotti, J., Yule, D. L., & Boase, J. (2020). *Genetic Species Identification of Coregonines from Juvenile Assessment and Commercial Fisheries in Lake Erie, Detroit River and St. Clair River*. Paper presented at the 13th International Coregonid Symposium, Bayfield, Wisconsin, USA.
- Svärdson, G. (1949). The coregonid problem. I. Some general aspects of the problem. *Report: Institute of Fresh-water Research, Drottningholm*, 29, 89–101.
- Svärdson, G. (1979). *Speciation of Scandinavian Coregonus [fresh water fish, Sweden]*. Drottningholm, Sweden: Report-Institute of Freshwater Research.
- Svärdson, G. (1998). Postglacial dispersal and reticulate evolution of Nordic coregonids. *Nordic Journal of Freshwater Research*, 74, 3–32.
- Todd, T. N. (1998). *Environmental modification of gillraker number in coregonine fishes*. Biology and management of Coregonid, fishes-1996, 305–315.
- Todd, T. N., & Smith, G. R. (1992). A review of differentiation in Great Lakes ciscoes. *Polskie Archiwum Hydrobiologii*, 39(3–4), 261–267.
- Todd, T. N., Smith, G. R., & Cable, L. E. (1981). Environmental and genetic contributions to morphological differentiation in ciscoes (Coregoninae) of the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(1), 59–67. <https://doi.org/10.1139/f81-008>
- Turgeon, J., & Bernatchez, L. (2003). Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp. (Teleostei: Salmonidae): Implications for the conservation of an evolutionary legacy. *Conservation Genetics*, 4(1), 67–81.
- Turgeon, J., Estoup, A., & Bernatchez, L. (1999). Species flock in the North American Great Lakes: Molecular ecology of lake Nipigon ciscoes (Teleostei: Coregonidae: Coregonus). *Evolution*, 53(6), 1857–1871. <https://doi.org/10.2307/2640446>
- Turgeon, J., Reid, S. M., Bourret, A., Pratt, T. C., Reist, J. D., Muir, A. M., & Howland, K. L. (2016). Morphological and genetic variation in Cisco (*Coregonus artedii*) and Shortjaw Cisco (*C. zenithicus*): Multiple origins of Shortjaw Cisco in inland lakes require a lake-specific conservation approach. *Conservation Genetics*, 17(1), 45–56.
- Vendrami, D. L. J., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., ... Hoffman, J. I. (2017). RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity. *Royal Society Open Science*, 4(2), 160548. <https://doi.org/10.1098/rsos.160548>
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17(19), 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., ... Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385), 357. <https://doi.org/10.1038/nature10824>
- Vuorinen, J. A., Bodaly, R. A., Reist, J. D., Bernatchez, L., & Dodson, J. J. (1993). Genetic and Morphological Differentiation between Dwarf and Normal Size Forms of Lake Whitefish (*Coregonus clupeaformis*)

- in Como Lake, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(1), 210–216. <https://doi.org/10.1139/f93-023>
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., ... Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22(3), 787–798. <https://doi.org/10.1111/mec.12023>
- Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci\*. *Conservation Genetics*, 7(2), 167–184. <https://doi.org/10.1007/s10592-005-9100-y>
- Waples, R. S., Antao, T., & Luikart, G. (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, 197(2), 769–780. <https://doi.org/10.1534/genetics.114.164822>
- Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 3(3), 244–262.
- Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, 117(4), 233–240. <https://doi.org/10.1038/hdy.2016.60>
- Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications*, 11(9), 1518–1526. <https://doi.org/10.1111/eva.12687>
- Weigel, D., Adams, J., Jepson, M., Waits, L., Caudill, C., & Weigel, D. (2018). Introgressive hybridization between native and hatchery-origin non-native steelhead (*Oncorhynchus mykiss*). *Aquatic Conservation Marine and Freshwater Ecosystems*, 29, <https://doi.org/10.1002/aqc.3028>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Wells, L., & McClain, A. (1973). *Lake Michigan: man's effects on native fish stocks and other biota Great Lakes Fishery Commission Technical Report No. 20*.
- Whiteley, A. R., Persaud, K. N., Derome, N., Montgomerie, R., & Bernatchez, L. (2009). Reduced sperm performance in backcross hybrids between species pairs of whitefish (*Coregonus clupeaformis*). *Canadian Journal of Zoology*, 87(7), 566–572. <https://doi.org/10.1139/Z09-042>
- Wilson, C. C., Stott, W., Miller, L., D'Amelio, S., Jennings, M. J., & Cooper, A. M. (2008). Conservation Genetics of Lake Superior Brook Trout: Issues, Questions, and Directions. *North American Journal of Fisheries Management*, 28(4), 1307–1320. <https://doi.org/10.1577/M05-190.1>
- Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, 186(S1), S74–S89. <https://doi.org/10.1086/682405>
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, 65(7), 1897–1911. <https://doi.org/10.1111/j.1558-5646.2011.01269.x>
- Yule, D. L., Addison, P. A., Evrard, L. M., Cullis, K. I., & Cholwek, G. A. (2009). *2008 spawning cisco investigations in the Canadian waters of Lake Superior*. US Geological Survey Technical Report 4-08.
- Yule, D., Stockwell, J., Evrard, L., Cholwek, G., & Cullis, K. (2006). *Comparison of commercial landings of cisco to acoustic estimates of abundance in Thunder Bay and Black Bay, Ontario*. Ann Arbor, MI: US Geological Survey, Great Lakes Science Center.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Ackiss AS, Larson WA, Stott W.

Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the

Laurentian Great Lakes. *Evol Appl*. 2020;13:1037–1054. <https://doi.org/10.1111/eva.12919>