Transcriptomic divergence predicts morphological and ecological variation underlying an adaptive radiation

Moisés A. Bernal^{1,2*}, Daniel L. Yule³, Wendylee Stott⁴, Lori Evrard³, Thomas E. Dowling⁵, Trevor J. Krabbenhoft^{2*}.

¹*Department of Biological Sciences, College of Science and Mathematics, Auburn University, Auburn, Alabama 36849, United States of America*

²*Department of Biological Sciences and RENEW Institute, University at Buffalo, Buffalo, NY 14260, United States of America*

³*U.S. Geological Survey, Great Lakes Science Center – Lake Superior Biological Station, 2800 Lake Shore Drive E., Ashland, WI 54806, United States of America*

⁴*Michigan State University CESU working for U.S. Geological Survey, Great Lakes Science Center, 1451 Green Road, Ann Arbor, MI 48105-2807, United States of America*

⁵*Wayne State University, Department of Biological Sciences, Detroit, Michigan, 48202, United States of America*

Keywords: salmonid, speciation, adaptation, morphometrics, osteology, selection, stable isotopes, hybridization.

*Corresponding authors:

Moisés A. Bernal (mab0205@auburn.edu)

Trevor J. Krabbenhoft (tkrabben@buffalo.edu)

Abstract

Groups of sympatric taxa with low inter-specific genetic differentiation, but considerable ecological differences, offer great opportunities to study the dynamics of divergence and speciation. This is the case of ciscoes (*Coregonus* spp.) in the Laurentian Great Lakes, which are characterized by a complex evolutionary history and are commonly described as having undergone an adaptive radiation. In this study, morphometrics, stable isotopes and transcriptome sequencing were used to study the relationships within the *Coregonus artedii* complex in western Lake Superior. We observed general concordance for morphological, ecological and genomic variation, but the latter was more taxonomically informative as it showed less overlap among species in multivariate space. Low levels of genetic differentiation were observed between individuals morphologically identified as *C. hoyi* and *C. zenithicus*, and we hypothesize this could be associated with recent hybridization between the two species. Transcriptome-based single nucleotide polymorphisms exhibited significant divergence for genes associated with vision, development, metabolism and immunity, among species that occupy different habitats. This study highlights the importance of using an integrative approach when studying groups of taxa with a complex evolutionary history, as individual-level analyses of multiple independent datasets can provide a clearer picture of the patterns and processes associated with the origins of biodiversity.

INTRODUCTION:

The speciation continuum offers exciting opportunities to understand the mechanisms driving the accumulation of genetic and ecological divergence among closely related groups. This is especially relevant when ecological opportunities can promote and/or maintain differentiation among such groups, even in the presence of gene flow (Arnold, 2006; Fitzpatrick, Gerberich, Kronenberger, Angeloni, & Funk, 2015). Hybridization during early stages of differentiation can complicate the identification of divergent groups, as selectively neutral alleles may be freely exchanged between forms (Hohenloe et al., 2013; Feulner & Seehausen, 2019). Conversely, genes associated with traits of reproductive or ecological advantage (Hench, Vargas, Höppner, McMillan, & Puebla, 2019; Meier, Marques, Wagner, Excoffier, & Seehausen, 2018; Richards & Martin, 2017) are expected to be far more resistant to gene flow among groups. This scenario of divergence can be further complicated by sporadic or cyclical environmental disturbances, which can disrupt the ecological barriers maintaining isolation between closely related forms (Feulner & Seehausen, 2019; Ohlberger, Mehner, Staaks, & Hölker, 2008), or affect the geographic barriers that hinder gene flow (Turgeon & Bernatchez, 2003; Wilson & Bernatchez, 1998). These patterns are commonly observed in lacustrine environments, such as the Laurentian Great Lakes (hereafter the Great Lakes), where multiple closely related groups/forms occur in sympatry (e.g., *Coregonus* spp., *Salvelinus* spp.).

Aquatic organisms in the Great Lakes are characterized by a complex evolutionary history as a result of Pleistocene glaciation events. Glaciers covered most areas currently associated with the Great Lakes during the Pleistocene, leading to the isolation of lakes that remained ice-free (Bailey & Smith, 1981). The retreat of the ice sheets, which started

approximately 14,000 years before present, led to the formation of lakes and rivers that promoted connectivity between the isolated refugia (Bailey & Smith, 1981). These cycles of vicariance followed by subsequent hypothesized admixture resulted in fish groups that have led to complex phenotypic and genomic patterns and complicated taxonomic relationships (Bernatchez & Dodson, 1991; Turgeon, Estoup, & Bernatchez, 1999; Wilson & Bernatchez, 1998). One example is the ciscoes in the genus *Coregonus* (subgenus *Leucichthys*), a group that contains several species characterized by distinct morphological and ecological traits. An early study of this group recognized eight distinct Great Lakes species in the *C. artedi* complex: *C. alpenae*, *C. artedi*, *C. johannae*, *C. kiyi*, *C. hoyi*, *C. nigripinnis*, *C. reighardi* and *C. zenithicus*, as well as several sub-species (Koelz, 1929). Later work has resulted in considerable debate over the systematics of the group, due to phenotypic and genetic similarity among some of the species (Todd, Smith, & Cable, 1981; Turgeon & Bernatchez, 2003; Turgeon, Estoup, & Bernatchez, 1999).

During the 20th century, human mediated effects such as overfishing, pollution, habitat degradation and invasive species directly affected the Great Lakes (Mills, Leach, Carlton, & Secor, 1994; Regier, Whillans, Christie, & Bocking, 1999). These changes led to population declines and extinctions of many coregonines (Smith, 1968). For example, it is thought that overfishing led to the extinction of six deep-water cisco forms in Lake Michigan (Eshenroder et al., 2016). Moreover, there has been a substantial decline in the abundance of deep-water forms in Lake Superior since the 1920s, especially *C. zenithicus* (Bronte et al., 2010; Hoff & Todd, 2004; Pratt, 2012). In many instances, population declines and extinctions have been followed by expansion of more abundant species into habitats previously occupied by other groups. Such is the case for *C. hoyi* which now occurs in deep waters that were once occupied by other species in

lakes Michigan and Huron (Bunnell et al., 2012). Studies in marine and freshwater fishes have shown that habitat degradation and other types of environmental change, along with fluctuations in population density can lead to blurring of ecological and behavioral barriers that keep closely related groups separated when they lack reproductive isolation (Seehausen, 2006; Taylor et al., 2006). Over time, the erosion of pre-mating mechanisms can lead to loss of taxonomic diversity via “speciation reversal” (Zhang, Thibert-Plante, Ripa, Svanbäck, & Brännström, 2019). It has also been suggested that reduction in population densities for one taxon can lead to increased rates of hybridization with other more abundant and closely related groups (i.e., Hubbs principle; Hubbs, 1955), as it becomes challenging for the rare species to find mates of their own kind. Hybridization can quickly exacerbate the loss of genetic diversity caused by the initial population declines (Arnold, Bulger, Burke, Hempel, & Williams, 1999). Previous studies have suggested that hybridization between closely related groups could have accelerated the extinction of cisco forms, representing a loss of biodiversity (Todd & Stedman, 1989). Given the sympatric nature of Lake Superior ciscoes and their low levels of genetic differentiation (Ackiss, Larson, & Stott, 2020; Turgeon & Bernatchez 2003; Turgeon, Estoup, & Bernatchez, 1999), it is important to understand the phenotypic and genetic differences that characterize the extant members of the *C. artedi* complex, to improve the understanding of their evolutionary history and to inform conservation efforts in Lake Superior.

Today, up to six species are thought to be present in Lake Superior (Eshenroder et al., 2016), and three are routinely caught with standard trawl or gillnet sampling: *C. artedi*, *C. hoyi* and *C. kiyi* (US Geological Survey, 2018). The most abundant is *C. kiyi* (Yule et al., 2013) and it tends to occupy bathymetric depths greater than 100m (Rosinski, Vinson, & Yule, 2020) and predominantly feeds on *Mysis diluviana* (Gamble, Hrabik, Stockwell, & Yule, 2011a;

Ahrenstorff et al., 2011). The second most abundant form is *C. artedi* (Yule et al., 2013) which is found near the lake surface, especially during spring and summer (Stockwell, Yule, Gorman, Isaac, & Moore, 2006; Yule et al., 2013; Rosinski, Vinson, & Yule, 2020). The primary diet of adult *C. artedi* consists of *Mysis*, cladocerans and calanoid copepods (Gamble, Hrabik, Stockwell, & Yule, 2011a, b; Ahrenstorff, Hrabik, Stockwell, Yule & Sass, 2011). *Coregonus hoyi* is the third most abundant form and can be associated with the lakebed during the day and found in waters of the hypolimnion at night (Yule et al., 2013). They primarily consume calanoid copepods and cladocerans as juveniles, while adults utilize a diverse diet of zooplankton, *Diporeia* spp. and *Mysis* (Gamble, Hrabik, Stockwell, & Yule, 2011b; Sierszen et al., 2014). *Coregonus zenithicus* is also extant but is thought to be extremely rare (Bronte et al., 2010; Hoff & Todd, 2004; but see Pratt, 2012), while the status of *C. nigripinnis* and *C. reighardi* is uncertain (Eshenroder et al, 2016; Todd & Smith, 1980). Previous studies in Lake Superior using stable isotopes over three time-frames (1897-1929, 1934-1966, and 1972-1998) concluded that *C. zenithicus* and *C. nigripinnis* had a high degree of overlap in both trophic position and their use of basal carbon resources, while *C. reighardi* was distinct from these other two species (Schmidt et al., 2009). Overall, it is estimated that *C. zenithicus* consumes primarily *Mysis*, *Diporeia* spp., calanoid copepods and cladocerans (reviewed by Eshenroder et al., 2016). However, due to their rarity, it has not been possible to include them in more recent studies.

Due to their considerable ecological differences, the *Coregonus artedi* species complex provides a unique opportunity to explore mechanisms that lead to the differentiation and maintenance of closely related sympatric forms. There have been many previous genetic studies of the *Coregonus* species complex in the Great Lakes, and most have relied on putatively neutral markers (i.e., mitochondrial DNA and microsatellites). Some studies have identified significant

differences associated with geographic regions but have failed to identify differences among species (Turgeon & Bernatchez, 2003; Turgeon, Estoup, & Bernatchez, 1999). A more recent study used single nucleotide polymorphisms (SNPs) generated by RADseq to identify genetic differentiation among forms in Lake Superior, but did not include morphological or ecological data (Ackiss, Larson, & Stott, 2020). Since morphological and ecological divergence is partially heritable in this group (Todd & Smith, 1980), we anticipate that there is underlying functional genetic divergence in some parts of the genome, which could be identified by comparing the transcriptomes of species in the *Coregonus artedi* complex. An initial study examined functional differences in opsin genes across the complex, and found evidence of adaptive molecular evolution in *rhodopsin*, consistent with differences in depth preferences among forms (Eaton et al. 2020), suggesting that functional genetic differences might also be present in other gene families.

In this study, we aim to understand if there is concordance between morphology, feeding habits, and genomic divergence among four species of *Coregonus* from Lake Superior: *C. artedi*, *C. kiyi*, *C. hoyi*, and *C. zenithicus*. Despite ongoing debates on whether *C. nigripinnis* and *C. reighardi* should be synonymized with *C. zenithicus* (Eshenroder et al., 2016; Todd & Smith, 1980), these groups were not included in the present study as they are very rare in Lake Superior. Specific questions for our study include: 1) What are the main morphological traits that characterize different species? 2) What are the main ecological traits that differentiate the deep-water ciscoes *C. hoyi* and *C. zenithicus*? 3) Which highly divergent genes are associated with ecological and morphological differences? By understanding the phenotypic, ecological, and genomic differences among these species, this study presents an integrative view of the divergence of fishes of the Laurentian Great Lakes.

159

160 **METHODS**

161 **Sample collections**

162 Individuals of *C. artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* were collected between May and
 163 November of 2015 from multiple sites throughout western Lake Superior (Figure 1, Table S1).
 164 Identification of individuals followed Koelz (1929) and Eshenroder et al. (2016). Extant forms of
 165 *C. hoyi* and *C. zenithicus* were identified based on length of the lower jaw (lower jaw is shorter
 166 for *C. zenithicus*; Eshenroder et al., 2016) and a cursory examination of the number of gill-rakers
 167 (*C. zenithicus* has 35-40 and *C. hoyi* 41-45; Eshenroder et al., 2016). Captured fish were
 168 identified to species, euthanized by pithing, and dissected to obtain tissues for RNA isolation. In
 169 total, eight individuals per species were included in all analyses, except for *C. zenithicus* (n=7;
 170 see Results). All sampling and handling of fish was done following the guidelines for the care
 171 and use of fishes by the American Fisheries Society (AFS 2014).

172

173 **Morphometrics**

174 After dissection for tissues that would be used in RNAseq analyses, fish were placed on ice until
 175 they could be photographed. Fishes were placed on a mesh on top of a wooden frame with a ruler
 176 for scale (mm). Attempts were made to position specimens in a natural position, with the mouth
 177 slightly opened (Eshenroder et al., 2016). Dorsal, adipose, caudal and anal fins were extended
 178 with pins to allow for their measurement. Pelvic and pectoral fins were positioned so the
 179 horizontal and distal ends were clearly visible. A card with the specimen number and species
 180 field identification was included in each photograph. A digital image of the whole body was
 181 taken, and each image was scrutinized to ensure that all landmarks needed for body

measurements were visible. Individuals were then frozen at -20°C to preserve tissues for the stable isotope analysis.

Digital images were used to obtain 11 body measurements described in Eshenroder et al. (2016): body depth (BDD), dorsal fin height (DOH), head length (HLL), maxillary length (MXL), orbital length (OOL), pelvic-anal distance (PAD), pectoral fin length (PCL), preorbital length (POL), pelvic fin length (PVL), pectoral-pelvic distance (PPD), and standard length (STL; Table S2). The lengths of the PCL and PVL were divided by PPD and PAD, respectively, providing fin-length-to-body-distance ratios. These ratios have been shown to be useful for distinguishing *Coregonus* species (Koelz, 1929; Eshenroder et al., 2016). The premaxillary angle (PMA) was also measured with a protractor but was not included in the PCA analysis because a cursory analysis showed PMA was taxonomically uninformative.

The image analysis software SigmaScan Pro V5.0.0 was used to obtain all measurements, calibrating the distances with the ruler included in each image. Body measurements (in mm) were transformed to their natural log. A principal component analysis (PCA) was used to capture the maximum amount of variation in just a few dimensions, using the *prcomp* function in R, with default settings (R Core Team 2013). To remove the size component from the morphometric measures, the covariance matrix was estimated, and the first principal component of these measures was used as a general measure of size (Dos Reis, Pessôa, & Strauss, 1990). Except for the two fin ratios, all transformed measurements were regressed against the first principal component to obtain size-free residuals that we analyzed in a second PCA.

Stable Isotope Analysis

We used stable isotope ratios ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) to estimate isotopic niches and trophic overlap of the identified forms. Individuals stored at -20°C were thawed and white muscle tissue was removed posterior to the dorsal fin. Skin was removed and the sample rinsed in deionized water, dried at $50\text{--}60^{\circ}\text{C}$, homogenized, and $0.5\text{--}1.0\text{mg}$ was packed into tin capsules. Samples were analyzed at the University of California—Davis Stable Isotope Facility (<http://stableisotopefacility.ucdavis.edu/>). To accurately interpret the $\delta^{13}\text{C}$ values it is important to normalize for lipid content, especially when $\text{C:N}_{\text{bulk}} > 4$ (Hoffman, Sierszen, & Cotter 2015). We applied the arithmetic-mass-balance normalization of Hoffman, Sierszen, & Cotter (2015) to all samples: $\delta^{13}\text{C}_{\text{lipid free}} = \delta^{13}\text{C}_{\text{bulk}} + (\Delta\delta^{13}\text{C}_{\text{lipid}} * (\text{C:N}_{\text{lipid free}} - \text{C:N}_{\text{bulk}})) / \text{C:N}_{\text{bulk}}$. For this equation, C:N_{bulk} was molar, $\text{C:N}_{\text{lipid free}} = 3.8$ and the lipid discrimination ($\Delta\delta^{13}\text{C}_{\text{lipid}}$) = -6.4‰ . The $\delta^{15}\text{N}$ values were not corrected to allow for detection of $\delta^{15}\text{N}$ enrichment, which occurs with increasing depth in Lake Superior benthos but not zooplankton (Sierszen et al., 2006; 2014).

Overlap of isotopic niches (*sensu* Newsome, Martinez del Rio, Bearhop, & Phillips, 2007) was examined using Stable Isotope Bayesian Ellipses, in R-SIBER v2.1.3 (Jackson, Inger, Parnell, & Bearhops, 2011). Corrected standard ellipse areas (SEAc) were developed so that a subsequently sampled data point would have a 95% probability of being encompassed by a given SEAc. This was accomplished by multiplying the semi-major and -minor axes by 2.45 (Jackson, Inger, Parnell, & Bearhop, 2011). The percentage of overlap between form ellipses were calculated by using the *maxLikOverlap* function in R (R Core Team, 2013).

Transcriptome sequencing, assembly and SNP analysis.

For the transcriptomic analyses, we collected tissues that are expected to play important roles in functional differentiation among *Coregonus* species: the right half of upper and lower jaws and

associated muscle and skin, first gill arch, eye, and brain. Tissue samples were obtained from the same individuals used in the morphological and dietary analyses immediately after pithing. These tissues are actively growing in adult fish, and therefore are expected to show variation based on morphological and ecological differences of the analyzed species. Collected tissues were preserved in RNAlater (Ambion, Inc.) and similar amounts were homogenized and pooled for the RNA extractions. Total RNA was extracted from each homogenate using TRIzol (Life Technologies, Inc.) and RNA Mini Isolation Kits (Ambion, Inc.) following manufacturers' instructions (including treatment with DNase). RNA quality was assessed with a Bioanalyzer 2100 and quantified using a Qubit 2.0 fluorometer (Life Technologies, Inc.). Library preparation was done at the RTSF Genomics Core facility at Michigan State University, using Illumina TruSeq RNA library kits, with Illumina Ribo-Zero Gold for rRNA depletion. Libraries were normalized and pooled for multiplex sequencing, and the pool was loaded on both lanes of an Illumina HiSeq 2500 Rapid Run flow cell (paired-end, 250bp). Removal of Illumina adaptors and quality filtering ($Q > 30$) was done with TrimGalore (Krueger, 2015). Resulting sequences were used to assemble a reference transcriptome for each species using Trinity (Grabherr et al., 2011), and only contigs longer than 300bp were maintained. Contigs were summarized using Transdecoder (Haas et al., 2013) which only retains samples with predicted open reading frames. CDHIT (Li & Godzik, 2006) was used to cluster contigs by similarity (95%), matching to the most similar cluster in the dataset. The completeness of *C. artemis* transcriptomes (Trinity vs. summarizing with Transdecoder and CDHIT) were assessed with the Benchmarking Universal Single Copy Ortholog database version 3 (BUSCO; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), using the Actinopterygii database (*actinopterygii_odb9*). This resulted in 4,049 complete BUSCOs (874 single-copy and 3,175 duplicated) and 200 missing for the

Trinity/CDHIT transcriptome, and 3,876 complete BUSCOs (977 single-copy and 2899 duplicated) and 303 missing BUSCOs for the Transdecoder transcriptome (Figure S1). Given that the Transdecoder approach greatly reduced the redundancy of reads while only minimally reducing the completeness of the transcriptome, this version was chosen for downstream analyses. Annotation of the four transcriptomes was done with BLAST search algorithm, using the Uniprot (accessed January 2019) and Trembl (accessed January 2019) databases as reference and an e-value of 1e-10 and below as cutoff. Additionally, transcripts were annotated using Blastn searches against cDNA sequences from six related fish genomes in Ensembl (Ensembl.org): *Danio rerio*, *Esox lucius*, *Hucho hucho*, *Salmo salar*, *Salmo trutta*. Default search parameters were used for Blastn searches except -evalue 1e-5 and -max_target_seqs 20. Blast hits with <70% similarity and <100 bp alignment were removed and the hit with the highest HSP of the remaining alignments was retained. Annotation to gene ontology (GO) was obtained with the UniprotKB Retrieve/IDmapping (available at: <https://www.uniprot.org/uploadlists/>), using the UniProt and Trembl gene names as query and exporting a table with the multiple GO terms per gene name.

To compare sequence variation within and among species, we called single nucleotide polymorphisms (SNPs) in the sequenced protein coding loci or adjacent untranslated regions (UTRs) for each individual. Paired-end sequences of all individuals were mapped to the transcriptome of *C. artemis*, using Bowtie2 (Langmead & Salzberg, 2012) with default parameters. The rate of mapping to the *C. artemis* transcriptome was >82% for all individuals. SNP calling was done using the BCFtools platform (Li et al., 2009), with the commands “mpileup” (-a 1, -C50, -Q25), and “call” (only keeping variant sites, excluding indels). The accessory script *vcfutils.pl* was used to retain sites with coverage higher than 15x, and *VCFtools*

(Danecek et al., 2011) was used to keep SNPs separated by 1,000bp, that were present in all individuals. This resulted in a matrix of 22,285 variant SNPs, with no missing data.

Observed (H_o) and expected (H_e) heterozygosity, number of alleles (N_a), inbreeding coefficient (F_{is}), and genetic diversity (θ) were estimated with “*basic.stats*” of the R package *Pegas* (Paradis, 2010) to quantify levels of genetic variation. The package *Genodive* (Meirmans & Tienderen, 2004) was used to estimate the Fixation Index (F_{st}) for pairwise comparisons between the different species. This approach includes the Bonferroni correction of p -values for multiple comparisons (5,000 iterations). *VCFtools* was used to estimate pairwise F_{st} by individual loci between the different groups.

In addition, two separate analyses were performed to assess the degree of concordance between genomic markers and the results of the morphological analyses. First, a PCA was performed with the R package *pcadapt* (Luu, Bazin, & Blum, 2017), using the transcriptome covariance matrix and compared to PCA results from the morphological data. Second, individuals were divided into groups based on their ancestry using a maximum likelihood approach with the program *Ohana* (Cheng, Mailund, & Nielsen, 2016). The analysis was run for 100 iterations (stable likelihoods were observed after 23 iterations), and resulting estimates were used to generate a population assignment plot using Python scripts included in *Ohana*. To determine if these lineages were genetically admixed, we implemented the three-population analysis (f_3 statistics; Reich, Thangaraj, Patterson, Price, & Singh, 2009) using Treemix (Pickrell et al., 2012). Here, the genetic composition of one coregonine species was compared to two others, and admixture was implied if the z -score was negative (Reich, Thangaraj, Patterson, Price, & Singh, 2009).

The program *Ohana* (Cheng, Racimo, & Nielsen, 2019) was also used to determine if

some of the observed SNPs may be under divergent selection among species. Potential candidate loci were considered to be under selection when the maximum likelihood algorithm detects large differences in allele frequencies among species (i.e., Log-likelihood ratios >10; Cheng, Racimo, & Nielsen, 2019). Individual candidate loci that exhibited elevated F_{st} and/or were identified as evolving under divergent selection with *Ohana* were examined to determine whether any GO categories were significantly overrepresented among highly divergent genes using a Mann-Whitney U test of ranks (GO-MWU) with Benjamini-Hochberg correction (Wright, Aglyamova, Meyer, & Matz, 2015; scripts available: https://github.com/z0on/GO_MWU). This analysis was performed for the highly differentiated genes between populations, as well as the genes that may be under divergent selection for all species, detected with *Ohana*. This test determines if there is a significant enrichment of GO categories among the top values of the distribution (i.e., highest F_{st}). Only categories that are represented by five or more transcripts were taken into consideration for the analyses, and the cutoff for false discovery rate was 10%. For each comparison, the enrichment was done separately for the domains Biological Process (BP), Cellular Component (CC) and Molecular Function (MF).

Because salmonids underwent an ancestral whole genome duplication ~88 Ma (Lien et al., 2016), there is a possibility that some of the identified SNPs are located in paralogous rather than orthologous loci (i.e., ohnologs). This could inflate estimates of divergence between species. In order to confirm the validity of the results, we filtered the transcripts to only include putatively orthologous sequences between the four target groups of our study and the more distantly related European whitefish *C. lavaretus* (NCBI SRA: SRR6321817-SRR6321824; Carruthers et al., 2018) using *Orthofinder* (Emms & Kelly, 2015). The resulting SNP matrix was used to re-estimate the pairwise F_{st} following the steps described above.

RESULTS

One individual of *C. zenithicus* always clustered with samples of *C. kiyi* for both genetic and morphological traits (eye diameter in particular; Figure S2). Upon further examination of the photographs, it was determined this individual was a *C. kiyi*, and due to this initial misidentification, this individual was excluded from the results below to avoid any confusion.

Morphology

Great Lakes coregonines are challenging to identify based on morphology. The PCA showed that variables with greatest influence on PC1 (Figure 2a; Table 1) were paired fin lengths (PCL and PVL), paired-fin length to body length ratios (PCL/PPD & PVL/PAD), and eye orbit size (OOL). *Coregonus kiyi* had the longest paired fins, *C. artedi* had the shortest, and *C. hoyi* and *C. zenithicus* were intermediate (Table 1). PCA scores clearly separated samples of *C. kiyi* and *C. artedi* and were most influenced by PVL/PAD, PCL and BDD. The morphometric analysis could not differentiate samples of *C. zenithicus* and *C. hoyi* included in our study, as they showed considerable overlap for the evaluated traits (Figure 2a).

Isotopic Niche Breadth and Trophic Overlap

Stable isotope measurements made it possible to differentiate among most of the species (Figure 2b). There was clear separation of convex hulls in isotopic space between the shallow-water *C. artedi* and deep-water *C. kiyi* (Figure 2b, Table 2). Samples of *C. kiyi* generally had the highest mean $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values, but a small $\delta^{13}\text{C}$ range (0.3) and $\delta^{15}\text{N}$ range (0.7), indicating this species had the narrowest isotopic niche (Table 2). Meanwhile, samples of *C. artedi* had the

lowest mean $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values (Table 2). Individuals identified as *C. hoyi* fell between *C. artedi* and *C. kiyi* in the biplot (Figure 2b). These results corroborate previous analyses that suggest *C. hoyi* has an opportunistic feeding strategy and feeds at intermediate depths (Sierszen et al., 2014, Rosinski et al., 2020). Our results showed that samples of *C. zenithicus* overlapped with *C. artedi* and *C. hoyi*.

The SEAc were smallest for *C. kiyi*, intermediate for *C. artedi* and *C. hoyi* and largest for *C. zenithicus* (Table 2; Figure S3). There was no overlap between the ellipses representing *C. artedi* and *C. kiyi*, and there was minimal overlap between these two species and *C. hoyi* (Table 3). This supports a scenario of strong niche partitioning between these three species. In sharp contrast, samples of *C. zenithicus* showed a high degree of overlap with the other forms, especially with *C. artedi* and *C. hoyi* (Figure S3).

Transcriptomics

Raw reads for all individuals are available in the Sequence Read Archive of NCBI, as part of BioProject PRJNA659559. After removal of low-quality reads and trimming adaptors, we obtained an average of 46 million reads per individual ($\text{SD} \pm 16,414,234$; Table S3). The transcriptome assembly with Trinity resulted in 1,235,811 assembled contigs for *C. artedi*, 1,504,437 for *C. hoyi*, 1,703,560 for *C. kiyi* and 1,645,826 for *C. zenithicus*. After summarizing with Transdecoder and CD-hit, the final transcriptomes resulted in 167,179 assembled contigs for *C. artedi*, 167,135 for *C. hoyi*, 137,947 for *C. kiyi* and 134,387 for *C. zenithicus* (Table S4). The resulting transcriptome is available in the Transcriptome Shotgun Assembly Sequence Database of NCBI (GIUL000000000). RNAseq reads from all species (eight individuals per species, seven for *C. zenithicus*) were mapped to the transcriptome of *C. artedi* (167,135

contigs). After quality filtering with BCFtools, the resulting matrix was comprised of 22,285 high-confidence SNPs, with no missing data.

The molecular diversity indexes for all 31 samples were $H_o = 0.27$, $H_e = 0.22$, $F_{is} = -0.26$, $N_a = 147,841$ and $\theta = 0.73$ (Table S5). Average F_{st} across all samples was 0.018, and G_{test} results ($p < 0.01$) indicate that the null hypothesis of panmixia was rejected, suggesting the presence of genetic differentiation among species. Pairwise F_{st} estimates showed significant, albeit low, levels of genetic differentiation between species (Table 4). The largest differences were found between *C. kiyi* and *C. zenithicus* ($F_{st} = 0.025$, $p < 0.01$, Figure 4S) and *C. artedi* vs. *C. kiyi* ($F_{st} = 0.024$, $p < 0.01$; Figure 3a). The lowest level of genetic divergence was found between *C. hoyi* and *C. zenithicus* ($F_{st} = 0.00$, $p > 0.788$; Figure 3b).

The smaller panel of SNPs derived from orthologous sequences from Orthofinder (7,898 SNPs) resulted in F_{st} values comparable to those estimated with the full dataset (Table S7). This confirms our initial observations of differentiation between the closely related groups, suggesting the significant F_{st} values are not simply derived from paralogous sequences. This is a relevant distinction to make in salmonids, given their duplicated genome.

The PCA showed concordant patterns of variation with morphology and isotopes, but with greater separation among species (Figure 2c). The largest axis of differentiation (PC1=5.44% of variance explained) was between samples of *C. kiyi* versus the rest of the groups (Figure 2c). Meanwhile PC2 explained the differences between *C. artedi* with the rest of the groups (PC2=4.58%). The SNP data showed complete separation between *C. artedi* and *C. zenithicus*, which was not observed for isotopes and morphology. As with other analyses, the PCA showed overlap between samples of *C. hoyi* and *C. zenithicus*. The PCA also showed overlap for individuals of *C. hoyi* and *C. kiyi*, and between *C. artedi* and *C. zenithicus* (Figure

2c). Meanwhile, no overlap was observed between *C. kiyi* and *C. artedi*.

The population assignment test was consistent with the presence of three groups, with clear separation among *C. artedi*, *C. kiyi*, and a cluster of *C. hoyi* and *C. zenithicus* (Figure 4a), consistent with the PCA results. When we repeated the analysis with $K=4$, the groups formed by individuals of *C. artedi* and *C. kiyi* remained cohesive (as with $K=3$), yet the samples of *C. hoyi* and *C. zenithicus* are split into two mixed groups (Figure 4b). For the estimates of $K=3$, evidence of admixture was observed between *C. hoyi* and *C. kiyi*, *C. artedi* and *C. zenithicus*, and *C. artedi* and *C. hoyi* (Figure 4a).

The three-population analysis revealed the group formed by *C. hoyi* and *C. zenithicus* is admixed, as the Z-scores ranged from -9.72 to -14.79 (Table S6). Even when we detected potential hybrids between other species with the PCA and assignment test (i.e., *C. hoyi* and *C. kiyi*, *C. artedi* and *C. zenithicus*), these comparisons showed no significant hybridization with the f_3 statistic. This is probably because admixture was only observed for one individual of the compared groups (Table S6).

The pairwise F_{st} comparisons between species also allowed detecting highly differentiated genes between species (Supplementary Data 1-6). Thus, the most differentiated genes for each comparison were: *Fork-head Box Q1a* ($F_{st}=0.74$) between *C. artedi* and *C. hoyi* (Figure S4); *Rhodopsin* ($F_{st}=0.83$) between *C. artedi* and *C. kiyi* (Figure 3a); *RCC1 Domain Containing 1* ($F_{st}=0.64$) for *C. artedi* and *C. zenithicus* (Figure S4); *Zinc Finger MYM-Type Containing 2* ($F_{st}=0.86$) for *C. hoyi* and *C. kiyi* (Figure S4); *Spectrin Beta, Non-Erythrocytic 2* ($F_{st}=0.54$) for *C. hoyi* and *C. zenithicus* (Figure 3b); *Vacuolar Protein Sorting 13 Homolog A* ($F_{st}=0.78$) for *C. kiyi* and *C. zenithicus* (Figure S4). The analysis of selection with Ohana revealed 68 genes with a Log-likelihood ratios > 10 (“divergent selection”) and 12 loci with Log-

likelihood ratios > 15 (“very high divergent selection”; Cheng et al., 2016; Supplementary Data 7). These could be associated with ecological and morphological functions, such as regulation of epithelial formation (*Desmoplakin*; *Keratin Type 1*), cell morphology and organization (*Zinc Finger MYM-type 2*; *Peroxiredoxin-1*), development of nervous system (*Glutathione-specific gamma-glutamylcyclotransferase 1*; *cAMP-responsive element modulator*), metabolism of carbohydrates (*Protein phosphatase 1 regulatory subunit 3C*), vision (*Rhodopsin*), and cellular organization during development (*Fork-head box protein F1*). In addition, we explored if there was overlap between the analysis of Ohana (68 genes) and the top 50 differentiated genes for each of the pairwise *Fst* comparisons (i.e. 6 pairwise comparisons; 300 genes in total). This comparison resulted in 47 genes that showed both high *Fst* and potential divergent selection across all ciscoes (Supplementary Data 8).

GO enrichment analyses with the Mann-Whitney Test for the pairwise *Fst* estimates showed 17 significant GO terms (FDR < 10%) across four different pairwise comparisons (Supplementary Data 9). The comparison of *C. artedi* vs. *C. zenithicus* showed the most enrichment, with seven terms for Biological Process, two for cellular component, and five for Molecular Function. This included the terms membrane organization (Biological Process), and secondary active transmembrane transporter activity (Molecular Function). Other comparisons with significant results (FDR < 10%) for only one category were: *C. artedi* vs. *C. hoyi* (Molecular Function: Rho GTPase binding), *C. hoyi* vs. *C. zenithicus* (Biological Process: cell chemotaxis, plasma membrane raft assembly); *C. kiyi* vs. *C. zenithicus* (Cellular Component: plasma membrane region, phosphatidylinositol 3-kinase complex).

DISCUSSION

Understanding the dynamics of speciation can be enhanced by studying taxa with differences in phenotypic and ecological traits, but modest genetic differentiation. This is the case in Great Lakes coregonines, which are characterized by having complex evolutionary origins, as they show variation in morphological traits due to plasticity, selection, and/or hybridization. For taxonomically-complex taxa such as coregonines, integrating new results with previously published studies is hindered by a long history of researchers either implicitly or explicitly applying different taxonomic frameworks, often in the absence of museum voucher specimens. Thus, it is often not possible to reconcile whether apparent differences across space, time, or taxa, are due to true biological differences or reflect the taxonomic identifications employed. In taxa where this complexity occurs, we suggest that researchers either (1) explicitly describe the methods and morphological characters used in assigning individuals to taxa, (2) include linked morphological/ecological, and genetic data from the same specimens, or (3) deposit voucher specimens in natural history museums to serve as an historical record (e.g., see Schmidt et al. 2009). With this in mind, the present study applied an integrative approach to characterize morphological, ecological, and genetic analyses to the same set of individuals to disentangle the differences found within this economically and ecologically relevant group of fishes. The linked genome-to-phenome approach applied in this study will allow for greater comparability across studies and more robust analyses of variation in Great Lakes ciscoes. Limitations of this approach include the necessarily small sample sizes, but with DNA sequencing costs diminishing over time, this issue should eventually be overcome. Overall, understanding the relationships within *Coregonus* is highly relevant today, as overfishing, habitat degradation and invasive species have led to declines in abundance and diversity for decades (Eshenroder et al., 2016).

Isotopic measures show agreement with previous studies:

Based on isotopic measurements, it appears that the main driving mechanism of dietary partitions is depth preference, which in turn influences prey availability. These findings confirm previous observations that muscle derived $\delta^{13}\text{C}$ concentrations in *C. kiyi* and *C. hoyi*/*C. zenithicus* are less depleted than in *C. artedi*. This reflects preference for pelagic habitat and primary consumption of limnetic zooplankton during summer for the latter (Sierszen et al., 2014; Rosinski et al., 2020). This result is expected given the near-obligate pelagic existence of adult individuals of *C. artedi* during summer stratification (Rosinski, Vinson, & Yule, 2020). Meanwhile, high average $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values of *C. kiyi* are consistent with a diet dominated by *Mysis diluviana* (Gamble, Hrabik, Stockwell, & Yule, 2011a; Ahrenstorff et al., 2011; Sierszen et al., 2014) and preference for deep waters (Rosinski et al., 2020). The intermediate position of *C. hoyi* suggest they occupy intermediate depths and are opportunistic feeders with a diet that is obtained from pelagic and benthic zones. This result is also consistent with previous studies that showed *C. hoyi* change their diet seasonally (*Diporeia* in spring, *Mysis diluviana* in summer, and *Daphnia* spp. and calanoid copepods in autumn; Gamble, Hrabik, Stockwell, & Yule, 2011b). When compared to the other species, the isotopic values of *C. zenithicus* (Figure 2b) suggest that some individuals in this group may be obtaining part of their nutrition from the pelagic zone, through consumption of zooplankton. Samples of *C. zenithicus* from this study had terminal mouths similar to *C. artedi*, which could explain the wide breadth of dietary components compared to the other groups. Both the present study and Schmidt et al. (2009) have *C. zenithicus* and *C. hoyi* occupying similar trophic positions, and both show *C. zenithicus* obtaining nutrition from more $\delta^{13}\text{C}$ depleted prey resources relative to *C. hoyi*. Yet, Schmidt et al. (2009) showed these two species occupied

unique isotopic space (especially for their Lake Superior samples of 1972-1998), while we obtained far greater overlap. Given these contrasting results, it is necessary to expand the studies on the dietary overlap between *C. hoyi* and *C. zenithicus* among different populations of Lake Superior.

Multiple lines of evidence support divergence of *C. artedi* and *C. kiyi*:

The forms that showed the highest levels of divergence across all measured traits were the shallow water *C. artedi* and the deep-water *C. kiyi*. One of the most striking morphological differences between the two is the size of the eye, which is much larger in *C. kiyi* relative to the rest of their bodies (Table 1). Genetic differences were also associated with vision, as two highly divergent genes between the two species were *Rhodopsin* (RHO; $F_{st}=0.86$) and *Homeodomain-interacting protein kinase 2* (HIPK2; $F_{st}=0.55$). *Rhodopsin* is a key photoreceptor in low light conditions, and differences in this gene have been previously reported for the two species (Eaton et al., 2020), as well as for other lineages exposed to dim-lit environments (Hill et al., 2019; Musilova et al., 2019; Schott, Refvik, Hauser, López-Fernández, & Chang, 2014). *Homeodomain-interacting protein kinase 2* is associated with tissue growth during development, including eye size and lens formation in mice (Inoue et al., 2010; Poon, Zhang, Lin, Manning, & Harvey, 2012). Additional loci that were highly differentiated between *C. artedi* and *C. kiyi* include: *Insulin Growth Factor 1 Receptor* (IGRF1; $F_{st}=0.87$) which is essential for embryonic growth and cell proliferation of vertebrates (Schlueter, Peng, Westerfield, & Duan, 2007); *Plasminogen Activator* (PLAT; $F_{st}=0.73$), associated with tissue formation and modification in fishes (Bader, et al. 2012); *Dystroglycan* (DAG1; $F_{st}=0.62$) which has many functions including muscle development in zebrafish (Parsons, Campos, Hirst, & Stemple, 2002); and *cAMP*-

responsive element modulator (CREM; $F_{st}=0.62$) important for neurogenesis and neuro-plasticity of vertebrates (Mioduszevska, Jaworski, & Kaczmarek, 2003). These are candidate genes that could exhibit adaptive genetic variants associated with habitat preferences of the two species, and the morphological differences found between them.

The PCL/PPD and PVL/PAD ratios were also considerably smaller in *C. artedi* compared to *C. kiyi*. Previous studies suggest that this is a convergent adaptation across lacustrine fishes, where large bodies and smaller fins provide advantages for maintaining buoyancy in pelagic environments with little swimming (i.e., *C. artedi*), while small bodies with longer fins favor swimming at depth (*C. kiyi*; Eshenroder, Sideleva, & Todd, 1999). Overall, the results from our study suggest concordance between morphological, genetic, and ecological differences for a shallow-water pelagic species (*C. artedi*) and the deep-water species (*C. kiyi*). Future studies are needed to confirm if there is a relationship between the aforementioned genes and morphological and ecological differences detected among the groups.

Lack of differentiation between *C. zenithicus* and *C. hoyi*:

One of the revealing findings of our study was the limited differentiation between samples initially identified as *C. zenithicus* and *C. hoyi*. At first sight, the isotopic analysis suggests *C. hoyi* are obtaining more nutrition from the benthic zone, but this signal could be confounded by seasonal variation. Further, samples identified as *C. zenithicus* and *C. hoyi* showed no significant differences in the overall F_{st} , and the f_3 -statistic indicated rampant admixture between the two taxa. Still, the pairwise comparisons of locus specific F_{st} showed differentiation for genes associated with nerve growth and neural development (*Spectrin Beta Non-Erythrocytic 2*, $F_{st}=0.54$; *Growth Associated Protein 43*, $F_{st}=0.47$; Benowitz & Routtenberg, 1997; Lise et al.,

2012), as well as cell proliferation in vertebrates (*Tensin 2*; $F_{st}=0.49$; Hafizi, Ibraimi & Dahbäck, 2005). These highly differentiated loci did not influence the low overall F_{st} as they represent a very small proportion of the sampled loci. Future studies across populations of *C. hoyi* and *C. zenithicus* are necessary to understand the full extent of the genomic divergence between the two species.

The key phenotypic traits used in our study to distinguish *C. zenithicus* and *C. hoyi* were lower jaw position (*C. zenithicus*: included or terminal; *C. hoyi*: extended) and gill raker counts (*C. zenithicus* mean $\pm$ SD = 39.5 ± 2.3 ; *C. hoyi* = 42.4 ± 2.1 ; Eshenroder et al., 2016), both of which are commonly used in the field to identify these species. Previous studies also suggest that both species can be differentiated through pre-maxilla angle, which are $\sim 40^\circ$ for *C. hoyi*, and between 60° - 65° for *C. zenithicus* (Eshenroder et al., 2016). The range of pre-maxillary angles for our samples was on average 35.88° (SD ± 4.05) for *C. hoyi*, and 35.52° (SD ± 3.14) for *C. zenithicus*. Thus, it is possible that we did not collect samples of *C. zenithicus* in our study and that all three measurements (i.e., lower jaw position, gill raker morphology, and pre-maxillary angles) should be employed when differentiating the two species.

Alternatively, it is possible that these two taxa are now forming a hybrid-swarm that is morphologically and genetically very similar in certain areas of the Great Lakes. This hypothesis stems from the recent collapse of *C. zenithicus* in Lake Superior. Koelz (1929) reported that *C. zenithicus* was the dominant deep-water cisco in Lake Superior in 1921-1922 representing 90% of the ciscoes caught in bottom set gillnets (Eshenroder et al., 2016). By 2001-2003 they represented only 4% of ciscoes caught in nearshore and offshore bottom trawl samples (Gorman & Todd 2007). Declines in *C. zenithicus* abundance, coupled with the propensity for *C. hoyi* to expand into areas previously occupied by other deep species (Bronte et al., 2010; Bunnell et al.,

2012), could have promoted hybridization. In our study, we also show evidence of admixture between 1 or 2 individuals *C. hoyi* and *C. kiyi*, as well as *C. zenithicus* and *C. artedi*. Previous studies have demonstrated that hybridization is possible between closely related coregonines in both North America and Europe (Hudson, Lundsgaard-Hansen, Lucek, Vonlanthen, & Seehausen, 2017; Hudson, Vonlanthen, & Seehausen, 2011; Kirtiklis & Jankun, 2006; Todd & Stedman, 1989), and successful crosses of *C. artedi* and *C. hoyi* have been made in captivity (W. Stott, personal communication). Further, hybridization has been suggested as one of the drivers of morphological changes in *C. hoyi* in Lake Huron after the decline of other species in the complex (Todd & Stedman, 1989). This scenario of speciation reversal in *C. zenithicus*, where morphological diversity could have been lost due to both human-mediated stressors and admixture with closely related species, could represent an example of how commercially and ecologically relevant lineages have been lost in recent decades. It remains to be determined if other populations of *C. zenithicus* and *C. hoyi* in Lake Superior are also admixed, given the limited number of samples included in the present study. This is especially relevant considering previous surveys have reported finding phenotypically distinct *C. zenithicus* in Canadian waters of Lake Superior (Pratt & Chong, 2012; Pratt, 2012).

Concordance among morphology, ecology and genetics:

This study exemplifies how analyzing the same specimens with multiple approaches can enhance our understanding of the relationships among the *C. artedi* complex. Our results showed concordance for morphological, ecological, and genetic divergence for three of the four species studied. The highest levels of differentiation for all the analyzed traits was observed for *C. kiyi*, which is the species typically found in deeper waters of Lake Superior. This is expected given that adaptation to life at depth has been reported as one of the main drivers of differentiation for

lacustrine fishes, including African cichlids (Albertson, 2008; Schlieven et al., 2001), grayling (Olson, Krabbenhoft, Hrabik, Mendsaikhan, & Jensen, 2019), sculpins of Lake Baikal (Kontula, Kirilchik, & Väinölä, 2003), lake trout in Great Lakes (Perreault-Payette et al., 2017), and European coregonines (Vonlanthen et al., 2009). In Lake Superior, our results suggest adaptation to depth has resulted in a larger eye and modifications of the visual system (e.g., *Rhodopsin*; Eaton et al. 2020), as well as a diet centered around the deep macro-invertebrate *Mysis*, when compared to its shallower counterparts.

This study also suggests that another driver of differentiation for *Coregonus* is the benthic/pelagic axis, which is also known to lead to the differentiation of lacustrine species (e.g., cichlids: Hulsey, Roberts, Loh, Rupp, & Streelman, 2013; perch: Svanbäck & Eklöv, 2004). This is evidenced by the results of stable isotope analysis, where *C. artedi* appears to have mostly a pelagic diet, while the remaining groups had a combination of pelagic/benthic elements (Sierszen et al., 2014). Observed dietary preferences could also be linked to the differences in the angle of the mouth, as having terminal mouths can be associated with feeding on pelagic zooplankton (*C. artedi*), while having a more angled mouth could provide advantages when feeding at depth (*C. kiyi* and *C. hoyi*; Etheridge, Bean, Maitland, Ballantyne, & Adams, 2012). It is important to highlight that feeding morphologies can change for this group based on competition and prey availability, as populations of *C. artedi* and *C. zenithicus* can show considerable differences in feeding structures across the Great Lakes depending on the presence/absence of congeners (Turgeon et al., 2016).

Overall, the results of our study suggest that differentiation across *Coregonus* could be promoted by habitat preferences (i.e., depth and benthic/pelagic habit) and dietary partitions, both of which are well known drivers of diversification of multiple fish groups (Bernardi, 2013).

Status of the Great Lakes and conservation of genetic resources:

Extensive efforts are being made to improve the conditions for organisms that inhabit the Great Lakes, including the reduction of contaminants (e.g., PCBs) and controlling densities of invasive species (e.g., Siefkes et al. 2013 and <https://www.epa.gov/greatlakes/lakewide-action-and-management-plans-great-lakes>) and reestablishment of native fishes (e.g., Muir et al. 2012). Along with these efforts, it is essential for managers to identify and monitor functional genomic variants associated with ecological and morphological differences of coregonines. This is critical as it allows for the maintenance of the evolutionary potential of the *C. artedi* complex across the Great Lakes. Further, evaluating individual genes and molecular pathways associated with ecological divergence across the Great Lakes will allow us to understand if some populations are at higher risk of speciation reversal due to population declines or changing environmental conditions. These approaches should be implemented across lakes that have different faunal compositions, and especially in lakes that have fewer sympatric species of *Coregonus* than Lake Superior (e.g., Lakes Michigan and Huron), as both of these factors are known to influence the amount of divergence among coregonines (e.g., Turgeon et al., 2016).

Conclusions:

In this study, we analyzed patterns of divergence among four species of *Coregonus* from Lake Superior and found concordance among morphological, ecological and genomic approaches. Based on the morphological analyses, we confirmed previous observations that eye size, fin length and length of the lower jaw explain most of the variation between species. There are clear contrasts in the dietary preferences of *C. artedi* compared to *C. kiyi*, while observed differences between *C. hoyi* and *C. zenithicus* appear to be largely influenced by seasonality. Transcriptomic

analyses were able to distinguish between three of the four species, and high levels of differentiation in genes associated with body shape, eye size, fin shape, vision, organization of the nervous system and early development. There was evidence of hybridization between samples morphologically identified as *C. hoyi* and *C. zenithicus* and future studies should assess the extent of this admixture in Lake Superior. Overall, the integrative approach applied in this study exemplifies how multiple lines of evidence can help elucidate the relationships among sympatric groups with complex evolutionary history, such Great Lakes *Coregonus* species.

Acknowledgements:

We thank the U.S. Geological Survey Research Vessel Kiyi Captain Joe Walters, First Mate Keith Peterson, and Engineer Charles Carrier. Special thanks to Ryan Menenbroeker who measured our specimens, Jean Adams for analyzing the morphometric data, and to Caroline Rosinski, Mark Vinson, and Andrew Muir for providing helpful insight. We would like to thank Tianying Lan, Nathan Backenstose, Katherine Eaton, and Jessie Pelosi for their help with the genetic analyses and data handling. Early drafts of this manuscript were improved with solicited reviews by Tom Pratt and Gary Longo. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Funding was provided by start-up funds to TED from Wayne State University and the Great Lakes Fishery Commission (Award no. 2018-KRA-44073).

REFERENCES

1. Ackiss, A. S., Larson, W. A., & Stott, W. (2020). Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes. *Evolutionary Applications*, 13(5), 1037–1054. doi: 10.1111/eva.12919

2. Albertson, R. C. (2008). Morphological Divergence Predicts Habitat Partitioning in a Lake Malawi Cichlid Species Complex. *Copeia*, 2008(3), 689–698. doi: 10.1643/CG-07-217
3. Arnold, M. L. (2006). *Evolution Through Genetic Exchange*. OUP Oxford.
4. American Fisheries Society (AFS). 2014. Guidelines for the use of fishes in research. American Fisheries Society, Bethesda, Maryland.
5. Arnold, M. L., Bulger, M. R., Burke, J. M., Hempel, A. L., & Williams, J. H. (1999). Natural hybridization: how low can you go and still be important? *Ecology*, 80(2), 371–381.
6. Ahrenstorff, T.D., T.R. Hrabik, J.D. Stockwell, D.L. Yule, and G.G. Sass. 2011. Seasonal dynamic diel vertical migrations of *Mysis diluviana*, coregonine fishes, and siscowet lake trout in the pelagia of western Lake Superior. *Transactions of the American Fisheries Society* 140:1504-1520.
7. Bager, R., Kristensen, T.K., Jensen, J.K., Szczur, A., Christensen, A., Andersen, L.M.M. Johansen, J.S., Larsen, N., Baatrup, E., Huang, M., Ploug, M., & Andreasen, A.A. (2012). Urokinase-type plasminogen activator-like proteases in teleosts lack genuine receptor-binding epidermal growth factor-like domains. *Journal of Biological Chemistry* 287 (33): 27526-27536.
8. 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.
9. Benowitz, L. I., & Routtenberg, A. (1997). GAP-43: an intrinsic determinant of neuronal development and plasticity. *Trends in Neurosciences*, 20(2), 84–91. doi: 10.1016/S0166-2236(96)10072-2
10. Bernardi, G. (2013). Speciation in fishes. *Molecular Ecology*, 22(22), 5487–5502. doi: 10.1111/mec.12494
11. Bernatchez, L., & Dodson, J. J. (1991). Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, 45(4), 1016–1035.
12. Bronte, C. R., Hoff, M. H., Gorman, O. T., Thogmartin, W. E., Schneeberger, P. J., & Todd, T. N. (2010). Decline of the Shortjaw Cisco in Lake Superior: The Role of Overfishing and Risk of Extinction. *Transactions of the American Fisheries Society*, 139(3), 735–748. doi: 10.1577/T09-019.1
13. Bunnell, D. B., Eshenroder, R. L., Krause, A. E., & Adams, J. V. (2012). Depth segregation of deepwater ciscoes (*Coregonus spp.*) in Lake Michigan during 1930–1932 and range expansion of *Coregonus hoyi* into deeper waters after the 1990s. *Advances in Limnology*, 63, 3–24.
14. Carruthers, M., Yurchenko, A. A., Augley, J. J., Adams, C. E., Herzyk, P., & Elmer, K. R. (2018). *De novo* transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics*, 19(1), 32. doi: 10.1186/s12864-017-4379-x
15. Cheng, J. Y., Mailund, T., & Nielsen, R. H. (2016). Ohana, a tool set for population genetic analyses of admixture components. doi: 10.1101/071233
16. Cheng, J. Y., Racimo, F., & Nielsen, R. (2019). Ohana: detecting selection in multiple populations by modelling ancestral admixture components. *BioRxiv*, 546408.
17. 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. doi: 10.1093/bioinformatics/btr330

18. Dos Reis, S. F., Pessôa, L. M., & Strauss, R. E. (1990). Application of size-free canonical discriminant analysis to studies of geographic differentiation. *Revista Brasileira de Genética*, 13(3), 509-520.
19. Eaton, K. M., Bernal, M. A., Backenstose, N. J., & Krabbenhoft, T. (2020). Nanopore amplicon sequencing reveals molecular convergence and local adaptation of opsin genes.
20. Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1), 157. doi: 10.1186/s13059-015-0721-2
21. Eshenroder, R. L., Sideleva, V. G., & Todd, T. N. (1999). Functional convergence among pelagic sculpins of Lake Baikal and deepwater ciscoes of the Great Lakes. *Journal of Great Lakes Research*, 25(4), 847-855.
22. 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. Great Lakes Fishery Commission.
23. Etheridge, E. C., Bean, C. W., Maitland, P. S., Ballantyne, S., & Adams, C. E. (2012). Discontinuous infraspecific variation in ecological and morphological traits has consequences for conservation of powan (*Coregonus lavaretus*) in Scotland. *Advances in Limnology*, 63, 505–517.
24. Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology*, 28(3), 615–629. doi: 10.1111/mec.14977
25. Fitzpatrick, S. W., Gerberich, J. C., Kronenberger, J. A., Angeloni, L. M., & Funk, W. C. (2015). Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, 18(1), 37–47. doi: 10.1111/ele.12388
26. Gamble, A.E., Hrabik, T.R., Stockwell, J.D., & Yule, D.L. (2011a). Trophic connections in Lake Superior part I: the offshore fish community. *Journal of Great Lakes Research*, 37, 541-549.
27. Gamble, A.E., Hrabik, T.R., Stockwell, J.D., & Yule, D.L. (2011b). Trophic connections in Lake Superior part II: the nearshore fish community. *Journal of Great Lakes Research*, 37, 550-560.
28. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, 29(7), 644.
29. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Lieber, M. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512.
30. Hafizi, S., Ibraimi, F., & Dahlbäck, B. (2005). C1-TEN is a negative regulator of the Akt/PKB signal transduction pathway and inhibits cell survival, proliferation, and migration. *The FASEB journal*, 19(8), 971-973.
31. Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O., & Puebla, O. (2019). Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution*, 3(4), 657–667. doi: 10.1038/s41559-019-0814-5
32. Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., Folkvord, A., ... Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences*, 116(37), 18473–18478. doi: 10.1073/pnas.1908332116

33. Hoff, M. H., & Todd, T. N. (2004). Status of the shortjaw cisco (*Coregonus zenithicus*) in Lake Superior. *Annales Zoologici Fennici*, 41(1), 147–154. Retrieved from JSTOR.
34. Hoffman, J. C., Sierszen, M. E., & Cotter, A. M. (2015). Fish tissue lipid-C: N relationships for correcting $\delta^{13}\text{C}$ values and estimating lipid content in aquatic food-web studies. *Rapid Communications in Mass Spectrometry*, 29(21), 2069–2077.
35. Hohenlohe, P.A., Day, M.D., Amish, S.J., Miller, M.R., Kamps-Hughes, N., Boyer, M.C., Muhlfeld, C.C., Allendorf, F.W., Johnson, E.A., & Luikart, G. (2013). Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* 22(11), 3002–3013.
36. Hubbs, C. L. (1955). Hybridization between Fish Species in Nature. *Systematic Biology*, 4(1), 1–20. doi: 10.2307/sysbio/4.1.1
37. 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. doi: 10.1111/eva.12446
38. Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2011). Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological Sciences*, 278(1702), 58–66. doi: 10.1098/rspb.2010.0925
39. Hulsey, C. D., Roberts, R. J., Loh, Y.-H. E., Rupp, M. F., & Streelman, J. T. (2013). Lake Malawi cichlid evolution along a benthic/limnetic axis. *Ecology and Evolution*, 3(7), 2262–2272. doi: 10.1002/ece3.633
40. Inoue, T., Kagawa, T., Inoue-Mochita, M., Isono, K., Ohtsu, N., Nobuhisa, I., ... Taga, T. (2010). Involvement of the Hipk family in regulation of eyeball size, lens formation and retinal morphogenesis. *FEBS Letters*, 584(14), 3233–3238. doi: 10.1016/j.febslet.2010.06.020
41. Jackson, A. L., Inger, R., Parnell, A. C., & Bearhop, S. (2011). Comparing isotopic niche widths among and within communities: SIBER—Stable Isotope Bayesian Ellipses in R. *Journal of Animal Ecology*, 80(3), 595–602.
42. Kirtiklis, L., & Jankun, M. (2006). Chromosome analysis in coregonid individuals in the interspecific hybridization zone. *Journal of Applied Ichthyology*, 22(5), 401–403. doi: 10.1111/j.1439-0426.2006.00759.x
43. Koelz, W. (1929). Coregonid fishes of the Great Lakes. *U.S. Bureau Fisheries Bulletin*, 43, 297–643.
44. Kontula, T., Kirilchik, S. V., & Väinölä, R. (2003). Endemic diversification of the monophyletic cottoid fish species flock in Lake Baikal explored with mtDNA sequencing. *Molecular Phylogenetics and Evolution*, 27(1), 143–155. doi: 10.1016/S1055-7903(02)00376-7
45. Krueger, F. (2015). Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, 516, 517.
46. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. doi: 10.1038/nmeth.1923
47. Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659.
48. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

49. Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... & Grammes, F. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200-205.
50. Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Parolin Schneckenberg, R., ... Németh, A. H. (2012). Recessive Mutations in SPTBN2 Implicate β -III Spectrin in Both Cognitive and Motor Development. *PLoS Genetics*, 8(12). doi: 10.1371/journal.pgen.1003074
51. 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.
52. Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Molecular Biology and Evolution*, 35(6), 1489–1506. doi: 10.1093/molbev/msy051
53. Meirmans, P. G., & Tienderen, P. H. V. (2004). genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792–794. doi: 10.1111/j.1471-8286.2004.00770.x
54. Mills, E. L., Leach, J. H., Carlton, J. T., & Secor, C. L. (1994). Exotic Species and the Integrity of the Great Lakes. *BioScience*, 44(10), 666–676. doi: 10.2307/1312510
55. Mioduszevska, B., Jaworski, J., & Kaczmarek, L. (2003). Inducible cAMP early repressor (ICER) in the nervous system – a transcriptional regulator of neuronal plasticity and programmed cell death. *Journal of Neurochemistry*, 87(6), 1313–1320. doi: 10.1046/j.1471-4159.2003.02116.x
56. Muir, A. M., Krueger, C. C., & Hansen, M. J. (2012). Re-establishing lake trout in the Laurentian Great Lakes: past, present, and future. *Great Lakes fishery policy and management: a binational perspective, 2nd edition*. Michigan State University Press, East Lansing, 533-588.
57. Musilova, Z., Cortesi, F., Matschiner, M., Davies, W. I. L., Patel, J. S., Stieb, S. M., ... Salzburger, W. (2019). Vision using multiple distinct rod opsins in deep-sea fishes. *Science*, 364(6440), 588–592. doi: 10.1126/science.aav4632
58. Newsome, S. D., Martinez del Rio, C., Bearhop, S., & Phillips, D. L. (2007). A niche for isotopic ecology. *Frontiers in Ecology and the Environment*, 5(8), 429-436.
59. Ohlberger, J., Mehner, T., Staaks, G., & Hölker, F. (2008). Temperature-related physiological adaptations promote ecological divergence in a sympatric species pair of temperate freshwater fish, *Coregonus spp.* *Functional Ecology*, 22(3), 501–508. doi: 10.1111/j.1365-2435.2008.01391.x
60. Olson, K. W., Krabbenhoft, T. J., Hrabik, T. R., Mendsaikh, B., & Jensen, O. P. (2019). Pelagic–littoral resource polymorphism in Hovsgol grayling *Thymallus nigrescens* from Lake Hovsgol, Mongolia. *Ecology of Freshwater Fish*, 28(3), 411–423. doi: 10.1111/eff.12464
61. Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419–420. doi: 10.1093/bioinformatics/btp696
62. Parsons, M. J., Campos, I., Hirst, E. M. A., & Stemple, D. L. (2002). Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos. *Development*, 129(14), 3505–3512.
63. 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.
64. Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., ... Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature Communications*, 3(1), 1–6. doi: 10.1038/ncomms2140
65. Poon, C. L. C., Zhang, X., Lin, J. I., Manning, S. A., & Harvey, K. F. (2012). Homeodomain-Interacting Protein Kinase Regulates Hippo Pathway-Dependent Tissue Growth. *Current Biology*, 22(17), 1587–1594. doi: 10.1016/j.cub.2012.06.075
66. Pratt, T. C. (2012). The distribution and abundance of deepwater ciscoes in Canadian waters of Lake Superior. *Adv. Limnol*, 63, 25–41.
67. Pratt, T. C., & Chong, S. C. (2012). Contemporary life history characteristics of Lake Superior deepwater ciscoes. *Aquatic Ecosystem Health & Management*, 15(3), 322–332. doi: 10.1080/14634988.2012.708639
68. R Core Team (2013). R: A language and environment for statistical computing.
69. Regier, H. A., Whillans, T. H., Christie, W. J., & Bocking, S. A. (1999). Over-fishing in the Great Lakes: the context and history of the controversy. *Aquatic Ecosystem Health & Management*, 2(3), 239–248. doi: 10.1080/14634989908656959
70. Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489–494. doi: 10.1038/nature08365
71. Richards, E. J., & Martin, C. H. (2017). Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLOS Genetics*, 13(8), e1006919. doi: 10.1371/journal.pgen.1006919
72. Rosinski, C. L., Vinson, M. R., & Yule, D. L. (2020). Niche Partitioning among Native Ciscoes and Nonnative Rainbow Smelt in Lake Superior. *Transactions of the American Fisheries Society*, 149(2), 184–203.
73. Schlieven, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T., & Tautz, D. (2001). Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology*, 10(6), 1471–1488. doi: 10.1046/j.1365-294X.2001.01276.x
74. Schlueter, P. J., Peng, G., Westerfield, M., & Duan, C. (2007). Insulin-like growth factor signaling regulates zebrafish embryonic growth and development by promoting cell survival and cell cycle progression. *Cell Death & Differentiation*, 14(6), 1095–1105. doi: 10.1038/sj.cdd.4402109
75. Schott, R. K., Refvik, S. P., Hauser, F. E., López-Fernández, H., & Chang, B. S. W. (2014). Divergent Positive Selection in Rhodopsin from Lake and Riverine Cichlid Fishes. *Molecular Biology and Evolution*, 31(5), 1149–1165. doi: 10.1093/molbev/msu064
76. Seehausen, O. (2006). Conservation: losing biodiversity by reverse speciation. *Current Biology*, 16(9), R334–R337.
77. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. doi: 10.1093/bioinformatics/btv351
78. Smith, S. H. (1968). Species Succession and Fishery Exploitation in the Great Lakes. *Journal of the Fisheries Research Board of Canada*, 25(4), 667–693. doi: 10.1139/f68-063
79. Siefkes, M. J., Steeves, T. B., Sullivan, W. P., Twohey, M. B., & Li, W. (2013). Sea lamprey control: past, present, and future. *Great Lakes Fisheries Policy and Management. Michigan State University Press, East Lansing*, 651–704.

80. Sierszen, M. E., Peterson, G. S., & Scharold, J. V. (2006). Depth-specific patterns in benthic planktonic food web relationships in Lake Superior. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(7), 1496-1503.
81. 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.
82. Stockwell, J.D., D.L. Yule, O.T. Gorman, E.J. Isaac, & S.A. Moore. 2006. Evaluation of bottom trawls as compared to acoustics to assess adult lake herring (*Coregonus artedii*) abundance in Lake Superior. *Journal of Great Lakes Research*, 32(2), 32280-292.
83. Svanbäck, R., & Eklöv, P. (2004). Morphology in perch affects habitat specific feeding efficiency. *Functional Ecology*, 18(4), 503–510. doi: 10.1111/j.0269-8463.2004.00858.x
84. Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2), 343–355.
85. Todd, T.N., & G.R. Smith. (1980). Differentiation in *Coregonus zenithicus* in Lake Superior. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 2228-2235.
86. 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. doi: 10.1139/f81-008
87. Todd, T. N., & Stedman, R. M. (1989). Hybridization of ciscoes (*Coregonus spp.*) in Lake Huron. *Canadian Journal of Zoology*, 67(7), 1679–1685. doi: 10.1139/z89-241
88. 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. doi: 10.1023/A:1021860910719
89. 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. doi: 10.1111/j.1558-5646.1999.tb04568.x
90. 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. doi: 10.1007/s10592-015-0759-4
91. U.S. Geological Survey. (2018). Compiled reports to the Great Lakes Fishery Commission of the annual bottom trawl and acoustics surveys for 2017. Wisconsin, USA. http://www.glfc.org/pubs/lake_committees/common_docs/CompiledReportsfromUSGS2017.pdf
92. Vonlanthen, P., Roy, D., Hudson, A. G., Largiadèr, C. R., Bittner, D., & Seehausen, O. (2009). Divergence along a steep ecological gradient in lake whitefish (*Coregonus sp.*). *Journal of Evolutionary Biology*, 22(3), 498–514. doi: 10.1111/j.1420-9101.2008.01670.x
93. Wilson, C. C., & Bernatchez, L. (1998). The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology*, 7(1), 127–132. doi: 10.1046/j.1365-294x.1998.00302.x

94. Wright, R. M., Aglyamova, G. V., Meyer, E., & Matz, M. V. (2015). Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics*, 16, 371. doi: 10.1186/s12864-015-1540-2
95. Yule, D.L., J.V. Adams, T.R. Hrabik, M.R. Vinson, Z. Woiak, and T.D. Ahrenstorff. (2013). Use of classification trees to apportion single echo detections to species: application to the pelagic fish community of Lake Superior. *Fisheries Research*, 140, 123-132.
96. Zhang, L., Thibert-Plante, X., Ripa, J., Svanbäck, R., & Brännström, Å. (2019). Biodiversity loss through speciation collapse: mechanisms, warning signals, and possible rescue. *Evolution*, 73(8), 1504-1516.

Data Accessibility:

The raw reads for all individuals and the assembled transcriptome of *Coregonus artedii* are available in NCBI as BioProject PRJNA659559. The bioinformatic scripts used for the transcriptomic analyses are available at github.com/evofish.

Author Contributions:

TED and TJK conceived the study; DY and LE collected the samples; DLY, LE, and TJK collected the data; MAB, DY, WS and TJK analyzed the data and interpreted the results; the manuscript was written by MAB and DY and all the co-authors contributed to the final version.

TABLES

Table 1. Summary of morphological measurements that were most useful for the differentiation of *Coregonus artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus*. Values in the table represent the average measures or ratios ($\pm$ SD). Measurement abbreviations are defined in Supplementary Table S2. The Principal Component Analysis loadings for PC1 and PC2 that exceeded ± 0.2 for one (or both) of the dimensions are also provided. Table abbreviations: PCL=Pectoral fin length, PPD=Pectoral-pelvic distance, PVL=Pelvic fin length, PAD=Pelvic-anal distance, OOL=Orbital length (eye), BDD=Body depth.

	Total length (mm)	PCL (mm)	PPD (mm)	PCL/PPD	PVL (mm)	PAD (mm)	PVL/PAD	OOL (mm)	BDD (mm)
<i>PC1</i>	-	0.6	-0.22	0.47	0.25	-0.15	0.4	0.24	-0.14
<i>PC2</i>	-	0.4	-0.21	0.07	-0.19	0.24	-0.73	0.14	-0.32
<i>C. artedi</i>	332 (21)	41.0 (3.4)	87.6 (5.2)	0.47 (0.05)	37.1 (2.5)	63.2 (3.7)	0.59 (0.03)	12.6 (0.9)	64.9 (3.7)
<i>C. hoyi</i>	243 (25)	32.4 (4.1)	64.0 (6.7)	0.51 (0.07)	30.5 (3.2)	42.2 (5.3)	0.73 (0.07)	11.3 (0.6)	45.6 (5.5)
<i>C. zenithicus</i>	261 (20)	31.9 (5.0)	68.9 (5.6)	0.46 (0.07)	32.1 (3.4)	46.8 (4.5)	0.69 (0.07)	11.5 (0.7)	51.3 (4.5)
<i>C. kiyi</i>	225 (16)	35.9 (3.3)	56.2 (4.2)	0.64 (0.03)	31.0 (3.4)	38.1 (4.4)	0.82 (0.04)	12.6 (0.7)	40.2 (5.7)

Table 2. The isotopic metrics of $\delta^{13}\text{C}_{\text{lipid free}}$ and $\delta^{15}\text{N}$ (‰), for *C. artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* collected from western Lake Superior (May-November 2015), and the corrected standard ellipse areas (SEAc).

	Sample size	Mean $\delta^{15}\text{N}$ (range)	Mean $\delta^{13}\text{C}$ (range)	SEAc
<i>C. artedi</i>	8	6.7 (0.9)	-25.9 (1.1)	2.33
<i>C. hoyi</i>	8	8.1 (1.5)	-25.3 (0.8)	2.41
<i>C. zenithicus</i>	7	7.8 (2.5)	-25.7 (1.0)	7.19
<i>C. kiyi</i>	8	8.7 (0.7)	-24.7 (0.3)	0.57

Table 3. Percentage of overlap in isotopic niches between pairings of *C. artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* based on the overlap of corrected standard ellipse areas. The values are the percentage of overlap of a given form in columns in the isotopic niche space of the other forms in rows.

Species	<i>C. artedi</i>	<i>C. hoyi</i>	<i>C. kiyi</i>	<i>C. zenithicus</i>
<i>C. artedi</i>	-	8.80	0.00	29.22
<i>C. hoyi</i>	8.53	-	24.66	29.23
<i>C. kiyi</i>	0.00	5.82	-	5.18
<i>C. zenithicus</i>	90.04	87.26	65.58	-

Table 4. Estimates of pairwise genetic differentiation (Weir-Cockerham F_{st}) among *C. artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* of Lake Superior (below diagonal) and corresponding p-values (above the diagonal).

Species	<i>C. artedi</i>	<i>C. hoyi</i>	<i>C. kiyi</i>	<i>C. zenithicus</i>
<i>C. artedi</i>	-	<0.001	<0.001	<0.001
<i>C. hoyi</i>	0.015	-	<0.001	0.788
<i>C. kiyi</i>	0.024	0.023	-	<0.001
<i>C. zenithicus</i>	0.017	0.000	0.025	-

FIGURES

Figure 1. (A) Collection sites for organisms analyzed in the study and (B) sympatric forms of Lake Superior coregonines: *Coregonus artedii*, *C. hoyi*, *C. kiyi* and *C. zenithicus*. Detailed information about sites are provided in Supplementary Table S1. Photographs by D.L Yule.

Figure 2. Principal Component Analysis (PCA) for (A) morphological measures; (B) bi-plot for isotope ratios; and (C) PCA of transcriptomic variance (22,285 SNPs) for four species of *Coregonus* from Lake Superior. *Coregonus artedii* (red), *C. hoyi* (blue), *C. kiyi* (green) and *C. zenithicus* (purple).

Figure 3. Histograms of pairwise F_{st} for individual genes in pairwise comparisons: between (A) the shallow-water *C. artedi* and deep-water *C. kiyi*, and (B) the two species with the lowest levels of divergence *C. zenithicus* and *C. hoyi*. The X-axis represents values of F_{st} (with binning of 0.05) and arrows represent the gene with the largest F_{st} (0.88 for *C. artedi* and deep-water *C. kiyi*; 0.54 for *C. zenithicus* and *C. hoyi*). The Y-axis represents frequency of individuals in log-scale. Photographs by D.L. Yule.

Figure 4. Admixture plots for *Coregonus artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* using 22,285 high-confidence SNPs, simulating (A) three groups ($K=3$) and (B) four groups ($K=4$).