Transcriptomic divergence predicts morphological and ecological variation underlying an adaptive radiation

Moisés A. Bernal[1,2*], Daniel L. Yule[3], Wendylee Stott[4], Lori Evrard[3], Thomas E. Dowling[5], 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 artedi* 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

67 approximately 14,000 years before present, led to the formation of lakes and rivers that promoted
68 connectivity between the isolated refugia (Bailey & Smith, 1981). These cycles of vicariance
69 followed by subsequent hypothesized admixture resulted in fish groups has led to complex
70 phenotypic and genomic patterns and complicated taxonomic relationships (Bernatchez &
71 Dodson, 1991; Turgeon, Estoup, & Bernatchez, 1999; Wilson & Bernatchez, 1998). One
72 example is the ciscoes in the genus *Coregonus* (subgenus *Leucichthys*), a group that contains
73 several species characterized by distinct morphological and ecological traits. An early study of
74 this group recognized eight distinct Great Lakes species in the *C. artedi* complex: *C. alpenae*, *C.*
75 *artedi*, *C. johannae*, *C. kiyi*, *C. hoyi*, *C. nigripinnis*, *C. reighardi* and *C. zenithicus*, as well as
76 several sub-species (Koelz, 1929). Later work has resulted in considerable debate over the
77 systematics of the group, due to phenotypic and genetic similarity among some of the species
78 (Todd, Smith, & Cable, 1981; Turgeon & Bernatchez, 2003; Turgeon, Estoup, & Bernatchez,
79 1999).

80 During the 20th century, human mediated effects such as overfishing, pollution, habitat
81 degradation and invasive species directly affected the Great Lakes (Mills, Leach, Carlton, &
82 Secor, 1994; Regier, Whillans, Christie, & Bocking, 1999). These changes led to population
83 declines and extinctions of many coregonines (Smith, 1968). For example, it is thought that
84 overfishing led to the extinction of six deep-water cisco forms in Lake Michigan (Eshenroder et
85 al., 2016). Moreover, there has been a substantial decline in the abundance of deep-water forms
86 in Lake Superior since the 1920s, especially *C. zenithicus* (Bronte et al., 2010; Hoff & Todd,
87 2004; Pratt, 2012). In many instances, population declines and extinctions have been followed by
88 expansion of more abundant species into habitats previously occupied by other groups. Such is
89 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

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

148 In this study, we aim to understand if there is concordance between morphology, feeding
149 habits, and genomic divergence among four species of *Coregonus* from Lake Superior: *C. artedi*,
150 *C. kiyi*, *C. hoyi*, and *C. zenithicus*. Despite ongoing debates on whether *C. nigripinnis* and *C.*
151 *reighardi* should be synonymized with *C. zenithicus* (Eshenroder et al., 2016; Todd & Smith,
152 1980), these groups were not included in the present study as they are very rare in Lake Superior.
153 Specific questions for our study include: 1) What are the main morphological traits that
154 characterize different species? 2) What are the main ecological traits that differentiate the deep-
155 water ciscoes *C. hoyi* and *C. zenithicus*? 3) Which highly divergent genes are associated with
156 ecological and morphological differences? By understanding the phenotypic, ecological, and
157 genomic differences among these species, this study presents an integrative view of the
158 divergence of fishes of the Laurentian Great Lakes.

METHODS

Sample collections

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

Morphometrics

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

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

184 Digital images were used to obtain 11 body measurements described in Eshenroder et al.
185 (2016): body depth (BDD), dorsal fin height (DOH), head length (HLL), maxillary length
186 (MXL), orbital length (OOL), pelvic-anal distance (PAD), pectoral fin length (PCL), preorbital
187 length (POL), pelvic fin length (PVL), pectoral-pelvic distance (PPD), and standard length (STL;
188 Table S2). The lengths of the PCL and PVL were divided by PPD and PAD, respectively,
189 providing fin-length-to-body-distance ratios. These ratios have been shown to be useful for
190 distinguishing *Coregonus* species (Koelz, 1929; Eshenroder et al., 2016). The premaxillary angle
191 (PMA) was also measured with a protractor but was not included in the PCA analysis because a
192 cursory analysis showed PMA was taxonomically uninformative.

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

202

203 **Stable Isotope Analysis**

We used stable isotope ratios ($\delta^{13}C$ and $\delta^{15}N$) 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 50-60°C, homogenized, and 0.5-1.0mg 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 $\delta^{13}C$ values it is important to normalize for lipid content, especially when $C:N_{bulk} > 4$ (Hoffman, Sierszen, & Cotter 2015). We applied the arithmetic-mass-balance normalization of Hoffman, Sierszen, & Cotter (2015) to all samples: $\delta^{13}C_{lipid\ free} = \delta^{13}C_{bulk} + (\Delta\delta^{13}C_{lipid} * (C:N_{lipid\ free} - C:N_{bulk})) / C:N_{bulk}$. For this equation, $C:N_{bulk}$ was molar, $C:N_{lipid\ free}$= 3.8 and the lipid discrimination ($\Delta\delta^{13}C_{lipid}$)= -6.4‰. The $\delta^{15}N$ values were not corrected to allow for detection of $\delta^{15}N$ 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

227 associated muscle and skin, first gill arch, eye, and brain. Tissue samples were obtained from the
228 same individuals used in the morphological and dietary analyses immediately after pithing.
229 These tissues are actively growing in adult fish, and therefore are expected to show variation
230 based on morphological and ecological differences of the analyzed species. Collected tissues
231 were preserved in RNAlater (Ambion, Inc.) and similar amounts were homogenized and pooled
232 for the RNA extractions. Total RNA was extracted from each homogenate using TRIzol (Life
233 Technologies, Inc.) and RNA Mini Isolation Kits (Ambion, Inc.) following manufacturers'
234 instructions (including treatment with DNase). RNA quality was assessed with a Bioanalyzer
235 2100 and quantified using a Qubit 2.0 fluorometer (Life Technologies, Inc.). Library preparation
236 was done at the RTSF Genomics Core facility at Michigan State University, using Illumina
237 TruSeq RNA library kits, with Illumina Ribo-Zero Gold for rRNA depletion. Libraries were
238 normalized and pooled for multiplex sequencing, and the pool was loaded on both lanes of an
239 Illumina HiSeq 2500 Rapid Run flow cell (paired-end, 250bp). Removal of Illumina adaptors
240 and quality filtering (Q > 30) was done with TrimGalore (Krueger, 2015). Resulting sequences
241 were used to assemble a reference transcriptome for each species using Trinity (Grabherr et al.,
242 2011), and only contigs longer than 300bp were maintained. Contigs were summarized using
243 Transdecoder (Haas et al., 2013) which only retains samples with predicted open reading frames.
244 CDHIT (Li & Godzik, 2006) was used to cluster contigs by similarity (95%), matching to the
245 most similar cluster in the dataset. The completeness of *C. artedi* transcriptomes (Trinity *vs.*
246 summarizing with Transdecoder and CDHIT) were assessed with the Benchmarking Universal
247 Single Copy Ortholog database version 3 (BUSCO; Simão, Waterhouse, Ioannidis, Kriventseva,
248 & Zdobnov, 2015), using the Actinopterygii database (*actinopterygii_odb9*). This resulted in
249 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 99 (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. artedi*, using Bowtie2 (Langmead & Salzberg, 2012) with default parameters. The rate of mapping to the *C. artedi* 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 (*f3 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 $\delta^{13}C$ and $\delta^{15}N$ values, but a small $\delta^{13}C$ range (0.3) and $\delta^{15}N$ range (0.7), indicating this species had the narrowest isotopic niche (Table 2). Meanwhile, samples of *C. artedi* had the

lowest mean $\delta^{13}C$ and $\delta^{15}N$ 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 (SD ± 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 (GIUL00000000). RNAseq reads from all species (eight individuals per species, seven for *C. zenithicus*) were mapped to the transcriptome of *C. artedi* (167,135

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

367 The molecular diversity indexes for all 31 samples were H_o= 0.27, H_e = 0.22, F_{is} = -0.26,
368 N_a = 147,841 and θ= 0.73 (Table S5). Average F_{st} across all samples was 0.018, and *Gtest* results
369 (p<0.01) indicate that the null hypothesis of panmixia was rejected, suggesting the presence of
370 genetic differentiation among species. Pairwise F_{st} estimates showed significant, albeit low,
371 levels of genetic differentiation between species (Table 4). The largest differences were found
372 between *C. kiyi* and *C. zenithicus* (F_{st}= 0.025, $p< 0.01$, Figure 4S) and *C. artedi* vs. *C. kiyi* (F_{st} =
373 0.024, $p< 0.01$; Figure 3a). The lowest level of genetic divergence was found between *C. hoyi*
374 and *C. zenithicus* (F_{st}= 0.00, $p>0.788$; Figure 3b).

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

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

388 2c). Meanwhile, no overlap was observed between *C. kiyi* and *C. artedi*.

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

396 The three-population analysis revealed the group formed by *C. hoyi* and *C. zenithicus* is
397 admixed, as the Z-scores ranged from -9.72 to -14.79 (Table S6). Even when we detected
398 potential hybrids between other species with the PCA and assignment test (i.e., *C. hoyi* and *C.*
399 *kiyi*, *C. artedi* and *C. zenithicus*), these comparisons showed no significant hybridization with the
400 *f3 statistic*. This is probably because admixture was only observed for one individual of the
401 compared groups (Table S6).

402 The pairwise *Fst* comparisons between species also allowed detecting highly
403 differentiated genes between species (Supplementary Data 1-6). Thus, the most differentiated
404 genes for each comparison were: *Fork-head Box Q1a* (*Fst*=0.74) between *C. artedi* and *C. hoyi*
405 (Figure S4); *Rhodopsin* (*Fst*=0.83) between *C. artedi* and *C. kiyi* (Figure 3a); *RCC1 Domain*
406 *Containing 1* (*Fst*=0.64) for *C. artedi* and *C. zenithicus* (Figure S4); *Zinc Finger MYM-Type*
407 *Containing 2* (*Fst*=0.86) for *C. hoyi* and *C. kiyi* (Figure S4); *Spectrin Beta, Non-Erythrocytic 2*
408 (*Fst*=0.54) for *C. hoyi* and *C. zenithicus* (Figure 3b); *Vacuolar Protein Sorting 13 Homolog A*
409 (*Fst*= 0.78) for *C. kiyi* and *C. zenithicus* (Figure S4). The analysis of selection with Ohana
410 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).

457

458 **Isotopic measures show agreement with previous studies:**

459 Based on isotopic measurements, it appears that the main driving mechanism of dietary partitions
460 is depth preference, which in turn influences prey availability. These findings confirm previous
461 observations that muscle derived $\delta^{13}C$ concentrations in *C. kiyi* and *C. hoyi*/*C. zenithicus* are less
462 depleted than in *C. artedi*. This reflects preference for pelagic habitat and primary consumption
463 of limnetic zooplankton during summer for the latter (Sierszen et al., 2014; Rosinski et al.,
464 2020). This result is expected given the near-obligate pelagic existence of adult individuals of *C.*
465 *artedi* during summer stratification (Rosinski, Vinson, & Yule, 2020*)*. Meanwhile, high average
466 $\delta^{13}C$ and $\delta^{15}N$ values of *C. kiyi* are consistent with a diet dominated by *Mysis diluviana* (Gamble,
467 Hrabik, Stockwell, & Yule, 2011a; Ahrenstorff et al., 2011; Sierszen et al., 2014*)* and preference
468 for deep waters (Rosinski et al., 2020). The intermediate position of *C. hoyi* suggest they occupy
469 intermediate depths and are opportunistic feeders with a diet that is obtained from pelagic and
470 benthic zones. This result is also consistent with previous studies that showed *C. hoyi* change
471 their diet seasonally (*Diporeia* in spring, *Mysis diluviana* in summer, and *Daphnia* spp. and
472 calanoid copepods in autumn; Gamble, Hrabik, Stockwell, & Yule, 2011b). When compared to
473 the other species, the isotopic values of *C. zenithicus* (Figure 2b) suggest that some individuals in
474 this group may be obtaining part of their nutrition from the pelagic zone, through consumption of
475 zooplankton. Samples of *C. zenithicus* from this study had terminal mouths similar to *C. artedi*,
476 which could explain the wide breath of dietary components compared to the other groups. Both
477 the present study and Schmidt et al. (2009) have *C. zenithicus* and *C. hoyi* occupying similar
478 trophic positions, and both show *C. zenithicus* obtaining nutrition from more $\delta^{13}C$ depleted prey
479 resources relative to *C. hoyi.* Yet, Schmidt et al. (2009) showed these two species occupied

480 unique isotopic space (especially for their Lake Superior samples of 1972-1998), while we
481 obtained far greater overlap. Given these contrasting results, it is necessary to expand the studies
482 on the dietary overlap between *C. hoyi* and *C. zenithicus* among different populations of Lake
483 Superior.
484
485 **Multiple lines of evidence support divergence of *C. artedi* and *C. kiyi*:**

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

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

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

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

517 One of the revealing findings of our study was the limited differentiation between samples
518 initially identified as *C. zenithicus* and *C. hoyi*. At first sight, the isotopic analysis suggests *C.*
519 *hoyi* are obtaining more nutrition from the benthic zone, but this signal could be confounded by
520 seasonal variation. Further, samples identified as *C. zenithicus* and *C. hoyi* showed no significant
521 differences in the overall F_{st}, and the *f3-statistic* indicated rampant admixture between the two
522 taxa. Still, the pairwise comparisons of locus specific F_{st} showed differentiation for genes
523 associated with nerve growth and neural development (*Spectrin Beta Non-Erythrocytic 2*,
524 F_{st}=0.54; *Growth Associated Protein 43*, F_{st}=0.47; Benowitz & Routtenberg, 1997; Lise et al.,

525 2012), as well as cell proliferation in vertebrates (*Tensin 2*; *Fst*=0.49; Hafizi, Ibraimi & Dahbäck,
526 2005). These highly differentiated loci did not influence the low overall F_{st} as they represent a
527 very small proportion of the sampled loci. Future studies across populations of *C. hoyi* and *C.*
528 *zenithicus* are necessary to understand the full extent of the genomic divergence between the two
529 species.

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

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

548 2012), could have promoted hybridization. In our study, we also show evidence of admixture
549 between 1 or 2 individuals *C. hoyi* and *C. kiyi*, as well as *C. zenithicus* and *C. artedi*. Previous
550 studies have demonstrated that hybridization is possible between closely related coregonines in
551 both North America and Europe (Hudson, Lundsgaard□Hansen, Lucek, Vonlanthen, &
552 Seehausen, 2017; Hudson, Vonlanthen, & Seehausen, 2011; Kirtiklis & Jankun, 2006; Todd &
553 Stedman, 1989), and successful crosses of *C. artedi* and *C. hoyi* have been made in captivity (W.
554 Stott, personal communication). Further, hybridization has been suggested as one of the drivers
555 of morphological changes in *C. hoyi* in Lake Huron after the decline of other species in the
556 complex (Todd & Stedman, 1989). This scenario of speciation reversal in *C. zenithicus*, where
557 morphological diversity could have been lost due to both human-mediated stressors and
558 admixture with closely related species, could represent an example of how commercially and
559 ecologically relevant lineages have been lost in recent decades. It remains to be determined if
560 other populations of *C. zenithicus* and *C. hoyi* in Lake Superior are also admixed, given the
561 limited number of samples included in the present study. This is especially relevant considering
562 previous surveys have reported finding phenotypically distinct *C. zenithicus* in Canadian waters
563 of Lake Superior (Pratt & Chong, 2012; Pratt, 2012).

564 **Concordance among morphology, ecology and genetics:**

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

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

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

591 Overall, the results of our study suggest that differentiation across *Coregonus* could be
592 promoted by habitat preferences (i.e., depth and benthic/pelagic habit) and dietary partitions,
593 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

617 analyses were able to distinguish between three of the four species, and high levels of
618 differentiation in genes associated with body shape, eye size, fin shape, vision, organization of
619 the nervous system and early development. There was evidence of hybridization between
620 samples morphologically identified as *C. hoyi* and *C. zenithicus* and future studies should assess
621 the extent of this admixture in Lake Superior. Overall, the integrative approach applied in this
622 study exemplifies how multiple lines of evidence can help elucidate the relationships among
623 sympatric groups with complex evolutionary history, such Great Lakes *Coregonus* species.
624

625 **Acknowledgements:**

626 We thank the U.S. Geological Survey Research Vessel Kiyi Captain Joe Walters, First Mate
627 Keith Peterson, and Engineer Charles Carrier. Special thanks to Ryan Menenbroeker who
628 measured our specimens, Jean Adams for analyzing the morphometric data, and to Caroline
629 Rosinski, Mark Vinson, and Andrew Muir for providing helpful insight. We would like to thank
630 Tianying Lan, Nathan Backenstose, Katherine Eaton, and Jessie Pelosi for their help with the
631 genetic analyses and data handling. Early drafts of this manuscript were improved with solicited
632 reviews by Tom Pratt and Gary Longo. Any use of trade, product, or firm names is for
633 descriptive purposes only and does not imply endorsement by the U.S. Government. Funding
634 was provided by start-up funds to TED from Wayne State University and the Great Lakes
635 Fishery Commission (Award no. 2018-KRA-44073).
636

637 **REFERENCES**

638
639 1. Ackiss, A. S., Larson, W. A., & Stott, W. (2020). Genotyping-by-sequencing illuminates
640 high levels of divergence among sympatric forms of coregonines in the Laurentian Great
641 Lakes. Evolutionary Applications, 13(5), 1037–1054. doi: 10.1111/eva.12919

642 2. Albertson, R. C. (2008). Morphological Divergence Predicts Habitat Partitioning in a Lake
643 Malawi Cichlid Species Complex. Copeia, 2008(3), 689–698. doi: 10.1643/CG-07-217
644 3. Arnold, M. L. (2006). Evolution Through Genetic Exchange. OUP Oxford.
645 4. American Fisheries Society (AFS). 2014. Guidelines for the use of fishes in research.
646 American Fisheries Society, Bethesda, Maryland.
647 5. Arnold, M. L., Bulger, M. R., Burke, J. M., Hempel, A. L., & Williams, J. H. (1999). Natural
648 hybridization: how low can you go and still be important? Ecology, 80(2), 371–381.
649 6. Ahrenstorff, T.D., T.R. Hrabik, J.D. Stockwell, D.L. Yule, and G.G. Sass. 2011. Seasonal
650 dynamic diel vertical migrations of *Mysis diluviana*, coregonine fishes, and siscowet lake
651 trout in the pelagia of western Lake Superior. Transactions of the American Fisheries Society
652 140:1504-1520.
653 7. Bager, R., Kristensen, T.K., Jensen, J.K., Szczur, A., Christensen, A., Andersen, L.MM
654 Johansen, J.S., Larsen, N., Baatrup, E., Huang, M., Ploug, M., & Andreasen, A.A. (2012).
655 Urokinase-type plasminogen activator-like proteases in teleosts lack genuine receptor-
656 binding epidermal growth factor-like domains. Journal of Biological Chemistry 287 (33):
657 27526-27536.
658 8. Bailey, R. M., & Smith, G. R. (1981). Origin and geography of the fish fauna of the
659 Laurentian Great Lakes basin. Canadian Journal of Fisheries and Aquatic Sciences, 38(12),
660 1539–1561.
661 9. Benowitz, L. I., & Routtenberg, A. (1997). GAP-43: an intrinsic determinant of neuronal
662 development and plasticity. Trends in Neurosciences, 20(2), 84–91. doi: 10.1016/S0166-
663 2236(96)10072-2
664 10. Bernardi, G. (2013). Speciation in fishes. Molecular Ecology, 22(22), 5487–5502. doi:
665 10.1111/mec.12494
666 11. Bernatchez, L., & Dodson, J. J. (1991). Phylogeographic structure in mitochondrial DNA of
667 the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations.
668 Evolution, 45(4), 1016–1035.
669 12. Bronte, C. R., Hoff, M. H., Gorman, O. T., Thogmartin, W. E., Schneeberger, P. J., & Todd,
670 T. N. (2010). Decline of the Shortjaw Cisco in Lake Superior: The Role of Overfishing and
671 Risk of Extinction. Transactions of the American Fisheries Society, 139(3), 735–748. doi:
672 10.1577/T09-019.1
673 13. Bunnell, D. B., Eshenroder, R. L., Krause, A. E., & Adams, J. V. (2012). Depth segregation
674 of deepwater ciscoes (*Coregonus spp.*) in Lake Michigan during 1930–1932 and range
675 expansion of Coregonus hoyi into deeper waters after the 1990s. Advances in Limnology, 63,
676 3–24.
677 14. Carruthers, M., Yurchenko, A. A., Augley, J. J., Adams, C. E., Herzyk, P., & Elmer, K. R.
678 (2018). *De novo* transcriptome assembly, annotation and comparison of four ecological and
679 evolutionary model salmonid fish species. BMC Genomics, 19(1), 32. doi: 10.1186/s12864-
680 017-4379-x
681 15. Cheng, J. Y., Mailund, T., & Nielsen, R. H. (2016). Ohana, a tool set for population genetic
682 analyses of admixture components. doi: 10.1101/071233
683 16. Cheng, J. Y., Racimo, F., & Nielsen, R. (2019). Ohana: detecting selection in multiple
684 populations by modelling ancestral admixture components. BioRxiv, 546408.
685 17. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin,
686 R. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. doi:
687 10.1093/bioinformatics/btr330

688 18. Dos Reis, S. F., Pessôa, L. M., & Strauss, R. E. (1990). Application of size-free canonical
689 discriminant analysis to studies of geographic differentiation. Revista Brasileira de Genética,
690 13(3), 509-520.
691 19. Eaton, K. M., Bernal, M. A., Backenstose, N. J., & Krabbenhoft, T. (2020). Nanopore
692 amplicon sequencing reveals molecular convergence and local adaptation of opsin genes.
693 20. Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome
694 comparisons dramatically improves orthogroup inference accuracy. Genome Biology, 16(1),
695 157. doi: 10.1186/s13059-015-0721-2
696 21. Eshenroder, R. L., Sideleva, V. G., & Todd, T. N. (1999). Functional convergence among
697 pelagic sculpins of Lake Baikal and deepwater ciscoes of the Great Lakes. *Journal of Great*
698 *Lakes Research*, *25*(4), 847-855.
699 22. Eshenroder, R. L., Vecsei, P., Gorman, O. T., Yule, D., Pratt, T. C., Mandrak, N. E., … Muir,
700 A. M. (2016). Ciscoes (*Coregonus*, subgenus *Leucichthys*) of the Laurentian Great Lakes and
701 Lake Nipigon. Great Lakes Fishery Commission.
702 23. Etheridge, E. C., Bean, C. W., Maitland, P. S., Ballantyne, S., & Adams, C. E. (2012).
703 Discontinuous infraspecific variation in ecological and morphological traits has
704 consequences for conservation of powan (*Coregonus lavaretus*) in Scotland. Advances in
705 Limnology, 63, 505–517.
706 24. Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of
707 sympatric whitefish species flocks. Molecular Ecology, 28(3), 615–629. doi:
708 10.1111/mec.14977
709 25. Fitzpatrick, S. W., Gerberich, J. C., Kronenberger, J. A., Angeloni, L. M., & Funk, W. C.
710 (2015). Locally adapted traits maintained in the face of high gene flow. Ecology Letters,
711 18(1), 37–47. doi: 10.1111/ele.12388
712 26. Gamble, A.E., Hrabik, T.R., Stockwell, J.D., & Yule, D.L. (2011a). Trophic connections in
713 Lake Superior part I: the offshore fish community. Journal of Great Lakes Research, 37, 541-
714 549.
715 27. Gamble, A.E., Hrabik, T.R., Stockwell, J.D., & Yule, D.L. (2011b). Trophic connections in
716 Lake Superior part II: the nearshore fish community. Journal of Great Lakes Research, 37,
717 550-560.
718 28. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... &
719 Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from
720 RNA-Seq data. Nature biotechnology, 29(7), 644.
721 29. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., …
722 Lieber, M. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the
723 Trinity platform for reference generation and analysis. Nature Protocols, 8(8), 1494–1512.
724 30. Hafizi, S., Ibraimi, F., & Dahlbäck, B. (2005). C1-TEN is a negative regulator of the
725 Akt/PKB signal transduction pathway and inhibits cell survival, proliferation, and
726 migration. The FASEB journal, 19(8), 971-973.
727 31. Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O., & Puebla, O. (2019). Inter-
728 chromosomal coupling between vision and pigmentation genes during genomic divergence.
729 Nature Ecology & Evolution, 3(4), 657–667. doi: 10.1038/s41559-019-0814-5
730 32. Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., Folkvord, A., …
731 Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish
732 rhodopsin. Proceedings of the National Academy of Sciences, 116(37), 18473–18478. doi:
733 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 δ13C 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

780 49. Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., ... & Grammes, F.
781 (2016). The Atlantic salmon genome provides insights into rediploidization. Nature,
782 533(7602), 200-205.
783 50. Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Parolin
784 Schnekenberg, R., … Németh, A. H. (2012). Recessive Mutations in SPTBN2 Implicate β-III
785 Spectrin in Both Cognitive and Motor Development. PLoS Genetics, 8(12). doi:
786 10.1371/journal.pgen.1003074
787 51. Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans
788 for selection based on principal component analysis. Molecular Ecology Resources, 17(1),
789 67–77.
790 52. Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018).
791 Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. Molecular Biology
792 and Evolution, 35(6), 1489–1506. doi: 10.1093/molbev/msy051
793 53. Meirmans, P. G., & Tienderen, P. H. V. (2004). genotype and genodive: two programs for
794 the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes, 4(4), 792–
795 794. doi: 10.1111/j.1471-8286.2004.00770.x
796 54. Mills, E. L., Leach, J. H., Carlton, J. T., & Secor, C. L. (1994). Exotic Species and the
797 Integrity of the Great Lakes. BioScience, 44(10), 666–676. doi: 10.2307/1312510
798 55. Mioduszewska, B., Jaworski, J., & Kaczmarek, L. (2003). Inducible cAMP early repressor
799 (ICER) in the nervous system – a transcriptional regulator of neuronal plasticity and
800 programmed cell death. Journal of Neurochemistry, 87(6), 1313–1320. doi: 10.1046/j.1471-
801 4159.2003.02116.x
802 56. Muir, A. M., Krueger, C. C., & Hansen, M. J. (2012). Re-establishing lake trout in the
803 Laurentian Great Lakes: past, present, and future. *Great Lakes fishery policy and*
804 *management: a binational perspective, 2nd edition. Michigan State University Press, East*
805 *Lansing*, 533-588.
806 57. Musilova, Z., Cortesi, F., Matschiner, M., Davies, W. I. L., Patel, J. S., Stieb, S. M., …
807 Salzburger, W. (2019). Vision using multiple distinct rod opsins in deep-sea fishes. Science,
808 364(6440), 588–592. doi: 10.1126/science.aav4632
809 58. Newsome, S. D., Martinez del Rio, C., Bearhop, S., & Phillips, D. L. (2007). A niche for
810 isotopic ecology. Frontiers in Ecology and the Environment, 5(8), 429-436.
811 59. Ohlberger, J., Mehner, T., Staaks, G., & Hölker, F. (2008). Temperature-related
812 physiological adaptations promote ecological divergence in a sympatric species pair of
813 temperate freshwater fish, *Coregonus spp*. Functional Ecology, 22(3), 501–508. doi:
814 10.1111/j.1365-2435.2008.01391.x
815 60. Olson, K. W., Krabbenhoft, T. J., Hrabik, T. R., Mendsaikhan, B., & Jensen, O. P. (2019).
816 Pelagic–littoral resource polymorphism in Hovsgol grayling *Thymallus nigrescens* from Lake
817 Hovsgol, Mongolia. Ecology of Freshwater Fish, 28(3), 411–423. doi: 10.1111/eff.12464
818 61. Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular
819 approach. Bioinformatics, 26(3), 419–420. doi: 10.1093/bioinformatics/btp696
820 62. Parsons, M. J., Campos, I., Hirst, E. M. A., & Stemple, D. L. (2002). Removal of
821 dystroglycan causes severe muscular dystrophy in zebrafish embryos. Development, 129(14),
822 3505–3512.
823 63. Perreault-Payette, A., Muir, A. M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P., &
824 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. Schliewen, 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 artedi*) 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 artedi*) 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

915 94. Wright, R. M., Aglyamova, G. V., Meyer, E., & Matz, M. V. (2015). Gene expression
916 associated with white syndromes in a reef building coral, *Acropora hyacinthus*. BMC
917 Genomics, 16, 371. doi: 10.1186/s12864-015-1540-2
918 95. Yule, D.L., J.V. Adams, T.R. Hrabik, M.R. Vinson, Z. Woiak, and T.D. Ahrenstorff. (2013).
919 Use of classification trees to apportion single echo detections to species: application to the
920 pelagic fish community of Lake Superior. Fisheries Research, 140, 123-132.
921 96. Zhang, L., Thibert-Plante, X., Ripa, J., Svanbäck, R., & Brännström, Å. (2019). Biodiversity
922 loss through speciation collapse: mechanisms, warning signals, and possible rescue.
923 Evolution, 73(8), 1504-1516.
924

925 Data Accessibility:
926
927 The raw reads for all individuals and the assembled transcriptome of *Coregonus artedi* are
928 available in NCBI as BioProject PRJNA659559. The bioinformatic scripts used for the
929 transcriptomic analyses are available at **github.com/evofish.**
930
931 Author Contributions:
932
933 TED and TJK conceived the study; DY and LE collected the samples; DLY, LE, and TJK
934 collected the data; MAB, DY, WS and TJK analyzed the data and interpreted the results; the
935 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 (±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)**
PC1	-	0.6	-0.22	0.47	0.25	-0.15	0.4	0.24	-0.14
PC2	-	0.4	-0.21	0.07	-0.19	0.24	-0.73	0.14	-0.32
C. artedi	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)
C. hoyi	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)
C. zenithicus	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)
C. kiyi	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 $\delta^{13}C_{lipid\ free}$ and $\delta^{15}N$ (‰), 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 $\delta^{15}N$ (range)**	**Mean $\delta^{13}C$ (range)**	**SEAc**
C. artedi	8	6.7 (0.9)	-25.9 (1.1)	2.33
C. hoyi	8	8.1 (1.5)	-25.3 (0.8)	2.41
C. zenithicus	7	7.8 (2.5)	-25.7 (1.0)	7.19
C. kiyi	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	*C. artedi*	*C. hoyi*	*C. kiyi*	*C. zenithicus*
C. artedi	-	8.80	0.00	29.22
C. hoyi	8.53	-	24.66	29.23
C. kiyi	0.00	5.82	-	5.18
C. zenithicus	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	*C. artedi*	*C. hoyi*	*C. kiyi*	*C. zenithicus*
C. artedi	-	<0.001	<0.001	<0.001
C. hoyi	0.015	-	<0.001	0.788
C. kiyi	0.024	0.023	-	<0.001
C. zenithicus	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 artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus*. Detailed information about sites are provided in Supplementary Table S1. Photographs by D.L Yule.

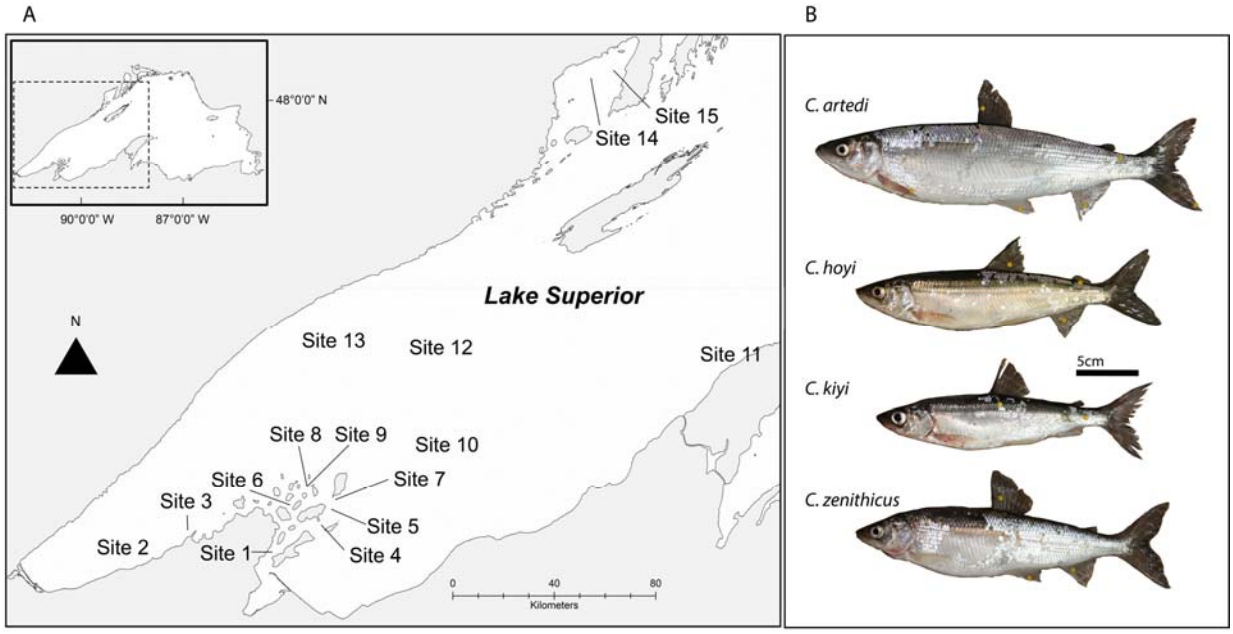

Figure 2. Principal Component Analysis (PCA) for (A) morphological measures; (B) bi-plot for isotope ratios; and (C) PCA of transcriptomic variance (22,285 SNPs) for four species of *Coregonus* from Lake Superior. *Coregonus artedi* (red), *C. hoyi* (blue), *C. kiyi* (green) and *C. zenithicus* (purple).

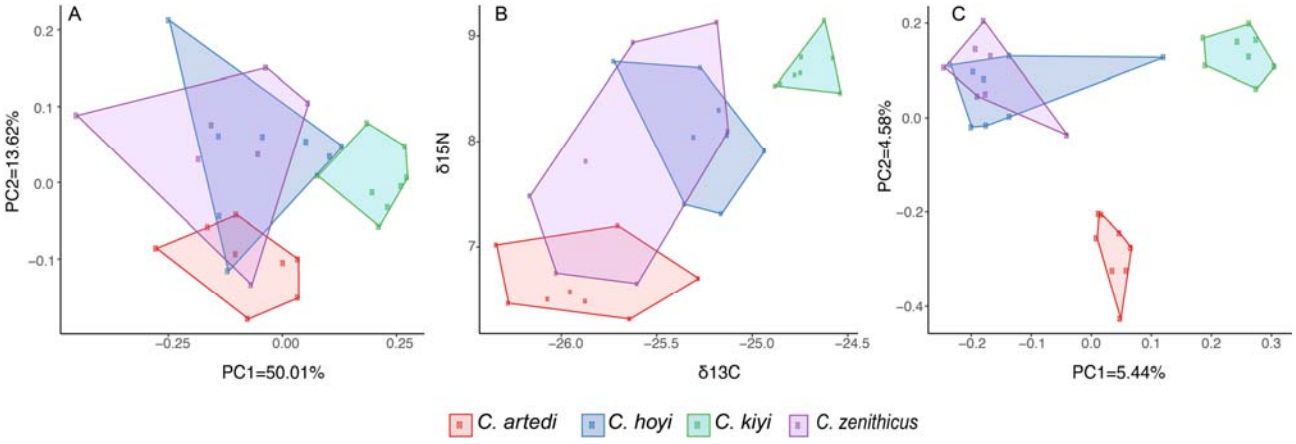

Figure 3. Histograms of pairwise F_{st} for individual genes in pairwise comparisons: between (A) the shallow-water *C. artedi* and deep-water *C. kiyi*, and (B) the two species with the lowest levels of divergence *C. zenithicus* and *C. hoyi*. The X-axis represents values of F_{st} (with binning of 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.

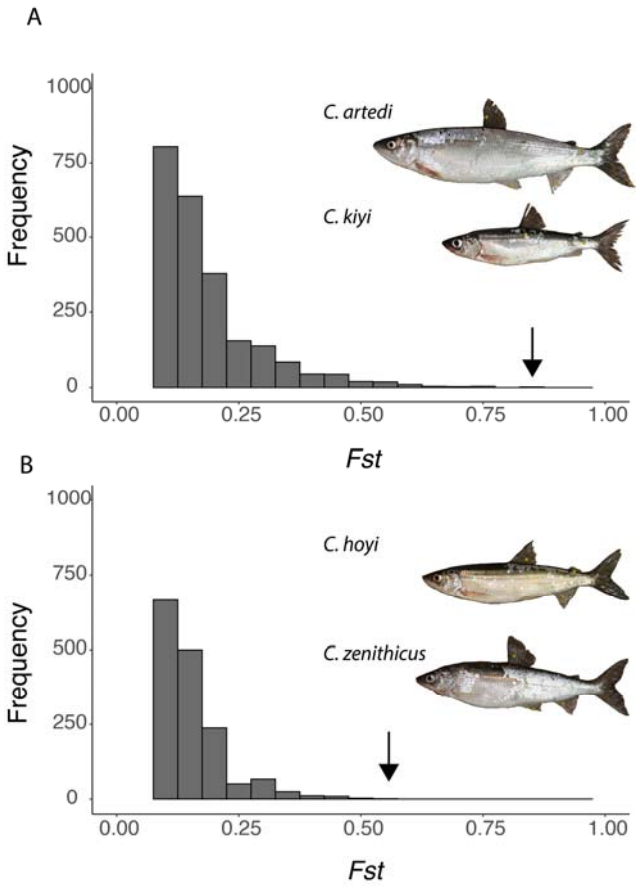

Figure 4. Admixture plots for *Coregonus artedi*, *C. hoyi*, *C. kiyi* and *C. zenithicus* using 22,285 high-confidence SNPs, simulating (A) three groups (*K*=3) and (B) four groups (*K*=4).