Genomics reveals identity, phenology and population demographics of larval ciscoes (*Coregonus artedi*, *C. hoyi*, and *C. kiyi*) in the Apostle Islands, Lake Superior

Hannah Lachance[a], Amanda S. Ackiss[b,c], Wesley A. Larson[d,e], Mark R. Vinson[f], and Jason D. Stockwell[a,1,*]

[a]Rubenstein Ecosystem Science Laboratory, University of Vermont, Burlington, Vermont 05401

[b]Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin - Stevens Point, 800 Reserve St. Stevens Point, WI 54481

[c]U.S. Geological Survey, Great Lakes Science Center, 1451 Green Road, Ann Arbor, MI 48105

[d]U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve St. Stevens Point, WI 54481

[e]National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, AK, 99801

[f]U.S. Geological Survey, Great Lakes Science Center, Lake Superior Biological Station, Ashland, WI 54806; Email: mvinson@usgs.gov

*corresponding author: jason.stockwell@uvm.edu

[1]*Given their role as Associate Editor, Jason Stockwell had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to Handling Editor name Wendylee Stott*

Abstract

We demonstrate, for the first time, the ability to reliably assign an assemblage of larval coregonines [Salmonidae Coregoninae] to shallow and multiple deepwater species. Larval coregonines from the Apostle Islands, Lake Superior, were genotyped using restriction site-associated DNA sequencing (RADseq) and were assigned to species using reference genotypes from adult corgonines from the same region. Of the 193 genotyped larvae, 101 were assigned as *Coregonus artedi* (average assignment probability = 97.6%), 57 were assigned as *C. kiyi* (average assignment probability = 95.5%), and 28 were assigned as *C. hoyi* (average assignment probability = 89.0%). *Coregonus artedi* were collected earliest in the season, followed by *C. kiyi*, and *C. hoyi*. Estimates of genetic diversity within each species provide a baseline for future monitoring in the Apostle Islands. Our success with species assignment indicates the promise of leveraging genomic data for larval coregonine identification, which could enable assessing and evaluating early life history dynamics and recruitment processes at the species level to the benefit of ongoing coregonine restoration and management efforts.

Key words: Restriction site-associated DNA sequencing; Early Life History; larval fish; Coregonus; Great Lakes

Introduction

The Laurentian Great Lakes were home to at least 11 coregonine species (Salmonidae Coregoninae) which filled many ecological niches, served many ecological functions, and fueled an economic engine during the 19th and 20th Centuries (Baldwin et al., 2002; Eshenroder et al., 2016; Koelz, 1929; Smith, 1964). However, only nine species remain, and most are restricted to Lake Superior with small populations in lakes Huron, Michigan and Ontario, and only lake whitefish (*Coregonus clupeaformis*) in Lake Erie (Eshenroder et al., 2016). The ecological and economic importance of coregonines has led to a growing and active interest in their restoration and rehabilitation in all the Great Lakes (Bronte et al., 2017; Zimmerman and Krueger 2009). For example, bloater (*C. hoyi*) are currently propagated in hatcheries and stocked in Lake Ontario to re-establish self-sustaining populations (OMNRF, 2015; Klinard et al., 2020), as are cisco (*C. artedi*) in outer Saginaw Bay, Lake Huron (Riley and Ebener, 2020).

Most coregonines can be visually identified at the adult life stage using morphometric attributes (Eshenroder et al., 2016; Koelz, 1929). Past research on wild-caught coregonine larvae relied on best available knowledge for species identification, including visual cues (Hinrichs 1979; Auer 1982) and assumptions about adult populations and distributions (e.g., Myers et al., 2008; Oyadomari and Auer 2008). However, recent work using genetic barcoding of the mitochondrial cytochrome C oxidase I (COI) gene and microsatellite DNA indicate our ability to visually distinguish lake whitefish from other coregonine species is poor (George et al., 2018; McKenna et al., 2020). Microsatellite DNA has been used successfully to discriminate between larval cisco and bloater (Claramunt et al., 2019; Stott et al., 2021), but no method has yet been developed to discriminate the multiple deepwater coregonine species from one another or from cisco and lake whitefish.

The inability to identify all coregonine species at early life stages limits our ability to study species' life histories during a critical part of their life cycle. If coregonine larvae cannot be identified, we cannot know when each species hatches, the habitats they occupy, or their growth and mortality rates. Such information is critical for assessing niche overlap among native and non-native species, and without it, efforts to restore and rehabilitate coregonine populations throughout the Great Lakes will be compromised (Bronte et al., 2017).

Initial genetic analyses of coregonines, based on microsatellites, indicated minimal genetic variation exists among species (Turgeon and Bernatchez, 2003). However, microsatellites represent a small portion of the genome. Advanced genomic sequencing techniques capture more of the genome, and therefore can better identify differences among species. Recent research employing restriction-site associated DNA sequencing (RADseq) illustrated the ability of genome-wide data to differentiate adult *C. artedi*, *C. hoyi* and *C. kiyi* (Ackiss et al. 2020).

In this exploratory study, we used RADseq to identify coregonine larvae collected in the Apostle Islands, Lake Superior, USA. Larvae were assigned to species using reference RADseq data from known adult coregonines previously genotyped from the Apostle Islands (Ackiss et al., 2020). Several demographic traits of each identified species were then assessed and compared as a preliminary evaluation of possible early life stage differences. Additional genetic analyses were conducted to provide information about the genetic diversity of each species, including effective population size (N_e), and observed and expected heterozygosity (H_o and H_e).

Methods and Materials

Sample collection

Samples were collected during a larval coregonine foraging study, with sampling protocols detailed in Lucke et al. (2020). Samples were collected weekly at 10 sites throughout the Apostle Islands, Lake Superior, between May 14 and July 25, 2018, during the day (see Fig. 1 in Lucke et al., 2020). Larval fish were collected with a bongo net equipped with 0.5 m diameter, 500 µm mesh conical nets (Myers et al., 2008). Each tow was conducted on a straight transect for 10 min with the net approximately 15 m behind the vessel, moving downwind, with the net frame visible at the surface. Tow speed averaged 4.6 km/h. Starting bathymetric depths of larval fish tows averaged 26.5 m (range: 2.3 - 67.7 m) and ending depths averaged 24.3 m (range: 2.3 - 69.5 m). At the completion of each tow, larvae were immediately counted and preserved in 90% ethanol.

Phenology and demographic measurements

Larvae were imaged and assessed on a dissecting microscope equipped with a camera (Olympus SZX7, Waltham, MA). *Coregonus* larvae were visually identified as lake whitefish or "other coregonines" based on visual keys (Auer, 1982) and genetic analyses were performed on the "other coregonines" larvae. Yolk sac condition (YSC) was assessed as "yolk sac and oil globule present", "oil globule only", or "absorbed".

The length of each larva was measured to the nearest 0.01 mm on the microscope images using the Leica Application Suite (Leica Microsystems Inc, Buffalo Grove, IL). Stomachs were removed from each individual for a diet study (Lucke et al., 2020) and the remaining body was preserved in 100% ethanol until DNA extraction. The instruments and bench space were sterilized between the processing of each larva.

DNA extraction and barcoding

Larvae were subset first by sampling week. Two early-season weeks, two mid-season weeks and four late-season weeks were chosen to ensure we had relatively equal sample sizes for the beginning, middle, and end of the sampling period. Within each of these weeks, length ranges from which to randomly sample larvae for our analyses were set to follow the primary cohort of coregonine larvae throughout the larval stage. Additionally, we selected several small individuals from the last three weeks of sampling, as their size suggested a possible new cohort of emerging larvae and perhaps a different species than the cohort that appeared over the first half of the sampling period.

DNA was extracted from 221 larvae using Qiagen DNeasy Blood and Tissue kits (Germantown, MD). The 221 larvae were genetically identified as lake whitefish or non-whitefish coregonines by analyzing the restriction fragment length polymorphisms of the cytochrome C oxidase I gene to identify lake whitefish larvae that may have been missed by our visual assessment (George et al., 2018). PCR was used to amplify the mitochondrial cytochrome C oxidase I gene using the following conditions: an initialization and denaturing step at 95°C for 2 min followed by 36 cycles of 95°C for 30 sec, 52°C for 40 sec, 68°C for 1 min, and a 10 min final elongation at 95°C. Twenty microliters of the PCR product was combined with 1.0 µl of the restriction enzyme *Eco0109I* and 4 µl of the *Eco0109I* buffer (New England BioLabs, Ipswich, MA). The restriction enzyme solution was run for 2 hrs at 37°C. The digested PCR product was then run next to a 100 bp ladder on a 2% gel at 100 V for at least 30 min so the species could be identified by the number and size of the DNA band(s) present.

Library preparation, sequencing, and genotyping

Restriction enzyme-associated DNA (RAD) libraries were prepared using the bestRAD protocol (Ali et al., 2016) and followed the methods of Ackiss et al. (2020). Whole genomic DNA was normalized, digested with the restriction enzyme *SbfI*, and barcoded by individual larva. Pooled libraries of 96 individuals were sonicated to ~300-500bp fragments, and non-target sequences were removed using Dynabeads™ M-280 Streptavidin magnetic beads (Invitrogen) to bind target fragments. Master library barcodes were then ligated followed by size selection and a 12-cycle PCR enrichment using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina®. Library products were quantified with a Qubit® 2.0 fluorometer and sent to Novogene (Sacramento, CA) for sequencing on two lanes of an Illumina® HiSeq 4000 (San Diego, CA). Ten additional larvae from this study were included in a separate library that was prepared concurrently and sent to Novogene as part of a larger sequencing run on an Illumina® NovaSeq S4.

Raw sequences of larval samples were quality controlled and filtered with Stacks (v.2.3d; Catchen et al., 2011, 2013) at the same time as 106 reference adult coregonines identified as *C. artedi*, *C. hoyi*, and *C. kiyi* from Ackiss et al. (2020) to ensure the best comparison for downstream species identification. Reads were demultiplexed using barcodes, filtered for presence of cut site, and trimmed to 140 bp using *process_radtags,* and reads with low quality scores were discarded (parameter flags: -e *SbfI* -c -q -r -t 140 --filter_illumina --bestrad). Sequences were aligned into groups of matching sequences assumed to represent homologous DNA regions using the bounded model in *ustacks* with a minimum depth of 3 to create a group, a maximum number of mismatches of 5, and calling haplotypes from secondary reads and gapped alignments disabled (parameter flags: --disable-gapped -m 3 -M 5 -H --max_locus_stacks 4 --

model_type bounded --bound_high 0.05), and a catalog of consensus loci was created using 126 ciscoes from the Great Lakes, including 51 individuals used in the creation of a *C. artedi* linkage map (Blumstein et al., 2020), 5 Apostle Islands larvae, and 5 adults of *C. artedi*, 5 adults of *C. hoyi,* 5 adults of *C. kiyi*, and 5 adults each of *C. nigripinnis* and *C. zenithicus* from various locations across Lake Superior with 3 mismatches allowed per sample locus and gapped alignments disabled (parameter flags: -n 3 --disable_gapped). Individual reads were matched to the catalog and variant sites were called using *gstacks*. *Populations* was used to write loci genotyped in at least 30% of individuals to a variant call format (vcf) file for further data filtering (parameter flag: -r 0.3).

After Stacks, VCFtools (v.0.1.15; Danecek et al., 2011) was used to filter out 1) loci that were missing in more than 20% of individuals (i.e. a genotyping rate of 80%), 2) individuals that were missing 50% of the loci, and 3) loci with a minor allele count (MAC) of less than 3. Because salmonids, including coregonines, have undergone a recent whole genome duplication, HDplot (McKinney et al., 2017) was used to identify putatively paralogous loci with a heterozygosity cutoff of > 0.55 and a read ratio deviation of > 5 or < -5, and flagged loci were removed with VCFtools. To account for linkage between multiple single nucleotide polymorphisms (SNPs) on a RAD locus, the SNP with the highest minor allele frequency (MAF) from each RAD tag was retained for downstream analysis. PGDSpider (v2.1.1.5: Lischer and Excoffier, 2012) was used to convert the final filtered vcf files to genepop files.

Population genomic analyses

The R (v 3.6.0; Team 2018) package 'assignPOP' (v.1.1.4; Chen et al., 2018) was used to identify the species of each larva by comparing the larval genotypes to reference adult genotypes for confirmed *C. artedi*, *C. hoyi*, and *C. kiyi* from the Apostle Islands used in Ackiss et al.

(2020). Before importing genotypes into R for assignment tests, missing data were imputed using the default settings in Beagle v5.1 (Browning et al., 2018). We used two resampling cross-validation methods to evaluate the assignment accuracy of the reference individuals. Monte-Carlo cross validation was performed using the assign.MC function using 50%, 70% and 90% of random individuals crossed by 10%, 25%, 50% and all loci with 50 iterations, with proportions of loci sampled randomly. We also performed k-fold cross-validation, which ensures all reference individuals are tested, with the assign.kfold function with k groups of 2, 3, 4, and 5 crossed by 10%, 25%, 50% and all loci. Both assignment accuracy tests used the support vector machine (svm) predictive model. Larvae were assigned to species with the assign.X function using reference genotypes from previously identified adults, the default criterion for retaining the number of principal components ("Kaiser-Guttman"), and the svm predictive model. Larvae assigned to species were then run in ADMIXTURE v1.3 (Alexander et al., 2009) with a K = 1-5 and a k-fold cross-validation (--cv=10) procedure to examine genetic clusters and to test for the presence of putative hybrids. If an individual had a q-score < 70% to the cluster to which it was assigned, it was considered a putative hybrid following Ackiss et al. (2020). 'Adegenet' (v.2.1.1; Jombart, 2008) was used for principal components analysis (PCA) to visualize allele frequency differences of assigned larvae.

The program NeEstimator (v.2.1; Do et al. 2014) was used to calculate effective population sizes (N_e) from larval samples using the LDNe method (Waples and Do, 2008) with random mating and a P_{crit} value of 0.05. To increase the accuracy of our estimates, loci were reduced to those that could be placed on the *C. artedi* linkage map (Blumstein et al. 2020), and linkage group information for each SNP was used to correct for physical linkage (Waples et al., 2016).

Estimates of genetic diversity, observed and expected heterozygosity (H_o and H_e, respectively), were calculated in GenoDive (v3; Meirmans, 2020).

Results

DNA extraction, barcoding, library prep, sequencing and genotyping

Of the 221 *Coregonus* larvae barcoded with COI, five were lake whitefish, 212 were non-lake whitefish coregonines, and four had a novel banding pattern (see Electronic Supplementary Material (ESM) Figures S1, S2). From the 212 non-lake whitefish coregonines, 202 individuals were chosen for RAD sequencing based on the time and length sub-setting described above. Sequencing resulted in more than 770 million reads with 3,812,906 average reads per individual. During processing, 19 of the 106 reference adults and 9 of the 202 larvae with sequence data were removed due to missing data above our filtering threshold. The majority of reference individuals dropped were *C. hoyi* and *C. kiyi*. The RAD preparations from these were from archived extractions of slightly older collections (2005), which likely resulted in more degraded DNA which does not perform well in RAD library preparation (Graham et al., 2015). The remaining 87 reference adults (*C. artedi*, n=55; *C. hoyi*, n=14; *C. kiyi*, n=18) and 193 larvae were genotyped at 32,021 SNPs for downstream analyses.

Population genomic analyses

Assignment accuracy was high, consistently averaging above 95% in *C. artedi* reference adults and 100% in *C. hoyi* and *C. kiyi* reference adults, even when the percentage of loci was reduced. Assignment accuracy plots are included in supplemental material (see ESM Figures S3, S4). Of the 193 genotyped larvae, 101 were assigned as *C. artedi* (average assignment probability = 97.6%), 57 were assigned as *C. kiyi* (average assignment probability = 95.5%), and

28 were assigned as *C. hoyi* (average assignment probability = 89.0%). Three was the most supported number of genetic clusters in ADMIXTURE (see ESM Figure S5). Seven larvae had q-scores < 70% and were identified as putative hybrids following Ackiss et al. (2020; Figs. 1, 2) - 5 larvae that had assigned as *C. artedi*, 1 as *C. hoyi,* and 1 as *C. kiyi*. The 7 putative hybrids appeared to be 2 *C. hoyi* x *C. kiyi* crosses and 5 *C. hoyi* x *C. artedi* crosses (Fig. 2).

Chromosome (linkage group) placement was available for 14,887 loci and was used to generate estimates of N_e without putative hybrids The estimated N_e from *C. artedi* and *C. hoyi* larvae was 16,847 (95% CI: 14,255-20,590) and 12,484 (95% CI: 8,153-26,634), respectively (Table 1). A value of 'infinite' was generated for *C. kiyi* larvae, which is generally an indication that N_e is large and the sample size was not adequate to produce an accurate estimate (Do et al., 2014; Waples and Do, 2010). Estimates of observed and expected heterozygosity (H_o, H_e) ranged from 0.211-0.238 and 0.225-0.258, respectively, with *C. hoyi* exhibiting the highest values of both.

Phenology and demographic measurements

A total of 605 coregonines were collected during the 11-week sampling period in 2018 (Fig. 3). The numbers collected varied weekly, with the most collected the week of June 11 (N = 189) and the fewest collected the week of July 23 (N = 7). All three species were captured across the range of sampling sites over the sampling period (Fig. 4), but they appeared to follow a successive pattern. *Coregonus artedi* appeared first, *C. kiyi* second, and *C. hoyi* third. *Coregonus artedi* appeared in the genotyped samples during the week of May 21 and declined after the week of June 18 (Figs. 3-5). *Coregonus kiyi* first appeared the week of May 28 and represented a higher proportion of identified individuals over the second half of the field sampling (Figs. 3-5).

Coregonus hoyi did not appear until mid-June and were most prevalent in mid-July (Figs. 3-5; Table 1).

Yolk sac classification provides an indication of hatch date (Oyadomari and Auer, 2007, 2008) and these data align with the species succession observed throughout the sampling period. *Coregonus artedi* larvae with yolk sac and oil globule were found during the weeks of May 21 and 28 and ranged in length from 9.2 mm to 12.9 mm (Fig. 5). Many *C. artedi* with just an oil globule were also found during the same two weeks and ranged in length between 10.4 and 13.6 mm. Only a single *C. artedi* individual was found with an oil globule on June 11; no larvae were genotyped for the week of June 4. For the remainder of the study period, only *C. artedi* with their yolk sac fully absorbed were identified, and they ranged in size from 12.6 to 19.6 mm (Fig. 5). Identified *C. hoyi* only had an oil globule (July 9 and 16, length range 8.9–10.3 mm) or a fully absorbed yolk sac (June 18 to July 23, length range 11.7–21.3 mm; Fig. 5). Similarly, identified *C. kiyi* had either an oil globule (May 28; length range 10.6–14.6 mm) or the yolk sac fully absorbed (May 28 to July 16; length range 12.3–21.7 mm; Fig. 5). Hybrids either had both yolk sac and oil globule (May 21 to May 28; lengths of 11.1 and 12.1 mm) or the yolk sac fully absorbed (July 9 to July 16; length range 16.4–18.8 mm).

Discussion

Our results offer excellent opportunities for investigating early life history dynamics of coregonines in the Laurentian Great Lakes basin. We demonstrated, for the first time, that an assemblage of shallow and multiple deepwater larval coregonine species can be identified to species using genomic data. The differentiation with extended sampling of the genome via RADseq allowed for the use of reference adult genotypes for larval assignment. The clustering

we observed in assigned larvae in a PCA was congruent with observed clustering in adults (Ackiss et al. 2020). Similar to adult collections, a handful of putative hybrids were also present in larval samples. The number of reference *C. artedi* used for larval assignment was considerably larger than the number of reference *C. hoyi*, which could impact the ability to categorize putative hybrids with intermediate allele frequencies. Additionally, putative hybrids, especially those grouped between deepwater *C. hoyi* and *C. kiyi* in the PCA, could represent rare deepwater species such as *C. zenithicus* or *C. nigripinnis*. Teasing apart whether putative hybrids represent true hybrids or rare deepwater species will be contingent on well-preserved tissue collections from rare species that can be used to establish reference allele frequency baselines. Future work using genetic assignment to identify coregonine larvae should aim for large numbers of reference individuals (n = 30-50) to improve baseline estimates of per population allele frequencies (Cornuet et al., 1999; Paetkau et al., 2004). Overall, the capacity to assign more than 96% of genotyped larvae to species with high confidence contributes to a growing body of genomics and transcriptomics research that suggests *C. artedi*, *C. hoyi* and *C. kiyi* are more genetically distinct than previously thought (Ackiss et al., 2020, Eaton et. al., 2021).

The ability to discern *C. artedi*, *C. hoyi* and *C. kiyi* provided an initial demographic and phenological baseline for larval coregonines in the Apostle Islands of Lake Superior, and a glimpse into the potential application for coregonine research in general. The successive appearance of each species in our samples suggested a progression of hatch times, starting with *C. artedi*, followed by *C. kiyi* and then *C. hoyi*. Our observed lengths of *C. artedi* and *C. hoyi* matched length-at-hatch from the laboratory (McCormick et al., 1971, Rice et al., 1987a), suggesting we began to capture each of these species soon after hatch. However, the general absence of yolk sacs for *C. hoyi* and *C. kiyi*, particularly in relation to their length ranges,

suggests several hypotheses: (1) *C. hoyi* and *C. kiyi* spawn in the Apostle Islands, where their embryos hatch, but the larvae remain below surface waters after hatch long enough to absorb much, if not all, of their yolk sac before moving to the surface; (2) *C. hoyi* and *C. kiyi* spawn, or at least the embryos hatch, outside of our sampling area and larvae are advected into the Apostle Islands after absorbing much, if not all, of their yolk sac; or (3) insufficient numbers of individuals were sequenced to identify small *C. hoyi* and *C. kiyi* with yolk sacs, due to our process of selecting individuals for genotyping under a limited budget. *Coregonus hoyi* historically spawned in February and March in the Apostle Islands (Dryer and Beil, 1968), larvae can remain in the hypolimnion for a period of 5-10 days after first feeding (Rice et al., 1987b), and can complete yolk absorption 8 days after hatch (Rice et al., 1987a,b), suggesting support for the first hypothesis, but not ruling out the second. Virtually nothing is known about *C. kiyi* spawning or early life history to provide any insight on either of the first two hypotheses for this species. The third hypothesis is viable as we were not able to genotype all individuals. In the future, age estimation (e.g., daily otolith rings) of identified larvae will better address such hypotheses about hatch dates as well as estimating mortality and growth rates (e.g., Oyadomari and Auer, 2008), all critical components for examining relationships between early life history and recruitment dynamics (Houde, 1987; Miller et al., 1988).

Length-frequency distributions showed larval coregonines generally grew as a cohort from the start of sampling until early July, with a majority of genotyped individuals assigned as *C. artedi* and *C. kiyi*. In July, however, we observed a broadening of the length distribution, including an increase in smaller individuals, indicating newly emerged cohort(s). Genotyping suggested the smallest individuals in July were *C. hoyi*, consistent with larval *C. hoyi* dynamics in Lake Michigan. Prolonged spawning of *C. hoyi* in Lake Michigan throughout the winter

months leads to multiple larval cohorts appearing between June and September with a broad length-frequency distribution (Rice et al., 1987b). Our results in Lake Superior suggest *C. hoyi* may also exhibit later and protracted spawning and larval hatching compared to *C. artedi* and *C. kiyi*. Consequently, we hypothesize that small coregonine larvae that appear in late summer and beyond in Lake Superior are likely to be *C. hoyi*.

RADseq data also provided baseline information on genomic diversity in larval *C. artedi, C. kiyi,* and *C. hoyi* in the Apostle Islands. Heterozygosity estimates in larvae were comparable to previous estimates from RADseq data in adults from the same region (Ackiss et al., 2020). Broad trends in estimated N_e for larvae in this region of Lake Superior were also similar to those in adult samples where *C. artedi* and *C. hoyi* exhibited comparable estimates to each other and lower estimates than *C. kiyi.* Here, comparable N_e estimates were found between larval *C. artedi* and *C. hoyi*, but a sample size of 57 larval *C. kiyi* was insufficient to generate an estimate of N_e. The inability to generate a finite estimate of N_e is suggestive of a large effective population size and is a common difficulty in marine populations with a large number of contributing breeders and high gene flow (Hare et al., 2011; Marandel et al., 2019). *Coregonus kiyi* is the most abundant *Coregonus* species in Lake Superior at depths >100 m with an estimated adult population size of 384 million in 2011 (Yule et al., 2013). Overall, N_e estimates were much larger for larvae than adults and, given the highly variable recruitment of coregonines (e.g., Axenrot and Degerman, 2015; Lepak et al., 2017; Stockwell et al., 2009), are likely a result of collecting larvae close to time of emergence before sweepstakes recruitment (Hedgecock, 1994) has largely impacted survival rates.

Moving forward, the genomics approach used here will be streamlined and standardized into a user-friendly and cost-effective tool. A genotyping-in-thousands by sequencing (GT-seq;

Campbell et al., 2015) panel of ~500 SNPs is in the final stage of optimization and will make efficient genetic identification of larval coregonines to species widely available, continuing the development of methods that are more reliable than visual identification (George et al., 2018). Such tools will open the door to novel insights on coregonine early life history at the species level, including hatch dates, spatial and temporal distributions, and growth and mortality rates in relation to environmental conditions and habitat. Our initial finding that larval community progression appears to be *C. artedi*, *C. kiyi*, and then *C. hoyi* in the Apostle Islands can be tested in other regions of Lake Superior, and could potentially be extended to inland lakes with diverse coregonine assemblages (e.g., Piette-Lauzière et al., 2019) to test more broadly for differences in early life history emergence patterns among coregonines. Ultimately, application of such tools will better elucidate how early life stages of coregonine species may play a role in their recruitment dynamics, and ultimately inform conservation and restoration efforts.

Acknowledgements

We thank Verena Lucke, Jay Glase, Lori Evrard, Kaitlyn Maines, and Peter Euclide for their assistance. Melissa Pespeni and Wendylee Stott provided valuable insights in the development of the ideas for this manuscript. We thank Wendylee Stott, Daniel Yule, and anonymous reviewers for comments on earlier versions of the manuscript. This work was funded by the USGS under Grant/Cooperative Agreement No. G17AC00042 to the University of Vermont. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The work was also funded by the following awards to HL: 2017-2018 Universities.com Women in STEM scholarship, the Heiser Educational Graduate Award, and the Norman S. Baldwin Fishery Science Scholarship.

Data Availability

RAD sequence data from reference adults were previously uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA):PRJNA600818. Larval RAD sequence data will be archived on the SRA through the publicly accessible Genomic Observatories Metadatabase (https://geome-db.org).

References

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. Evol. Appl. 13, 1037–1054.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C., Miller, M.R., 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics 202, 389-400.

Auer, N.A., 1982. Identification of larval fishes of the Great Lakes basin with emphasis on the Lake Michigan drainage. Great Lakes Fish. Comm., Special Publication 82-3, Ann Arbor, Michigan.

Axenrot, T., Degerman, E., 2015. Year-class strength, physical fitness and recruitment cycles in Vendace (*Coregonus albula*). Fish. Res. 173, 61–69.

Baldwin, N.A., Saalfeld, R.W., Dochoda, M.R., Buettner, H.J., Eshenroder, R.L., 2002. Commercial fish production in the Great Lakes 1867–2000. Great Lakes Fishery Commission, Ann Arbor, Michigan. Available: www.glfc.org. (September 2006).

Blumstein, D.M., Campbell, M.A., Hale, M.C., Sutherland, B.J., McKinney, G.J., Stott, W., Larson, W.A., 2020. Comparative genomic analyses and a novel linkage map for cisco (*Coregonus artedi*) provide insights into chromosomal evolution and rediploidization across salmonids. G3: Genes, Genomes, Genetics, 10, 2863-2878.

Bronte, C.R., Bunnell, D.B., David, S.R., Gordon, R., Gorsky, D., Millard, M.J., Read, J., Stein, R.A., Vaccaro, L., 2017. Report from the Workshop on Coregonine Restoration Science. U.S. Geological Survey. Available from: https://pubs.usgs.gov/of/2017/1081/ofr20171081.pdf.

Browning B.L., Zhou, Y., Browning, S.R., 2018. A one-penny imputed genome from next generation reference panels. Am. J. Hum. Genet. 103, 338-348

Campbell, N.R., Harmon, S.A., Narum, S.R., 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Res. 15, 855-867.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics 1, 171-182.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124-3140.

Chen, K.Y., Marschall, E.A., Sovic, M.G., Fries, A.C., Gibbs, H.L., Ludsin, S.A., 2018. assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. Meth. Ecol. Evol. 9, 439-446.

Claramunt, R.M., Smith, J., Donner, K., Povolo, A., Herbert, M.E., Galarowicz, T., Claramunt, T.L., Deboe, S., Stott, W., Jonas, J.L., 2019. Resurgence of cisco (*Coregonus artedi*) in Lake Michigan. J. Great Lakes Res. 45, 821–829.

Cornuet, J. M., Piry, S., Luikart, G., Estoup, A., & Solignac, M., 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics, 153, 1989-2000.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158.

Do, C., Waples, R.S., Peel, D., Macbeth, G., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Res. 14, 209-214.

Dryer, W.R., Beil, J., 1968. Growth changes of the bloater (*Coregonus hoyi*) of the Apostle Islands region of Lake Superior. Trans. Am. Fish. Soc. 97, 146-158.

Eaton, K.M., Bernal, M.A., Backenstose, N.J., Yule, D.L., Krabbenhoft, T.J., 2021. Nanopore amplicon sequencing reveals molecular convergence and local adaptation of rhodopsin in Great Lakes salmonids. Genome Biol. Evol. 13, evaa237.

Eshenroder, R.L, Vecsei, P., Gorman, O.T., Yule, D.L., Pratt, T.C., Mandrak, N.E., Bunnell, D.B., Muir, A.M., 2016. Ciscoes (*Coregonus*, subgenus *Leucichthys*) of the Laurentian Great Lakes and Lake Nipigon [online]. Available from: www.glfc.org/pubs/misc/Ciscoes_of_the_Laurentian_Great_Lakes_and_Lake_Nipigon.pdf [accessed 11 January 2017].

George, E.M., Hare, M.P., Crabtree, D.L., Lantry, B.F., Rudstam, L.G., 2018. Comparison of genetic and visual identification of cisco and lake whitefish larvae from Chaumont Bay, Lake Ontario. Can. J. Fish. Aquatic Sci. 75, 1329-1336.

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J.A., Pierson, T., Rogers, S.M., Wilson, J.Y., 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol. Ecol. Res. 15, 1304-1315.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., Ruegg, K., Palstra, F., 2011. Understanding and estimating effective population size for practical application in marine species management. Conserv. Biol. 25, 438-449. https://doi.org/10.1111/j.1523-1739.2010.01637.x

Hedgecock, D., 1994. Does variance in reproductive success limit effective population size of marine organisms? Pp. 122–134 in A. Beaumont, ed. Genetics and evolution of aquatic organisms. Chapman and Hall, London.

Hinrichs, M.A., 1979. A description and key of the eggs and larvae of five species of fish in the subfamily Coregoninae. Masters Thesis, University of Wisconsin-Stevens Point..
Houde, E.D., 1987. Fish early life dynamics and recruitment variability. Am. Fish. Soc. Symp. 2, 17–29.
Jombart, T., 2008. "adegenet: a R package for the multivariate analysis of genetic markers." Bioinformatics 24, 1403-1405. doi: 10.1093/bioinformatics/btn129.
Klinard, N.V., Matley, J.K., Halfyard, E.A., Connerton, M., Johnson, T.B., Fisk, A.T., 2020. Post-stocking movement and survival of hatchery-reared bloater (*Coreognus hoyi*) reintroduced to Lake Ontario. Freshw. Biol. 65, 1073-1085.
Koelz, W.N., 1929. Coregonid fishes of the Great Lakes. Bulletin of the United States Bureau of Fisheries 43, 297-643.
Lepak, T.A., Ogle, D.H., Vinson, M.R., 2017. Age, year-class strength variability, and partial age validation of kiyis from Lake Superior. N. Am. J. Fish. Manage. 37, 1151–1160.
Lischer, H.E.L, Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 22, 298-299.
Lucke, V.S., Stewart, T.R., Vinson, M.R., Glaze, J.D., Stockwell, J.D., 2020. Larval *Coregonus* spp. diets and zooplankton community patterns in the Apostle Islands, Lake Superior. J. Great Lakes Res. 46, 1391-1401.
Marandel, F., Lorance, P., Berthelé, O., Trenkel, V.M., Waples, R.S., Lamy, J.-B., 2019. Estimating effective population size of large marine populations, is it feasible? Fish. Fish., 20, 189– 198. https://doi.org/10.1111/faf.12338
McCormick, J.H., B.R. Jones, and R.F. Syrett., 1971. Temperature requirements for growth and survival of larval ciscos (*Coregonus artedii*). J. Fish. Board Can. 28, 924-927.
McKenna, Jr., J.E., Stott, W., Chalupnicki, M., Johnson, J.H., 2020. Spatial segregation of cisco (*Coregonus artedi*) and lake whitefish (*C. clupeaformis*) larvae in Chaumont Bay, Lake Ontario. J. Great Lakes Res. 46, 1485-1490.
McKinney, G.J., Waples, R.K., Seeb, L.W., Seeb, J.E., 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by- sequencing data from natural populations. Mol. Ecol. Res. 17, 656-669.
Meirmans, P.G., 2020. GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. Mol. Ecol. Res. 20, 1126-1131.
Miller, T.J., Crowder, L.B., Rice, J.A., Marschall, E.A., 1988. Larval size and recruitment mechanisms in fishes: toward a conceptual framework. Can. J. Fish. Aquat. Sci. 45, 1657–1670.
Myers, J.T., Stockwell, J.D., Jones, M.L., Yule, D.L., Black, J.E., 2008. Evaluating sampling strategies for larval cisco (*Coregonus artedi*). J. Great Lakes Res. 34, 245–252.
OMNRF (Ontario Ministry of Natural Resources & Forestry), 2015. *Stocking strategy for the Canadian waters of Lake Ontario*. Lake Ontario Management Unit, Ontario Ministry of Natural Resources & Forestry, Picton, ON, Canada.
Oyadomari, J.K., Auer, N.A., 2007. Influence of rearing temperature and feeding regime on otolith increment deposition in larval ciscoes. Trans. Am. Fish. Soc. 136, 766-777.
Oyadomari, J.K., Auer, N.A., 2008. Transport and growth of larval cisco (*Coregonus artedi*) in the Keweenaw Current region of Lake Superior. Can. J. Fish. Aquatic Sci. 65,1447-1458.

Paetkau, D., Slade, R., Burden, M. and Estoup, A., 2004, Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol. Ecol. 13, 55-65. https://doi.org/10.1046/j.1365-294X.2004.02008.x

Piette-Lauzière, G., Bell, A.H., Ridgway, M.S., Turgeon, J., 2019. Evolution and diversity of two cisco forms in an outlet of glacial Lake Algonquin. Ecol. Evol, 9, 9654–9670.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rice, J.A., Crowder, L.B., Binkowski, F.P., 1987a. Evaluating potential sources of mortality for larval bloater (*Coregonus hoyi*): starvation and vulnerability to predation. Can. J. Fish. Aquatic Sci. 44, 467-472.

Rice, J.A., Crowder, L.B., Holey, M., 1987. Exploration of mechanisms regulating larval survival in Lake Michigan bloater: a recruitment analysis based on characteristics of individual larvae. Trans. Am. Fish. Soc. 116, 703-718.

Smith, S.H., 1964. Status of the deepwater cisco population of Lake Michigan. Trans. Am. Fish. Soc. 93, 155-163.

Stockwell, J.D., Ebener, M.P., Black, J.A., Gorman, O.T., Hrabik, T.R., Kinnunen, R.E., Mattes, W.P., Oyadomari, J.K., Schram, S.T., Schreiner, D.R., Seider, M.J., Sitar, S.P., Yule, D.L., 2009. A synthesis of cisco recovery in Lake Superior: implications for native fish rehabilitation in the Laurentian Great Lakes. N. Am. J. Fish. Manage. 29, 626–652.

Stott, W., MacDougall, T., Roseman, E.F., Lenart, S., Chiotti, J., Yule, D.L., Boase, J., 2021. Genetic species identification of coregonines from juvenile assessment and commercial fisheries in Lake Erie, Detroit River and St. Clair River. Advanc. Limnol. 66, 000-000.

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. Conserv. Genet. 4, 67-81.

Waples, R.S., Do, C.H.I., 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Mol. Ecol. Res. 8, 753–756.

Waples, R.S., Do, C.H.I., 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol. Appl. 3, 244-262.

Waples, R.K., W.A. Larson, and R.S. Waples. 2016. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity 117, 233–240.

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. Fish. Res. 140, 123-132.

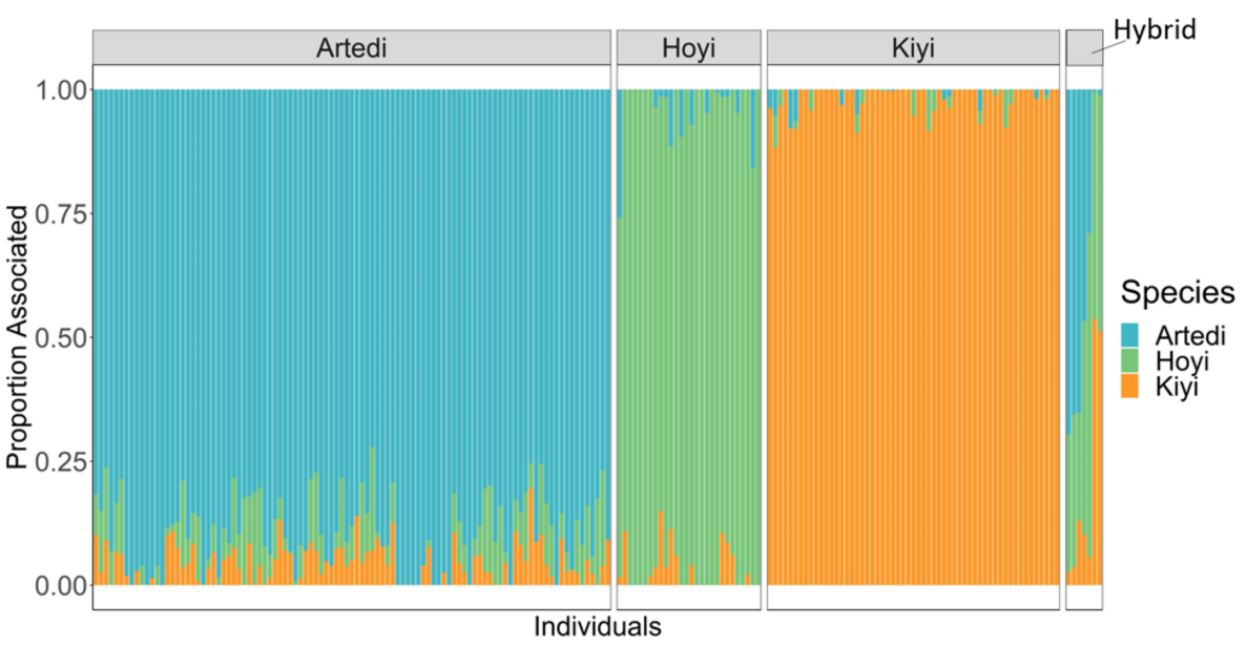
Zimmerman, M.S. and Krueger, C.C., 2009. An ecosystem perspective on re-establishing native deepwater fishes in the Laurentian Great Lakes. N. Am. J. Fish. Manage. 29, 1352-1371.

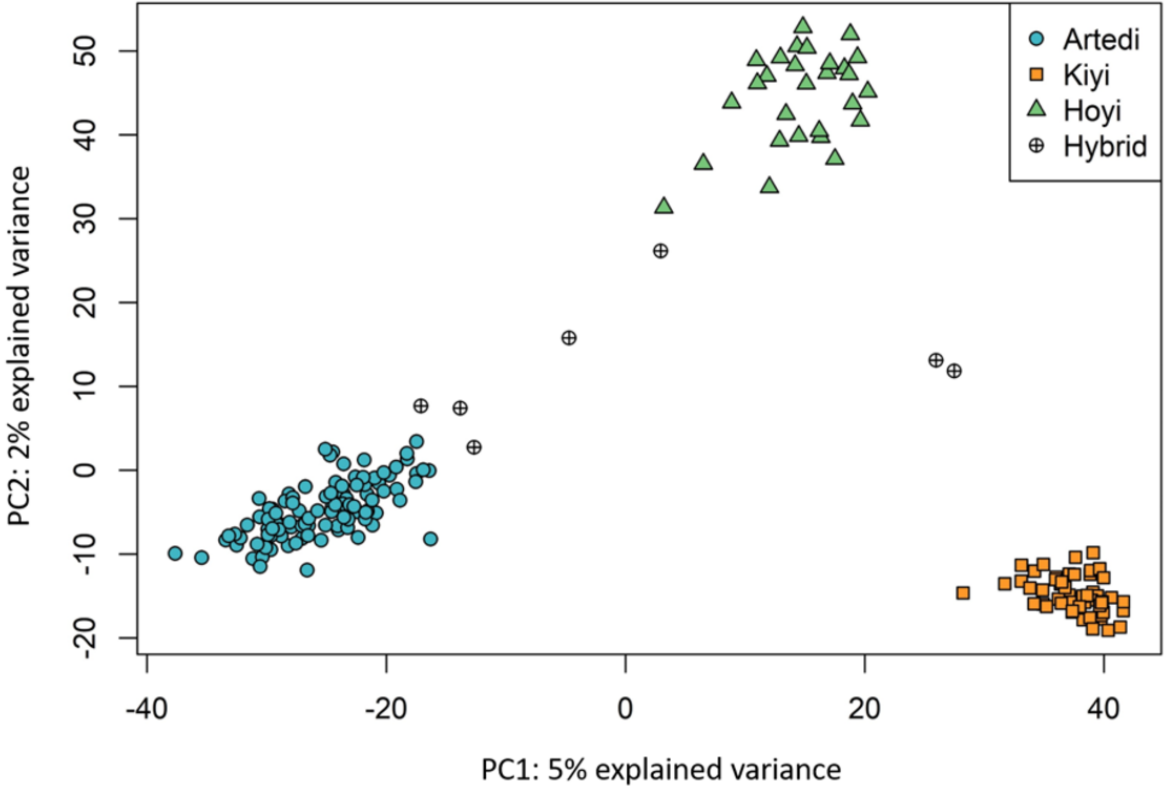
Table 1. Species-specific data for 193 larvae that were sequenced from the spring and summer 2018 sampling in the Apostle Islands, Lake Superior. The number of larvae assigned to each species (N), length range, weeks (start date) of collection, effective population size (N_e), observed heterozygosity (H_o), and expected heterozygosity (H_e) for each species is reported. An 'Inf' value indicates the sample size was not adequate to generate a N_e estimate. "NA" indicates not applicable.

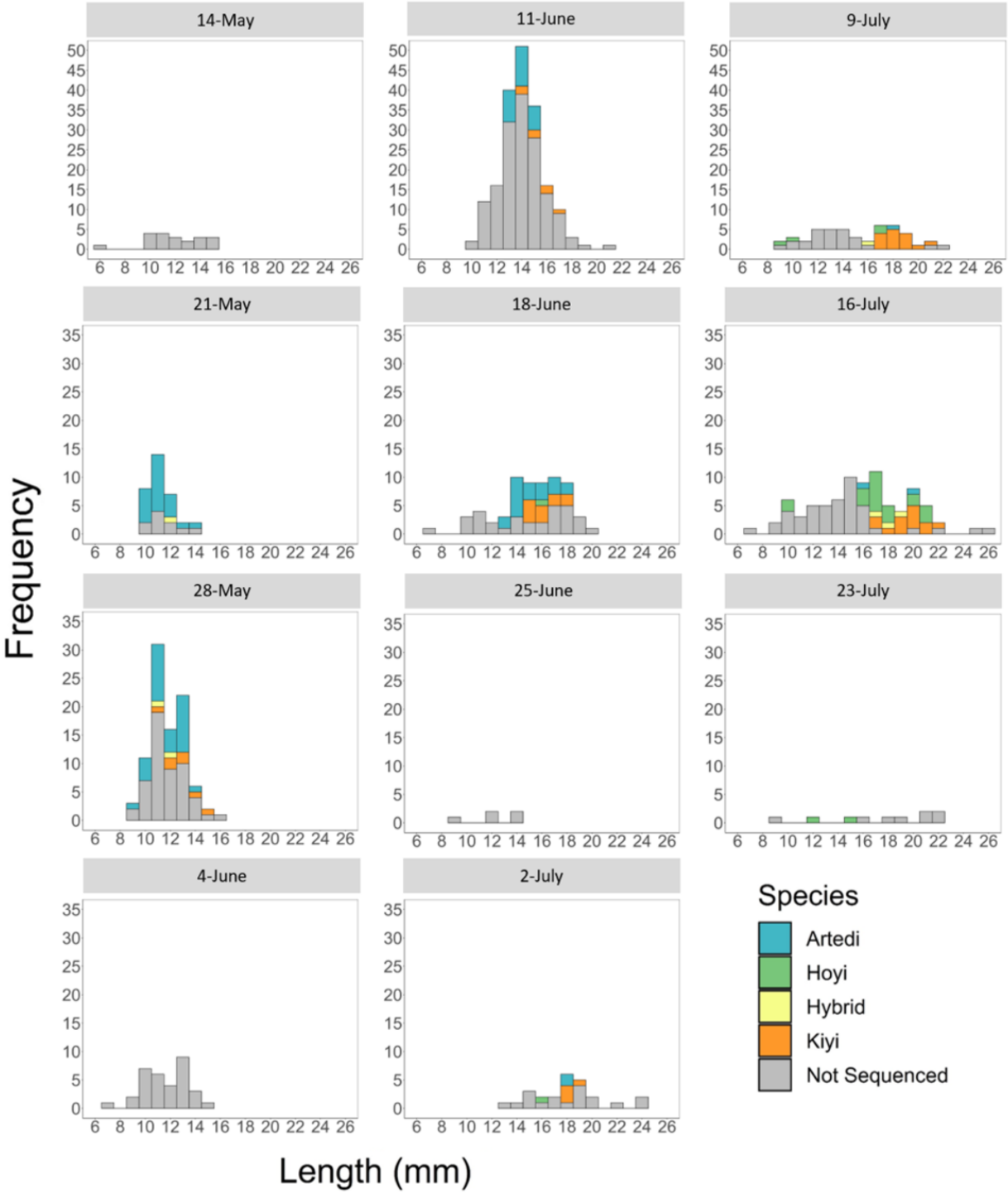
Species	**N**	**Length range (mm)**	**Weeks of collection**	N_e **(95% CI)**	H_o	H_e
C. artedi	101	9.2 - 19.6	May 21-Jul 16	16,847 (14,255-20,590)	0.215	0.229
C. kiyi	57	10.6 - 21.7	May 28-Jul 16	Inf (Inf – Inf)	0.211	0.225
C. hoyi	28	8.9 - 21.3	Jun 18-Jul 23	12,484 (8,153-26,634)	0.238	0.258
Putative hybrid	7	11.1 - 18.8	May 21-Jul 16	NA	NA	NA

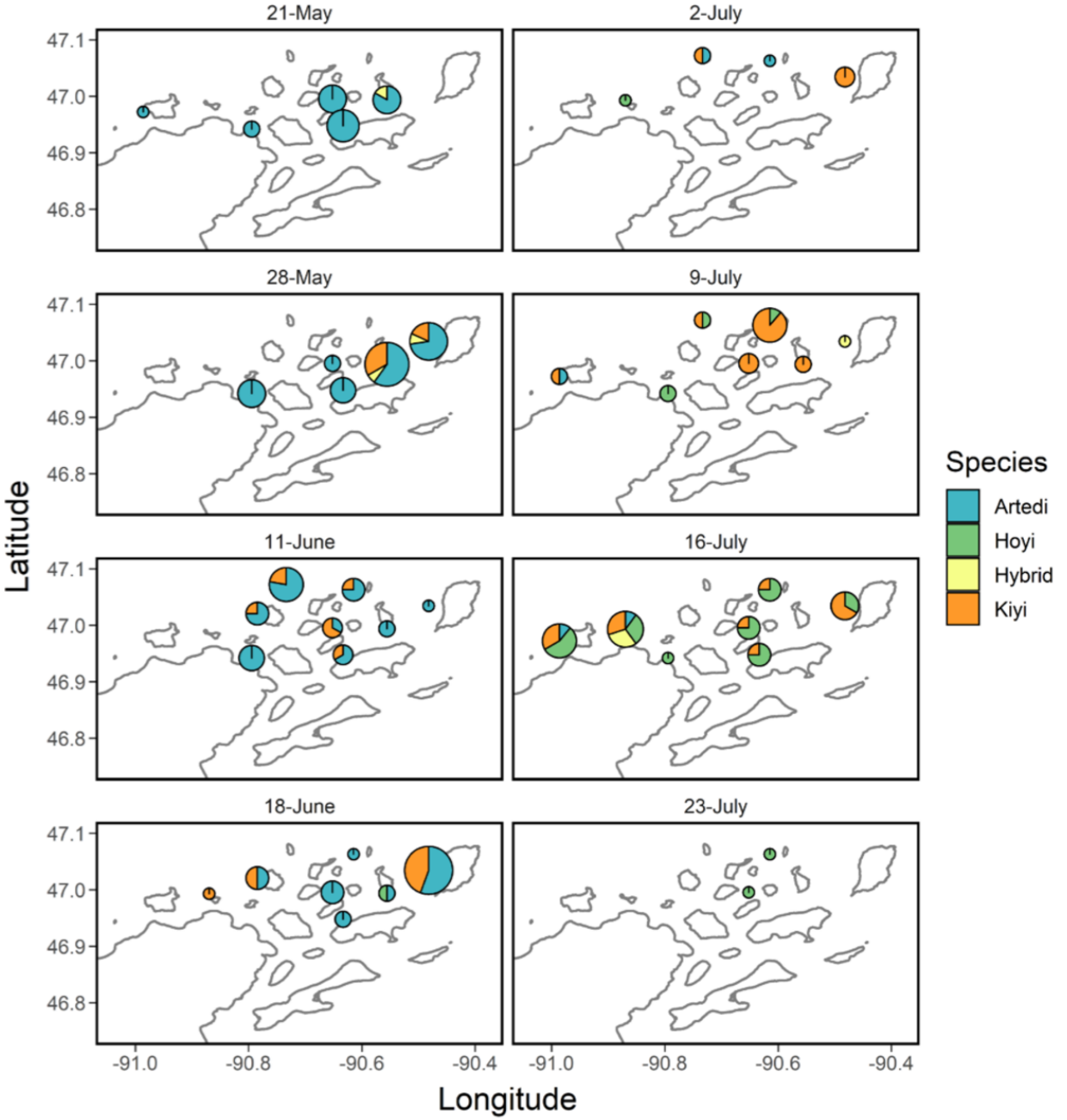
List of Figures

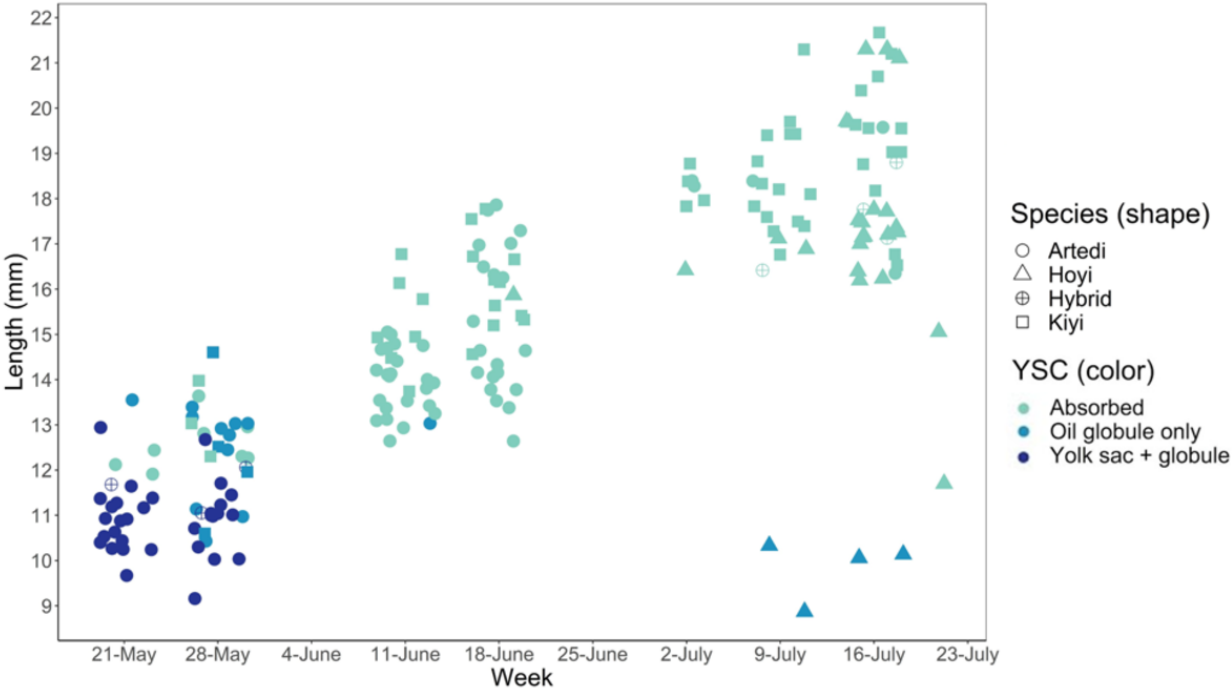
Fig. 5. Yolk sac condition (YSC) of 193 genotyped coregonine larvae collected in the Apostle Islands, Lake Superior, between the weeks of May 21 and July 23, 2018. Symbols represent species and colors represent YSC. Symbols are horizontally displaced within weeks for easier viewing.


Artedi
Hoyi
Kiyi
Hybrid
1.00
0.75
0.50
0.25
0.00
Proportion Associated
Individuals
Species
Artedi
Hoyi
Kiyi


Artedi
Kiyi
Hoyi
Hybrid
PC2: 2% explained variance
PC1: 5% explained variance
-20
-10
0
10
20
30
40
50
-40
-20
0
20
40


14-May
11-June
9-July
21-May
18-June
16-July
28-May
25-June
23-July
4-June
2-July
Frequency
Length (mm)
Species
Artedi
Hoyi
Hybrid
Kiyi
Not Sequenced


21-May
2-July
28-May
9-July
11-June
16-July
18-June
23-July
47.1
47.0
46.9
46.8
-91.0
-90.8
-90.6
-90.4
Latitude
Longitude
Species
Artedi
Hoyi
Hybrid
Kiyi


Length (mm)
22
21
20
19
18
17
16
15
14
13
12
11
10
9
21-May
28-May
4-June
11-June
18-June
25-June
2-July
9-July
16-July
23-July
Week
Species (shape)
Artedi
Hoyi
Hybrid
Kiyi
YSC (color)
Absorbed
Oil globule only
Yolk sac + globule