Life history implications of kinship structure in an Atlantic herring schooling aggregation

James Kho[1], Gregory McCracken[1], Jenni McDermid[2], and Daniel Ruzzante[1]

[1]Dalhousie University
[2]Fisheries and Oceans Canada Gulf Region

February 22, 2024

Abstract

Kinship in natural marine fish systems has been little explored in part due to limitations of available genomic approaches. This is especially true for very abundant and commercially exploited marine pelagic fish with broadcast spawning. Recent advances in genomics have however, facilitated an improved understanding of population structure in marine systems at fine geographic scales. Here, we identify kinship structure in an aggregation of Atlantic herring (Clupea harengus) juveniles. We identified 11 half-sib pairs using a suite of 92 sequenced microsatellite DNA markers in a sample of N=1391 herring individuals comprising two cohorts (ages 0+ and 1+) collected from the 2018 Fall research survey of the Southern Gulf of St. Lawrence. In addition, we were also able to detect a full-sib pair using the combination of half-sib pairs and whole mitochondrial DNA (mtDNA) analysis. This study suggests the presence of kinship structure in a juvenile herring aggregation three months or more post hatching with implication for our understanding of herring early life history.

[1] Department of Biology, Dalhousie University, Halifax, NS, Canada

[2] Fisheries and Oceans Canada, Moncton, NB, Canada

ORCID IDs:

JK: https://orcid.org/0000-0002-6993-355X

GRM: https://orcid.org/0000-0002-1701-1529

JLM: https://orcid.org/0000-0002-2842-3502

DER: https://orcid.org/0000-0002-8536-8335

Corresponding author e-mail:james.kho@dal.ca

Co-author emails: gregory.mccracken@dal.ca, jenni.mcdermid@dfo-mpo.gc.ca, daniel.ruzzante@dal.ca

Abstract:

Kinship in natural marine fish systems has been little explored in part due to limitations of available genomic approaches. This is especially true for very abundant and commercially exploited marine pelagic fish with broadcast spawning. Recent advances in genomics have however, facilitated an improved understanding of population structure in marine systems at fine geographic scales. Here, we identify kinship structure in an aggregation of Atlantic herring (*Clupea harengus*) juveniles. We identified 11 half-sib pairs using a suite of 92 sequenced microsatellite DNA markers in a sample of N=1391 herring individuals comprising two

cohorts (ages 0+ and 1+) collected from the 2018 Fall research survey of the Southern Gulf of St. Lawrence. In addition, we were also able to detect a full-sib pair using the combination of half-sib pairs and whole mitochondrial DNA (mtDNA) analysis. This study suggests the presence of kinship structure in a juvenile herring aggregation three months or more post hatching with implication for our understanding of herring early life history.

Key words: Genetics, Kinship, Atlantic Herring, *Clupea harengus* , microsatellite, mtDNA

Introduction

Insight into kin structure in the wild is essential towards a robust understanding of demography and population dynamics (Bravington, Grewe, and Davies, 2016; Waples and Anderson, 2017; Ruzzante et al., 2019). Failure to consider the presence of closely related individuals in studies on population structure and demography can under some circumstances lead to violation of assumptions and, occasionally to downwardly biased estimates of population parameters (Waples and Anderson, 2017). Identification of kinship structure in wild populations should therefore ideally be considered when conducting population level analyses. Despite their importance, kinship is not usually included as part of the analysis but is assumed not to influence inferences (Kamel and Grosberg, 2013). Kin structure in marine systems in particular is often overlooked due to the assumption that gamete and larval dispersal coupled with large open populations would overwrite any form of kinship structure (Christie et al., 2010; Kamel and Grosberg, 2013). This assumption and the lack of genetic tools have thus far prevented further investigation into this area of study (Christie et al., 2010; Schunter et al., 2014). Here we focused on Atlantic herring (*Clupea harengus*) to better understand the implications of kinship in a highly abundant and widely dispersed pelagic marine system.

Herring is considered a model species for studies examining adaptation and natural selection in pelagic marine fish systems due to their large population size and low genetic differentiation at neutral loci (Lamichhaney et al., 2012; Pettersson et al., 2019). Their predictable spawning behavior makes them a good representative species for studies on kinship structure. The species is widely distributed on both sides of the North Atlantic (Barrio et al., 2016; Kerr et al., 2019) and plays important ecological and economic roles as food for larger fish, marine mammals, and seabirds, and as subject of profitable fisheries in both Europe and North America, respectively (Barrio et al., 2016; Benoît and Rail, 2016; McDermid et al., 2018; Wilson et al., 2018). Herring migrate extensively during their lifetime but are known to return to their natal areas for spawning (McQuinn, 1997; Ruzzante et al., 2006; Melvin et al., 2009). They are iteroparous broadcast spawners and generally spawn once a year in schooling aggregations with multiple spawning waves during a spawning season (Sinclair and Tremblay, 1984; McPherson et al., 2003; Melvin et al., 2009; Wilson et al., 2018). In some regions, two seasonal spawning components are known to coexist: Spring spawning and Fall spawning herring. Herring lay their eggs on the bottom of the sea floor (Hare and Richardson, 2014; McDermid et al., 2018). Upon hatching, their larvae remain grouped while staying close to the spawning grounds as they migrate vertically (Sinclair and Iles, 1989).

Recent studies revealed significant genetic differentiation among spawning seasons and locations thought to correspond to ocean temperature (Ruzzante et al. 2006; Lamichhaneys et al., 2017; Kerr et al., 2019; Han et al., 2020; Fuentes-Pardo, 2020). Genetic studies on early life stages have shaped much of our knowledge on the mechanisms of the population structure in this system but they have also allowed us to incorporate aspects of their life history (Stephenson and Power, 1988; McPherson et al., 2004; Bang et al., 2006; Ruzzante et al., 2006; Sinclair and Power, 2015; Berg et al., 2019). For instance, larval retention on spawning grounds plays an important part in keeping the integrity of the herring local metapopulation system (Sinclair and Iles, 1989; McQuinn, 1997). Despite extensive genetic studies on this system, few have looked at the mitochondrial DNA (mtDNA) of herring.

Here we used mtDNA for parentage analysis as mtDNA inheritance is strictly maternal in most vertebrates (Brown, 2008), making it valuable in kinship analysis. Studies on the mtDNA of herring have primarily

focused on using them as markers for population structure (Teacher et al., 2012), but little is known about the ancestral inheritance and mtDNA mutation rate. Estimating the mtDNA mutation rates is difficult without interspecific comparisons, but the assessment of mtDNA sequence differences between closely related individuals have been employed as a solution (Haag-Liautard et al., 2008). The incorporation of mtDNA data with kinship analysis using a suite of nuclear markers improves the confidence towards identifying sibship structure.

Our goal is to examine kinship structure among schooling Atlantic herring. We sequenced 92 microsatellite DNA markers on N= 1391 Atlantic herring juveniles collected from a single schooling aggregation in the Gulf of St. Lawrence. While no kinship structure was expected given the presumed abundance of Atlantic herring in the Gulf of St. Lawrence, we identified 11 half-sib pairs among 980 age 0+ herring juveniles. We then sequenced the whole mtDNA genome of all individuals involved in the 11 presumed half-sib pairs. This analysis revealed a candidate full-sib pair with mutation sites on a potentially adaptive region of the mtDNA.

Materials and Methods

Atlantic herring sampling and age determination

We extracted DNA from a total of n=2045 Atlantic herring collected over a period of 2 days in September 2018 in Cumberland Strait in the Southern Gulf of St. Lawrence (Figure 1) as part of the 2018 Fisheries and Oceans Canada research vessel trawl survey. Tissue samples were stored in 95% EtOH.

Fork length and otolith rings were used to assign ages with individuals $<$ 15 cm long assigned to age class 0+ and those between 17-21 cm assigned to age class 1+ (Figure S1). Otolith growth rings were then used to cross reference these assignments and identify the spawning season (Jones, 1986). Out of the 2045 herring, 1391 fish were 0+ Spring spawned and 654 fish were 1+ Fall spawned.

Molecular methodology and microsatellite markers development

DNA was extracted with the Glassmilk protocol modified from Elphinstone, Hinten, Anderson, and Nock (2003). Randomly chosen samples from each plate were then run through a 1.0% agarose gel to assess DNA quality. Out of the 2045 herrings extracted, 1656 individuals were retained post quality control. Extracted DNA was stored at -20°C until sequencing.

Microsatellite markers were designed from whole genome sequence data of a Gulf of St. Lawrence herring individual using MSATCOMMANDER v1.0 (Faircloth, 2008). A total of 192 microsatellites were designed and tested on 125 herring collected from throughout the Gulf of St. Lawrence (sequenced on Illumina MiSeQ). MEGASAT (Zhan et al., 2017) was used to demultiplex and score alleles (minimum depth per sample per locus at 50 reads). When necessary, allele scores were adjusted manually within MEGASAT. A total of 118 loci were retained after MEGASAT. Microchecker 2.2.3 (Van Oosterhout et al., 2004), GenAlEx 6.5 (Peakall and Smouse, 2006), and Arlequin v. 3.5.2.2 (Excoffier et al., 2005) were then used to assess the heterozygosity and presence of missing data and/or null alleles. Details regarding reagents and the quality control of microsatellites can be found in McCracken et al. (2014a, b). A total of 94 loci distributed across all 26 chromosomes were retained for use in downstream analyses (Table S1).

All 1656 herring individuals were sequenced on the Illumina HiSeq 2500 – High Throughput for these 94 microsatellite loci. Subsequently, 2 more loci were removed due to missing data bringing down the final number of microsatellite DNA markers to 92. After further quality control using FastQC (Babraham Bioinformatics), GenAlEx, and MEGASAT to address missing data and low read depth, we retained 1391 herring individuals for kinship analysis.

Identifying kinship structure from herring samples using *CKMRsim*

The R (Rstudio Desktop v.1.3.1093 with R v.4.0.1) package *CKMRsim*(Anderson, 2016) was used to detect kinship structure within and across the cohorts. *CKMRsim* employs allele frequencies to estimate the power for pairwise relationship inference. It does this by evaluating the false positive (FPRs) and false negative rates (FNRs), and outputs log-likelihood ratios for inferring the relationship of interest. The log-likelihood ratio can be visually represented by a density plot showing the statistical power needed to confidently identify a particular sibship pair. The output summary table contains the FNRs, FPRs, standard error, number of nonzero weights (Num-nonzero wts), andλ^*. λ^*reflects the log-likelihood ratios to be used for determining the number of Half-Sib or Full-Sib Pairs (HSPs or FSPs). Choosing the most appropriate λ^* involves setting a threshold to balance the FNR and FPR values thereby minimizing the error margin of the kinship estimates. This is done by calculating a reference false positive rate (Anderson, 2016):

Reference FPR = $(\frac{1}{10\ (N\ *(N-1))})$ (1)

where N is the number of individuals examined and [N * (N-1)] is the number of comparisons. This reference FPR is used as a threshold to confidently distinguish related pairs from unrelated individuals.

Mitochondrial sequencing and phylogenetic analysis

We sequenced the mtDNA from 19 individuals: 18 individuals identified as members of a half-sib pair and an extra individual chosen as an outgroup using Mitochondrial DNA Isolation Kit (ab65321, Abcam). These samples were then sent for whole mtDNA sequencing to The Centre of Applied Genomics DNA Sequencing Facility (Hospital for Sick Kids, University of Toronto) on Illumina HiSeq 2500. The discovery of several single base differences between members of a presumed full-sib pair (where no differences were expected, see below) led to a resequencing effort to discount sequencing error. MultiQC and Trimmomatic v.0.39 (Bolger et al., 2014) were used, respectively for quality control and trimming of Illumina adapters. All individuals were retained post quality control. Data were aligned to the complete mitochondrial genome of *Clupea harengus* (NC_009577) reference sequence using Clustal Omega algorithm in Geneious Prime 2021.2 (Biomatters Ltd.). Neighbour joining phylogenetic trees were created with the aligned sequences using Geneious Tree Builder tool with the global alignment with free end gaps setting. To address potential sequencing errors, we used the function binom.test from the R package *stats* v.4.0.2 (R Core Team, 2020) to model the expected number of base pairs due to sequencing error based on the observed base pair (bp) differences between two individuals. The resulting probability is then evaluated using a binomial distribution based on the reference sequencing error for the Illumina HiSeq 2500 of 0.112% ± 0.544% (Stoler and Nekrutenko, 2021).

Results

Estimating kinship pairs using *CKMRsim*

Of the *N*= 1391 Atlantic herring retained, *N*= 980 were Age 0+ and *N*= 411 were Age 1+. Age 0+ herring are spring spawned and Age 1+ herring are fall spawned (Table 1). No reliable HSP and FSP relationships were detected between individuals in the age 1+ cohort nor across cohorts between any of the age 0+ and age 1+ individuals (see ESM). Given these results, we focused on the age 0+ cohort for downstream analyses.

Table 2 shows the FPRs and FNRs for statistically distinguishing half-sib pairs from unrelated individuals within the age 0+ cohort. The reference FPR for this cohort is FPR=1.042e-07. This value corresponds to λ^* [?] 12.6 (Table 2). We considered λ^*= 12 as a threshold to estimate the number of half-sib pairs in the Age 0+ cohort to reduce the effect of FNRs (Figure 2). There are 11 half-sib pairs in this cohort with a logl_ratio $>$ 12 that are formed by 18 unique individuals (Table 3). Out of these 18 individuals, two individuals, Ch250, and Ch290 are half-sibs of two individuals each, and a third individual, Ch258 is a half-sib of three individuals, Ch290, Ch226, and Ch534 (Table 3), which in turn, are not identified as half-sibs to each other.

An individual cannot be related as a half-sib with three individuals who are not related to each other; at least one of these presumed half-sib pairs must be a full-sib pair.

The λ^* value for FSP was estimated using the same reference FPR value of 1.042e-07 but based on a different FNR and FPR table that takes into account the ability to distinguish full-sib instead of half-sib pairs (Table 4). The FPR values generated using the age 0+ cohort did not pass the reference FPR values. Upon extending the FNR query to cover a wider range, a $\lambda^*$of 12.6 was found to correspond to the reference FPR value, but this value lies within the range of the HSP's λ^*value (Figure 2). Consequently, while seven candidate FSPs are identified with a logl_ratio [?] 12.6, statistical power is insufficient for distinguishing true FSPs from HSPs; this is evidenced by the presence of true HSPs being identified as candidate FSPs in Table 5 and the overlap between full-sibs and half-sibs logl_ratio curves shown in Figure 2. Nevertheless, the 6^{th} pair in the list of candidate FSPs are individuals Ch226 and Ch290. Therefore, while statistical power is insufficient to distinguish HSP from FSP, the only way that the pattern described above for Ch258 can be true is if two of the other individuals are related as full-sibs. Individual Ch258 is half-sib with Ch226, Ch290, and Ch534 while Ch226 and Ch290 are full-sib to each other. We next asked the question whether the various HSP identified above were paternally or maternally related by sequencing their mitogenome.

Phylogenetic analysis of the sibship pairs

As expected, the individual chosen as an outgroup did not pair with any of the 11 HSPs (Figure 3).

The two individuals that make up the candidate FSP (Ch226 and Ch290) instead, were expected to exhibit the same mtDNA haplotype and thus, to cluster closest to each other. Indeed, they are the only pair identified above as kin that are linked at a terminal branch (Figure 3, Box). Interestingly, despite being full-sibs the per base analysis of their mtDNA revealed 33 bp differences between the two individuals. To examine whether these base pair differences are real or due to sequencing artefacts, we first generated a binomial distribution (Figure S4) to model the expected number of per base variation in a 16700 bp long sequence due to sequencing error using the reference error rate for Illumina HiSeq 2500 of 0.1% +- 0.544% (Stoler and Nekrutenko, 2021). The mean and median of the distribution are both 17 bp (Figure S4). We then used a binomial distribution (R package *stats*) to estimate the probability of observing 33 bp variation based on the reference error percentage and sequence length. The resulting probability of observing 33 bp differences due to sequencing error is 0.2% ($P<.0001$). This error rate falls within the range of one standard deviation to the reference sequencing error rate which suggests that the 33 bp differences observed in the FSP are potential sequencing artefacts. Despite this evidence, some of the base pair differences were also found in other individuals, suggesting that they may be true mutations. To examine this possibility, we re-sequenced the two individuals in the FSP on the MiSeq. At least 15 of the 33 bp differences were detected in the same position and state, and an extra bp difference was found in the same position but with a different base conversion (Table S2). This suggests that at least 15-16 bp are likely single nucleotide polymorphisms (SNPs) that took place within a single generation, but this result warrants further investigation. We were unable to examine the remaining 16 bp differences due to poor coverage in the second resequencing effort.

Out of the 33 base pair differences detected in the FSP, 12 were detected in other individuals not involved in the FSP. Five of the 12 variations were detected in one individual, four were detected in two individuals, one was detected in three individuals, and two variations were detected in five individuals not involved in the FSP suggesting some of them may be true mutations (Table S2). Most of the putative mutation sites were in the D-Loop at 5 bp, followed by *ND2* gene at 2 bp and the remaining 5 bps are distributed across the mitogenome including within genes such as *CytB, ND5,* and *COX1* at 1bp per gene. Interestingly, all but five of the sites are silent mutations with D-Loop having three missense and one nonsense mutations, and *COX1* gene having one missense mutation. Eight out of the 15 putative SNPs that reappeared in the resequencing data were silent mutations and six were found in other individuals. Out of the seven non-silent mutations, only one was found in other individuals (Table S2). The full-sib pair (Ch226 and Ch290) exhibits the lowest number of mutations compared to the HSPs, the mutation sites between which range from 62 bp to 131 bp (Table 5). For the 11 HSPs, the relative number of substitutions per region were also assessed in Table 6.

In general, the D-Loop and *ND5* regions have the highest number of mutations while *COX1* has the lowest, and this is consistent across the 11 pairs. The largest gap in number of mutations within pairs is also found between the presumed FSP and the HSP exhibiting the lowest number of mutations (33 vs. 62, i.e., gap = 29 bp; Table 5 and Figure S3). Mutations are not localized but are spread out across 3-4 regions in the mtDNA. With the exception of the FSP (Ch226 and Ch290), no other HSP was paired at a terminal branch suggesting the pairs are all likely related through their male parents. The average number of nucleotide substitutions per site suggest minor genetic changes between individuals within most of the branches (>100 bp differences between tips). Several pairs have strong bootstrap support value such as Ch891 and Ch588.

Additional neighbour joining trees based exclusively on the D-loop,*CytB* , *COX1* and *ND5* regions of the mtDNA were generated to provide a comparison across regions of varying mutation rates. The *ND5* tree was the most similar to the whole mtDNA tree and suggests that the majority of the relationship structure in Figure 3 is due to this region in the mtDNA. This finding is in concordance with the observation of mutations sites in the Ch226 and Ch290 pair being found mostly in this mtDNA region. Overall, all four trees showed less information regarding relatedness compared to Figure 3 (Figure S2). The D-Loop tree, for example, suggests many individuals share a common ancestor including several half-sib pairs such as Ch232-Ch250 and Ch549-Ch933; however, we know from the point mutation analysis that these variable sites are distributed across the mtDNA and therefore the whole mtDNA is a more suitable representation.

Discussion

We detected 11 half-sib pairs from 980 spring spawned age 0+ Atlantic herring individuals using 92 microsatellite loci. This finding coupled with the fact that the 0+ herring were at least 3 months old when collected, imply that herring schooling aggregations in part comprise groups of related individuals for at least 3 months after hatching. Our ability to detect kinship among the 0+ spring spawned cohort is of relevance given the fact that the spawning stock biomass of spring spawning herring in the southern Gulf of St. Lawrence declined nearly fivefold from 1995 to 2002 and has remained at low biomass for the past 20 years (McDermid et al., 2018; Turcotte et al., 2021a; Turcotte et al., 2021b). The fact that kinship structure was found in a highly connected metapopulation system such as Atlantic herring suggests that the population size may be relatively low. We note, however, that no direct estimate of population size is feasible with our within-cohort data because of the potential influence of a sweepstakes-like variance in reproductive success (i.e. Hedgecock effect) (Hedgecock and Pudovkin, 2011; Waples and Feutry, 2022). Three individuals were detected as kin to more than one partner, one of them in fact was related as a half-sib to three partners, two of whom were identified as full-sibs. We observed 33 bp differences between the mtDNA sequences of the two individuals identified as full-sibs which is within the range of sequencing error as determined by the binomial tests. Re-sequencing of the FSP revealed, however, that at least 15 of these bp differences may be real SNPs, but most are silent mutations. Regardless, this is the lowest number of base variations observed among kinship pairs in this study. Analysis of the mtDNA of the HSPs also reveals little connection between individuals within a given pair, suggesting the potential for paternal ancestry across the half-sibs. Below we discuss the implications of these results both for the spring spawning component of the Gulf of St. Lawrence Atlantic herring and more generally for our understanding of mutation rate in the mitochondrial genome.

Three of the 18 individuals identified as members of a half-sib pair appeared related as a half-sib to multiple individuals, two of them appeared as half-sib to two individuals each, and the third (Ch 258) appeared as a half-sib to three individuals (Ch226, Ch290, and Ch534) but these three were not identified as half-sibs to each other. This case is only possible if two out of the three individuals are full-sibs. Based on our subsequent analyses, individuals Ch226 and Ch290 were indeed identified as potential full-sibs. Therefore, even though from a statistical point of view there is not enough power to confidently distinguish true full-sib pairs from half-sib pairs using 92 polymorphic microsatellite markers and *CKMRsim* , we can infer that Ch226 and Ch290 are likely full-sib pairs given the half-sib relationships between these four individuals. There are 33 bp differences between the FSP with the majority appearing to be silent mutations, suggesting

minimal effect on the function of the genes. The higher number of bp differences ([?]62bp) observed between HSP individuals suggest they are likely related through paternal ancestry.

The variable nature of the D-Loop is well documented (Iguchi et al., 1997; Sharma et al., 2005; Xiao et al., 2009); therefore, some degree of polymorphism is to be expected in this region. Compared to the non-coding D-Loop region, mutation sites within a coding region such as the *ND* genes that code for NADH dehydrogenase complex may have a larger impact on the individual (Hauser, Turan, and Carvalho, 2001). The*ND5* gene specifically codes for NADH dehydrogenase subunit 5 and together with *ND2* and *ND4* genes, form a transmembrane protein pump complex homologous to Na^+/H^+ antiporters (Yu et al., 2010; Teacher et al., 2012); *ND5/6* genes have been shown to exhibit high levels of differentiation in clupeids (Hauser et al., 2001). A mutation in this region will likely affect the ability of proton pumps in the mitochondria and influence the oxidative phosphorylation mechanism of the cell (da Fonseca et al., 2008; Teacher et al., 2012; Consuegra et al., 2015). The critical biological functions of these regions are likely the reason why only silent mutations are observed in these regions between the FSP. Studies have also suggested the*ND5* gene may be under positive selection across taxa including mammals (de Fonseca et al., 2008), and other fish species such as Pacific salmon (Garvin et al., 2011). In addition, a study using whole mtDNA genome of Atlantic herring from the Baltic Sea revealed evidence for positive selection in *ND5* along with *ND2* and*ND4* genes but found weak correlation between genetic diversity and temperature and salinity (Teacher et al., 2012). Herring has a reported mutation rate of 2.0 x 10^{-9} per base per generation (Feng et al., 2017) for germ cells but the mutation rate in mtDNA is not known. Here we report a mtDNA mutation rate of 7.19 x 10^{-4} mutations per generation using the number of mutations from the full-sib pair which corresponds to a single generation. Out of the individual mtDNA regions, D-Loop appears to have the highest mutation rate at 4.78 x 10^{-3} which agrees with its variable nature caused by less stringent DNA repair mechanisms (Iguchi et al., 1997; Sharma et al., 2005; Xiao et al., 2009). It is important to note that these rates were estimated using one full-sib pair only, and additional samples may provide more precise estimates.

The detection of one full-sib and several half-sib pairs within a large schooling aggregation of juvenile herring suggests juvenile herring remain together for several months after hatching. This finding contradicts the assumption that no kinship structure can be detected in a marine system especially in a broadcast spawner like herring (Christie et al., 2010; Kamel and Grosberg, 2013). The age 0+ herring collected in this study were spring spawned fish that were collected in the fall. Based on this timeline, we would expect to find little to no kinship structure because of the nature of broadcast spawning and the length of time post hatch. While herring larvae can adjust their bearing across the water column, they are not strong enough to swim against the ocean currents and are likely to be swept away from their original spawning site (Fortier and Leggett, 1983; Sinclair and Tremblay, 1984; Stephenson and Power, 1988). The retention of larval herring in a spawning ground has been previously documented (Sinclair and Iles, 1989; McQuinn, 1997), but the degree of relatedness among larvae was poorly understood possibly due to sampling and technological limitations. Here, our results suggest that related herring individuals still congregate together within a more close-knit aggregation than previously thought. To our knowledge, this study is the first to examine kinship structure in wild caught herring at the juvenile stages.

The neighbour joining tree suggests a high number of mutational differences in mtDNA haplotypes between members of any of the HSPs. This suggests the HSPs are paternally related. As broadcast spawners, Atlantic herring females lay blankets of eggs that sink to the ground on coastal waters while male herring release their sperm in the water column (Evans and Geffen, 1998). The chances of one male fertilizing multiple eggs of the same female are likely low due to the large number of males in the vicinity as well as the environmental factors that affect the sperm. This ultimately results in a very low chance to detect sibship in this system without sampling many individuals. However, the number of base pair differences between the lowest number of mutations in HSP (62 bp) and the candidate FSP (33 bp) is the largest (at least 29 bp assuming some bp differences are the result of sequencing error) across all pairs. The relatively large gap in the number of mutations between the candidate FSP and all other HSPs suggests that Ch226-Ch290 are indeed full-sibs.

Acknowledgements

We thank Fisheries and Oceans Canada for sample collection. Sequencing was funded by NSERC Strategic Partnership Grants for Projects and NSERC Discovery grant to DER. JK was supported by NSERC SPGP and NSERC Discovery grant to DER and Dalhousie's Faculty of Graduate Studies Scholarship. We would like to thank Eric Anderson for his guidance with *CKMRsim* , Tony Eindfelt for his help with the mitogenome analysis, Sarah Salisbury for helping with the binomial analysis, and Dr. Paul Bentzen for the suggestions on the data analysis. We would also like to thank Ian Patterson, Samuel Prystupa, and Mallory Van Wyngaarden for aiding lab work. We would like to thank The Center of Applied Genomics (Hospital for Sick Kids, University of Toronto) for their sequencing services and the Marine Gene Probe Lab at Dalhousie University for the lab facilities. Finally, we would like to thank the Editor and anonymous reviewers for their valuable contributions to the study.

References

Anderson, E. C. (2016). CKMRsim. https://doi.org/10.5281/zenodo.820162

Anthony, V. C., & Fogarty, M. J. (1985). Environmental effects on recruitment, growth, and vulnerability of Atlantic herring (Clupea harengus harengus) in the Gulf of Maine region. *Canadian Journal of Fisheries and Aquatic Sciences* , *42* (S1), s158-s173.

Bang, A., Gronkjaer, P., Clemmesen, C., & Hoie, H. (2006). Parental effects on early life history traits of Atlantic herring (Clupea harengus L.) larvae. *Journal of Experimental Marine Biology and Ecology* , 334(1), 51-63.

Barrio, A. M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H. E., ... & Andersson, L. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *elife* , 5, e12081.

Benoit, H. P., & Rail, J. F. (2016). *Principal predators and consumption of juvenile and adult Atlantic Herring (Clupea harengus) in the southern Gulf of St. Lawrence* . Canadian Science Advisory Secretariat.

Berg, F., Slotte, A., Andersson, L., & Folkvord, A. (2019). Genetic origin and salinity history influence the reproductive success of Atlantic herring. *Marine Ecology Progress Series* , 617, 81-94.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* , *30* (15), 2114-2120.

Bravington, M. V., Grewe, P. M., & Davies, C. R. (2016). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* , *7* (1), 1-8.

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Statistical Science* , *31* (2), 259-274.

Brown, K. H. (2008). Fish mitochondrial genomics: sequence, inheritance and functional variation. *Journal of Fish Biology* , 72(2), 355-374.

Christie, M. R., Johnson, D. W., Stallings, C. D., & Hixon, M. A. (2010). Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. *Molecular ecology* , *19* (5), 1042-1057.

Consuegra, S., John, E., Verspoor, E., & de Leaniz, C. G. (2015). Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genetics Selection Evolution* , 47(1), 1-10.

da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC genomics* , 9(1), 1-22.

Elphinstone, M. S., Hinten, G. N., Anderson, M. J., & Nock, C. J. (2003). An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* , *3* (2), 317-320.

Evans, J. P., & Geffen, A. J. (1998). Male characteristics, sperm traits, and reproductive success in winter-spawning Celtic Sea Atlantic herring, Clupea harengus. *Marine Biology* , 132(2), 179-186.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics* , *1* , 117693430500100003.

Faircloth, B. C. (2008). MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular ecology resources* , *8* (1), 92-94.

Feng, C., Pettersson, M., Lamichhaney, S., Rubin, C. J., Rafati, N., Casini, M., ... & Andersson, L. (2017). Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate.*Elife* , 6, e23907.

Fortier, L., & Leggett, W. C. (1983). Vertical migrations and transport of larval fish in a partially mixed estuary. *Canadian Journal of Fisheries and Aquatic Sciences,* 40(10), 1543-1555.

Fuentes Pardo, A. P. (2020). *Adaptive and neutral genetic variation in spring-and fall-spawning herring (Clupea harengus L.) in the Northwest Atlantic* (Doctoral dissertation).

Garvin, M. R., Bielawski, J. P., & Gharrett, A. J. (2011). Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PloS one* , 6(9), e24127.

Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., & Keightley, P. D. (2008). Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. *PLoS biology* , 6(8), e204.

Hauser, L., Turan, C., & Carvalho, G. R. (2001). Haplotype frequency distribution and discriminatory power of two mtDNA fragments in a marine pelagic teleost (Atlantic herring, Clupea harengus). *Heredity* , 87(6), 621-630.

Han, F., Jamsandekar, M., Pettersson, M. E., Su, L., Fuentes-Pardo, A. P., Davis, B. W., Bekkevold, D., Berg, F., Casini, M., Dahle, G., Farrell, E. D., Folkvord, A., & Andersson, L. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *ELife* , 9, e61076.

Hare, J.A., and D.E. Richardson. Chapter 15- The use of early life stages in stock identification studies. Stock Identification Methods, edited by S.X. Cadrin, L.A. Kerr, and S. Mariani. Academic Press. 2014. Pp. 329-364.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science* , 87(4), 971-1002.

Iguchi, K. I., Tanimura, Y., & Nishida, M. (1997). Sequence divergence in the mtDNA control region of amphidromous and landlocked forms of ayu.*Fisheries science* , 63(6), 901-905.

Jones, C. (1986). Determining age of larval fish with the otolith increment technique. *Fishery Bulletin* , 84(1), 91-104.

Kamel, S. J., & Grosberg, R. K. (2013). Kinship and the evolution of social behaviours in the sea. *Biology letters* , *9* (6), 20130454.

Kerr, Q., Fuentes-Pardo, A. P., Kho, J., McDermid, J. L., & Ruzzante, D. E. (2019). Temporal stability and assignment power of adaptively divergent genomic regions between herring (Clupea harengus) seasonal spawning aggregations. *Ecology and evolution* , *9* (1), 500-510.

Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., Singh, R., Ruzzante, D. E., & Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences* , *114* (17), E3452-E3461.

McCracken, G. R., Wilson, K. L., Paterson, I., Perry, R., Keefe, D., & Ruzzante, D. E. (2014). Development of 17 novel microsatellite markers for the longnose sucker (Catostomus catostomus) and successful cross-specific amplification of 14 previously developed markers from congeneric species. *Conservation Genetics Resources* , 6(2), 329-332.

McCracken, G. R., Wilson, K. L., Brewis, H. T., McBride, M. C., Paterson, I., Perry, R., Keefe, D., & Ruzzante, D. E. (2014). Development of 26 novel microsatellite makers for the round whitefish (Prosopium cylindraceum) and successful polymorphic cross-specific amplification of seven previously developed salmonid markers.*Conservation genetics resources* , 6(4), 1023-1026.

McDermid, J., Swain, D. P., Turcotte, F., Robichaud, S. A., & Surette, T. (2018). *Assessment of the NAFO Division 4T southern Gulf of St. Lawrence Atlantic herring (Clupea harengus) in 2016 and 2017* . Canadian Science Advisory Secretariat.

McPherson, A. A., O'Reilly, P. T., & Taggart, C. T. (2004). Genetic differentiation, temporal stability, and the absence of isolation by distance among Atlantic herring populations. *Transactions of the American Fisheries Society* , 133(2), 434-446.

McPherson, A. A., Stephenson, R. L., & Taggart, C. T. (2003). Genetically different Atlantic herring Clupea harengus spawning waves.*Marine Ecology Progress Series* , 247, 303-309.

McQuinn, I. H. (1997). Metapopulations and the Atlantic herring. Reviews in fish biology and fisheries, 7(3), 297-329.

Melvin, G. D., Stephenson, R. L., & Power, M. J. (2009). Oscillating reproductive strategies of herring in the western Atlantic in response to changing environmental conditions. *ICES Journal of Marine Science* , 66(8), 1784-1792.

Peakall, R. O. D., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular ecology notes* , *6* (1), 288-295.

Pettersson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., ... & Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection. *Genome research* , 29(11), 1919-1928.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* , *4* (3), 535-538.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ruzzante, D. E., Mariani, S., Bekkevold, D., Andre, C., Mosegaard, H., Clausen, L. A., ... & Carvalho, G. R. (2006). Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings of the Royal Society B: Biological Sciences* , 273(1593), 1459-1464.

Ruzzante, D. E., McCracken, G. R., Forland, B., MacMillan, J., Notte, D., Buhariwalla, C., Flemming, J. M., & Skaug, H. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods in Ecology and Evolution* , *10* (9), 1445-1453.

Ruzzante, D. E., Taggart, C. T., & Cook, D. (1996). Spatial and temporal variation in the genetic composition of a larval cod (Gadus morhua) aggregation: cohort contribution and genetic stability.*Canadian Journal of Fisheries and Aquatic Sciences* , 53(12), 2695-2705.

Schunter, C., Pascual, M., Garza, J. C., Raventos, N., & Macpherson, E. (2014). Kinship analyses identify fish dispersal events on a temperate coastline. *Proceedings of the Royal Society B: Biological Sciences* , *281* (1785), 20140556.

Sharma, H., Singh, A., Sharma, C., Jain, S. K., & Singh, N. (2005). Mutations in the mitochondrial DNA

D-loop region are frequent in cervical cancer. *Cancer cell international* , 5, 34. https://doi.org/10.1186/1475-2867-5-34

Sinclair, M., & Power, M. (2015). The role of "larval retention" in life-cycle closure of Atlantic herring (Clupea harengus) populations.*Fisheries Research* , 172, 401-414.

Sinclair, M., & Tremblay, M. J. (1984). Timing of spawning of Atlantic herring (Clupea harengus harengus) populations and the match–mismatch theory. *Canadian Journal of Fisheries and Aquatic Sciences* , 41(7), 1055-1065.

Sinclair, M., & Iles, T. D. (1989). Population regulation and speciation in the oceans. *ICES Journal of Marine Science,* 45(2), 165-175.

Stephenson, R. L., & Power, M. J. (1988). Semidiel vertical movements in Atlantic herring Clupea harengus larvae: A mechanism for larval retention?. *Marine ecology progress series* . Oldendorf, 50(1), 3-11.

Stoler, N., & Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. *NAR genomics and bioinformatics* , 3(1), lqab019.

Teacher, A. G., Andre, C., Merila, J., & Wheat, C. W. (2012). Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC evolutionary biology* , 12(1), 1-14.

Turcotte, F., Swain, D. P., & McDermid, J. L. (2021). *NAFO 4TVn Atlantic herring population models: from Virtual Population Analysis to Statistical Catch-at-Age estimating time-varying natural mortality* . Canadian Science Advisory Secretariat= Secretariat canadien de consultation scientifique.

Turcotte, F., Swain, D. P., McDermid, J. L., & DeJong, R. A. (2021). *Assessment of the NAFO Division 4TVn Southern Gulf of St. Lawrence Atlantic Herring (Clupea Harengus) in 2018-2019* . Canadian Science Advisory Secretariat (CSAS).

Waldman, B. (1988). The ecology of kin recognition. *Annual review of ecology and systematics* , 19(1), 543-571.

Waples, R. S., & Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: a cautionary view.

Waples, R. S., & Feutry, P. (2022). Close-kin methods to estimate census size and effective population size. *Fish and Fisheries* , 23(2), 273-293.

Wilson, C. J., Murphy, H. M., Bourne, C., Pepin, P., & Robert, D. (2018). Feeding ecology of autumn-spawned Atlantic herring (Clupea harengus) larvae in Trinity Bay, Newfoundland: is recruitment linked to main prey availability?. *Journal of Plankton Research* , 40(3), 255-268.

Xiao, Y., Zhang, Y., Gao, T., Yanagimoto, T., Yabe, M., & Sakurai, Y. (2009). Genetic diversity in the mtDNA control region and population structure in the small yellow croaker Larimichthys polyactis.*Environmental biology of fishes* , 85(4), 303-314.

Yu, J. N., Azuma, N., Yoon, M., Brykov, V., Urawa, S., Nagata, M., ... & Abe, S. (2010). Population genetic structure and phylogeography of masu salmon (Oncorhynchus masou masou) inferred from mitochondrial and microsatellite DNA analyses. *Zoological science* , 27(5), 375-385.

Zhan, L., Paterson, I. G., Fraser, B. A., Watson, B., Bradbury, I. R., Nadukkalam Ravindran, P., ... & Bentzen, P. (2017). MEGASAT: automated inference of microsatellite genotypes from sequence data. *Molecular ecology resources* , *17* (2), 247-256.

Data Accessibility

Microsatellite files and metadata is available on Dryad Digital Repository with DOI: https://doi.org/10.5061/dryad.xd2547djs

Author contributions

This manuscript is part of JK's PhD thesis under DER's supervision. All authors contributed towards the design of the study. Atlantic herring were processed by JK and JLM. Genetic data were generated and analyzed by JK and GRM. The manuscript was written primarily by JK with input from DER, and contributions by JLM and GRM.

Tables and Figures

Table 1. General summary of the Atlantic herring samples collected from Gulf of St. Lawrence in 2018 divided according to age class.

Age class/ cohorts	N	Length (cm)	Spawning season	N (post filtering)
Age 0+	1391	6 – 15	Spring	980
Age 1+	654	17 - 21	Fall	411

Table 2. Estimates of false negatives and false positive rates for the 0+ cohort comparing half-sibs to unrelated individuals. The underlined values in the False positive rates and λ*columns straddle the reference false positive-rate value used as a threshold (1.042E-07), but λ* = 12 was chosen in downstream analyses to further minimize the effects of false negatives.

	False negative rates	False positive rates	Standard error	Num-nonzero wts	λ*
1	0.01	3.73E-05	2.30E- 6	19800	4.95
2	0.05	1.03E-06	3.38E- 8	19000	9.84
3	0.1	*1.34E-07*	3.54E- 9	18000	*12.3*
4	0.11	*1.043e-07*	2.77e-09	17792	*12.56*
5	0.2	9.81E-09	2.18E-10	16000	15.3
6	0.3	1.21E-09	2.51E-11	14000	17.6

Table 3. 11 half-sib pairs detected for Age0+ GSL herring data (n = 980), all with logl_ratio value greater than 12. Individuals that have multiple half-sibs in the list are in red. Ch258 is paired with three individuals: Ch290, Ch226, and Ch534, denoted with an asterisk next to their ID. In all cases 92 microsatellite loci were employed

	Individual 1	**Individual 2**	**logl_ratio**
1	Ch884	Ch891	27.5
2	Ch549	Ch933	16.9
3	Ch258*	Ch290*	16.4
4	Ch250	Ch290	16
5	Ch232	Ch588	15.5
6	Ch633	Ch764	15.5

7	Ch634	Ch987	15.2
8	Ch258*	Ch226*	14.9
9	Ch250	Ch774	14
10	Ch1115	Ch676	12.6
11	Ch258*	Ch534*	12.2

Table 4. Estimates of false negatives and false positive rates for the 0+ cohort comparing full-sibs to unrelated individuals. None of the false-positive rates listed below passed the reference false positive-rate value used as a threshold (1.042E-07) indicating there is insufficient statistical power to confidently identify FSPs.

	False negative rates	False positive rates	Standard error	Num-nonzero wts	λ*
1	0.01	3.344e-21	3.331e-22	19800	41.2
2	0.05	1.267e-24	6.526e-26	19000	50.3
3	0.10	1.529e-26	6.265e-28	18000	55.2
4	0.20	5.227e-29	1.728e-30	16000	61.3
5	0.30	7.016e-31	2.118e-32	14000	65.8

Table 5. 17 full-sib pairs with the corresponding logl_ratio detected for Age0+ GSL herring data using 92 loci (n = 980). Cells in bold indicate pairs that were identified previously as HSPs. The pair in red is the first candidate full-sib pair that was also deduced from the HSP analysis involving Ch258 with these two individuals and Ch534. The last column shows the corresponding number of mutations between the two individuals based on the mtDNA alignment of HSPs and the FSP.

	Individual 1	**Individual 2**	**logl_ratio**	**Number of mutations (Bp)**
1	**Ch884**	**Ch891**	**25.9**	**109**
2	**Ch258**	**Ch290**	**19.3**	**126**
3	**Ch250**	**Ch290**	**16.5**	**109**
4	**Ch258**	**Ch226**	**15.2**	**131**
5	**Ch549**	**Ch933**	**13.7**	**81**
6	Ch226	Ch290	13.0	33
7	**Ch232**	**Ch588**	**12.6**	**62**
8	**Ch633**	**Ch764**	**12.0**	**73**
9	Ch250	Ch774	8.4	-
10	Ch637	Ch984	7.7	-
11	**Ch258**	**Ch534**	**4.8**	**67**
12	Ch463	Ch964	4.7	-
13	**Ch1115**	**Ch676**	**4.6**	**116**
14	Ch748	Ch764	4.2	-
15	Ch980	Ch981	2.4	-
16	Ch250	Ch258	1.5	-
17	Ch634	Ch731	1.0	-
Other half-sib pairs	Other half-sib pairs			
	Ch250	Ch774	-	116
	Ch634	Ch987	-	96

Table 6. List of sibship pairs and the corresponding number of mutations (bp) based on the mtDNA alignment. The total number of mutations (Whole mtDNA) is shown alongside the number of mutations

for the D-Loop, *COX1*, *CytB*, *ND2*, and *ND5*regions. For the half-sibs (2-12), the percent proportion of the number of mutations by the length of the region is shown below each base substitution values. The mutation rate for the full-sib pair (Ch226-290) is estimated by dividing the number of mutations by the total length of the respective mtDNA regions (in bp).

	Sibship pairs	**Whole mtDNA (16,700bp)**	**D-Loop (1,047bp)**	COX1 **(1,551bp)**	CytB **(1,141bp)**	ND2 **(1,045bp)**	ND5 **(1,836bp)**
1	Ch226-Ch290	12	5	1	1	2	1
	Mutation rate (per generation)	$7.19x10^{-4}$	$4.78x10^{-3}$	$6.45x10^{-4}$	$8.76x10^{-4}$	$1.91x10^{-3}$	$5.45x10^{-4}$
2	Ch232-Ch588	62 (0.37%)	13 (1.24%)	0 (0%)	10 (0.88%)	6 (0.57%)	12 (0.65%)
3	Ch258-Ch534	67 (0.40%)	10 (0.96%)	1 (0.06%)	8 (0.70%)	8 (0.77%)	12 (0.65%)
4	Ch633-Ch764	73 (0.44%)	18 (1.72%)	7 (0.45%)	11 (0.96%)	19 (1.82%)	16 (0.87%)
5	Ch549-Ch933	81 (0.49%)	12 (1.15%)	1 (0.06%)	2 (0.18%)	8 (0.77%)	11 (0.60%)
6	Ch634-Ch987	96 (0.57%)	13 (1.24%)	2 (0.13%)	5 (0.44%)	11 (1.05%)	18 (0.98%)
7	Ch884-Ch891	109 (0.65%)	17 (1.63%)	5 (0.32%)	12 (1.05%)	12 (1.15%)	21 (1.14%)
8	Ch250-Ch290	109 (0.65%)	17 (1.63%)	9 (0.58%)	13 (1.14%)	13 (1.24%)	18 (0.98%)
9	Ch1115-Ch676	116 (0.69%)	16 (1.53%)	7 (0.45%)	16 (1.40%)	16 (1.53%)	17 (0.93%)
10	Ch250-Ch774	116 (0.69%)	19 (1.81%)	6 (0.39%)	14 (1.23%)	11 (1.05%)	20 (1.09%)
11	Ch258-Ch290	126 (0.75%)	15 (1.43%)	8 (0.52%)	17 (1.49%)	11 (1.05%)	23 (1.25%)
12	Ch258-Ch226	131 (0.78%)	18 (1.72%)	5 (032%)	13 (1.14%)	10 (0.96%)	24 (1.31 %)

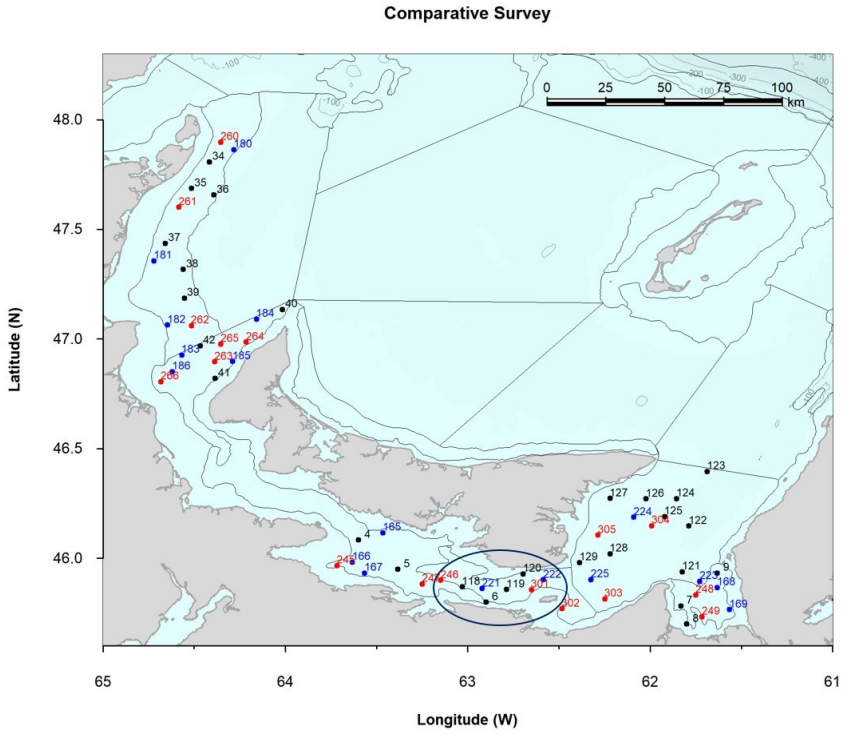


Figure 1. Map of the Fisheries and Oceans Canada 2018 trawl survey in Southern Gulf of St. Lawrence. The herring juveniles were collected along Cumberland strait from sites 118, 119, and 120.

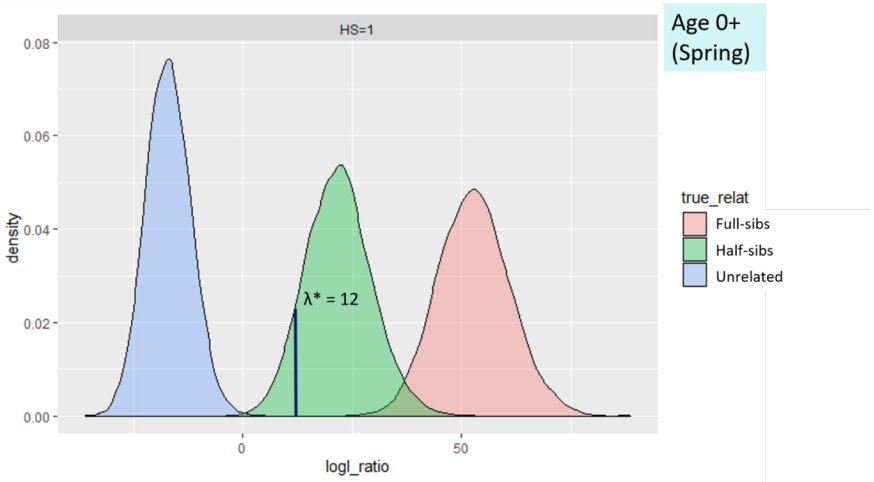


Figure 2. Log-likelihood of full-sibs vs. half-sibs vs. unrelated individuals for Age0+ data (n = 980). The black solid line indicates the $\lambda^*= 12$ obtained from the *CKMRsim* ; any HSPs detected using these markers with logl_ratio > 12 can be considered true pairs. The absence of an overlap between the curves for unrelated individuals and HSP suggest no concern for true unrelated pairs to be misidentified as HSPs. This is contrary to the overlapping full-sibs and half-sibs curve which suggests a lack of power in distinguishing between the two sibships category.

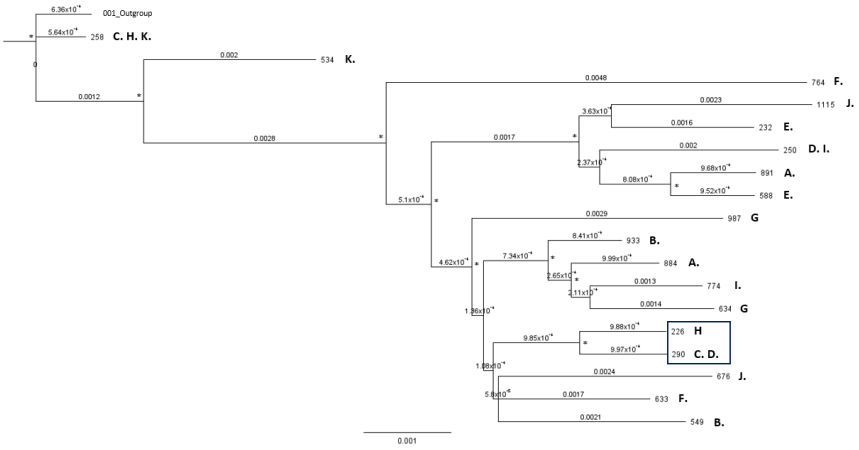


Figure 3. Neighbour joining phylogenetic tree generated from 19 Gulf of St. Lawrence Atlantic herring individuals. The tree was created using Geneious Tree Builder (global alignment with free end gaps) from Geneious Prime 2021.2. The sample IDs are accompanied by letters that indicate the half-sib pairs assignment based on *CKMRsim* results. The values on the branches show average nucleotide substitution per site. Significant bootstrap support values (90-100) are shown next to each branch division with an asterisk. The box indicates a candidate full-sib pair.

Supporting material

Table S1. List of microsatellite ID in order of chromosome number and base pair position. The *Clupea harengus* Ch_v2.0.2 genome (GenBank GCA_900700415.2) was used as reference.

Microsatellite ID	Chromosome ID	BP position
CHR155	1	32.4
CHR111	1	70.2
CHR13	1	173.4
CHR117	1	198.3
CHR92	2	20.0
CHR150	2	87.2
CHR40	2	109.7
CHR137	2	264.9
CHR67	3	33.5
CHR80	3	262.1
CHR110	3	291.2
CHR63	4	38.4
CHR79	5	18.1
CHR147	5	66.0
CHR153	5	126.3
CHR181	5	145.5
CHR5	5	176.6
CHR37	5	176.8
CHR190	5	191.0
CHR54	5	197.6
CHR156	6	74.3

Microsatellite ID	Chromosome ID	BP position
CHR138	6	122.8
CHR143	6	148.9
CHR164	6	270.5
CHR131	7	224.5
CHR94	7	270.8
CHR90	7	286.6
CHR53	8	231.8
CHR65	9	134.3
CHR22	9	207.0
CHR162	10	22.7
CHR142	10	23.2
CHR16	10	24.7
CHR89	10	230.3
CHR175	10	247.4
CHR171	10	275.8
CHR86	11	69.0
CHR95	11	118.6
CHR91	11	167.6
CHR122	11	203.6
CHR133	11	234.7
CHR52	11	251.6
CHR64	12	28.5
CHR172	12	174.2
CHR31	12	216.7
CHR127	13	80.4
CHR1	13	106.8
CHR43	13	114.9
CHR108	13	177.0
CHR167	13	183.7
CHR184	13	194.2
CHR30	13	235.1
CHR68	13	273.9
CHR123	14	95.1
CHR88	14	159.9
CHR114	14	191.2
CHR192	15	13.8
CHR176	15	54.8
CHR62	15	163.5
CHR38	15	203.7
CHR141	15	260.3
CHR81	16	17.6
CHR124	16	224.2
CHR32	16	226.4
CHR121	17	113.6
CHR152	17	272.3
CHR107	18	57.4
CHR159	18	117.9
CHR146	18	132.6
CHR161	18	202.5
CHR149	19	154.8

Microsatellite ID	Chromosome ID	BP position
CHR151	19	197.3
CHR87	19	197.3
CHR185	19	225.1
CHR17	20	132.1
CHR139	21	70.0
CHR188	21	83.2
CHR50	21	187.6
CHR144	21	246.2
CHR59	22	89.7
CHR69	22	98.0
CHR120	22	113.6
CHR109	22	131.7
CHR173	23	4.6
CHR47	23	141.7
CHR174	23	143.4
CHR56	24	82.3
CHR96	25	87.0
CHR119	25	95.2
CHR58	25	111.2
CHR134	25	132.5
CHR183	26	100.4

Table S2. Summary table of the 33 base pair variations across the FSPs (Ch226 and Ch290) showing the location in the genome, gene, sample ID exhibiting the mutation, the type of mutation, whether the mutation is present in other samples, and whether the base pair variation was confirmed by resequencing on the MiSeq. For the last two columns, "Y" is yes, "N" is no, and "N/A" is not able to be confirmed via re-sequencing (i.e. due to lack of coverage).

No.	**BP location**	**Gene**	**Sample ID**	**Type of mutation**	**Present in other samples?**	**Present in**
1	2454	I-rRNA	226	Silent mutation	N	N/A
2	3408	ND1	290	Silent mutation	Y (1 other)	Y
3	4290	ND2	226	Silent mutation	N	N/A
4	4323	ND2	290	Silent mutation	Y (2 others)	Y
5	4494	ND2	290	Silent mutation	N	Y
6	4644	ND2	226	Silent mutation	N	Y
7	4989	ND2	290	Silent mutation	Y (1 other)	Y
8	5227	tRNA-Ala	290	Non silent Point mutation	N	Y
9	5555	COX1	290	Non silent Point mutation	N	Y
10	6140	COX1	290	Non silent Point mutation	Y (2 others)	Y
11	9232	COX3	290	Silent mutation	Y (1 other)	Y
12	10441	ND4	226	Silent mutation	N	Y
13	10462	ND4	226	Silent mutation	N	N/A
14	10879	ND4	226	Silent mutation	N	N/A
15	11161	ND4	226	Silent mutation	N	N/A
16	12186	ND5	290	Non silent Point mutation	N	Y
17	12291	ND5	226	Non silent Point mutation	N	N/A
18	12471	ND5	226	Nonsense point mutation	N	N/A
19	12780	ND5	226	Non silent Point mutation	N	N/A
20	13332	ND5	290	Non silent Point mutation	N	N/A

No.	BP location	Gene	Sample ID	Type of mutation	Present in other samples?	Present in
21	13665	ND5	290	Nonsense point mutation	N	Y
22	13672	ND5	226	Silent mutation	Y (2 others)	N/A
23	14216	ND6	290	Nonsense point mutation	N	N/A
24	14386	CYTB	290	Silent mutation	N	N*
25	15396	CYTB	290	Non silent Point mutation	N	N/A
26	15420	CYTB	226	Non silent Point mutation	N	Y
27	15435	CYTB	290	Non silent Point mutation	N	N/A
28	15744	D-Loop	290	Non silent Point mutation	Y (5 others)	N/A
29	15791	D-Loop	226	Silent mutation	N	N/A
30	15852	D-Loop	226	Non silent Point mutation	Y (2 others)	N/A
31	15937	D-Loop	226	Nonsense point mutation	N	Y
32	16088	D-Loop	290	Silent mutation	Y (5 others)	Y
33	16468	D-Loop	226	Non silent Point mutation	Y (3 others)	N/A

***N:** the base pair variation was detected in the same location but had a different base pair conversion.

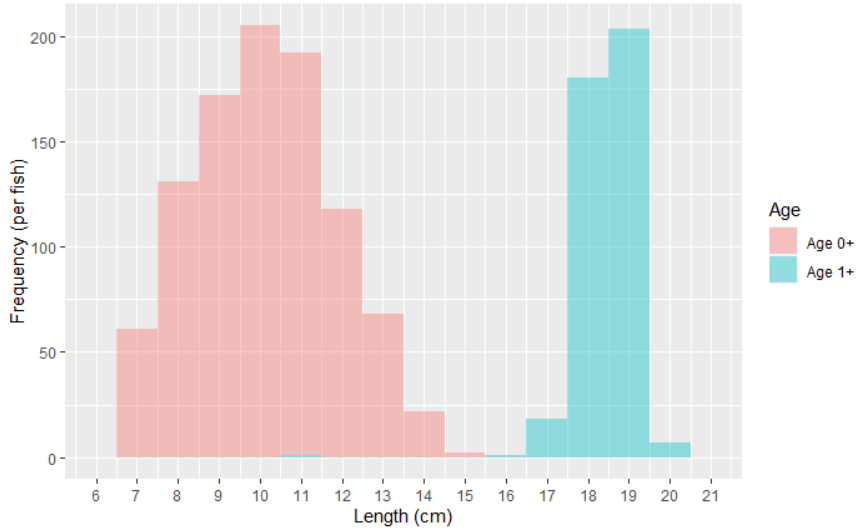
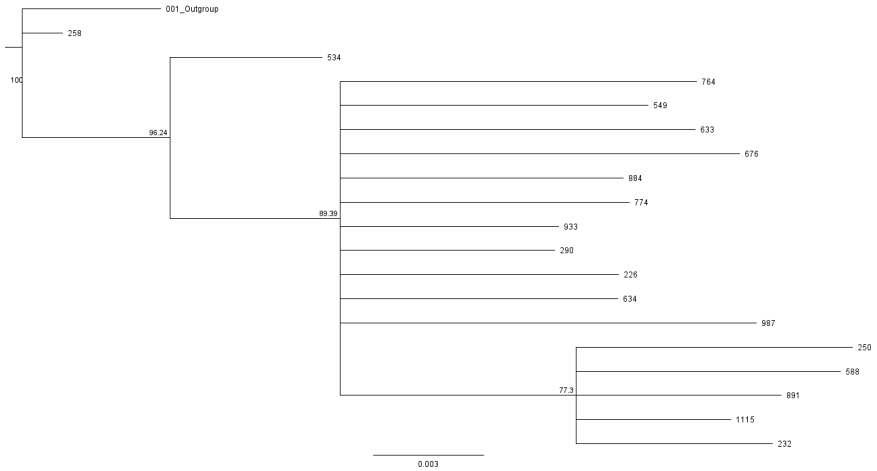
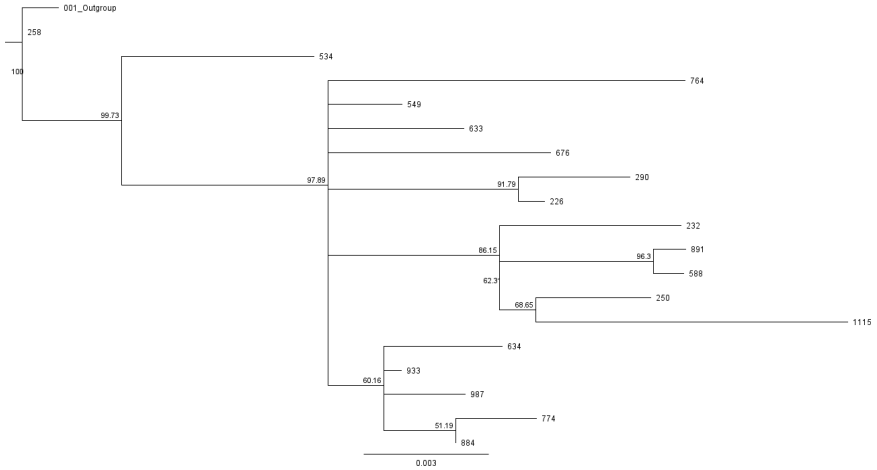


Figure S1 . Frequency of herring for each length category (cm) for age 0+ (N=980) and 1+ (N=411) Gulf of Saint Lawrence samples. Based on otolith analysis, the age 0+ are Spring and the age 1+ are Fall spawned fish.

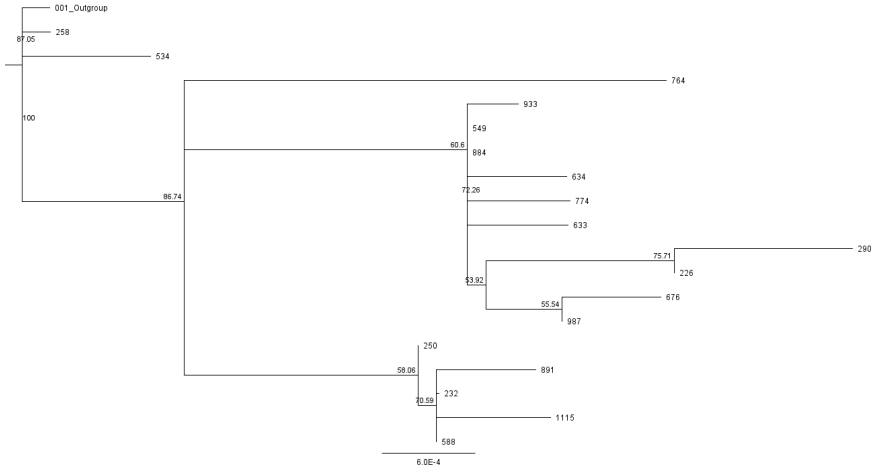
A. D-Loop Tree



B. *CytB* Tree



C. *COX1* Tree



D. *ND5* Tree

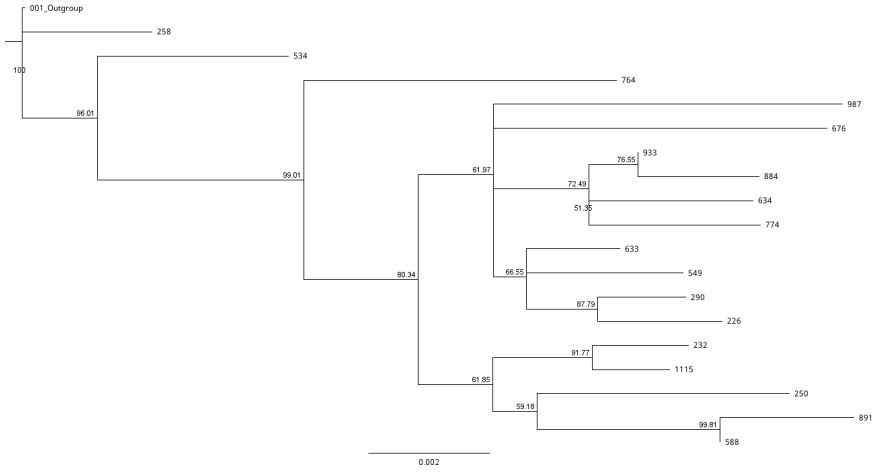


Figure S2. Neighbour joining phylogenetic tree generated from 19 Gulf of St. Lawrence Atlantic herring individuals on **(A.)**D-Loop, **(B.)** *CytB* , **(C.)** *COX1* ,**(D.)** *ND5* regions of the mtDNA. The tree was created using Geneious Tree Builder (global alignment with free end gaps) from Geneious Prime 2021.2. The values on the branches show average nucleotide substitution per site. Significant bootstrap support values (90-100) are shown next to each branch division with an asterisk. The box indicates a candidate full-sib pair.

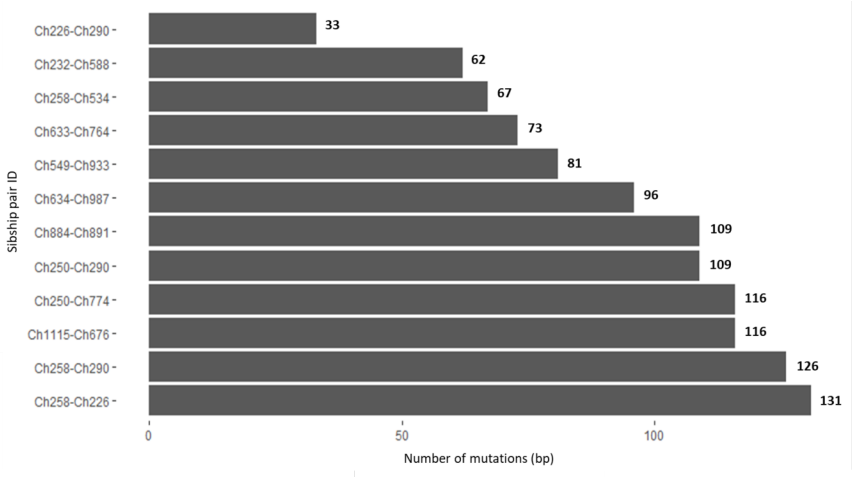


Figure S3. Histogram showing the number of mutations (bp) for full-sib (orange) and half-sib (blue) pairs. The largest difference between the number of mutations is between the full-sib and the half-sib pairs at 29 bp.

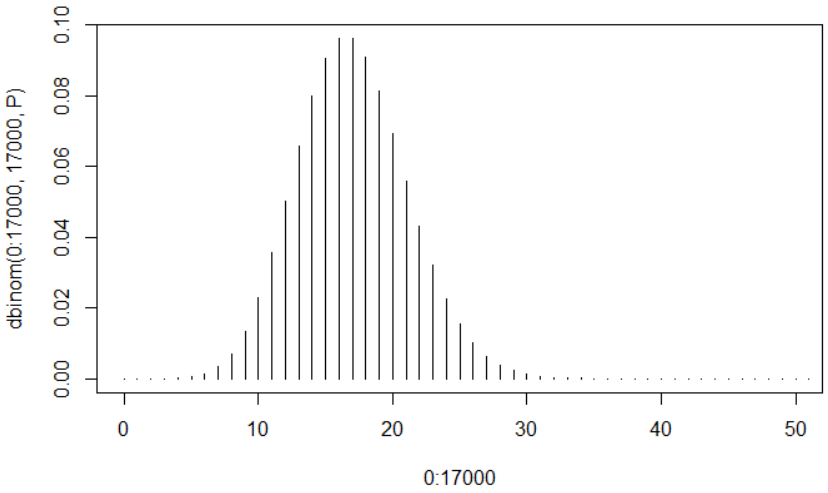


Figure S4. A bar graph showing the binomial distribution of expected per base variation due to sequencing error of a 16700 bp long sequence based on a 0.1% error rate. The x-axis shows the number of expected per base variation and the y axis shows the probability of sequencing error causing the corresponding number of base pair variation; the median of the curve is at 17 bp.

Half-sib pairs for Age 1+, Full-sib pairs analysis and kinship across cohorts

The same power analysis was conducted on the Age 1+ dataset to detect HSPs. Despite being able to identify 23 HSPs in the Age 1+ dataset, the smaller sample size of N = 411 meant significantly fewer pairwise comparisons than the Age 0+ dataset and therefore insufficient power to confidently identify these pairs as true half-sibs. This is reflected in the logl_ratio values of the Age 1+ HSPs being lower than the reference threshold lambda star value.