Transcriptomics of epidermal mucus as a nonlethal method to compare gene expression variation among fish populations

Nicolette Andrzejczyk[1], Lee Hrenchuk[2], Vince Palace[2], and Daniel Schlenk[1]

[1]University of California Riverside
[2]IISD Experimental Lakes Area

September 28, 2020

Abstract

Although transcriptomic analysis of wild organisms is a powerful tool to understand molecular differences among populations, most methods require the use of lethal sampling. In fish, the use of epidermal mucus is a promising method for development of nonlethal sampling tools. Previous studies have shown that mRNA is dynamically regulated in fish epidermal mucus following stressor exposure, suggesting that mucus is reflective of molecular changes occurring within the organism in response to its environment. The aim of the study was to determine whether transcriptomics of mucus could discern molecular differences among populations of lake trout. In order to do so, mucus was collected and sequenced from four geographically-distinct lake trout (Salvelinus namaycush) populations at the IISD Experimental Lakes Area. Principal component analysis (PCA) and hierarchical clustering of read data showed that each lake trout population had unique transcriptomic profiles, suggesting that RNA sequencing of mucus is able to discern molecular differences among fish populations. Furthermore, differential gene expression analysis identified regulation of immune-related transcripts and viral gene expression transcripts among populations. PCA and a mixed linear model of water quality parameters indicated that environmental variables accounted for transcriptomic variation among populations. However, 32% of transcriptomic variance was unaccounted for by the mixed linear model, suggesting that other variables may influence transcription, such as epigenetics and presence of pathogens. Overall, results indicate that RNA sequencing of epidermal mucus is an effective, nonlethal method to study transcriptional differences among fish populations and may be especially useful for studies of endangered species.

Introduction

In both laboratory and field experiments, lethal techniques have traditionally been employed in the collection of wildlife tissue for analysis of gene expression, contaminant loading, stable isotopes, and other endpoints of interest. However, there has been pressure in the scientific community to develop nonlethal sampling techniques as sacrifice of wildlife raises concerns, particularly in the study of threatened species (Bennett et al., 2016; Sloman, Bouyoucos, Brooks, & Sneddon, 2019). To minimize impact on animal welfare, nonlethal analyses of fish health have been increasingly used in scientific studies, including blood sampling (Brumbaugh, Schmitt, & May, 2005; Rodríguez-Jorquera et al., 2019), fin clipping (Imbery et al., 2019; Sanderson et al., 2009), tissue biopsies (Baker, Blanchfield, Paterson, Flett, & Wesson, 2004; Henderson, Stevens, & Lee, 2016; Razavi et al., 2019), and collection of scales (Fincel, Vandehey, & Chipps, 2012; Lake, Ryba, Serbst, & Libby, 2006). The use of fish epidermal mucus for nonlethal assessment of animal health has also become of more interest in the past decade (Brinchmann, 2016; Bulloch et al., 2020; Cordero, Brinchmann, Cuesta, Meseguer, & Esteban, 2015; Dzul-Caamal et al., 2016; Greer et al., 2019; Provan et al., 2013; Rajan, Lokesh, Kiron, & Brinchmann, 2013; Ren, Zhao, Su, Peatman, & Li, 2015). As collection of epidermal mucus is minimally-invasive, inexpensive, rapid, and requires little training, mucus-based analyses are a promising route for furthering the development of nonlethal sampling techniques in fish.

Epidermal mucus has several biological and ecological roles that influence fish health and interaction with the

external environment (Reverter, Tapissier-Bontemps, Lecchini, Banaigs, & Sasal, 2018). Mucus is secreted by epidermal goblet cells, forming a protective layer over the epithelial cells of fish skin. Among other functions, mucus is crucial for ionic and osmotic regulation, respiration, locomotion, and protection against pollutants, pathogens, and abrasion (Shephard, 1994). The multiple roles of epidermal mucus can be attributed to its physical properties as well as its biologically-active composition of various RNA, DNA, lipids, carbohydrates, metabolites, and microbiota (Brinchmann, 2016; Reverter et al., 2018). Moreover, the chemical composition of mucus is altered following exposure to biotic and abiotic stressors. For example, aquaculture-based studies in fish have identified alterations in the composition and abundance of proteins, especially those with immune-related functions, following food deprivation, overcrowding, and bacterial infection (Cordero, Morcillo, Cuesta, Brinchmann, & Esteban, 2016; Easy & Ross, 2009; Guardiola, Cuesta, Arizcun, Meseguer, & Esteban, 2014; Provan et al., 2013; Rajan et al., 2013). Exposure to environmental contaminants can also induce changes in mucus composition, such as alterations in protein profiles of gilthead seabream (*Sparus aurata*) mucus exposed to heavy metals (Guardiola et al., 2015), increases of antioxidant response in mucus of oil-exposed dusky splitfin (*Goodea gracilis*) (Dzul-Caamal et al., 2016), and potential changes in mucosal F_2-isoprostanes following oxidative stress (Bulloch et al., 2020). Thus, analysis of mucus may be a useful tool for monitoring fish health.

In addition to proteomic profiles, expression patterns of mRNA in fish mucus can also provide insight on organism health following various stress conditions. Following *Deepwater Horizon* slick oil exposure, the mucosal transcriptome of juvenile mahi-mahi (*Coryphaena hippurus*) showed differential expression of transcripts involved in immune response, cardiotoxicity, and calcium homeostasis which are phenotypes commonly associated with crude oil or polycyclic aromatic hydrocarbon exposure (Greer et al., 2019). Moreover, using quantitative polymerase chain reaction (qPCR), differential expression of immune-related genes was detected in mucus of channel catfish (*Ictalurus punctatus*) following bacterial challenge (Ren et al., 2015). Results from previous studies suggest that the mucus transcriptome captures gene expression changes associated with biotic and abiotic stressors and, thus, is a viable method to assess fish health. However, mRNA-based analysis of epidermal mucus has yet to be tested in the field and may be a promising nonlethal method for studying wild fish populations.

Transcriptomic analysis of wild populations is a powerful tool in molecular ecology as it can be used to assess gene expression variation among populations and to explore mechanisms by which populations respond to environmental stimuli (Alvarez, Schrey, & Richards, 2015; Meier et al., 2014; Oleksiak, Churchill, & Crawford, 2002; Richards, Rosas, Banta, Bhambhra, & Purugganan, 2012; Whitehead et al., 2012). More specifically, transcriptome-wide assessments have been crucial in understanding mechanistic responses to different environmental conditions and how those responses vary in different wild populations. For example, Whitehead et al. 2010 compared transcriptomic responses in killifish (*Fundulus heteroclitus*) to polychlorinated biphenyl (PCB) within a pollutant-tolerant population and a nearby sensitive population. PCB exposure within the sensitive population caused significant transcriptomic changes that were linked to developmental deformities, whereas physiological and transcriptomic responses within the tolerant population where comparatively negligible. Through comparative transcriptomics, PCB tolerance was likely attributed to the blockade of the aryl hydrocarbon receptor (AHR) signaling pathway in the tolerant population, a key mediator in PCB toxicity. Although transcriptomics studies are useful in providing insight on global molecular differences among populations, they often rely on lethal sampling which may make it difficult to study threatened species.

While mucus has proven to be useful as a diagnostic metric in laboratory studies, efficacy of mucus-based gene expression analysis to evaluate environmental stressors or population differences has yet to be assessed in wild fish populations. To determine whether mucus can be used to nonlethally assess transcriptomic differences in geographically-distinct fish populations, RNA sequencing was conducted on epidermal mucus collected from four lake trout (*Salvelinus namaycush)* populations residing in boreal lakes at the IISD Experimental Lakes Area (IISD-ELA) research station in Northwestern Ontario, Canada. Through sequencing, differential gene expression analysis will elucidate molecular mechanisms that vary among lake trout populations due to ecological differences or epigenetic factors. Transcriptomics of epidermal mucus may be useful for assessing effects of environmental contaminants, varying ecological conditions, epigenetic modifications, and other

variables among fish populations without the use of lethal sampling.

Methods

Sampling Region

The IISD Experimental Lakes Area (IISD-ELA), located in Northwestern Ontario, Canada, is a series of 58 small boreal lakes and their watersheds which have been set aside for scientific research since 1968 (https://www.iisd.org/ela/). Being located in a sparsely populated region, the lakes at IISD-ELA are mostly unaffected by anthropogenic influences. For the present study, Lakes 260, 223, 224, and 373 were selected for comparison as they have extensive, long-term databases for abiotic and biological components of the ecosystem (Figure 1). L223 and L224 are connected by a shallow ˜300 m stream flowing from L224 into L223; however, there is no movement of lake trout populations between the lakes (personal communication, Michael D. Rennie, Lakehead University). Furthermore, L223 and L224 are located about 3 km from L260 and about 8 km from L373, with L260 and L373 being about 6 km apart. All four of the lakes are located within the same quaternary watershed (Figure 1). Key hydrologic characteristics for the lakes are listed in Table 1a. In terms of area, L260 is the largest of the four lakes. However, L224 and L373 are about 33% greater in volume and are deeper than L260 and L223. Overall, L224 and L373 are comparable in terms of volume and depth, whereas L260 and L223 contain less volume and are shallower.

Collection of Samples

Epidermal mucus was collected from seven lake trout captured in each of L260, L223, L224, and L373 in late September/early October 2017. Fish were captured from a boat using hook and line and, upon capture, were placed in separate bins of fresh lake water until processing (held < 5 min). Retrieval and post-capture air exposure were minimized to reduce physiological stress (Cooke et al., 2013). Water was renewed for each fish to prevent cross-contamination between individuals. Prior to collecting mucus, lake trout were placed in a solution of 0.1g/L pH buffered (7.0) tricaine methanesulfonate (MS-222) mixed in lake water until fin movement ceased and fish were unresponsive to light pressure on the caudal fin. Total length, fork length, weight, and sex of the individuals were recorded. Sex was determined through external morphological characteristics if possible, or by expressing a small volume of gametes through gentle pressure on the abdomen. Mucus was collected by gently scraping the skin with a stainless steel spatula, avoiding displacement of scales, to aggregate the mucus along the body axis in the anterior to posterior direction. The aggregated mucus was aspirated using a syringe, dispensed into a microcentrifuge tube, placed on crushed dry ice, and stored at -80°C until analyzed. All tools and work area were cleansed with 95% ethanol between fish to prevent cross-contamination of mucus.

Sample collection, cDNA library generation, and Illumina sequencing

Total RNA was extracted from bulk mucus samples ($n = 4$ per lake) using the RNeasy Lipid Tissue Mini Kit (Qiagen, Qiagen, Germantown, MD) following the manufacturer's protocol. Total RNA concentrations were quantified using an Invitrogen Qubit 4 Fluorometer (Invitrogen, Carlsbad, CA) and a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Prior to cDNA library construction, RNA integrity was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

Sequencing cDNA libraries were prepared using the NEBNext? Ultra II Directional RNA Library Prep Kit for Illumina? (New England Biolabs, Ipswich, MA) with the NEBNext(r) rRNA Depletion Kit using 500 ng of total RNA as input. NEBNext(r) Multiplex Oligos for Illumina(r) Index Primer Sets were used to enable multiplexing of cDNA libraries. Final cDNA libraries were assessed on an Agilent 2100 Bioanalyzer to verify proper fragment sizes and to quantify final library concentrations. All cDNA libraries ($n = 16$) were pooled and the pooled sample was assessed on an Agilent 2100 Bioanalyzer prior to sequencing. Pooled, multiplexed libraries underwent 1 x 75 bp sequencing on one lane on an Illumina NextSeq 500 at the Institute for Integrative Genome Biology at the University of California, Riverside.

De novo transcriptome assembly and quality assessment

Quality of raw read data was assessed with FastQC (Andrews, 2010) v.0.11.7 prior to downstream analysis. Removal of adapter sequences and poor quality reads was performed using Trimmomatic (Bolger, Lohse, & Usadel, 2014) v.0.36 with the following parameters: ILLUMINACLIP: TruSeq3-SE.fa:2:30:10 LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:15 MINLEN: 36. Read quality was again assessed with FastQC following trimming. Trimmed read files from all samples were concatenated and used to build a *de novo* transcriptome using Trinity (Haas et al., 2013) v.2.8.4 with in silico read normalization and a default *k-mer* size of 25. Alignment statistics were used to assess accuracy and completeness of the assembly. To further assess completeness of the *de novo*transcriptome, BUSCO (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) v.3.0.2 was used to quantify the percentage of complete, fragmented, and missing single copy orthologs by aligning the transcriptome to highly-conserved single copy orthologs within the 'eukaryota_odb10' BUSCO dataset. As these single copy orthologs are evolutionarily-conserved, it is expected that a well-assembled transcriptome will have a high percentage of complete single copy orthologs matching the BUSCO dataset.

Functional annotation of the de novo transcriptome

Functional annotation of the Trinity assembly was performed using Trinotate (Bryant et al., 2017) v.3.1.1. The assembled putative genes were searched against several databases, including the NCBI non-redundant protein database (http://www.ncbi.nlm.nih.gov/), the UniProtKB/Swiss-Prot database (http://www.ebi.ac.uk/uniprot/), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/), the Gene Ontology (GO) knowledgebase (http://www.geneontology.org/), and the EMBL EggNOG database (http://eggnogdb.embl.de), using NCBI-BLASTx v.2.8.1+ reporting only the top hit (-max_target_seqs 1). Next, TransDecoder v.5.0.2 (http://transdecoder.github.io) was used to extract open read frames (ORFs) from the assembly and to predict which ORFs are likely to be coding sequences. The extracted ORFs were searched against the UniProtKB/Swiss-Prot database using NCBI-BLASTp v.2.8.1+ reporting only the top hit. Using extracted ORFs as input, HMMR (Finn, Clements, & Eddy, 2011) v.3.1b2 was used to identify functional protein domains from the Pfam domain database, TMHMM (Krogh, Karsson, von Heijne, & Sonnhammer, 2001) v.2.0c was used to predict transmembrane regions, and SignalP (Petersen, Brunak, von Heijne, & Nielsen, 2011) v.4.1c was used to predict signal peptide cleavage sites.

Hierarchical Clustering and Principle Component Analysis

Using Bowtie2 (Langmead & Salzberg, 2012) and RSEM (Li & Dewey, 2011) v.1.3.1, trimmed reads for each sample were aligned back to the annotated Trinity assembly and transcript abundances were quantified. In order to visualize transcriptome-wide similarities among samples, unsupervised hierarchical clustering was used to generate a sample-to-sample Euclidian distance heat map using the RSEM count matrix that underwent a regularized-logarithm transformation. Principal component analysis (PCA) was also performed on the transformed RSEM count matrix. Unsupervised hierarchical clustering was used to produce a heat map of the top 500 most-variable transcripts among all samples to further visualize sample similarities. The top 500 most-variable transcripts were extracted from the transformed count matrix using the*genefilter* (Gentleman, Carey, Huber, & Hahne, 2019) R package and were used as input for hierarchical clustering.

Differential Gene Expression and Pathway Analysis

The non-transformed transcript count matrix was input into DESeq2 (Love, Huber, & Anders, 2014) v.1.23.6 to determine differentially expressed transcripts among each lake-by-lake comparison at p [?] 0.05 after Benjamini-Hochberg false discovery rate (FDR) correction. Differentially-expressed transcripts that were annotated were input into ToppGene Suite (Kaimal, Bardes, Tabar, Jegga, & Aronow, 2009) to identify altered KEGG pathways and GO terms among lakes.

qPCR Validation

cDNA was created for L223 and L373 mucus samples using the Promega Reverse Transcription System (Promega, Madison, WI) with 500 ng of total RNA as input. Melt curve analysis and agarose gel elec-

trophoresis were performed to confirm primer specificity. The reactions for qPCR consisted of 10 µL SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA), 0.2 µM forward and reverse primers, and 2 µL of cDNA in a final volume of 20 µL. Samples were run in triplicate for each gene on a CFX Connect Real-Time PCR Detection System (Bio-Rad, Hercules, CA) with the following thermal cycler conditions: 95 °C for 10 min, 40 cycles of 95 °C for 15 s, 58 °C for 30 s, and 72 °C for 30 s. Relative expression among L223 and L373 samples was determined by the 2^{-Ct} method (Livak & Schmittgen, 2001) with β-actin as the normalizing gene. Primer information is listed in Additional File 1: Table S1.

Characterization of lake systems

Lake trout populations in L260, L223, L224, and L373 are sampled in autumn of each year for various parameters including abundance, growth, and condition. Total length and weight for lake trout sampled in 2017 were extracted from the IISD-ELA database. Fulton's condition factor (K) was calculated for each individual using the formula K = 100 x weight/length^3. Statistical differences in average total length, weight, and condition factor among each lake-by-lake comparison were assessed using a one-way analysis of variance (ANOVA) followed by a Tukey's Honest Significant Difference (HSD) test using a significance threshold of adj-p [?] 0.05.

As zooplankton are an important food source for lake trout, zooplankton abundances by species for L260, L223, L224, and L373 in late September/early October 2017 were extracted from the IISD-ELA database. For each lake, total abundance of zooplankton, species richness (no. of species), Shannon Weiner diversity indices (H'), and Shannon equitability indices (E_H) were calculated. The Shannon Weiner diversity index measures biodiversity, accounting for both species richness and evenness, and was calculated using $H' = -[?]p_i \ln p_i$, where p_i is the proportion of total individuals found in species i. The Shannon equitability index measures evenness of the community and was calculated using $E_H = H' / H_{max}$, where H_{max} is ln(species richness). Equitability is a value between 0 and 1, with 1 being complete evenness.

Since 1968, data for water quality, biological variables, and atmospheric conditions have been collected at the IISD-ELA. Data from September and October 2017 were extracted from the database for each target lake, which included the following environmental parameters: temperature (°C), dissolved oxygen (mg/L), conductivity (S/m), pH, alkalinity (µEq/L), chlorophyll a (µg/L), total dissolved nitrogen (µg/L), total dissolved phosphorus (µg/L), suspended carbon (µg/L), suspended phosphorus (µg/L), suspended nitrogen (µg/L), ammoniacal nitrogen (NH_3-N; µg/L), and nitrate-nitrogen (NO_3-N; µg/L). For consistency, values for each parameter were extracted for the epilimnion and metalimnion in each lake and averaged to obtain representative values for each lake. Statistical differences in average water quality parameters among each lake-by-lake comparison were assessed using a one-way ANOVA followed by a Tukey's Honest Significant Difference (HSD) test using a significance threshold of adj-p [?] 0.05. Using the aforementioned variables for each lake, PCA was used to reduce multidimensionality of water quality data and to graphically visualize similarity in lakes based on water quality characteristics. As the water quality parameters possess different units, all values were scaled to a unitless form of zero mean and variance of one prior to analysis in order to make the variables comparable. Results of the PCA were also used to examine the relationships between variables and the magnitude of importance of each variable on the principal components. Unsupervised hierarchical clustering of water quality parameters for each lake was used to further visualize similarities among lakes.

Water quality effects on transcriptome

A linear mixed model was used to assess percentages of transcriptional variance accounted for by selected environmental parameters. The linear mixed model was conducted using the R package *variancePartition*(Hoffman & Schadt, 2016) v.1.17.6 and included lake as a random effect (categorical) as well as dissolved oxygen, conductivity, pH, alkalinity, chlorophyll a, suspended phosphorus, suspended carbon, and total dissolved nitrogen as fixed effects (continuous). Gene expression, or the transcriptome-wide count data, was the response variable of interest. Not all environmental parameters were included in the model due to collinearity of certain variables. For example, total dissolved nitrogen, NO_3-N, and NH_3-N are positively correlated

which may produce misleading results and overestimated the contribution of these variables; thus, covariates were dropped. *variancePartition* fit a linear mixed model that jointly considered the contribution of all variables on the expression of each gene in the normalized transcriptome-wide gene count matrix. Using a multiple regression model, *variancePartition* assessed the effect of each individual variable on gene expression while correcting for all other variables included in the model (see *variancePartition*documentation for further statistical details).

Results

Lake trout sampling

Seven lake trout were collected by hook and line angling from each of the four target lakes in late September/early October of 2017. Weight and total length of collected lake trout were comparable among L223 (703.7 ± 30.27 g, 449.7 ± 6.0 mm), L224 (641.3 ± 54.32 g, 437.6 ± 10.96 mm), L260 (644.3 ± 87.99 g; 425.9 ± 19.77 mm), and L373 (574.4 ± 40.22 g, 420.1 ± 9.55 mm). Full characterization of collected lake trout is available in Additional File 1: Table S2.

Sequencing output and de novo transcriptome assembly

Illumina sequencing generated a total number of 405,026,380 single-end raw reads across the 16 FASTQ files accounting for 80.4 GB of data (Additional File 1: Table S3). A total of 175,910,591 reads remained following trimming of adapters and poor quality reads and removal of rRNA (Additional File 1: Table S3). Average reads per sample after trimming and rRNA removal was 10,994,412 reads. Thus, rRNA was a large proportion of total raw reads, as was observed in a previous RNA sequencing study of epidermal mucus (Greer et al., 2019). The remaining 175 million total reads were used for *de novo* transcriptome assembly. Trinity assembled 268,935 genes with an N50 contig length of 810 bp and an average contig length of 589 bp based on the longest isoform per gene (Additional File 1: Table S4). Of the 268,935 Trinity genes, 89,209 (33%) were annotated by BLASTX (Additional File 1: Table S4). Analysis of transcriptome completeness with BUSCO indicated 67% complete, 28.6% fragmented, and 4.4% missing single copy orthologs, indicating a fair reconstruction of the transcriptome. The incidence of fragmented single copy orthologs is likely due to the fact that RNA in epidermal mucus was partially degraded (RIN < 7) which may also explain the relatively low percentage of transcripts receiving a BLASTX annotation.

Transcriptomic variation among lakes

In order to visualize similarities in transcriptomes among lake trout populations, hierarchical clustering and PCA was conducted on transcriptome-wide count data. Results of sample-to-sample Euclidian distance clustering indicated an approximate clustering of L223 and L373 populations and clustering of L260 and L223 populations within separate clades (Figure S1). Transcriptomic differences among the lakes were further elucidated by the PCA of transcriptome-wide count data, which clearly showed that samples clustered by lake from which they were collected (Figure 2A). Again, L224 and L373 lake trout clustered more closely together, whereas L260 and L223 lake trout clustered apart from all other lakes. The eigenvalues of the two first principal components represented 53% of the total variance (PC1 33%; PC2 20%). There were no patterns based on sex of the lake trout in the PCA results, suggesting transcriptomic patterns were most likely driven by lake (Figure S2).

Hierarchical clustering of the top 500 most-variable transcripts revealed similar results, with individuals clustering into clearly defined clades by lake (Figure 2B). As with the PCA results, L224 and L373 lake trout clustered together, whereas L260 and L223 each clustered apart from the other lakes. These data suggest that individuals from L224 and L373 were most similar in terms of transcriptomic profiles, with L260 and L223 being the most different from the other lakes. Results of the hierarchical clustering and PCA overall suggest that there were clear transcriptomic differences among individuals collected from different lakes and that these differences can be detected nonlethally in epidermal fish mucus.

Differential Gene Expression and Pathway Analysis

There were significant differences in gene expression among the four lakes (Figure 3). The greatest numbers of differentially expressed transcripts were found among the comparisons between L223 and L260 (6,708 transcripts, 1,208 annotated transcripts), L224 and L260 (8,276 transcripts, 1,012 annotated transcripts), L260 and L373 (7,190 transcripts, 879 annotated transcripts), and L223 and L224 (8,123 transcripts, 1,629 annotated transcripts) (Figure 3). The lowest numbers of differentially expressed transcripts were among the comparisons between L223 and L373 (4,659 transcripts; 872 annotated transcripts) and L224 and L373 (4,098 transcripts, 426 annotated transcripts) (Figure 3). Again, the relatively low degree of annotation of transcripts was likely due to the high fragmentation of RNA transcripts in mucus, a result of RNAse present in the mucus (Brinchmann, 2016). A full list of all annotated differentially expressed transcripts for each lake-by-lake comparison may be found in Additional File 1: Table S5. Only transcripts which were annotated were used for further pathway analysis.

Due to the high number of differentially expressed transcripts, pathway analyses were used to understand biological pathways and functions that differ among lake trout populations. A list of all GO molecular functions and biological processes and KEGG pathways for each lake-by-lake comparison can be found in Additional File 1: Table S6. There was a high prevalence of immune-related pathways among the top ten biological process GO terms for most lake-by-lake comparisons (Table 2). Comparing L223 and L224, a notable observation was the presence of pathways related to viral infection within the top ranked GO biological functions, including 'viral process' (rank 11), 'viral gene expression' (rank 8), and 'viral transcription' (rank 9). Furthermore, several KEGG pathways among L223 and L224 were related to pathogens, such as 'Salmonella infection' (rank 9), 'Vibrio choerae infection' (rank 5), 'epithelial cell signaling in Helicobacter pylori infection' (rank 10), bacterial invasion of epithelial cells' (15), and 'shigellosis' (rank 13). Comparing L223 and L260, 'viral process' (rank 2) and 'viral gene expression' (rank 52) were also identified as altered GO biological functions as well as the KEGG pathways 'Salmonella infection' (rank 3), 'Shigellosis' (rank 7), 'hepatitis C' (rank 14), 'viral carcinogenesis' (rank 16), and others (Additional File 1: Table S6). As for the comparison among L233 and L373, several of the top GO biological pathways were related to cell adhesion and cell migration involved in the inflammatory response, such as 'biological adhesion' (rank 1), 'cell adhesion' (rank 2), 'cell-cell adhesion' (rank 4), 'T cell activation' (rank 12), and 'T cell differentiation' (rank 15) (Table 2; Additional File 1: Table S6). Comparison of L223 with all other lakes identified several immune- and infection-related pathways, suggesting L223 lake trout may be exposed to pathogens. Top GO biological functions among L224 and L373, the most transcriptionally-similar lakes, were mainly related to interleukin-1 beta (Table 2). Finally, comparisons of L260 with L224 and L373 both identified a high prevalence of pathways related to cell death and apoptosis among top GO biological pathways, such as 'regulation of cell death' and 'regulation of apoptotic process' (Table 2).

As immune- and pathogen-related pathways were identified within the top GO biological processes among L223 and all other lakes, immune-related transcripts differentially regulated in L223 were visualized based on normalized count data (Figure 4). Many immune- and stress-related transcripts were expressed to a greater degree in L223 lake trout compared to all other populations, among them members of the GTPase of immunity-associated protein family, cytokines, interferons, and heat shock proteins (Figure 4). Furthermore, many transcripts involved in viral transcription were found to be consistently upregulated in L223 relative to all other lakes (Figure 5). Aside from immune transcripts, other notable gene expression patterns included that of cytosolic phospholipase A2 gamma (*PLA2G4C*), which was the top differentially-expressed transcript and strongly upregulated (fold change > 7) in L260 lake trout in comparison with all other lakes. Up-regulator of cell proliferation (*URGCP*) was among the top differentially expressed transcripts in all lake-by-lake comparisons.

qPCR confirmation

qPCR was used to verify the expression of select genes that were differentially expressed in the RNA sequencing analysis. qPCR analysis confirmed the expression patterns that were observed in the RNA sequencing results (Figure S3).

Abiotic and biotic variables of lakes

There were significant differences in average total length and weight among lake trout collected throughout 2017 annual sampling from L260, L223, L224, and L373 (ANOVA, p [?] 0.05) (Table 1b; Figure S4). Average total length and weight of lake trout in Lake 260 were significantly greater than average total length and weight of lake trout in L223, L224, and L373 (Tukey HSD, adj-p [?] 0.0001 for all comparisons). Furthermore, average total length and weight of lake trout in Lake 223 were greater than average total length and weight of lake trout in L224 (Tukey HSD, adj-p [?] 0.002 for both comparisons). There were no significant differences in average total length and weight among L224 and L373 and among L223 and L373 (Tukey HSD, adj-$p>$ 0.05 for all comparisons). Although there were differences in length and weight, average condition factors did not significantly differ among populations (ANOVA, $p = 0.1$) (Table 1b; Figure S4).

Total zooplankton abundance ranked from highest to lowest was as follows: L223 > L260 > L224 > L373 (Table 1c). L223 and L260 also had the highest zooplankton species richness (38 species), whereas species richness was lower in L224 (29 species) and L373 (26 species) (Table 1c). The most abundant zooplankton within all lakes were rotifers (e.g. *Keratella cochlearis* ,*Kellicottia bostoniensis* , *Polyarthra remata* ,*Polyarthra vulgaris* , *Kellicottia longspina*) and copepods (e.g. *cyclopoida nauplii*) (Figure S5). L373 had the highest H' index of 2.50, followed by L260 with 2.08, then L223 with 1.93, and finally L224 with 1.67 (Table 1c). L373 and L260 also had the greatest evenness of zooplankton species with equitability index values of 0.77, followed by L223 and L224 with equitability values of 0.53 and 0.50, respectively. Thus, L373 and L260 contain the most diverse, evenly-distributed zooplankton communities. Abundance of all species across all lakes can be found in Additional File 1: Table S7.

Average water quality parameters are presented in Table 1d and visualized in Figure S6. There were no significant differences in average suspended phosphorus, suspended nitrogen, suspended carbon, and total dissolved phosphorus among lakes (Tukey HSD, adj-$p>$ 0.05 for all comparisons). However, L223 and L260 had significantly higher levels of total dissolved nitrogen than L224 and L373 (Tukey HSD, adj-p [?] 0.0001 for both comparisons). Chlorophyll-a levels were significantly lower in L224 compared to Lakes 223 and 260 (Tukey HSD, adj-$p > 0.03$ for both comparisons), while chlorophyll-a levels did not significantly differ among L223, L260, and L373 (Tukey HSD, adj-$p > 0.05$ for all comparisons). However, there was a visible trend of increased chlorophyll-a in L223 and L260 (Figure S6). Average dissolved oxygen levels in L260 were significantly lower compared to all other lakes (Tukey HSD, adj-p [?] 0.0001 for all comparisons). Average alkalinity was significantly higher in L373 compared to all other lakes (Tukey HSD, adj-p [?] 0.0001 for all comparisons), whereas alkalinity was significantly lower in L224 compared to all other lakes (Tukey HSD, adj-p [?] 0.0001 for all comparisons). Furthermore, average pH was significantly higher in L373 as compared to L224 and L260 (Tukey HSD, adj-p [?] 0.01 for all comparisons) but pH did not significantly differ among L223, L224 and L260 (Tukey HSD, adj-$p>$ 0.05 for all comparisons). Average temperature was comparable among the lakes; however, water temperature of L373 was significantly lower than in L223 and L224 (Tukey HSD, adj-p [?] 0.03 for all comparisons) whereas temperature did not significantly differ among L223, L224, and L260 (Tukey HSD, adj-$p>$ 0.05 for all comparisons). Finally, average conductivity was significantly lower in L224 compared to all other lakes (Tukey HSD, adj-p [?] 0.02 for all comparisons).

PCA of 11 variables for the four lakes was used to identify variation in water quality characteristics. Proportions of variance explained by PC1, PC2, and PC3 were 50.39%, 29.53%, and 20.08%, respectively. PC1 and PC2 explained the maximum amount of total variance (79.92%). Variable loadings on each PC are presented in Additional File 1: Table S8 and are graphically represented in a PCA biplot (Figure 6). PC1 had strong positive loadings from suspended nitrogen (0.42), suspended carbon (0.38), conductivity (0.38), and chlorophyll a (0.36) (Figure 2; Additional File 1: Table S8). The highest scoring lakes for PC1 were L223 (1.18) and L260 (1.68), while the PC1 score was weakest for L373 (0.61) and was strongly negative for L224 (-3.47). Furthermore, the strong negative PC1 score of L224 and strong negative PC1 loading of total dissolved phosphorus (-0.39) suggest that L224 had high total dissolved phosphorus relative to lakes scoring positively on PC1 and that total dissolved phosphorus likely drove the variation of L224 from other lakes.

PC2 had strong positive loadings from pH (0.55), alkalinity (0.42), and dissolved oxygen (0.34) as well as negative loading from chlorophyll-a (-0.24) (Figure 6; Additional File 1: Table S8). L223 was the only lake

with a positive PC2 score (2.49) out of the lakes of interest (Figure 6). Overall, PCA suggested that L260 and L223 were most similar in terms of water quality characteristics while L224 and L373 varied from all other lakes (Figure 6). Hierarchical clustering analysis confirmed that L223 and L260 were most closely related based on water quality characteristics whereas L224 and L373 clustered apart from L223 and L260 (Figure S7).

Relationships of environmental variables to mucus transcriptomic profiles

A linear mixed model was used to quantify the percent variance in transcriptome-wide gene expression that could be explained by water quality variables. The variables included in the model explained a median of 68.1% across all genes in the transcriptome-wide count matrix, leaving a median residual of 31.9% variance unexplained (Figure 7). Conductivity, chlorophyll a, alkalinity, and dissolved oxygen explained the greatest proportion of variance within the transcriptome (Figure 3).

Discussion

Mucus transcriptome differs among lakes

Following exposure to crude oil, Greer et al.(2019) demonstrated that sequencing of mahi-mahi mucus was effectively able to distinguish transcriptomes of oil-exposed and non-exposed individuals. Our results from the present study expand the utility of mucus as a nonlethal sampling tool by demonstrating that sequencing of RNA in mucus can distinguish transcriptomic differences among geographically-distinct populations (Figure 2). In terms of transcriptomic similarity, individuals from L224 and L373 were most similar whereas individuals from L260 and L223 clustered apart from all other lakes (Figure 2). It should be noted that the lakes of interest in the present study are located in close proximity to one another and within the same ecozone and, thus, vary only slightly in terms of ecological conditions. Even so, unique transcriptome profiles by lake were captured within lake trout mucus indicating that sequencing of mucus is able to robustly capture gene expression changes among fish populations. Therefore, analysis of mucus may be a viable tool for ecological studies interested in assessing gene expression differences among fish populations. As mucus sampling is nonlethal and minimally-invasive, this method could be especially useful for assessing threatened and endangered populations without inducing further perturbations. Moreover, differential gene expression analysis and subsequent gene ontology analysis can be used to further elucidate transcriptomic relationships among populations and identify abiotic and biotic factors driving population differences.

Differentially regulated genes and pathways among lakes

Differential gene expression analysis and pathway analysis was used to understand the underlying biological meaning of transcriptional differences among lake trout populations. Among each lake-by-lake comparison, there was a high prevalence of transcripts and gene ontology biological processes related to the immune system, suggesting an active response to stress in lake trout populations (Addition File 1: Table S5; Table S6) (Tort, 2011). Transcripts that were commonly differentially regulated among most lake-by-lake comparisons included members of the GTPase of immunity-associated protein family (*GIMAP2* ,*GIMAP4* , *GIMAP7* , *GIMAP8* , etc.), chemokine transcripts (*CCL2* , *CCL3* , *CCL12* , *CCL20* , etc.), interferon-related transcripts (*IRF1* , *IRF3* ,*IFI44* , etc.), and others (Figure 4). Expression of immune-related genes is not surprising as fish epidermal mucus is full of various immune relevant molecules (Brinchmann, 2016) and differential regulation of immune-related genes in mucus has been identified in previous studies (Greer et al., 2019; Ren et al., 2015). Aside from immune response transcripts, expression of viral RNA was also observed among lake trout populations.

RNA sequencing of epidermal mucus suggested the presence of retroviral infection in the sampled lake trout populations. Presence of retroviruses among lakes was evidenced by the expression of transcripts involved in transcription and replication of viral RNA, including transcripts coding for pol polyprotein, gag-pol polyprotein, replicase polyprotein, RNA-directed RNA polymerase, and RNA-dependent RNA polymerase (Additional File 1: Table S5) (Ahlquist, 2002; Fodor, 2013; Lepa & Siwicki, 2011). Interestingly, the transcript *gag-pol*encodes the gag-pol polyprotein of the Walleye dermal sarcoma virus (WDSV) (UniProt

ID: POL_WDSV), an exogenous retrovirus inducing dermal lesions and sarcomas in walleye (*Sander vitreus*) (Holzschu et al., 1995; K. Xu et al., 2013). Gag, a major structural protein, and pol, the reverse transcriptase, are essential proteins encoded within retroviral genomes and within the WDSV genome specifically (Holzschu et al., 1995; Katz, 1994; Rovnak & Quackenbush, 2010). Elevated expression of viral RNA has been identified in dermal sarcomas of WDSV-infected fish (Poulet, Vogt, Bowser, & Casey, 1995). While WDSV is specific to walleye, only six tumorigenic piscine retroviruses have been fully or partially sequenced and, thus, it is likely that the *gag-pol*transcript detected in the present study is endogenous to a closely related retrovirus that has not yet been sequenced (Quackenbush et al., 2001).

Following infection, cyclins encoded by WDSV induce cells to proliferate abnormally, resulting in lesions and tumors (LaPierre, Casey, & Holzschu, 1998). Interestingly, *URGCP* was consistently among the top differentially regulated transcripts among all lake-by-lake comparisons. *URGCP* stimulates cyclin to accelerate cell growth and promote tumor formation, and overexpression of *URGCP* has been implicated in virus-induced carcinomas in humans (Dodurga et al., 2014; Lale Satiroglu Tufan et al., 2002). As lymphocytes have been shown to aggregate around WSDV dermal lesions, the differential regulation of transcripts involved in lymphocyte development, trafficking, and function (e.g. *GIMAP2* , *GIMAP4* , *GIMAP7* ,*GIMAP8, CCL2* , *CCL3* , *CCL12* , *CCL20* ,*IL1B* , etc.) may also indicate antiviral response to epidermal lesions (Ciucci & Bosselut, 2014; Filen & Lahesmaa, 2010; Reyes-Cerpa et al., 2012). Differential regulation of interferon-related transcripts (*IRF1* , *IRF3* , *IFI44* , etc.) further suggests antiviral response as interferon regulatory factors play a critical role in the antiviral immune response in fish and transcription of interferon regulatory factors, interferons, and interferon-stimulated genes has been shown to be upregulated following viral infection in fish (Collet & Secombes, 2002; Huang et al., 2015; Robertsen, 2018; Yao, Huang, Fan, Kong, & Wang, 2012; Zou & Secombes, 2011). Overall, expression of WDSV-specific *gag-pol* , transcripts involved in cell proliferation, and immune transcripts associated with antiviral response suggest that viral carcinogenesis may be present in the studied lake trout populations.

Although transcripts for viral RNA transcription were differentially regulated among all lake-by-lake comparisons, these transcripts were only consistently upregulated within L223 lake trout, providing compelling evidence that prevalence of viral infection may be greatest within L223 (Figure 5). Comparison of L233 lake trout to all other lakes indicated that *URGCP* was consistently among the top ten transcripts differentially expressed in L223 and was strongly upregulated in L223 lake trout (FC $>$ 5). Along with this, viral-related processes were among the top altered GO biological processes in L223 comparisons, including viral gene expression, viral transcription, and viral process among L223 and L224 and among L223 and L260 (Table 2). While immune response was evident among most lakes, immune-related transcripts were most consistently upregulated in L223 lake trout and upregulation of certain transcripts was unique to L223 (e.g. interleukin-8 (*CXCL8*), interleukin 1 beta (*IL1B*), etc.) (Figure 4).

Aside from immune-specific transcripts, transcripts coding for various heat shock proteins were upregulated in lake trout from L223 compared to all other lakes, including *HSP90AA1* , *hsp90a.1* ,*hsp90ab1* , and *HSPA12A* (Figure 4; Additional File 1: Table S5). In fish, heat shock proteins are expressed in response to a range of biotic and abiotic stressors, such as environmental contaminants, heat and cold shock, food deprivation, and infectious pathogens (Basu et al., 2002; Cara, Aluru, Moyano, & Vijayan, 2005; Iwama, Thomas, Forsyth, & Vijayan, 1998; Roberts, Agius, Saliba, Bossier, & Sung, 2010). Given the evidence for viral transcription in L223 lake trout, the upregulation of heat shock protein transcripts may be in response to pathogens. For example, Hsp90 isoforms have been shown to be upregulated in several fish species following viral infection and bacterial challenge (Y. M. Chen et al., 2010; Wei, Gao, Wang, & Xu, 2013; Xie, Song, Weng, Liu, & Liu, 2015). Upregulation of Hsp90 alpha and beta proteins has also been observed in epidermal mucus of sea-lice infected Atlantic salmon (*Salmo salar* L.) (Provan et al., 2013). Along with this, Hsp90 is one of the most frequently observed host chaperone for viruses, aiding in synthesis, localization, and folding of viral proteins (Geller, Taguwa, & Frydman, 2012; Hu & Seeger, 1996; Kampmueller & Miller, 2005; Nagy, Wang, Pogany, Hafren, & Makinen, 2011). As dependence of viruses on Hsp90 appears to be nearly universal (Geller et al., 2012), the upregulation of Hsp90 isoforms in L223 may further suggest viral response. However, these results are not conclusive as Hsp90 is responsive to multiple different stressors. For

example, the expression of heat shock proteins may also be due to heat stress, as L223 is relatively shallow and the available optimal habitat for lake trout is limited during the stratified season; this phenomenon is reflected in the fact that L223 contained the greatest water temperatures, though this difference was not significantly different from L224 and L260 water temperatures (Table 1). Heat stress coupled with a higher load of pathogens may explain this together, though further research is needed to elucidate the upregulation of heat shock proteins in L223 lake trout.

Discussion of *PLA2G4C* , encoding cytosolic phospholipase A2 gamma (cPLA2γ), is warranted as it was consistently among the top upregulated gene in L260. While the function of cPLA2γ specifically has not been characterized in fish, it is known that cPLA2γ a lipolytic enzyme that catalyzes the release of arachidonic acid from membrane phospholipids in humans (Murakami, Masuda, & Kudo, 2003). In fish, arachidonic acid is a precursor for a wide spectrum of eicosanoids involved in hemodynamic regulation, immune and inflammatory responses, reproduction, renal and neural functions (Tocher, 2003). It has been suggested that cPLA2γ specifically is involved in the inflammatory response, as *PLA2G4C* was upregulated in rodents following pathogen infection and also has been shown to contribute to formation, replication, and assembly of hepatitis C virus (Brown et al., 2008; S. Xu et al., 2012). Due to the number of biological processes affected by arachidonic acid release, further work will be required to identify the downstream pathways associated with *PLA2G4C* upregulation in lake trout from L260.

Abiotic and biotic variable comparison among lakes

Lake trout from L260 had the greatest average total length and weight, followed by lake trout from L223. Although there were significant differences in average length and weight among lake trout populations, condition factor did not significantly differ suggesting a similar degree of health among populations. In terms of zooplankton populations, L260 and L223 contained the most abundant and species rich populations. As zooplankton represent a significant source of food for lake trout, the greatest abundance of zooplankton in L260 and L223 may explain the trend of larger individuals residing in those lakes. Moreover, L373 and L260 contained the most diverse, evenly-distributed zooplankton communities as suggested by Shannon diversity and equitability indices. Diversity and equitability were lower in L223 likely because the zooplankton community here was dominated by rotifer species (*Keratella cochlearis* , *Kellicottia bostoniensis* ,*Polyarthra remata* , *Polyarthra vulgaris* , *Kellicottia longspina*) (Figure S5). High rotifer dominance in lakes has shown to increase bacterial abundance through top-down control of flagellates (Fermani et al., 2013) and, thus, the observed pathogen responses in L223 lake trout may potentially be explained by greater rotifer abundance and, consequently, bacterial abundance. However, more research is needed to establish a causative link among rotifer and bacterial abundances at IISD-ELA.

Average water quality parameters were input into PCA to elucidate similarities and differences among the four lakes based on their water quality characteristics. L260 and L223 were most similar in terms of water quality parameters, whereas L373 and L224 deviated from all other lakes (Figure 6). L260 and L223 had high scores on PC1, which received highest positive loadings from suspended nitrogen, suspended carbon, conductivity, and chlorophyll a. As nitrogen and carbon are important for zooplankton productivity (Brett, Arhonditsis, Chandra, & Kainz, 2012; Loick-Wilde et al., 2016) and chlorophyll-a is an indicator of productivity in aquatic environments, PC1 represents lakes with high productivity. The high positive PC1 scores for L223 and L260 suggest they had higher productivity relative to the other lakes, which is further supported by the fact that L260 and L223 had the greatest zooplankton, chlorophyll-a, and total dissolved nitrogen levels (Table 1). PC1 scores were weakest for L373 and were strongly negative for L224, suggesting L373 and L224 were less productive than L223 and L260. Lake 224 scored strongly on PC2 which had strong positive loadings from pH, alkalinity, and dissolved oxygen (Figure 6). These PCA results further demonstrate that Lake 224 was lower productivity as high pH levels limit phytoplankton growth (C. Y. Chen & Durbin, 1994). Furthermore, the positive loading of dissolved oxygen and negative loading of chlorophyll-a on PC2 is indicative of oligotrophic conditions.

Overall, results of the PCA suggest that L260 and L223 were most similar in terms of water quality characteristics likely driven by their higher productivity (Figure 6). Moreover, L224 and L373 varied from all

other lakes, with variability likely being driven by more alkaline conditions in L373 and higher total dissolved phosphorus content in L224 (Figure 6). Unlike L260 and L223, water quality parameters and PCA results suggest that L373 and L224 were lower productivity lakes. Hierarchical clustering analysis confirmed that L223 and L260 were most similar based on water quality whereas L224 and L373 clustered apart from L223 and L260 (Figure S7). The divergence of L224 and L373 from the other lakes is likely primarily a function of water residence time which is driven by volume and catchment area. Thus, the smaller catchment area:volume ratio of L224 and L373 is likely a driver of the lower productivity of these lakes.

Relationships between environmental variables and mucus transcriptomic profiles

As the transcriptome responds to environmental changes, differences and similarities in environmental variables may explain mucus transcriptome variation among lake trout populations (Oomen & Hutchings, 2017). The high degree of transcriptomic similarity among L224 and L373 populations could potentially be explained by the fact that both lakes were fairly oligotrophic whereas L260 and L223, while still oligotrophic, tended to be more productive. However, comparison of results from the water quality PCA and transcriptome PCA suggests there are environmental factors not accounted for in the study. For example, while L260 and L373 were most similar in their water quality characteristics (Figure 6), lake trout from those lakes deviated from each other in terms of their transcriptomic profiles (Figure 2). Therefore, it is likely that there were other abiotic and biotic variables influencing gene expression changes other than those taken into account in the water quality PCA, such as pathogen loading, epigenetics, or food availability.

While PCA was used to characterize similarity in water quality profiles among lakes, a linear mixed model was used to quantify the percent variance in transcriptome-wide gene expression explained by those water quality variables. As the linear mixed model accounted for 68.1% of transcriptional variance, results suggest that variation in water quality contributes to transcriptional differences among lake trout populations (Figure 7). Conductivity, chlorophyll a, alkalinity, and dissolved oxygen explained the greatest proportion of variance within the transcriptome, suggesting these variables had the greatest influence on gene expression differences among lakes (Figure 7). Variables identified by the linear mixed model as accounting for the most transcriptomic variation among lake trout populations were also among the variables identified in the PCA as driving inter-lake water quality differences. However, a median of 31.9% of transcriptomic variance was not explained by the water quality variables in the model. Therefore, both the water quality PCA and linear mixed model suggest that the variables considered in the present study did not fully explain transcriptional variance and, thus, other variables could be driving transcriptomic differences among populations albeit similar water quality conditions. As suggested by differential gene expression data, an additional source of transcriptional variance among populations may stem from presence of pathogens. For example, despite their fairly similar water quality profiles, the divergence in transcriptomic profiles of L223 and L260 lake trout may be partially explained by the strong immune response in L223. Overall, it is challenging to robustly quantify the effects of environmental variables on gene expression patterns as transcription may be affected by even small changes in environment (Alvarez et al., 2015).

Aside from environmental factors, epigenetic modifications are another potential source of gene expression variation among populations, with studies showing that genes may be transgenerationally dysregulated following exposure to stressors in prior generations. Transgenerational alterations in the transcriptome are well-illustrated in a study which examined responses to increased temperature across generations in common reef fish (*Acanthochromis polyacanthus*) (Veilleux et al., 2015). Compared to controls, fish that were transgenerationally exposed to higher temperatures showed upregulation of metabolic, immune, and stress-responsive genes, likely an adaptive mechanism to maintain performance and cope with higher temperatures. Epigenetic modifications are highly possible in lake trout populations at IISD-ELA as L260 and L223 have been used for whole-lake experiments in the past. In an effort to understand population-level impacts of environmental estrogens, L260 was subject to whole-lake ethynylestradiol additions for three years starting in 2001, resulting in feminization and a near extirpation of fathead minnow (*Pimephales promelas*) as well as population declines of lake trout (Kidd et al., 2007; Palace et al., 2009). To simulate the effects of acid precipitation on freshwater lakes, sulfuric acid was added to L223 over a three year period from 1976 to 1978

(Schindler, Wagemann, Cook, Ruszczynski, & Prokopowich, 1980), resulting in disappearance of fathead minnow, increased lake trout embryonic mortality and deformities, and decreased survival of lake trout due to emaciation (Kennedy, 1980; Mills, Chalanchuk, Mohr, & Davies, 1987; Nero & Schindler, 1983; Schindler & Turner, 1982). Interestingly, L224 and L373, which have not been subject to experimental modifications, had highly similar transcriptomic profiles, whereas the experimentally-modified L260 and L223 diverge from all other lakes (Figure 2); these observations suggest that epigenetic modifications in response to past stressful conditions may contribute to the observed transcriptional variations.

As the transcriptome reflects both short-term plastic responses and transgenerational plasticity (Oomen & Hutchings, 2017), it is difficult to adequately discern whether transcriptional variation among populations stems from environmental changes or transgenerational alterations. Although concrete identification of explanatory variables requires further development, RNA sequencing of mucus provides meaningful information on transcriptomic variation among fish populations and can be used to identify molecular differences driving variation in a nonlethal manner.

Application of method

The results of the present study demonstrate that RNA sequencing of epidermal mucus can generate a wealth of information on molecular differences among wild fish populations without the use of lethal sampling. Although the populations considered in the present study were located in closely-situated, ecologically-similar lakes, our method was still able to clearly separate individual transcriptomic profiles by lake. Sequencing of mucus could be useful in comparative transcriptomic studies of populations residing in ecologically different conditions, such as studying ecologically-divergent populations residing in two different habitat types or comparing a reference population to one at a polluted site. Moreover, using mucus for gene expression analysis can be applied to various freshwater and marine species as mRNA-based mucus analysis has been successful in mahi-mahi (Greer et al., 2019), channel catfish (Ren et al., 2015), largemouth bass (unpublished), and now lake trout. As mucus sampling is nonlethal, transcriptomic studies of mucus would be of most value to studies concerned with threatened and endangered fish species. mRNA-based analysis of mucus can be used in conjunction with other nonlethal sampling methods, such as blood sampling for biomarkers of interest, as well as individual- and population-level metrics (e.g. through eDNA sampling (Adams et al., 2019)) to gain insight on health at multiple levels of biological organization.

Data Accessibility

Raw read data has been deposited in FASTQ format in the NCBI Sequence Read Archive database under the BioProject ID PRJNA665724.

Author Contributions

N.E.A. contributed to study design, collected samples, conducted bioinformatics and statistical analyses, and wrote the manuscript. V.P. contributed to study design, collected samples, provided field support, and edited the manuscript. L.H. collected samples, provided field support, and edited the manuscript. D.S. contributed to study design, provided field support, and edited the manuscript. All authors contributed final refinements to the manuscript.

Acknowledgements

Thank you Sonya Michaleski and Lauren Hayhurst for providing field support.

Figures and Tables

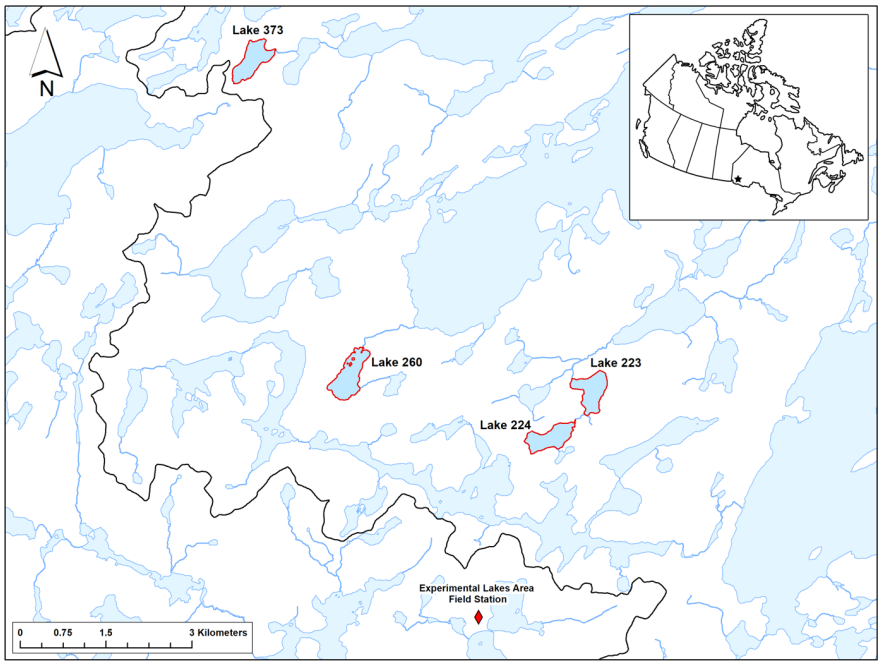

Figure 1. Map of the IISD Experimental Lakes Area, located in northwestern Ontario, Canada, indicating the locations of the four lakes used in the present study. Black line denotes quaternary watershed boundaries. Data source: Ontario Integrated Hydrology Data (Ontario Ministry of Natural Resources and Forestry - Provincial Mapping Unit).

Table 1. For each lake, (a) hydrological characteristics, (b) average metrics for lake trout collected in 2017 (± standard error of mean), (c) zooplankton population metrics, including total abundance, richness, Shannon Weiner diversity index (H'), and Shannon's equitability (E_H), and (d) average water quality parameters (± standard error of mean) (TDN = total dissolved nitrogen, TDP = total dissolved phosphorus).

		Lake 260	**Lake 223**	**Lake 224**	**Lake 373**
(a)					
	Area (m^2)	332,328	268,951	261,450	273,653
	Volume (m^3)	1,975,971	1,891,984	3,066,672	3,107,474
	Maximum Depth (m)	15.7	14.4	27.3	21.2
	Average Depth (m)	5.9	7.0	11.7	11.4
(b)					
	Total Length (mm)	472.1 ± 2.7	436.8 ± 1.5	426.0 ± 2.2	420.1 ± 9.5
	Weight (g)	845.4 ± 13.2	662.2 ± 7.1	612.4 ± 10.2	574.4 ± 40.2
	Condition Factor	0.7911 ± 0.0038	0.7894 ± 0.0041	0.7797 ± 0.0042	0.7603 ± 0.0176
	n	193	248	189	7
(c)					
	Abundance (count/L)	174.43	224.80	145.31	75.89
	Richness (# species)	38	38	29	26
	H' Index	2.08	1.93	1.67	2.50
	E_H	0.77	0.53	0.50	0.77
(d)					
	Αλκαλινιτψ (μΕχ/Λ)	133.1 ± 2.16	126.6 ± 0.08	106.7 ± 1.57	174.6 ± 1.25
	Conductivity (S/m)	17.86 ± 0.86	21.32 ± 2.60	12.22 ± 0.12	18.97 ± 1.37

	Lake 260	Lake 223	Lake 224	Lake 373
pH	6.82 ± 0.08	7.02 ± 0.06	6.971 ± 0.06	7.22 ± 0.03
Dissolved Oxygen (mg/L)	8.04 ± 0.32	11.31 ± 0.33	10.73 ± 0.25	10.78 ± 0.26
Temperature (°C)	11.15 ± 0.53	12.85 ± 0.59	11.92 ± 0.43	10.21 ± 0.45
ἠλοροπηψλλ-α (μγ/Λ)	3.82 ± 1.01	4.08 ± 1.36	1.52 ± 0.15	2.22 ± 0.22
Συσπενδεδ ἀρβον (μγ/Λ)	421.1 ± 19.38	453.8 ± 43.01	364.1 ± 21.94	442.3 ± 22.04
Συσπενδεδ Νιτρογεν (μγ/Λ)	44.65 ± 3.32	44.15 ± 4.96	33.46 ± 3.56	40.8 ± 4.76
Συσπενδεδ Πηοσπηορυς (μγ/Λ)	3.5 ± 0.29	2.75 ± 0.25	2.6 ± 0.16	3.13 ± 0.23
ΤΔΠ (μγ/Λ)	2.38 ± 0.21	2.25 ± 0.09	3.22 ± 0.40	2.11 ± 0.10
ΤΔΝ (μγ/Λ)	246.3 ± 5.31	230 ± 1.96	198.7 ± 2.82	199.5 ± 3.61
NO_3-N (μγ/Λ)	1.75 ± 0.85	1.75 ± 0.48	1.4 ± 0.31	0.5 ± 0.33
NH_3-N (μγ/Λ)	0.5 ± 0.29	6.75 ± 0.75	4.7 ± 0.52	1.13 ± 0.74

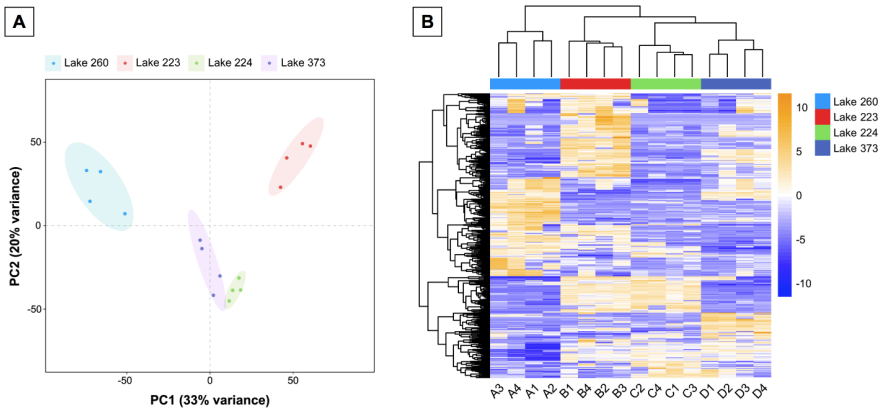

Figure 2. A) Principal component analysis of transcriptome-wide count data with individual lake trout colored by lake and B) unsupervised hierarchical clustering of the top 500 most variable transcripts in epidermal mucus of lake trout collected from Lakes 260, 223, 224, and 373 at the IISD Experimental Lakes Area.

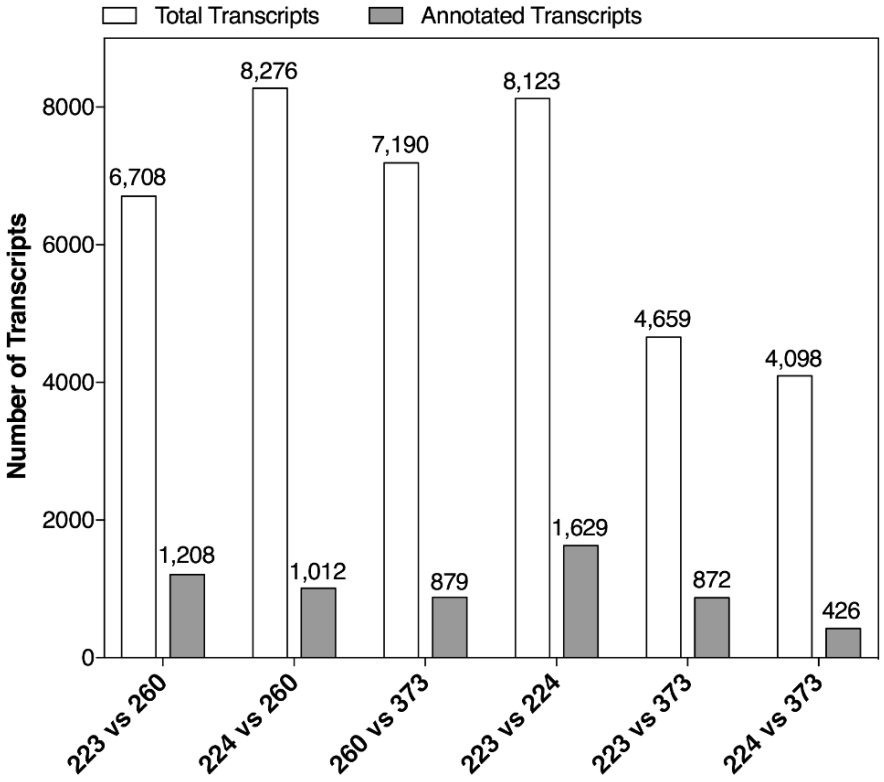

Figure 3. The number of differentially regulated transcripts identified by RNA sequencing in epidermal mucus collected from lake trout populations residing in different lakes. The total number of differentially regulated transcripts is presented for each lake-by-lake comparison as well as the subset of transcripts which were functionally annotated by BLASTX.

Table 2. The top ten gene ontology biological processes among each lake-by-lake comparison, including the FDR adjusted-p values and number of transcripts associated with the pathway.

Gene Ontology ID	**Pathway Name**	**FDR adj-p**	**# Tran**
L223 vs L224	**L223 vs L224**	**L223 vs L224**	**L223 vs**
GO:0006613	cotranslational protein targeting to membrane	1.58E-45	61
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	2.21E-44	59
GO:0045047	protein targeting to ER	3.52E-40	59
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	3.52E-40	59
GO:0072599	establishment of protein localization to endoplasmic reticulum	3.20E-39	59
GO:0070972	protein localization to endoplasmic reticulum	4.68E-37	62
GO:0006413	translational initiation	1.34E-36	70
GO:0019080	viral gene expression	1.35E-33	69
GO:0019083	viral transcription	7.79E-33	66
GO:0006612	protein targeting to membrane	1.41E-32	68
L223 vs L373	**L223 vs L373**	**L223 vs L373**	**L223 vs**
GO:0022610	biological adhesion	2.56E-11	94
GO:0007155	cell adhesion	2.67E-11	93
GO:0007010	cytoskeleton organization	1.31E-08	92
GO:0098609	cell-cell adhesion	2.87E-07	59
GO:0000902	cell morphogenesis	1.18E-06	68

GO:0043085	positive regulation of catalytic activity	1.18E-06	84
GO:0032989	cellular component morphogenesis	3.04E-06	71
GO:0034330	cell junction organization	3.04E-06	29
GO:0010256	endomembrane system organization	3.73E-06	42
GO:0001775	cell activation	3.89E-06	80
L223 vs L260	**L223 vs L260**	**L223 vs L260**	**L223 vs**
GO:0044403	symbiotic process	3.83E-15	93
GO:0016032	viral process	9.66E-15	88
GO:0044419	interspecies interaction between organisms	1.31E-14	94
GO:0010941	regulation of cell death	9.35E-12	142
GO:0016050	vesicle organization	2.82E-11	131
GO:0007010	cytoskeleton organization	4.53E-11	124
GO:0072657	protein localization to membrane	5.69E-11	69
GO:0034097	response to cytokine	9.23E-11	103
GO:0042981	regulation of apoptotic process	1.48E-10	128
GO:0043067	regulation of programmed cell death	2.18E-10	129
L224 vs L373	**L224 vs L373**	**L224 vs L373**	**L224 vs**
GO:0032691	negative regulation of interleukin-1 beta production	5.96E-04	7
GO:0032611	interleukin-1 beta production	5.96E-04	10
GO:0032692	negative regulation of interleukin-1 production	8.33E-04	7
GO:0032651	regulation of interleukin-1 beta production	1.00E-03	9
GO:0032612	interleukin-1 production	1.00E-03	10
GO:0032652	regulation of interleukin-1 production	2.78E-03	9
GO:0019226	transmission of nerve impulse	4.29E-03	8
GO:0000904	cell morphogenesis involved in differentiation	5.43E-03	25
GO:0050713	negative regulation of interleukin-1 beta secretion	7.60E-03	4
GO:0007626	locomotory behavior	9.31E-03	12
L260 vs L224	**L260 vs L224**	**L260 vs L224**	**L260 vs**
GO:0034097	response to cytokine	2.84E-08	82
GO:0071345	cellular response to cytokine stimulus	7.39E-06	71
GO:0007623	circadian rhythm	2.24E-05	25
GO:0061919	process utilizing autophagic mechanism	2.24E-05	40
GO:0006914	autophagy	2.24E-05	40
GO:0010941	regulation of cell death	2.65E-05	98
GO:0007010	cytoskeleton organization	3.55E-05	86
GO:0010942	positive regulation of cell death	5.03E-05	53
GO:0031331	positive regulation of cellular catabolic process	6.44E-05	32
GO:0048511	rhythmic process	1.03E-04	29
L260 vs 373	**L260 vs 373**	**L260 vs 373**	**L260 vs**
GO:0042981	regulation of apoptotic process	8.77E-08	94
GO:0010941	regulation of cell death	8.77E-08	100
GO:0043067	regulation of programmed cell death	8.77E-08	94
GO:0007009	plasma membrane organization	8.77E-08	27
GO:0046903	secretion	1.12E-07	94
GO:0044403	symbiotic process	2.64E-07	58
GO:0070201	regulation of establishment of protein localization	3.74E-07	56
GO:0016032	viral process	3.78E-07	55
GO:0070268	cornification	4.22E-07	18
GO:0044419	interspecies interaction between organisms	4.22E-07	59

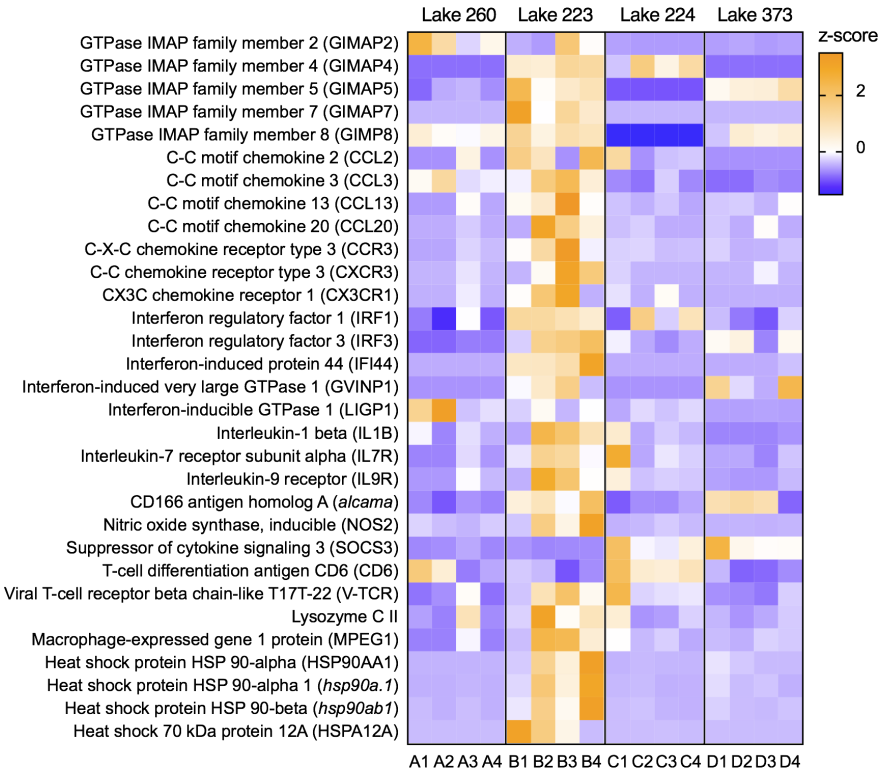

Figure 4. Relative expression of immune-related transcripts in epidermal mucus of lake trout collected from Lakes 260, 223, 224, and 373 at the IISD Experimental Lakes Area. Expression data are presented as the normalized count data, centered and scaled, from each individual in order to provide direct comparison of gene expression levels due to lack of a control group. Thus, orange represents higher count level of a particular transcript, whereas blue suggests lower count level. Transcripts shown were differentially-expressed in at least one lake-by-lake comparison.

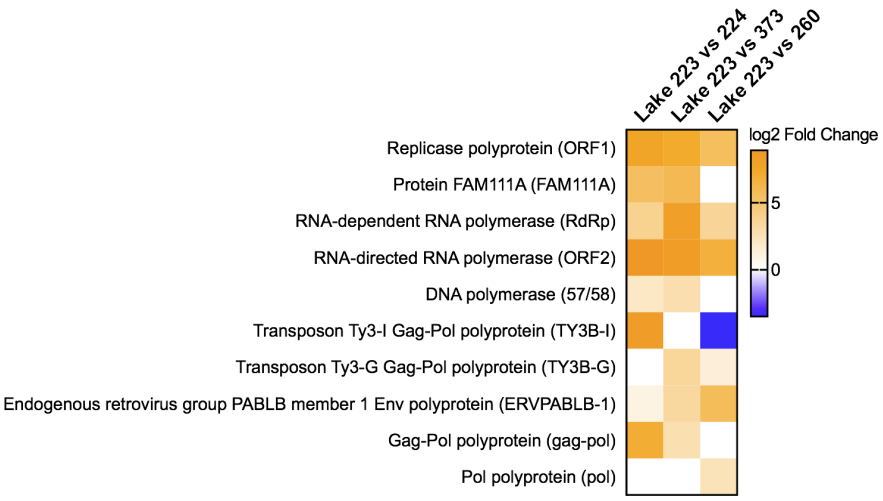

Figure 5. Differential expression of transcripts involved in viral gene transcription in lake trout from Lake 223 compared to Lakes 224, 373, and 260.

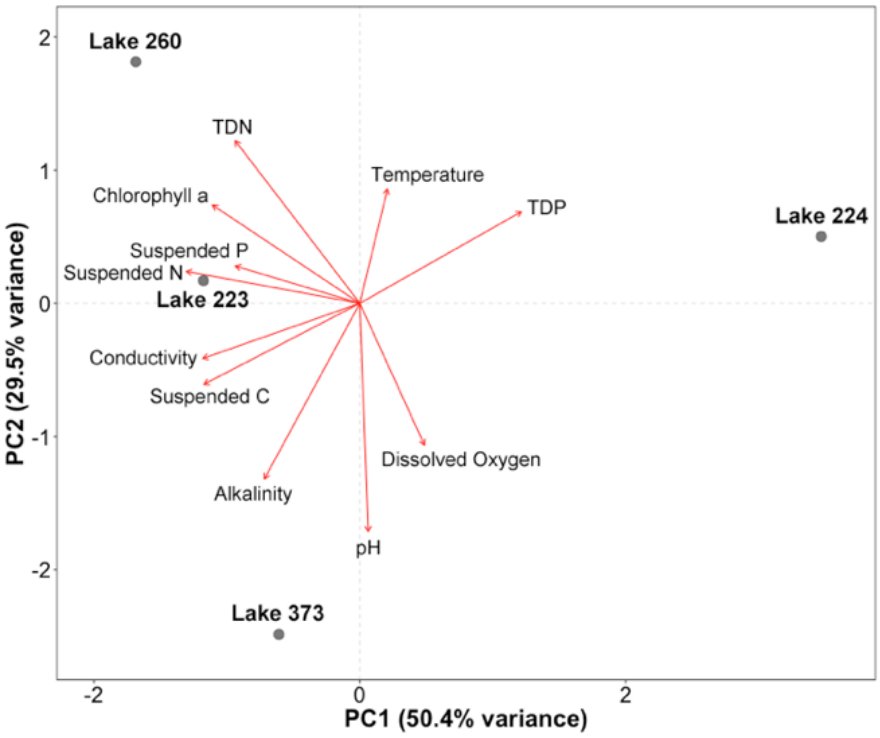

Figure 6. Principal component analysis of average water quality parameters for Lakes 260, 223, 224, and 373 with arrows indicating the loadings of each water quality variable.

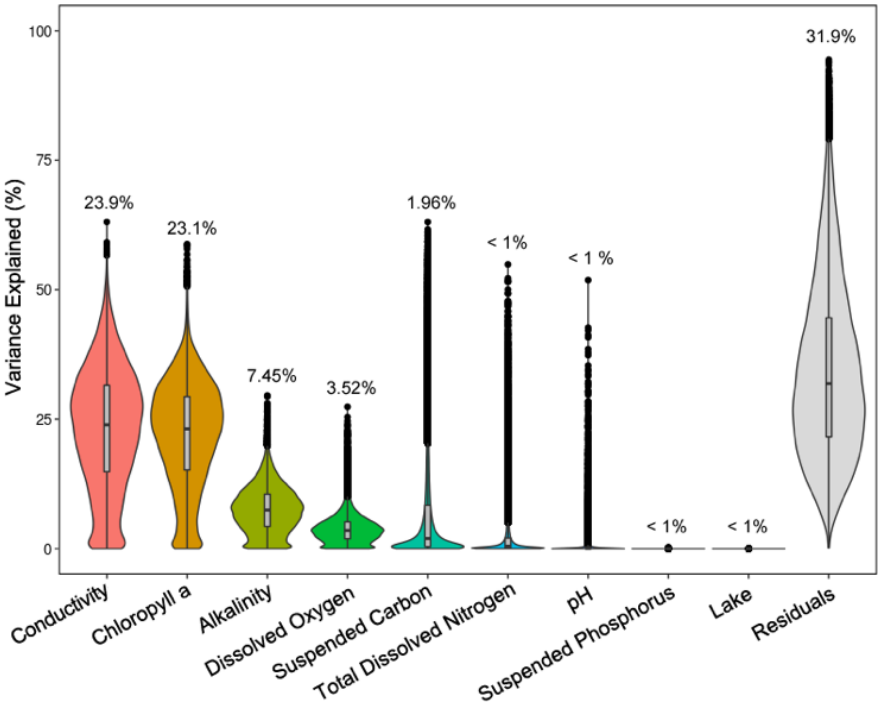

Figure 7. Results of the linear mixed model indicating the median variance explained by water quality

parameters across all genes within the lake trout transcriptome-wide read count data. Each variable explains a certain percent of variation across all genes after correcting for all other variables.

References

Adams, C. I. M., Knapp, M., Gemmell, N. J., Jeunen, G.-J., Bunce, M., Lamare, M. D., & Taylor, H. R. (2019). Beyond Biodiversity: Can Environmental DNA (eDNA) Cut It as a Population Genetics Tool? *Genes* , *10* (3), 192. https://doi.org/10.3390/genes10030192

Ahlquist, P. (2002). RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing. *Science* , *296* (5571), 1270–1273. Retrieved from http://science.sciencemag.org/

Alvarez, M., Schrey, A. W., & Richards, C. L. (2015). Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Molecular Ecology* , *24* (4), 710–725. https://doi.org/10.1111/mec.13055

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Baker, R. F., Blanchfield, P. J., Paterson, M. J., Flett, R. J., & Wesson, L. (2004). Evaluation of Nonlethal Methods for the Analysis of Mercury in Fish Tissue. *Transactions of the American Fisheries Society* , *133* (3), 568–576. https://doi.org/10.1577/t03-012.1

Basu, N., Todgham, A. E., Ackerman, P. A., Bibeau, M. R., Nakano, K., Schulte, P. M., & Iwama, G. K. (2002). Heat shock protein genes and their functional significance in fish. *Gene* , *295* (2), 173–183. https://doi.org/10.1016/S0378-1119(02)00687-X

Bennett, R. H., Ellender, B. R., Mäkinen, T., Miya, T., Pattrick, P., Wasserman, R. J., … Weyl, O. L. F. (2016). Ethical considerations for field research on fishes. *Koedoe* , *58* (1), 1–15. https://doi.org/10.4102/koedoe.v58i1.1353

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* , btu170.

Brett, M. T., Arhonditsis, G. B., Chandra, S., & Kainz, M. J. (2012). Mass flux calculations show strong Allochthonous support of freshwater Zooplankton production is unlikely. *PLoS ONE* , *7* (6), 1–9. https://doi.org/10.1371/journal.pone.0039508

Brinchmann, M. F. (2016). Immune relevant molecules identified in the skin mucus of fish using -omics technologies. *Mol. BioSyst.* ,*12* (7), 2056–2063. https://doi.org/10.1039/C5MB00890E

Brown, J. K., Knight, P. A., Thornton, E. M., Pate, J. A., Coonrod, S., Miller, H. R. P., & Pemberton, A. D. (2008). Trichinella spiralis induces de novo expression of group IVC phospholipase A2 in the intestinal epithelium. *International Journal for Parasitology* ,*38* (2), 143–147. https://doi.org/10.1016/j.ijpara.2007.10.002

Brumbaugh, W. G., Schmitt, C. J., & May, T. W. (2005). Concentrations of cadmium, lead, and zinc in fish from mining-influenced waters of northeastern Oklahoma: Sampling of blood, carcass, and liver for aquatic biomonitoring. *Archives of Environmental Contamination and Toxicology* , *49* (1), 76–88. https://doi.org/10.1007/s00244-004-0172-3

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., … Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports* , *18* (3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Bulloch, P., Schur, S., Muthumuni, D., Xia, Z., Johnson, W., Chu, M., … Tomy, G. T. (2020). F2-isoprostanes in Fish mucus: A new, non-invasive method for analyzing a biomarker of oxidative stress. *Chemosphere* , *239* .

Cara, J. B., Aluru, N., Moyano, F. J., & Vijayan, M. M. (2005). Food-deprivation induces HSP70 and HSP90 protein expression in larval gilthead sea bream and rainbow trout. *Comparative Biochemistry and Physiology*

- *Part B: Biochemistry and Molecular Biology* , *142* (4), 426–431.

Chen, C. Y., & Durbin, E. G. (1994). Effects of pH on the growth and carbon uptake of marine phytoplankton. *Marine Ecology Progress Series* , *109* (1), 83–94. https://doi.org/10.3354/meps109083

Chen, Y. M., Kuo, C. E., Wang, T. Y., Shie, P. S., Wang, W. C., Huang, S. L., ... Chen, T. Y. (2010). Cloning of an orange-spotted grouper Epinephelus coioides heat shock protein 90AB (HSP90AB) and characterization of its expression in response to nodavirus. *Fish and Shellfish Immunology* , *28* (5–6), 895–904. https://doi.org/10.1016/j.fsi.2010.02.004

Ciucci, T., & Bosselut, R. (2014). Gimap and T cells: a matter of life or death. *European Journal of Immunology* , *44* (2), 348–351.

Collet, B., & Secombes, C. J. (2002). Type I-interferon signalling in fish. *Fish and Shellfish Immunology* , *12* (5), 389–397. https://doi.org/10.1006/fsim.2001.0405

Cooke, S. J., Donaldson, M. R., O'connor, C. M., Raby, G. D., Arlinghaus, R., Danylchuk, A. J., ... Suski, C. D. (2013). The physiological consequences of catch-and-release angling: Perspectives on experimental design, interpretation, extrapolation and relevance to stakeholders. *Fisheries Management and Ecology* , *20* (2–3), 268–287. https://doi.org/10.1111/j.1365-2400.2012.00867.x

Cordero, H., Brinchmann, M. F., Cuesta, A., Meseguer, J., & Esteban, M. A. (2015). Skin mucus proteome map of European sea bass (Dicentrarchus labrax). *Proteomics* , *15* (23–24), 4007–4020. https://doi.org/10.1002/pmic.201500120

Cordero, H., Morcillo, P., Cuesta, A., Brinchmann, M. F., & Esteban, M. A. (2016). Differential proteome profile of skin mucus of gilthead seabream (Sparus aurata) after probiotic intake and/or overcrowding stress. *Journal of Proteomics* , *132* , 41–50. https://doi.org/10.1016/j.jprot.2015.11.017

Dodurga, Y., Gundogdu, G., Tekin, V., Koc, T., Satiroglu-Tufan, N. L., Bagci, G., & Kucukatay, V. (2014). Valproic acid inhibits the proliferation of SHSY5Y neuroblastoma cancer cells by downregulating URG4/URGCP and CCND1 gene expression. *Molecular Biology Reports* ,*41* (7), 4595–4599. https://doi.org/10.1007/s11033-014-3330-3

Dzul-Caamal, R., Salazar-Coria, L., Olivares-Rubio, H. F., Rocha-Gómez, M. A., Girón-Pérez, M. I., & Vega-López, A. (2016). Oxidative stress response in the skin mucus layer of Goodea gracilis (Hubbs and Turner, 1939) exposed to crude oil: A non-invasive approach. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology* . https://doi.org/10.1016/j.cbpa.2016.05.008

Easy, R. H., & Ross, N. W. (2009). Changes in Atlantic salmon (Salmo salar) epidermal mucus protein composition profiles following infection with sea lice (Lepeophtheirus salmonis). *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics* . https://doi.org/10.1016/j.cbd.2009.02.001

Fermani, P., Diovisalvi, N., Torremorell, A., Lagomarsino, L., Zagarese, H. E., & Unrein, F. (2013). The microbial food web structure of a hypertrophic warm-temperate shallow lake, as affected by contrasting zooplankton assemblages. *Hydrobiologia* , *714* (1), 115–130. https://doi.org/10.1007/s10750-013-1528-3

Filen, S., & Lahesmaa, R. (2010). GIMAP Proteins in T-Lymphocytes.*Journal of Signal Transduction* , *2010* , 268589. https://doi.org/10.1155/2010/268589

Fincel, M. J., Vandehey, J. A., & Chipps, S. R. (2012). Non-lethal sampling of walleye for stable isotope analysis: A comparison of three tissues. *Fisheries Management and Ecology* , *19* (4), 283–292. https://doi.org/10.1111/j.1365-2400.2011.00830.x

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* , *39* , W29–W37.

Fodor, E. (2013). The RNA polymerase of influenza A virus: mechanisms of viral transcription and replication. *Acta Virologica* , *57* , 113–122. https://doi.org/10.4149/av

Geller, R., Taguwa, S., & Frydman, J. (2012). Broad action of Hsp90 as host chaperone required for viral replication. *Biochimica et Biophysica Acta* , *1823* (3), 698–706. https://doi.org/10.1097/MCA.0000000000000178.Endothelial

Gentleman, R., Carey, V., Huber, W., & Hahne, F. (2019). genefilter: methods for filtering genes from high-throughput experiments. *R Package Version 1.68.0* .

Greer, J. B., Andrzejczyk, N. E., Mager, E. M., Stieglitz, J. D., Benetti, D., Grosell, M., & Schlenk, D. (2019). Whole-Transcriptome Sequencing of Epidermal Mucus as a Novel Method for Oil Exposure Assessment in Juvenile Mahi-Mahi (Coryphaena hippurus).*Environmental Science & Technology Letters* , *6* , 538–544. https://doi.org/10.1021/acs.estlett.9b00479

Guardiola, F. A., Cuesta, A., Arizcun, M., Meseguer, J., & Esteban, M. A. (2014). Comparative skin mucus and serum humoral defence mechanisms in the teleost gilthead seabream (Sparus aurata). *Fish and Shellfish Immunology* , *36* (2), 545–551. https://doi.org/10.1016/j.fsi.2014.01.001

Guardiola, F. A., Dioguardi, M., Parisi, M. G., Trapani, M. R., Meseguer, J., Cuesta, A., ... Esteban, M. A. (2015). Evaluation of waterborne exposure to heavy metals in innate immune defences present on skin mucus of gilthead seabream (Sparus aurata). *Fish and Shellfish Immunology* . https://doi.org/10.1016/j.fsi.2015.02.010

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* , *8* (8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Henderson, T. F., Stevens, S. Y., & Lee, C. J. (2016). Assessing the suitability of a non-lethal biopsy punch for sampling fish muscle tissue. *Fish Physiol Biochem* , *42* , 1521–1526. https://doi.org/10.1007/s10695-016-0237-z

Hoffman, G. E., & Schadt, E. E. (2016). variancePartition: Interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* , *17* . https://doi.org/10.1186/s12859-016-1323-z

Holzschu, D. L., Martineau, D., Fodor, S. K., Vogt, V. M., Bowser, P. R., & Casey, J. W. (1995). Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus.*Journal of Virology* , *69* (9), 5320–5331. https://doi.org/10.1128/jvi.69.9.5320-5331.1995

Hu, J., & Seeger, C. (1996). Hsp90 is required for the activity of a hepatitis B virus reverse transcriptase. *Proceedings of the National Academy of Sciences of the United States of America* ,*93* (3), 1060–1064. https://doi.org/10.1073/pnas.93.3.1060

Huang, Y., Huang, X., Cai, J., OuYang, Z., Wei, S., Wei, J., & Qin, Q. (2015). Identification of orange-spotted grouper (Epinephelus coioides) interferon regulatory factor 3 involved in antiviral immune response against fish RNA virus. *Fish and Shellfish Immunology* ,*42* (2), 345–352. https://doi.org/10.1016/j.fsi.2014.11.025

Imbery, J. J., Buday, C., Miliano, R. C., Shang, D., Round, J. M., Kwok, H., ... Helbing, C. C. (2019). Evaluation of Gene Bioindicators in the Liver and Caudal Fin of Juvenile Pacific Coho Salmon in Response to Low Sulfur Marine Diesel Seawater-Accommodated Fraction Exposure.*Environmental Science and Technology* , *53* (3), 1627–1638. research-article. https://doi.org/10.1021/acs.est.8b05429

Iwama, G. K., Thomas, P. T., Forsyth, R. B., & Vijayan, M. M. (1998). Heat shock protein expression in fish. *Reviews in F* , *8* , 35–56.

Kaimal, V., Bardes, E., Tabar, S., Jegga, A., & Aronow, B. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* , *37* , W305–W311. https://doi.org/https://doi.org/10.1093/nar/gkp427

Kampmueller, K. M., & Miller, D. J. (2005). The Cellular Chaperone Heat Shock Protein 90 Facilitates Flock House Virus RNA Replication in Drosophila Cells. *Journal of Virology* , *79* (11), 6827–6837. https://doi.org/10.1128/jvi.79.11.6827-6837.2005

Katz, R. (1994). The Retroviral Enzymes. *Annual Review of Biochemistry* , *63* (1), 133–173. https://doi.org/10.1146/annurev.biochem.63.1.133

Kennedy, L. A. (1980). Teratogenesis in lake trout (Salvelinus namaycush) in an experimentally acidified lake. *Canadian Journal of Fisheries and Aquatic Sciences* , *37* (12), 2355–2358. https://doi.org/10.1139/f80-282

Kidd, K. A., Blanchfield, P. J., Mills, K. H., Palace, V. P., Evans, R. E., Lazorchak, J. M., & Flick, R. W. (2007). Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences of the United States of America* ,*104* (21), 8897–8901. https://doi.org/10.1073/pnas.0609568104

Krogh, A., Karsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* , *305* (3), 567–580.

Lake, J. L., Ryba, S. A., Serbst, J. R., & Libby, A. D. (2006). Mercury in fish scales as an assessment method for predicting muscle tissue mercury concentrations in largemouth bass. *Archives of Environmental Contamination and Toxicology* , *50* (4), 539–544. https://doi.org/10.1007/s00244-005-5052-y

Lale Satiroglu Tufan, N., Lian, Z., Liu, J., Pan, J., Arbuthnot, P., Kew, M., ... Feitelson, M. A. (2002). Hepatitis Bx antigen stimulates expression of a novel cellular gene, URG4, that promotes hepatocellular growth and survival. *Neoplasia* , *4* (4), 355–368. https://doi.org/10.1038/sj.neo.7900241

Langmead, B., & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* , *9* (4), 357–359.

LaPierre, L. A., Casey, J. W., & Holzschu, D. L. (1998). Walleye Retroviruses Associated with Skin Tumors and Hyperplasias Encode Cyclin D Homologs. *Journal of Virology* , *72* (11), 8765–8771. https://doi.org/10.1128/jvi.72.11.8765-8771.1998

Lepa, A., & Siwicki, A. K. (2011). Retroviruses of wild and cultured fish. *Polish Journal of Veterinary Sciences* . https://doi.org/10.2478/V10181-011-0106-8

Li, B., & Dewey, C. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* , *12* (323). https://doi.org/10.1186/1471-2105-12-323

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCT Method.*Methods* , *25* (4), 402–408. https://doi.org/https://doi.org/10.1006/meth.2001.1262

Loick-Wilde, N., Weber, S. C., Conroy, B. J., Capone, D. G., Coles, V. J., Medeiros, P. M., ... Montoya, J. P. (2016). Nitrogen sources and net growth efficiency of zooplankton in three Amazon River plume food webs. *Limnology and Oceanography* , *61* (2), 460–481. https://doi.org/10.1002/lno.10227

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* , *15* (550). https://doi.org/https://doi.org/10.1186/s13059-014-0550-8

Meier, K., Hansen, M. M., Normandeau, E., Mensberg, K. L. D., Frydenberg, J., Larsen, P. F., ... Bernatchez, L. (2014). Local adaptation at the transcriptome level in brown trout: Evidence from early life history temperature genomic reaction norms. *PLoS ONE* ,*9* (1). https://doi.org/10.1371/journal.pone.0085171

Mills, K. H., Chalanchuk, S. M., Mohr, L. C., & Davies, I. J. (1987). Responses of fish populations in Lake 223 to 8 years of experimental acidification. *Canadian Journal of Fisheries and Aquatic Sciences* , *44* (SUPPL. 1), 114–125. https://doi.org/10.1139/f87-287

Murakami, M., Masuda, S., & Kudo, I. (2003). Arachidonate release and prostaglandin production by group IVC phospholipase A2 (cytosolic phospholipase A2γ). *Biochemical Journal* , *372* (3), 695–702. https://doi.org/10.1042/BJ20030061

Nagy, P. D., Wang, R. Y., Pogany, J., Hafren, A., & Makinen, K. (2011). Emerging picture of host chaperone and cyclophilin roles in RNA virus replication. *Virology* . https://doi.org/10.1016/j.virol.2010.12.061

Nero, R. W., & Schindler, D. W. (1983). Decline of Mysis relicla During the Acidification of Lake 223. *Canadian Journal of Fisheries and Aquatic Sciences* , *40* , 1095–1911.

Oleksiak, M. F., Churchill, G. A., & Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nature Genetics* , *32* (2), 261–266. https://doi.org/10.1038/ng983

Oomen, R. A., & Hutchings, J. A. (2017). Transcriptomic responses to environmental change in fishes: Insights from RNA sequencing.*Facets* , *2* (2), 610–641. https://doi.org/10.1139/facets-2017-0015

Palace, V. P., Evans, R. E., Wautier, K. G., Mills, K. H., Blanchfield, P. J., Park, B. J., ... Kidd, K. A. (2009). Interspecies differences in biochemical, histopathological, and population responses in four wild fish species exposed to ethynylestradiol added to a whole lake. *Canadian Journal of Fisheries and Aquatic Sciences* ,*66* (11), 1920–1935. https://doi.org/10.1139/F09-125

Petersen, T., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions.*Nature Methods* , *8* (10), 785–786.

Poulet, F. M., Vogt, V. M., Bowser, P. R., & Casey, J. W. (1995). In situ hybridization and immunohistochemical study of walleye dermal sarcoma virus (WDSV) nucleic acids and proteins in spontaneous sarcomas of adult walleyes (Stizostedion vitreum). *Veterinary Pathology* ,*32* (2), 162–172. https://doi.org/10.1177/030098589503200210

Provan, F., Jensen, L. B., Uleberg, K. E., Larssen, E., Rajalahti, T., Mullins, J., & Obach, A. (2013). Proteomic analysis of epidermal mucus from sea lice-infected Atlantic salmon, Salmo salar L. *Journal of Fish Diseases* , *36* (3), 311–321. https://doi.org/10.1111/jfd.12064

Quackenbush, S. L., Rovnak, J., Casey, R. N., Paul, T. A., Bowser, P. R., Sutton, C., & Casey, J. W. (2001). Genetic Relationship of Tumor-Associated Piscine Retroviruses. *Marine Biotechnology* ,*3* (0), S088–S099. https://doi.org/10.1007/s10126-01-0030-5

Rajan, B., Lokesh, J., Kiron, V., & Brinchmann, M. F. (2013). Differentially expressed proteins in the skin mucus of Atlantic cod (Gadus morhua) upon natural infection with Vibrio anguillarum. *BMC Veterinary Research* , *9* . https://doi.org/10.1186/1746-6148-9-103

Razavi, N. R., Halfman, J. D., Cushman, S. F., Massey, T., Beutner, R., Foust, J., ... Cleckner, L. B. (2019). Mercury concentrations in fish and invertebrates of the Finger Lakes in central New York, USA.*Ecotoxicology* . https://doi.org/10.1007/s10646-019-02132-z

Ren, Y., Zhao, H., Su, B., Peatman, E., & Li, C. (2015). Expression profiling analysis of immune-related genes in channel catfish (Ictalurus punctatus) skin mucus following Flavobacterium columnare challenge.*Fish and Shellfish Immunology* , *46* (2), 537–542. https://doi.org/10.1016/j.fsi.2015.07.021

Reverter, M., Tapissier-Bontemps, N., Lecchini, D., Banaigs, B., & Sasal, P. (2018). Biological and Ecological Roles of External Fish Mucus: A Review. *Fishes* , *3* (4), 41. https://doi.org/10.3390/fishes3040041

Reyes-Cerpa, S., Maisey, K., Reyes-Lopez, F., Toro-Ascuy, D., Sandino, A. ., & Imarai, M. (2012). Fish Cytokines and Immune Response. In H. Türker (Ed.), *New Advances and Contributions to Fish Biology* . InTechOpen. https://doi.org/http://dx.doi.org/10.5772/53504

Richards, C. L., Rosas, U., Banta, J., Bhambhra, N., & Purugganan, M. D. (2012). Genome-wide patterns of Arabidopsis gene expression in nature. *PLoS Genetics* , *8* (4). https://doi.org/10.1371/journal.pgen.1002662

Roberts, R. J., Agius, C., Saliba, C., Bossier, P., & Sung, Y. Y. (2010). Heat shock proteins (chaperones) in fish and shellfish and their potential role in relation to fish health: A review. *Journal of Fish Diseases* , *33* (10), 789–801. https://doi.org/10.1111/j.1365-2761.2010.01183.x

Robertsen, B. (2018). The role of type I interferons in innate and adaptive immunity against viruses in Atlantic salmon.*Developmental and Comparative Immunology* . https://doi.org/10.1016/j.dci.2017.02.005

Rodríguez-Jorquera, I. A., Colli-Dula, R. C., Kroll, K., Jayasinghe, B. S., Parachu Marco, M. V., Silva-Sanchez, C., ... Denslow, N. D. (2019). Blood Transcriptomics Analysis of Fish Exposed to Perfluoro Alkyls Substances: Assessment of a Non-Lethal Sampling Technique for Advancing Aquatic Toxicology Research. *Environmental Science and Technology* , *53* (3), 1441–1452. https://doi.org/10.1021/acs.est.8b03603

Rovnak, J., & Quackenbush, S. L. (2010). Walleye dermal sarcoma virus: Molecular biology and oncogenesis. *Viruses* , *2* (9), 1984–1999. https://doi.org/10.3390/v2091984

Sanderson, B. L., Tran, C. D., Coe, H. J., Pelekis, V., Steel, E. A., & Reichert, W. L. (2009). Nonlethal Sampling of Fish Caudal Fins Yields Valuable Stable Isotope Data for Threatened and Endangered Fishes.*Transactions of the American Fisheries Society* , *138* (5), 1166–1177. https://doi.org/10.1577/t08-086.1

Schindler, D. W., & Turner, M. A. (1982). Biological, Chemical and Physical Responses of Lakes to Experimental Acidification. In H. C. Martin (Ed.), *Long-Range Transport of Airborne Pollutants* (pp. 259–271). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-009-7966-6_19

Schindler, D. W., Wagemann, R., Cook, R. B., Ruszczynski, T., & Prokopowich, J. (1980). Experimental Acidification of Lake 223, Experimental Lakes Area: Background Data and the First Three Years of Acidification. *Canadian Journal of Fisheries and Aquatic Sciences* , *37* (3), 342–354. https://doi.org/10.1139/f80-048

Shephard, K. L. (1994). Functions for fish mucus. *Reviews in Fish Biology and Fisheries* , *4* (4), 401–429. https://doi.org/10.1007/BF00042888

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V, & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ,*31* (19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sloman, K. A., Bouyoucos, I. A., Brooks, E. J., & Sneddon, L. U. (2019). Ethical considerations in fish research. *Journal of Fish Biology* , *94* (4), 556–577. https://doi.org/10.1111/jfb.13946

Tocher, D. R. (2003). Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science* , *11* (2), 107–184. https://doi.org/10.1080/713610925

Tort, L. (2011). Stress and immune modulation in fish.*Developmental and Comparative Immunology* , *35* (12), 1366–1375. https://doi.org/10.1016/j.dci.2011.07.002

Veilleux, H. D., Ryu, T., Donelson, J. M., Van Herwerden, L., Seridi, L., Ghosheh, Y., ... Munday, P. L. (2015). Molecular processes of transgenerational acclimation to a warming ocean. *Nature Climate Change* , *5* (12), 1074–1078. https://doi.org/10.1038/nclimate2724

Wei, T., Gao, Y., Wang, R., & Xu, T. (2013). A heat shock protein 90 β isoform involved in immune response to bacteria challenge and heat shock from Miichthys miiuy. *Fish and Shellfish Immunology* ,*35* (2), 429–437.

Whitehead, A., Dubansky, B., Bodinier, C., Garcia, T. I., Miles, S., Pilley, C., ... Galvez, F. (2012). Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes.*Proceedings of the National Academy of Sciences of the United States of America* , *109* (50), 20298–20302. https://doi.org/10.1073/pnas.1109545108

Xie, Y., Song, L., Weng, Z., Liu, S., & Liu, Z. (2015). Hsp90, Hsp60 and sHsp families of heat shock protein genes in channel catfish and their expression after bacterial infections. *Fish and Shellfish Immunology* , *44* (2), 642–651.

Xu, K., Zhang, T. T., Wang, L., Zhang, C. F., Zhang, L., Ma, L. X., ... Zhang, Z. Y. (2013). Walleye dermal sarcoma virus: Expression of a full-length clone or the rv-cyclin (orf a) gene is cytopathic to the host and human tumor cells. *Molecular Biology Reports* ,*40* (2), 1451–1461. https://doi.org/10.1007/s11033-012-2188-5

Xu, S., Pei, R., Guo, M., Han, Q., Lai, J., Wang, Y., ... Chen, X. (2012). Cytosolic Phospholipase A2 Gamma Is Involved in Hepatitis C Virus Replication and Assembly. *Journal of Virology* ,*86* (23), 13025–13037. https://doi.org/10.1128/jvi.01785-12

Yao, C. L., Huang, X. N., Fan, Z., Kong, P., & Wang, Z. Y. (2012). Cloning and expression analysis of interferon regulatory factor (IRF) 3 and 7 in large yellow croaker, Larimichthys crocea. *Fish and Shellfish Immunology* . https://doi.org/10.1016/j.fsi.2012.02.015

Zou, J., & Secombes, C. J. (2011). Teleost fish interferons and their role in immunity. *Developmental and Comparative Immunology* ,*35* (12), 1376–1387. https://doi.org/10.1016/j.dci.2011.07.001