

# Tissue-specific transcriptomes reveal potential mechanisms of microbiome heterogeneity in an ancient fish

Matt J. Thorstensen <sup>\*</sup>, Alyssa M. Weinrauch, William S. Bugg, Ken M. Jeffries and W. Gary Anderson

Department of Biological Sciences, University of Manitoba, 212B Biological Sciences Building, 50 Sifton Road, Winnipeg, MB R3T 2N2, Canada

\*Corresponding author: Tel: +204 999 6765; Fax: +204 474 7604; Email: [matt.thorstensen@gmail.com](mailto:matt.thorstensen@gmail.com)

Citation details: Thorstensen, M.J., Weinrauch, A.M., Bugg, W.S. *et al.* Tissue-specific transcriptomes reveal potential mechanisms of microbiome heterogeneity in an ancient fish. *Database* (2023) Vol. 2023: article ID baad055; DOI: <https://doi.org/10.1093/database/baad055>

## Abstract

The lake sturgeon (*Acipenser fulvescens*) is an ancient, octoploid fish faced with conservation challenges across its range in North America, but a lack of genomic resources has hindered molecular research in the species. To support such research, we created a transcriptomic database from 13 tissues: brain, esophagus, gill, head kidney, heart, white muscle, liver, glandular stomach, muscular stomach, anterior intestine, pyloric cecum, spiral valve and rectum. The transcriptomes for each tissue were sequenced and assembled individually from a mean of 98.3 million ( $\pm 38.9$  million SD) reads each. In addition, an overall transcriptome was assembled and annotated with all data used for each tissue-specific transcriptome. All assembled transcriptomes and their annotations were made publicly available as a scientific resource. The non-gut transcriptomes provide important resources for many research avenues. However, we focused our analysis on messenger ribonucleic acid (mRNA) observations in the gut because the gut represents a compartmentalized organ system with compartmentalized functions, and seven of the sequenced tissues were from each of these portions. These gut-specific analyses were used to probe evidence of microbiome regulation by studying heterogeneity in microbial genes and genera identified from mRNA annotations. Gene set enrichment analyses were used to reveal the presence of photoperiod and circadian-related transcripts in the pyloric cecum, which may support periodicity in lake sturgeon digestion. Similar analyses were used to identify different types of innate immune regulation across the gut, while analyses of unique transcripts annotated to microbes revealed heterogeneous genera and genes among different gut tissues. The present results provide a scientific resource and information about the mechanisms of compartmentalized function across gut tissues in a phylogenetically ancient vertebrate.

Database URL: [https://figshare.com/projects/Lake\\_Sturgeon\\_Transcriptomes/133143](https://figshare.com/projects/Lake_Sturgeon_Transcriptomes/133143)

## Introduction

The lake sturgeon (*Acipenser fulvescens*) is an octoploid, ancient fish with conservation challenges across its range in North America (1). Molecular resources for lake sturgeon can thus support wide-ranging research questions about fundamental biology relevant to its conservation. However, such research has been hampered by the limited molecular resources available for studying the species. While a microsatellite panel and genotyping by sequencing have been used for population genetic research (2–5), microsatellites are not as informative as genomic approaches for individual genotype information and may miss patterns of admixture and hierarchical structure (6, 7). Sequencing resources such as reference transcriptomes or a well-annotated genome would thus enable more thorough molecular research, but the lack of sequence data for some species has complicated the development of new assays for research on stress responses. While some work has been conducted using specific primers developed to assay messenger ribonucleic acid (mRNA) abundance

in the species (8–10), the lack of a publicly available lake sturgeon transcriptome and genome has hindered molecular physiology and environmental deoxyribonucleic acid (DNA) work (11).

Furthermore, the early divergence of sturgeons from other teleosts (12, 13) makes representative species such as lake sturgeon useful for studying questions about vertebrate evolution. For example, the pyloric cecum was first studied by Aristotle, who hypothesized about storage, fermentation and digestive functions and ceca in fish digestive tracts, which were then later determined to increase the gut surface area for digestion and absorption (14, 15). Sturgeons represent the first evolutionary appearance of fused ceca with increased surface area (16, 17), making these fish an important group for understanding the evolution of vertebrate digestive organs and function. An important caveat is that the presence and function of the lake sturgeon pyloric cecum should not be viewed as a basal state for Actinopterygii, given that positive selection in certain genes has been observed in other sturgeons

Received 13 October 2022; Revised 16 June 2023; Accepted 23 July 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

and paddlefishes and has potentially occurred in lake sturgeon as well (18–20). Therefore, while the organ cannot be used as a basal state against which to compare teleosts because it is inappropriate to *a priori* rule out positive selection in pyloric cecum-related genes in the lake sturgeon, it can be used as a representative tissue to study the evolution of vertebrate digestion (21).

Alternatively, the lake sturgeon may be useful for studying vertebrate digestion from a whole-organism perspective, as gut microbiomes have been described in a wide variety of organisms, including insects, fishes and humans (22–25). The lake sturgeon microbiome has been linked to its physiological state, providing evidence for host–microbe interactions (26–28). With mRNA sequencing, nearly all mRNA transcripts from a tissue can be assembled and annotated regardless of species. The genus of a given transcript annotation can be inferred by using taxonomic information from transcriptome annotations—even if that genus is within bacteria or archaeobacteria. Therefore, RNA sequencing in the lake sturgeon can be used to study gut microbiome heterogeneity, with implications for the evolution of gut microbiome regulation across vertebrates. While the community structure of the microbiome is heavily influenced by environmental factors (29), the hypothetical presence of heterogeneity in genera and genes in the microbial community across the lake sturgeon gut would be consistent with the presence of tissue-specific microbial regulatory mechanisms. One caveat is that microbiome community assembly can be stochastic based on colonization events (30). Nevertheless, fish may have been the first group of vertebrate microbiome hosts to evolve an innate capacity for microbiome regulation (29). Ancient bony fish such as the lake sturgeon are thus valuable for studying the dynamics between host and microbiome.

Multi-tissue and tissue-specific approaches to transcriptome assembly allow for more systematic and in-depth analyses than typical transcriptomics, which may use one tissue for more focused investigations. For example, tissue-specific analyses revealed specialization in tissues and stages in cell division, photosynthesis, auxin transport, stress responses and secondary metabolism in the tomato (*Solanum pimpinellifolium*) (31). In a livebearing fish, *Poeciliopsis prolifica*, multiple tissues were used with RNA-seq data to investigate placental evolution, where the abundance of clusters of transcripts was associated with different tissues (32). In the Atlantic salmon (*Salmo salar*), a blood-specific transcriptome was compared to other tissue transcriptomes to identify genes and gene ontology (GO) terms unique to blood (33). Thus, transcriptome assembly with multiple tissues would provide a stronger resource for molecular research than single-tissue approaches.

One concern about transcriptomics in the lake sturgeon is that assembly may be affected by the octoploid status of the species (1). For instance, in situations where a transcriptome must be assembled without a reference genome, polyploid species are vulnerable to homeologs and distinct sequences with high similarity, such as repetitive sequences, which decrease the accuracy of the final assembly (34). One solution is to create transcriptomes specific to different conditions that may isolate different gene isoforms (34), a strategy consistent with the benefits of a multi-tissue, tissue-specific approach. In addition, different transcriptome assembly programs had varying success at accurately assembling polyploid transcriptomes, and careful selection of an assembly program such as

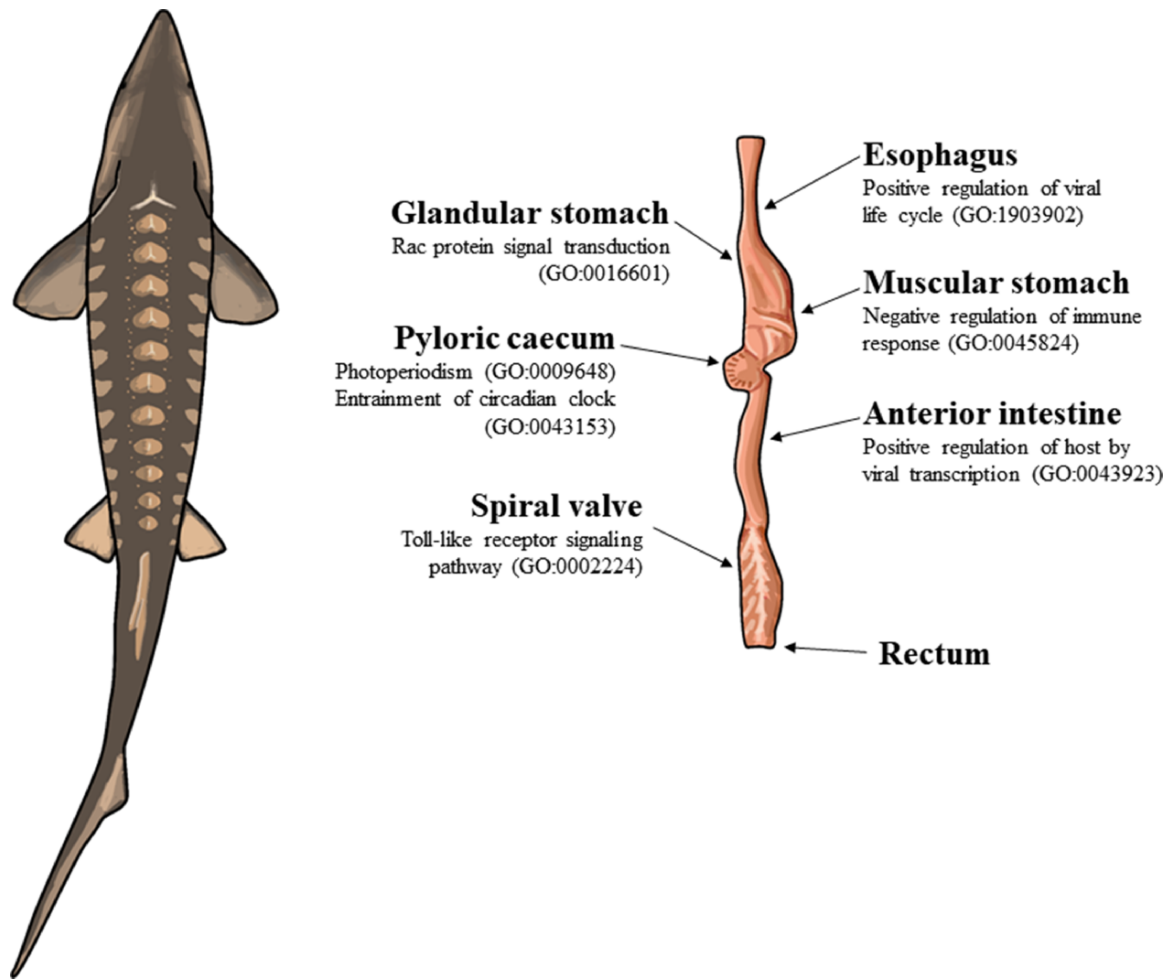
Trinity can at least partly address challenges introduced by high ploidies (34–37).

In this study, we assembled, annotated and analyzed the transcriptomes of 13 tissues in the lake sturgeon, sequenced with short-read mRNA sequencing (i.e. Illumina). The tissues were the brain, esophagus, gill, head kidney, heart, white muscle, liver, glandular stomach, muscular stomach, anterior intestine, pyloric cecum, spiral valve and rectum (Figure 1). In addition, all data used to assemble each of the 13 tissue-specific transcriptomes were assembled and annotated as an overall transcriptome. All transcriptomes and annotations were made publicly available for use as a scientific resource ([https://figshare.com/projects/Lake\\_Sturgeon\\_Transcriptomes/133143](https://figshare.com/projects/Lake_Sturgeon_Transcriptomes/133143)). The brain, gill, head kidney, heart, white muscle and liver transcriptomes have broad potential for studying many aspects of sturgeon biology. However, among the tissue-specific transcriptomes analyzed, the gut tissue transcriptomes represent components of an organ system with region-specific functions throughout (17). Because the seven gut tissue transcriptomes were the major anatomically distinct regions of the lake sturgeon gut, these transcriptomes were used in more focused analyses and discussed in greater detail. Transcriptome annotations were analyzed in two ways: an exploratory approach using GO terms (38) to identify unexpected transcript presence within and among tissues and a guided approach informed by prior knowledge, where we used the lake sturgeon as a representative ancient fish to investigate different facets of vertebrate digestion. We focused on the pyloric cecum because the tissue first appears in sturgeons in terms of increased surface area for digestion and absorption (15, 16). In addition, the overall patterns of biological processes such as immune regulation across the gut and signatures of the lake sturgeon microbiome along different gut tissues were investigated. We observed circadian rhythm genes in the pyloric cecum, various types of innate immune regulation across gut tissues and heterogeneity of bacteria- and archaea-associated transcripts in the lake sturgeon gut.

## Materials and methods

### Sampling and sequencing

Lake sturgeon of ~2 years old, ~100 g in size and unknown sex were sampled haphazardly from recirculating holding tanks and euthanized with an overdose of buffered tricaine methanesulfonate solution (MS-222; Sigma-Aldrich). They were the F<sub>1</sub> crosses from wild individuals sampled for gametes from the Winnipeg River in Manitoba, Canada. Thus, the fish used in the present study spent their entire lives in a hatchery facility. These individuals were fed once per day, and feeding had last been done 18–20 hours prior to sampling. Lake sturgeon sampled for gut tissue were fed on commercial trout pellets (EWOS Pacific), while those sampled for other tissues were fed on bloodworms. Sampling was conducted with aeration provided and at ~15°C to minimize stress. Tissues from each of the gill, liver, brain, head kidney, white muscle and heart were extracted from the sturgeon and stored in RNAlater (Thermo Fisher), and the PureLink RNA Mini Kit (Thermo Fisher) was used for RNA extractions following the manufacturer's protocols. By contrast, gut samples (esophagus, glandular stomach, muscular stomach, anterior intestine, pyloric cecum, spiral valve and rectum) were rinsed with 1X phosphate buffered saline to remove the remaining food content and were immediately placed in TRIzol and extracted



**Figure 1.** The illustration of a lake sturgeon (*Acipenser fulvescens*) and the gut tissues used for transcriptome assemblies in the present study. Beneath most gut tissues are representative, significant ( $q < 0.05$ ), GO terms unique to the tissue identified with enrichR. The GO terms present in the esophagus, glandular stomach, muscular stomach, anterior intestine and spiral valve represent possible innate immune system processes specific to each gut tissue in the present transcriptomes. The GO terms present in the pyloric caecum were processes related to circadian rhythms, unique to the tissue among the transcriptomes analyzed.

the same day following the manufacturer's protocol (Thermo Fisher) to limit RNA degradation that can be associated with digestive enzymes and gut bacteria naturally present in the tissues (39). For all tissues, equal amounts of RNA were pooled from three fish for sequencing. While pooled sequencing is less robust than individual sequencing, it is an effective strategy for balancing cost and statistical power in an experiment (40). Animal rearing and sampling were conducted following guidelines established by the Canadian Council for Animal Care and approved by the Animal Care Committee at the University of Manitoba under Protocol #F15-007.

Total RNA was sent to the Centre d'expertise et de services Génome Québec, Montreal, Quebec (<http://gqinnovationcenter.com>), where 250 ng of total RNA per tissue was used with the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs). Stranded complementary DNA libraries were created with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Fish were sequenced for 100 base pair paired-end reads on one lane of a NovaSeq 6000 (Illumina). A mean of 98.3 million ( $\pm 38.9$  million SD) reads were sequenced for each tissue (Table 1).

### Transcriptome assembly and annotation

Trinity v2.14.0 was used for transcriptome assembly (37, 41), while Trinotate v3.2.2 was used for transcriptome annotation (42–48). The databases used with Trinotate were blastp v2.10.0+ and blastx v2.10.0+ with Uniprot release 2022\_02, HMMER 3.3.2, SignalP v4.1f and TMHMM v2.0c (42–46). Except where specified, default options were used for all bioinformatic steps. Both Trinity and Trinotate were used on mRNA sequencing data from each tissue separately as tissue-specific transcriptomes and with all data from the 13 tissues as an overall transcriptome. Trinity was used for its effectiveness at assembling polyploid genomes (34). Transcriptome annotations were filtered for transcripts with  $E$ -values  $< 1 \times 10^{-6}$  and bit scores  $> 50$  (49). While these annotations included transcripts annotated to both bacteria and viruses (see below), these transcripts were not filtered out from the initial annotations because they may be related to the lake sturgeon microbiome. DIAMOND v2.1.6 was used with blastx to provide a database of annotations with the well-assembled sterlet (*Acipenser ruthenus*) genome (19, 50). This program was chosen for its performance and sensitivity in homology

**Table 1.** Assembly and annotation statistics for each of 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*)

Tissue	Number of reads	Number of putative genes	Number of transcripts	Number of transcripts with annotations	Number of unique annotated genes	Number of transcripts annotated to sterlet
Brain	52 583 855	116 196	185 242	120 970	13 252	63 957
Gill	64 243 963	165 299	250 456	103 501	12 848	77 281
Head kidney	51 005 951	104 228	174 924	147 225	12 363	63 531
Heart	61 776 431	115 141	161 496	58 872	10 709	45 670
Liver	52 606 238	68 560	98 998	47 739	9 817	35 194
White muscle	65 292 771	64 372	93 273	50 022	10 143	35 963
Esophagus	133 351 646	136 859	251 276	78 883	13 319	75 265
Glandular stomach	149 907 748	130 979	243 955	77 889	13 263	74 514
Muscular stomach	138 604 614	120 933	221 499	69 953	12 807	68 679
Pyloric cecum	146 607 316	118 211	220 417	71 752	12 836	69 171
Anterior intestine	114 334 754	105 674	197 080	66 478	12 388	64 211
Spiral valve	112 481 602	122 567	225 357	71 406	13 004	69 937
Rectum	134 824 388	127 892	244 401	80 825	13 580	76 398
Mean	98 278 559.77	115 147.00	197 567.23	80 424.23	12 333.00	63 059.31
Standard deviation	38 868 780.36	25 504.38	51 480.67	27 125.92	1 215.21	14 120.44
Overall	1 277 621 277	484 570	770 984	326 367	18 290	114 865

The number of reads represents the number of short-read mRNA sequences used in the assembly, the number of putative genes refers to the number of gene models assembled in Trinity and the number of transcripts represents the total number of transcripts identified. The number of transcripts with annotations and the number of unique annotated genes refer to annotations performed with Trinotate and associated programs. The number of transcripts annotated to sterlet (*Acipenser ruthenus*) refers to the number of transcripts with homologous sterlet annotations at  $E$ -values  $<1 \times 10^{-6}$  and bit scores  $>50$ . Means and standard deviations among all 13 transcriptomes are reported at the bottom of the table. The tissue-specific transcriptomes are labeled by tissue, while the transcriptome labeled 'Overall' refers to an assembly that included all data used to make the 13 tissue-specific transcriptomes.

searches. Consistent with the non-specific annotations, transcripts were filed for  $E$ -values  $<1 \times 10^{-6}$  and bit scores  $>50$ , and best blast hits were assessed with minimum  $E$ -values, maximum bit scores and maximum percent identities among multiple hits. BUSCO v5.1.2 was used to assess transcriptome completeness with respect to the Actinopterygii odb10 dataset (51). Specifically, BUSCO was run in transcriptome mode (`-mode transcriptome`) on each transcriptome. Each completeness metric was exported from the BUSCO results and visualized with the fishualize v0.2.3 package in the statistical computing environment R v1.1.2 (52, 53). To assess divergence in terms of mutation distance, Mash v1.1 was used to make pairwise comparisons between each transcriptome (54). Because evolutionary distance is not expected among transcriptomes from lake sturgeon sampled from a single population, the distances measured thus represent isoforms and paralogs among the gene models in the assembled transcriptomes.

### Annotation analyses

The statistical computing environment R v1.1.2 and R package Tidyverse v1.3.1 were used throughout functional analyses of lake sturgeon transcriptomes (53, 55). A gene set enrichment analysis was used to identify GO terms for each transcriptome using enrichR v3.0 with the Biological Process 2021, Molecular Function 2021 and Cellular Component 2021 databases (38). Only GO terms with a false discovery rate ( $q$ )  $<0.05$  were retained as significantly enriched. The R package UpSetR v1.4.0 was used to assess the uniqueness of GO terms among tissues (56). This assessment of uniqueness among tissues was repeated in an analysis excluding peripheral tissues and specific to the esophagus, glandular stomach, muscular stomach, pyloric cecum, anterior intestine, spiral

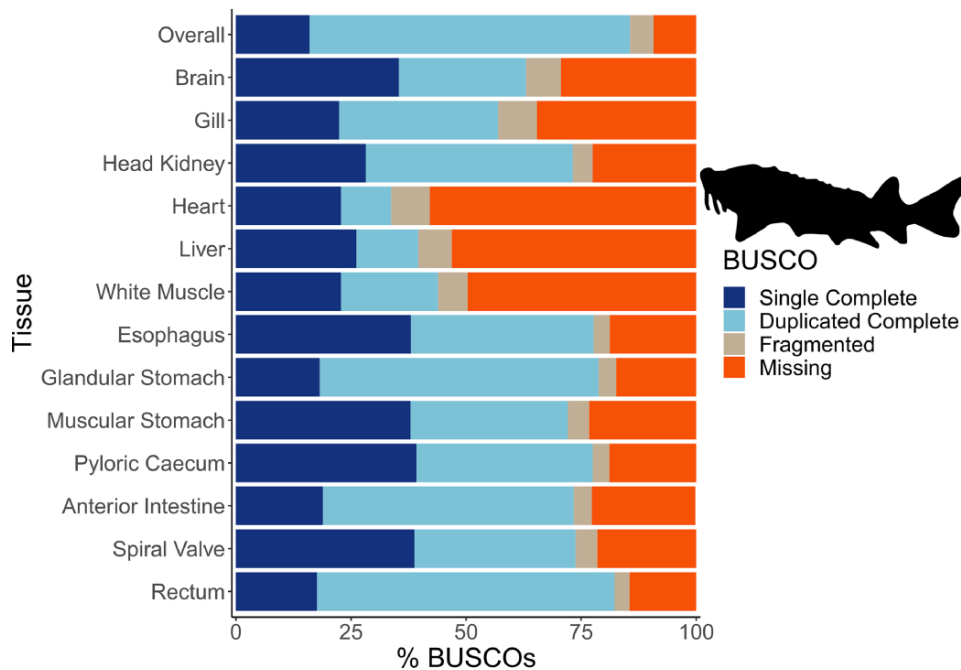
valve and rectum to identify patterns specific to lake sturgeon gut tissues.

Principal component analysis (PCA) was used to visualize differentiation among different tissues with respect to the present genes and GO terms. The overall transcriptome was excluded from PCA to explore the variance among the tissue-specific transcriptomes. A PCA with *prcomp* in the base R environment was used on a table of the presence or absence of gene names within each of the 13 tissues, along with separate PCAs for each of the Biological Process 2021, Molecular Function 2021 and Cellular Component 2021 GO databases. Specifically, the function *prcomp* was used on the presence-absence tables used to create UpSetR plots (see earlier). These PCAs were used to visualize differentiation among the different transcriptomes. The overall transcriptome was generally excluded from annotation comparisons as analyses of uniqueness among tissues focused on the set of tissue-specific transcriptomes. However, the number of unique genes was assessed in the overall transcriptome to identify the genes that were potentially missing or unannotated from each tissue-specific transcriptome but resolved with all data used in one assembly. The genes unique to the overall transcriptome were analyzed for GO terms with the same databases as the tissue-specific transcriptomes.

### Microbial analyses

The lake sturgeon microbiome was investigated by removing transcripts annotated to eukaryotes or viruses from the transcriptomes of all 13 tissues, using the transcriptome annotations described previously. Therefore, only transcripts annotated to bacteria or archaea remained from each transcriptome. A gene set enrichment analysis was performed on the microbe-annotated transcripts as in the previous section,





**Figure 2.** The transcriptome completeness assessed with BUSCO. Single complete represents orthologs that were present singly in a transcriptome, while duplicated complete represents orthologs duplicated in the transcriptome that matched the BUSCO profile. Fragmented orthologs were present in the transcriptomes, but not within the expected range of alignments in the BUSCO profile. Missing orthologs were those present in the BUSCO profile but missing in the transcriptome completely. The tissue-specific transcriptomes are labeled by tissue, while the transcriptome labeled 'Overall' refers to an assembly that included all data from the 13 tissues. The BUSCO profile used in the present analysis was the Actinopterygii odb10 dataset. The lake sturgeon icon and colors used were from the fishualize package in R.

but no significant GO terms were identified. UpSet plots were used to identify the uniqueness in annotated genes and microbial genera present among tissues.

## Results

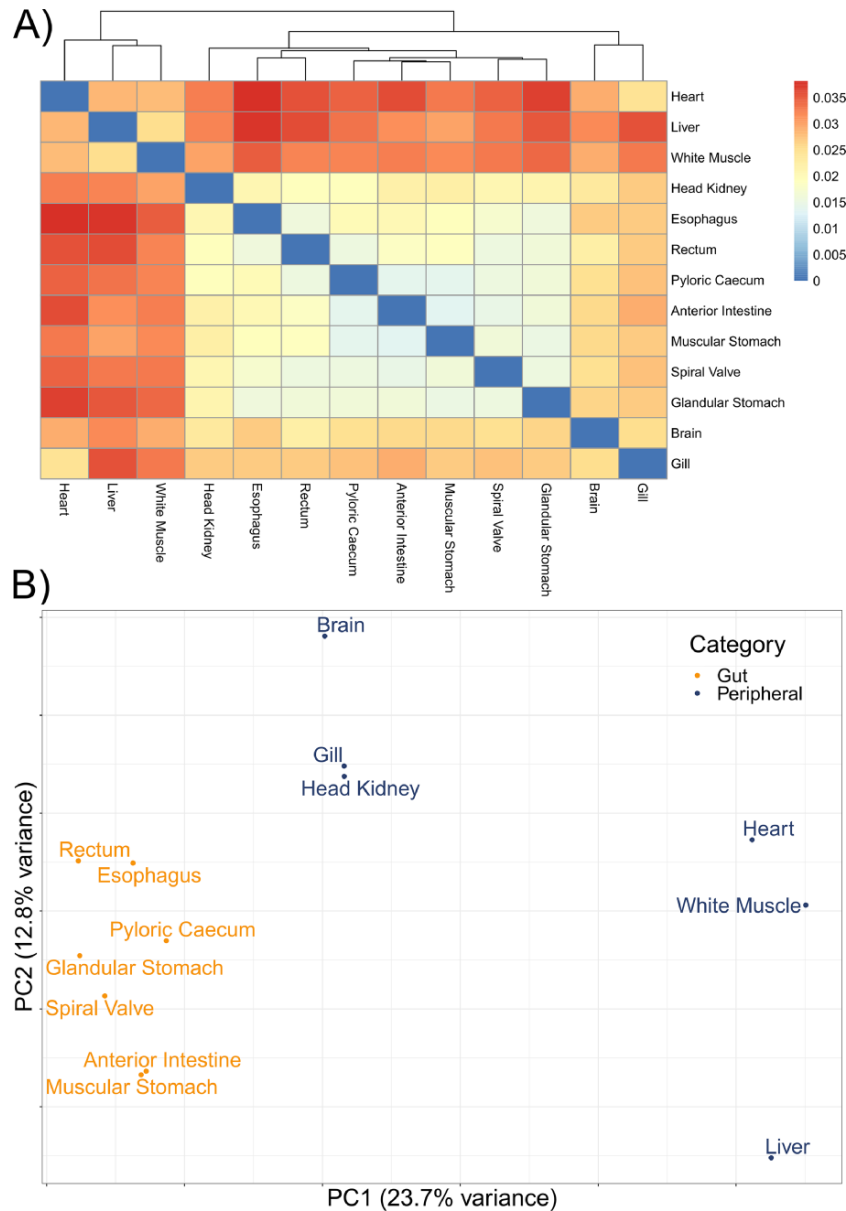
### Transcriptome assembly and annotation

The mean number of putative genes from the Trinity assemblies was 115 147 ( $\pm 24 405$  SD), while the mean number of transcripts was 197 567 ( $\pm 51 481$  SD) (Table 1). After filtering out transcripts with annotations with bit scores  $< 50$  and  $E$ -values  $> 1 \times 10^{-6}$ , the number of transcripts remaining with annotations among the tissue-specific transcriptomes was a mean of 76 319 ( $\pm 19 737$  SD), representing a mean of 12 350 ( $\pm 1217$  SD) unique genes with annotations (Table 1). A mean of 63 059 ( $\pm 14 120$  SD) transcripts were annotated to the sterlet genome among the tissue-specific transcriptomes and 114 865 transcripts in the overall transcriptome. As with other annotation databases in the present study, these results have been shared in an online repository ([https://figshare.com/projects/Lake\\_Sturgeon\\_Transcriptomes/133143](https://figshare.com/projects/Lake_Sturgeon_Transcriptomes/133143)). Overall, 3065 genes were identified in the analysis of uniqueness that included the overall and tissue-specific transcriptomes although uniqueness from gene sets for the tissue-specific transcriptomes was skewed downward (Supplementary Figure S1). Tissue-specific transcriptome completeness ranged from a minimum of 33.7% (heart) to a maximum of 82.2% (rectum) (overall mean  $65.1\% \pm 15.7\%$  SD) (Figure 2). Divergence among tissue-specific transcriptomes, as assessed using Mash, was statistically significant in each pairwise comparison ( $P < 0.05$ ) although distances between transcriptomes were greatest

between gut tissues and the heart, liver and white muscle tissues (Figure 3A).

### Annotation analyses

A mean of 832 ( $\pm 123$  SD) biological process GO terms were identified among the 13 tissue-specific transcriptomes (Table 2). Among the biological process GO terms, 367 were shared across all tissues, but a substantial number were also unique to individual tissues such as 101 GO terms in the liver and 71 in the heart (Figure 4; Supplementary Tables S1–S13). Qualitatively similar patterns of shared GO terms among all tissues, with substantial numbers of terms unique to each tissue, were observed in the molecular function and cellular component databases (Supplementary Figures S2 and S3). A similar pattern of uniqueness was also present among biological process GO terms in the gut tissues. Here, 464 terms were shared among all gut tissues, but smaller numbers of terms were unique to individual tissues, such as 59 unique to the anterior intestine and 46 unique to the pyloric caecum (Supplementary Figure S4). For molecular function and cellular component, a mean of 131 ( $\pm 19$  SD) and 150 ( $\pm 21$  SD) GO terms were identified, respectively. Molecular function and cellular component GO terms were qualitatively similar in patterns of uniqueness both when considering all 13 tissues (Supplementary Tables S1–S13) and among gut-only tissues (Supplementary Tables S14–S20). No significant GO terms were identified from the overall transcriptome, possibly because Fisher's exact test used in enrichR was implemented for experimental designs, as opposed to surveys of gene presence (57). Five molecular function GO terms were significant in a search of



**Figure 3.** The divergence among 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*). (A) A heatmap of pairwise distances between transcriptomes assessed with Mash, where higher Mash distances correspond to greater evolutionary divergence between the transcriptomes. Because no evolutionary divergence is expected for transcriptomes from one population of one species, these distances represent isoforms and paralogs of gene models. Higher values indicate more divergence. (B) A PCA of the present and absent genes in the 13 transcriptomes, performed with *prcomp* in R. Gut and peripheral tissues were distinguished for visualization, where gut tissues were the esophagus, glandular stomach, muscular stomach, anterior intestine, spiral valve and rectum, while peripheral tissues were the brain, gill, head kidney, heart, white muscle and liver. The distinction in color between gut and peripheral tissues is only for visualization and was not used to categorize data *a priori* in the PCA.

the 3065 genes unique to the overall transcriptome among all transcriptomes analyzed, while no biological process or cellular component terms were significant. The significant molecular function gene ontology terms were as follows: peptide alpha-*N*-acetyltransferase activity (GO:0004596; combined score = 329), peptide *N*-acetyltransferase activity (GO:0034212; combined score = 255), phosphatidate phosphatase activity (GO:0008195; combined score = 251), lipid phosphatase activity (GO:0042577; combined score = 173) and lysine *N*-methyltransferase activity (GO:0016278; combined score = 38). PCAs revealed differentiation between the liver transcriptome and those from other tissues in each

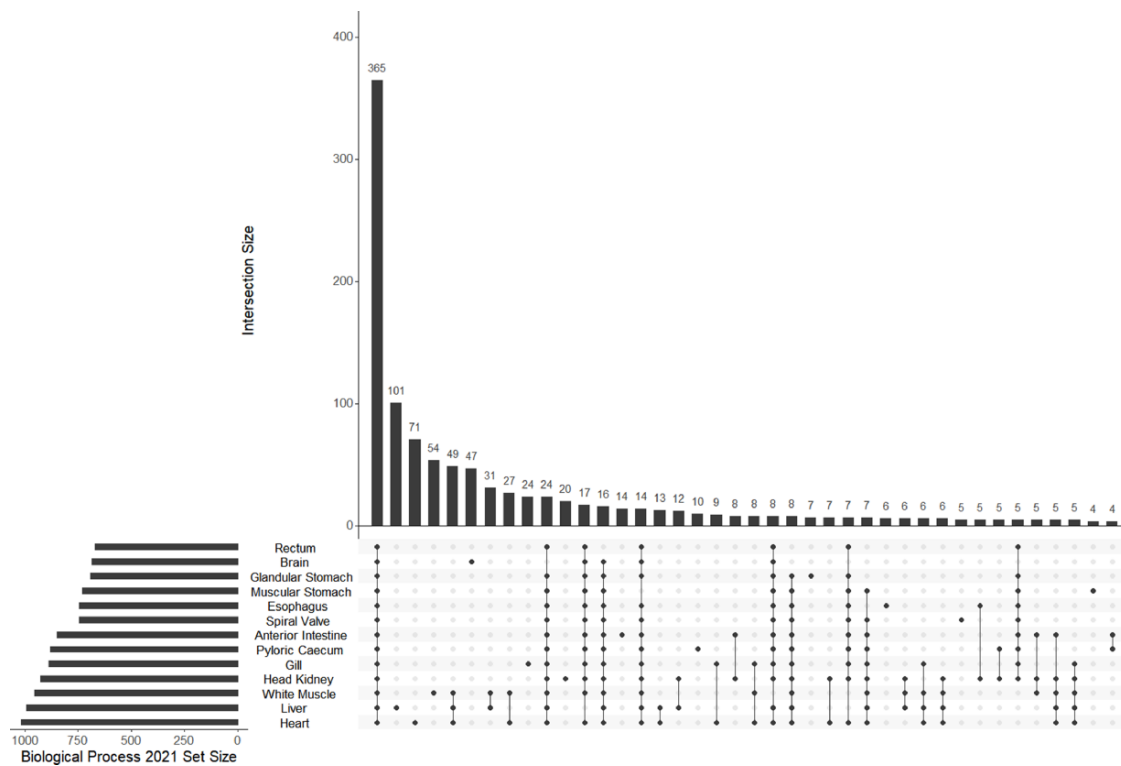
comparison, especially in a PCA of terms in the cellular components GO database (Figure 3B; Supplementary Figure S5). Gut tissues tended to cluster together compared to other tissues in each PCA, but because gut and peripheral tissues were extracted using separate protocols, some differentiation between the two groups of tissues may be a technical artifact.

In the pyloric caecum, the GO terms photoperiodism (GO:0009648) and entrainment of the circadian clock by photoperiod (GO:0043153) were uniquely present and related to periodicity (Supplementary Table S11). Patterns of tissue-specific immune regulation were uniquely

**Table 2.** GO terms and microbial information for each of 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*). GO terms were identified by using all unique annotated genes in a transcriptome with enrichR. The Biological Process 2021, Molecular Function 2021, and Cellular Component 2021 databases were searched for the GO analyses, where only significant terms ( $q < 0.05$ ) were retained for downstream analyses. Present microbial genera and genes were identified using annotation information from Trinotate

Tissue	Biological process GO terms	Molecular function GO terms	Cellular component GO terms	Microbial genera	Microbial genes
Brain	684	111	152	25	31
Gill	888	145	149	24	33
Head kidney	924	149	152	25	28
Heart	1 017	148	174	16	22
Liver	992	164	183	21	32
White muscle	954	152	186	12	16
Esophagus	742	131	138	56	105
Glandular stomach	691	110	131	45	83
Muscular stomach	729	117	135	75	177
Pyloric cecum	879	126	157	30	55
Anterior intestine	848	136	145	38	67
Spiral valve	744	119	124	65	169
Rectum	671	98	118	47	84
Mean	827.92	131.23	149.54	36.85	69.38
Standard deviation	118.55	18.93	20.51	18.77	51.43

GO terms were identified by using all unique annotated genes in a transcriptome with enrichR. The Biological Process 2021, Molecular Function 2021 and Cellular Component 2021 databases were searched for the GO analyses, where only significant terms ( $q < 0.05$ ) were retained for downstream analyses. The present microbial genera and genes were identified using annotation information from Trinotate.



**Figure 4.** The UpSet plot of shared and unique GO terms from each of 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*). The GO terms presented here are from the Biological Process 2021 database, significant at  $q < 0.05$ . The R package UpSetR was used to visualize these data.

present in several gut tissues. Rac protein signal transduction (GO:0016601) was uniquely present in the glandular stomach and may also represent a part of the innate immune system with its role in neutrophil recruitment (Supplementary Table S8). Negative regulation of immune response (GO:0045824) and autophagy of peroxisomes (GO:0030242) were unique to the muscular stomach

(Supplementary Table S9). Positive regulation of the host by viral transcription (GO:0043923) was uniquely present in the anterior intestine (Supplementary Table S10). Toll-like receptor 9 signaling pathway (GO:0034162), toll-like receptor signaling pathway (GO:0002224) and cellular response to interleukin-12 (GO:0071349) were each uniquely present in the spiral valve, consistent with a role for the tissue in the





between the two groups of tissues may be attributed to different RNA extraction methods used. Nevertheless, the distinction between the two groups is consistent with differences in physiological function. The peripheral tissue transcriptomes enable a variety of research questions on the lake sturgeon, such as liver and gill often used in work exploring the vertebrate stress response (63, 64). Tissues such as the brain, heart, head kidney and white muscle can be informative for developmental questions, with potential connections to nutrition and stress among other biological processes (63, 65–68). Meanwhile, the seven gut transcriptomes represent the major anatomically distinct regions of the gut. Because the gut encompasses an organ system with distinct compartmentalization of function (17), the seven gut transcriptomes provide an opportunity to study distinct digestion-related mechanisms. Gut transcriptomics was predicted to accelerate research on intestinal pathogen responses, dietary manipulations and osmoregulatory challenges (25), and the present data contribute to a longstanding body of work investigating the physiological mechanisms of the vertebrate gut (14, 15). We thus provide all 13 tissue-specific transcriptomes and 1 overall transcriptome as a scientific resource from this study but focus on discussing observations among the gut tissues.

### Circadian rhythm transcripts in the pyloric cecum

The pyloric cecum is a tissue of interest because it is absent in Agnatha and Chondrichthyes but is present in Actinopterygii (16, 17, 21). Given the sturgeon's status as an ancient actinopterygian, sister to the rest of the clade with a split ~318 million years ago (12, 69), they represent an early evolutionary appearance of the pyloric cecum because all sturgeons have pyloric ceca. Moreover, phylogenetic analyses of the origin of Acipenseriformes imply that the pyloric cecum is also ~318 million years old (69). Analyses of tissue-specific transcriptome annotations revealed notable patterns of transcript presence within pyloric cecum, which may have implications for different mechanisms of lake sturgeon digestion. For instance, GO terms photoperiodism (GO:0009648) and entrainment of the circadian clock by photoperiod (GO:0043153) were unique to the pyloric cecum. In addition, among the core clock genes of *clock*, *bmal1*, *per* (1, 2 and 3) and *cry* (1 and 2), all expressed in teleost pineal organs (70), *clock*, *cry1* and *cry2* were present in the lake sturgeon pyloric cecum. *Cry1* and *cry2*, which are photoreceptors with important roles in circadian rhythms, were also transcribed in the brain, liver, heart, retina, muscle, spleen, gill and intestine of European seabass (*Dicentrarchus labrax*), where rhythmic expression was observed in the brain and liver (71).

While the circadian rhythm-related genes observed with the present data likely differ evolutionarily from those in other vertebrates because of the ancient origin of Acipenseriformes (69), the present annotations were filtered for both bit scores and *E*-values (49). Therefore, the transcripts annotated to circadian rhythm-related genes are at least largely similar in sequence to those genes observed in other species. Notably, the clock-related GO terms observed in the present data were unique to the pyloric cecum, but individual genes were present among other tissues such as the brain. GO terms used in the present analyses were filtered for significance from Fisher's exact test (57). Therefore, transcriptomes with annotated clock genes but without enriched GO terms present

represent those with too few genes within the clock-related GO terms to be significant. The present results do not contradict the prior work that identified clock genes in other tissues but do provide novel findings of core clock genes in the pyloric cecum that may be related to feeding periodicity.

Feeding periodicity has been observed in numerous fish species (e.g. *Merlagius merlangus* (72); *Limanda limanda* (73)), across herbivores, detritivores, insectivores, zooplanktivores and macrophyte feeders (74). Because feeding periodicity is phylogenetically and ecologically widespread among fishes, we predict that physiological digestive mechanisms may contribute to the phenomenon. A circadian rhythm in metabolic rate was observed in lake sturgeon exposed to a 12-hour light–dark cycle, where metabolism was highest at sunrise (75). The lake sturgeon used for the present study was also fed predictably, once a day with a 11:13-hour light–dark cycle. Given the dual observations that feeding periodicity exists across fish species (74) and the presence of several core clock genes in the lake sturgeon pyloric cecum, we developed alternative hypotheses that may address underlying mechanisms of periodicity in the lake sturgeon, which may be applicable to other fishes (76).

First, we hypothesized that physiological mechanisms of digestion periodicity may be regulated by diel circadian clock rhythms in the pyloric cecum. Therefore, we predict diel fluctuations in transcript abundance of *clock*, *cry1* and *cry2* along with other circadian rhythm-related genes only in the pyloric cecum of laboratory-held lake sturgeon consistent with feeding times and their light–dark cycle. Specific roles for physiological mechanisms of digestion periodicity could include intestinal motility, intestinal function, innate immunity, microbiome regulation or cell proliferation (77–84). Alternatively, we hypothesized that the circadian rhythm genes observed in the pyloric cecum may represent a part of a whole-gut circadian response wave consistent with a phenomenon hypothesized in lab mice (85). That is, a whole-gut circadian response may have been initiated at feeding and was observed by chance in the pyloric cecum by sampling individuals ~18–20 hours after feeding. Therefore, from this hypothesis, we predict diel fluctuations of transcript abundance of *clock*, *cry1* and *cry2* in the pyloric cecum and other gut tissues. More posterior gut tissues, such as the spiral valve, may therefore show expression of the three predicted genes but chronologically later than the pyloric cecum. By contrast, the glandular and muscular stomachs may show evidence of this circadian response wave earlier in time than the pyloric cecum because of their positions prior to the cecum in the gut. While the first hypothesis about physiological functions of digestion as regulated by the pyloric cecum focuses on one gut tissue, the hypotheses are not necessarily mutually exclusive. A circadian response wave may pass through different gut tissues, but the pyloric cecum may play key roles in downstream processes from the circadian wave. This circadian response wave may be entrained by food intake times if it is consistent with mammalian physiology (86). Therefore, a timepoint- and tissue-specific approach is needed to test these hypotheses.

### Microbial observations

As lake sturgeon represents an early stage of actinopterygian gut evolution, several observations in the present database of tissue-specific transcriptomes were notable. One example is the presence of transcripts related to innate immunity

in the spiral valve. As gut tissues may be in contact with food and potentially associated pathogens from the external environment, innate immunity and immune responses involved in digestion may help to protect the fish from food- or environment-related pathogens (87). Cartilaginous fishes have gut-associated lymphoid tissues in the spiral valve (88, 89), consistent with the present observations of innate immune-related transcripts in the lake sturgeon spiral valve. GO terms related to toll-like receptor signaling pathways and cellular responses to interleukin were unique to the spiral valve in lake sturgeon, while terms related to innate immune function were also present in the muscular stomach, glandular stomach and anterior intestine. The GO term positive regulation of viral life cycle is not an immune response in itself, but its unique presence in the esophagus provides some evidence for the necessity of an innate immune response in other gut tissues. These heterogeneous signals of host–microbiome interactions along the gut are consistent with the evidence of host–microbiome interactions from 16S rRNA and DNA sequencing from microbial samples and the lake sturgeon spiral valve (26).

The tissue-specific transcriptome database enabled analyses of bacterial and archaeal transcripts as well as genera across different gut tissues. Many of these microbial transcripts and genera were unique to different gut sections, and thus, we concluded that the microbiome is likely heterogeneous across the lake sturgeon gut based on evidence of microbe presence unique to each tissue summarized later. A caveat is that the presence of microbial genera and transcripts in certain tissues may be attributed to contamination, such as in the brain, white muscle and heart, along with transcripts or genera shared among all tissues (7 genes and 8 genera identified as shared among all 13 tissues) (90). However, patterns of microbial presence were consistent with microbiome regulation and tissue-specific function for gut tissues. For example, among microbial genera unique to each tissue, the muscular stomach had the greatest number present (15), followed by the spiral valve (12) and the anterior intestine (6). A qualitatively similar pattern was found with transcripts of genes annotated to bacteria and archaea unique to each tissue, with the muscular stomach (114), spiral valve (101), esophagus (50), glandular stomach (23), rectum (18), and pyloric cecum (14) all supporting unique microbial communities. These results demonstrate that the greatest number of unique microbial genera and genes were identified in gut tissues as opposed to tissues outside of the gut. In addition, microbe-annotated genes and genera were largely unique among tissues, which is inconsistent with potential contamination from tools or extraction kits as similar types of microbes or microbial genomic material would then be observed among multiple tissues (90). A second caveat is that poly(A) tail isolation was used during library preparation for mRNA sequencing of the present data. This isolation was useful for characterizing eukaryotic transcripts but likely biased prokaryote-annotated transcripts toward those that were destabilized and in the process of RNA turnover (91). Thus, the results discussed here must be interpreted as a partial view of microbial transcriptional activity in the lake sturgeon gut. Nevertheless, the present results are consistent with a heterogeneous microbial community with tissue-specific mRNA transcription in the lake sturgeon gut.

Other work identified microbial community shifts in the lake sturgeon spiral valve in response to a failure to transition

diets and with feeding cessation (28). Similarly, spiral valve microbiome community composition changed in response to exposure to common antibiotics, drugs and chemicals used in lake sturgeon aquaculture (27). Therefore, gut microbiome community composition is dynamically connected to the physiological state of lake sturgeon (26). Because unique patterns of microbial genera and genes were found in both the spiral valve and other gut tissues in the present data, the analyses of gut transcriptomes demonstrate that host–microbiome interactions may occur along much of the lake sturgeon gut and that the interactions may be spatially heterogeneous and specific to different gut tissues. Thus, the transcriptomes used here may support work in the lake sturgeon that resolves spatially distinct mechanisms of host–microbiome interactions.

## Conclusions

In the present study, 13 tissue-specific transcriptomes and 1 overall transcriptome were presented as a resource for lake sturgeon research. The overlap of GO terms was analyzed among tissues. While shared patterns indicated consistent transcriptomic functions among tissues, the presence of unique GO terms showed that sequencing transcriptomes from multiple tissues enabled research questions that would otherwise be intractable. Moreover, the analysis of unique GO terms among tissues revealed the presence of transcribed genes related to photoperiodicity in the pyloric cecum, an observation consistent with a role for periodicity in digestive physiology in an ancient fish. Transcripts involved in innate immune function were found in the spiral valve and other gut tissues, which provide evidence in support of a prior hypothesis about the emergence of innate immunity in the gut of cartilaginous fishes and are consistent with specialization in immune function across gut tissues. An analysis of genes annotated to bacteria and archaea indicated potentially heterogeneous microbiota and microbial functions along different gut tissues, consistent with specialization in immune function and microbiome regulation along the lake sturgeon gut. As lake sturgeons are representative of sturgeon and paddlefish's status as ancient fishes, they constitute an early stage in the differentiation of several gut tissues. Studying this early stage in the differentiation of the gut as an organ system with distinct functions provided insights into digestive function, immunity and microbiome regulation. These results are a resource for lake sturgeon research and also provide information about the mechanisms of compartmentalized function across gut tissues.

## Supplementary material

Supplementary material is available at *Database* online.

## Data availability

The lake sturgeon transcriptomes and annotations are available on Figshare ([https://figshare.com/projects/Lake\\_Sturgeon\\_on\\_Transcriptomes/133143](https://figshare.com/projects/Lake_Sturgeon_on_Transcriptomes/133143)). Codes used in the analyses of the data in the present study are available on GitHub ([https://github.com/BioMatt/lakesturgeon\\_transcriptomes](https://github.com/BioMatt/lakesturgeon_transcriptomes)). Raw mRNA sequencing reads have been deposited at the National Center for Biotechnology Information Sequence Read Archive

(accession number: PRJNA949556; <https://www.ncbi.nlm.nih.gov/sra/PRJNA949556>).

## Contribution statement

M.J.T. performed bioinformatic analyses on the data and wrote the initial draft of the manuscript, with assistance from A.M.W. and W.S.B. All authors edited the manuscript. W.S.B. and A.M.W. conducted fish rearing, sampling, and mRNA extractions. In addition, all authors designed and conceived the study. W.G.A. and K.M.J. provided funding and supervision.

## Funding

Natural Sciences and Engineering Research Council of Canada (NSERC)/Manitoba Hydro Industrial Research Chair (to W.G.A.); NSERC Discovery Grants (#05348 to W.G.A., #05479 to K.M.J.).

## Conflict of interest

The authors declare no conflicts of interest.

## Acknowledgements

We thank Dr Robert Syme and François Lefebvre from the Canadian Centre for Computational Genomics for their support with mRNA sequencing. Evelien de Greef illustrated the lake sturgeon and gut, and Hamza Amjad assisted with RNA extractions. Many bioinformatic steps were enabled by the opportunity to use computing resources provided by the Digital Research Alliance of Canada.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## References

- Fontana, F., Bruch, R.M., Binkowski, F.P. *et al.* (2004) Karyotype characterization of the lake sturgeon, *Acipenser fulvescens* (Rafinesque 1817) by chromosome banding and fluorescent in situ hybridization. *Genome*, **47**, 742–746.
- Welsh, A. and May, B. (2006) Development and standardization of disomic microsatellite markers for lake sturgeon genetic studies. *J. Appl. Ichthyol.*, **22**, 337–344.
- McQuown, E., Gall, G.A.E. and May, B. (2002) Characterization and inheritance of six microsatellite loci in lake sturgeon. *Trans. Am. Fish. Soc.*, **131**, 299–307.
- Welsh, A.B., Blumberg, M. and May, B. (2002) Identification of microsatellite loci in lake sturgeon, *Acipenser fulvescens*, and their variability in green sturgeon, *A. medirostris*. *Mol. Ecol. Notes*, **3**, 47–55.
- Whitaker, J.M., Price, L.E., Boase, J.C. *et al.* (2020) Detecting fine-scale population structure in the age of genomics: a case study of lake sturgeon in the Great Lakes. *Fish. Res.*, **230**, 105646.
- Bohling, J., Small, M., Von Barga, J. *et al.* (2019) Comparing inferences derived from microsatellite and RADseq datasets: a case study involving threatened bull trout. *Conserv. Genet.*, **20**, 329–342.
- Lemopoulos, A., Prokkola, J.M., Uusi-Heikkilä, S. *et al.* (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—implications for brown trout conservation. *Ecol. Evol.*, **9**, 2106–2120.
- Bugg, W.S., Yoon, G.R., Schoen, A.N. *et al.* (2020) Effects of acclimation temperature on the thermal physiology in two geographically distinct populations of lake sturgeon (*Acipenser fulvescens*). *Conserv. Physiol.*, **8**, coaa087.
- Bugg, W.S., Jeffries, K.M. and Gary Anderson, W. (2021) Survival and gene expression responses in immune challenged larval lake sturgeon. *Fish Shellfish Immunol.*, **112**, 1–7.
- Bugg, W.S., Yoon, G.R., Brandt, C. *et al.* (2021) The effects of population and thermal acclimation on the growth, condition and cold responsive mRNA expression of age-0 lake sturgeon (*Acipenser fulvescens*). *J. Fish Biol.*, **99**, 1912–1927.
- Yusishen, M.E., Eichorn, F.-C., Anderson, W.G. *et al.* (2020) Development of quantitative PCR assays for the detection and quantification of lake sturgeon (*Acipenser fulvescens*) environmental DNA. *Conserv. Genet. Resour.*, **12**, 17–19.
- Gardiner, B.G. (1984) Sturgeons as living fossils. In: Eldredge N, Stanley SM (eds) *Living Fossils, Casebooks in Earth Sciences*. Springer New York, New York, NY, pp. 148–152.
- Choudhury, A. and Dick, T.A. (2003) The historical biogeography of sturgeons (Osteichthyes: Acipenseridae): a synthesis of phylogenetics, palaeontology and palaeogeography. *J. Biogeogr.*, **25**, 623–640.
- Aristotle (1937) *Parts of animals. Movement of animals. Progression of animals*. Translated by Peck AL and Forster ES. Harvard University Press, Cambridge, pp. 1–560.
- Buddington, R.K. and Diamond, J.M. (1986) Aristotle revisited: the function of pyloric caeca in fish. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 8012–8014.
- Buddington, R.K. and Diamond, J.M. (1987) Pyloric caeca of fish: a “new” absorptive organ. *Am. J. Physiol.*, **252**, G65–G76.
- Wilson, J.M. and Castro, L.F.C. (2010) Morphological diversity of the gastrointestinal tract in fishes. *Fish Physiol.*, **30**, 1–55.
- Wang, D., Zhong, L., Wei, Q. *et al.* (2010) Evolution of MHC class I genes in two ancient fish, paddlefish (*Polyodon spathula*) and Chinese sturgeon (*Acipenser sinensis*). *FEBS Lett.*, **584**, 3331–3339.
- Du, K., Stöck, M., Kneitz, S. *et al.* (2020) The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.*, **4**, 841–852.
- Cheng, P., Huang, Y., Lv, Y. *et al.* (2021) The American paddlefish genome provides novel insights into chromosomal evolution and bone mineralization in early vertebrates. *Mol. Biol. Evol.*, **38**, 1595–1607.
- Argyriou, T., Clauss, M., Maxwell, E.E. *et al.* (2016) Exceptional preservation reveals gastrointestinal anatomy and evolution in early actinopterygian fishes. *Sci Rep*, **6**, 18758.
- Clark, R.I. and Walker, D.W. (2018) Role of gut microbiota in aging-related health decline: insights from invertebrate models. *Cell. Mol. Life Sci.*, **75**, 93–101.
- Holt, C.C., Bass, D., Stentiford, G.D. *et al.* (2021) Understanding the role of the shrimp gut microbiome in health and disease. *J. Invertebr. Pathol.*, **186**, 107387.
- Pratte, Z.A., Perry, C., Dove, A.D.M. *et al.* (2022) Microbiome structure in large pelagic sharks with distinct feeding ecologies. *Animal Microbiome*, **4**, 17.
- Martin, A.M., Sun, E.W., Rogers, G.B. *et al.* (2019) The influence of the gut microbiome on host metabolism through the regulation of gut hormone release. *Front. Physiol.*, **10**, 428.
- Abdul Razak, S. and Scribner, K.T. (2020) Ecological and ontogenetic components of larval lake sturgeon gut microbiota assembly, successional dynamics, and ecological evaluation of neutral community processes. *Appl. Environ. Microbiol.*, **86**, e02662–19.



27. Abdul Razak,S., Bauman,J.M., Marsh,T.L. *et al.* (2022) Changes in lake sturgeon gut microbiomes relative to founding origin and in response to chemotherapeutant treatments. *Microorganisms*, **10**, 1005.
28. Abdul Razak,S., Valentine,S., Marsh,T. *et al.* (2022) Compositional dynamics of gastrointestinal tract microbiomes associated with dietary transition and feeding cessation in lake sturgeon larvae. *Microorganisms*, **10**, 1872.
29. Sullam,K.E., Essinger,S.D., Lozupone,C.A. *et al.* (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol. Ecol.*, **21**, 3363–3378.
30. Jones,E.W., Carlson,J.M., Sivak,D.A. *et al.* (2022) Stochastic microbiome assembly depends on context. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2115877119.
31. Pattison,R.J., Csukasi,F., Zheng,Y. *et al.* (2015) Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. *Plant Physiol.*, **168**, 1684–1701.
32. Jue,N.K., Foley,R.J., Reznick,D.N. *et al.* (2018) Tissue-specific transcriptome for *Poeciliopsis prolifica* reveals evidence for genetic adaptation related to the evolution of a placental fish. *G3*, **8**, 2181–2192.
33. Andrew,S.C., Primmer,C.R., Debes,P.V. *et al.* (2021) The Atlantic salmon whole blood transcriptome and how it relates to major locus maturation genotypes and other tissues. *Mar. Genomics*, **56**, 100809.
34. Voshall,A. and Moriyama,E.N. (2020) Next-generation transcriptome assembly and analysis: impact of ploidy. *Methods*, **176**, 14–24.
35. He,B., Zhao,S., Chen,Y. *et al.* (2015) Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genom.*, **16**, 65.
36. Madritsch,S., Burg,A. and Sehr,E.M. (2021) Comparing de novo transcriptome assembly tools in di- and autotetraploid non-model plant species. *BMC Bioinform.*, **22**, 146.
37. Grabherr,M.G., Haas,B.J., Yassour,M. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652.
38. Kuleshov,M.V., Jones,M.R., Rouillard,A.D. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
39. Heumüller-Klug,S. (2015) Degradation of intestinal mRNA: a matter of treatment. *WJG*, **21**, 3499–3508.
40. Takele Assefa,A., Vandesompele,J. and Thas,O. (2020) On the utility of RNA sample pooling to optimize cost and statistical power in RNA sequencing experiments. *BMC Genom.*, **21**, 312.
41. Haas,B.J., Papanicolaou,A., Yassour,M. *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, **8**, 1494–1512.
42. Finn,R.D., Clements,J. and Eddy,S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.
43. Punta,M., Coghill,P.C., Eberhardt,R.Y. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
44. Petersen,T.N., Brunak,S., von Heijne,G. *et al.* (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.
45. Krogh,A., Larsson,B., von Heijne,G. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
46. Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
47. Kanehisa,M., Goto,S., Sato,Y. *et al.* (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
48. Powell,S., Szklarczyk,D., Trachana,K. *et al.* (2012) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.*, **40**, D284–D289.
49. Pearson,W.R. (2013) An introduction to sequence similarity (“homology”) searching. *Curr. Protoc. Bioinform.*, **42**, 3–1.
50. Buchfink,B., Reuter,K. and Drost,H.-G. (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods*, **18**, 366–368.
51. Simão,F.A., Waterhouse,R.M., Ioannidis,P. *et al.* (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.
52. Schiettekatte,N., Brandl,S. and Casey,J. (2019) fishualize: color palettes based on fish species.
53. R Core Team. (2022) R: A language and environment for statistical computing.
54. Ondov,B.D., Treangen,T.J., Melsted,P. *et al.* (2016) Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.*, **17**, 132.
55. Wickham,H., Averick,M., Bryan,J. *et al.* (2019) Welcome to the Tidyverse. *J. Stat. Softw.*, **4**, 1686.
56. Conway,J.R., Lex,A. and Gehlenborg,N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.
57. Chen,E.Y., Tan,C.M., Kou,Y. *et al.* (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.*, **14**, 128.
58. Bugg,W.S., Thorstensen,M.J., Marshall,K.E. *et al.* (2023) Elevated temperatures reduce population-specific transcriptional plasticity in developing lake sturgeon (*Acipenser fulvescens*). *Mol. Ecol.*, **32**, 4044–4062.
59. Connon,R.E., Jeffries,K.M., Komoroske,L.M. *et al.* (2018) The utility of transcriptomics in fish conservation. *J. Exp. Biol.*, **221**, jeb148833.
60. Dammerman,K.J., Steibel,J.P. and Scribner,K.T. (2016) Increases in the mean and variability of thermal regimes result in differential phenotypic responses among genotypes during early ontogenetic stages of lake sturgeon (*Acipenser fulvescens*). *Evol. Appl.*, **9**, 1258–1270.
61. Loepky,A.R., Belding,L.D., Quijada-Rodriguez,A.R. *et al.* (2021) Influence of ontogenetic development, temperature, and pCO<sub>2</sub> on otolith calcium carbonate polymorph composition in sturgeons. *Sci. Rep.*, **11**, 13878.
62. Kukurba,K.R. and Montgomery,S.B. (2015) RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.*, **11**, 951–969.
63. Aluru,N. and Vijayan,M.M. (2009) Stress transcriptomics in fish: a role for genomic cortisol signaling. *Gen. Comp. Endocrinol.*, **164**, 142–150.
64. Jeffries,K.M., Teffer,A., Michaleski,S. *et al.* (2021) The use of non-lethal sampling for transcriptomics to assess the physiological status of wild fishes. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, **256**, 110629.
65. Ju,Z., Karsi,A., Kocabas,A. *et al.* (2000) Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene*, **261**, 373–382.
66. Carney,S.A., Chen,J., Burns,C.G. *et al.* (2006) Aryl hydrocarbon receptor activation produces heart-specific transcriptional and toxic responses in developing zebrafish. *Mol. Pharmacol.*, **70**, 549–561.
67. Mazurais,D., Darias,M., Zambonino-Infante,J.L. *et al.* (2011) Transcriptomics for understanding marine fish larval development. *Can. J. Zool.*, **89**, 599–611.
68. Sarropoulou,E., Kotoulas,G., Power,D.M. *et al.* (2005) Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. *Physiol. Genomics*, **23**, 182–191.
69. Shen,Y., Yang,N., Liu,Z. *et al.* (2020) Phylogenetic perspective on the relationships and evolutionary history of the Acipenseriformes. *Genomics*, **112**, 3511–3517.
70. Saha,S., Singh,K.M. and Gupta,B.B.P. (2019) Melatonin synthesis and clock gene regulation in the pineal organ of teleost fish

- compared to mammals: similarities and differences. *Gen. Comp. Endocrinol.*, **279**, 27–34.
71. Del Pozo, A., Vera, L.M., Sánchez, J.A. *et al.* (2012) Molecular cloning, tissue distribution and daily expression of cry1 and cry2 clock genes in European seabass (*Dicentrarchus labrax*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.*, **163**, 364–371.
  72. Pedersen, J. (2000) Food consumption and daily feeding periodicity: comparison between pelagic and demersal whiting in the North Sea. *J. Fish Biol.*, **57**, 402–416.
  73. Gwyther, D. and Grove, D.J. (1981) Gastric emptying in *Limanda limanda* (L.) and the return of appetite. *J. Fish Biol.*, **18**, 245–259.
  74. Welianje, W.S., Amarasinghe, U.S., Moreau, J. *et al.* (2006) Diel feeding periodicity, daily ration and relative food consumption in some fish populations in three reservoirs of Sri Lanka. *Aquat. Living Resour.*, **19**, 229–237.
  75. Svendsen, J.C., Genz, J., Anderson, W.G. *et al.* (2014) Evidence of circadian rhythm, oxygen regulation capacity, metabolic repeatability and positive correlations between forced and spontaneous maximal metabolic rates in lake sturgeon *Acipenser fulvescens*. *PLoS One*, **9**, e94693.
  76. Betts, M.G., Hadley, A.S., Frey, D.W. *et al.* (2021) When are hypotheses useful in ecology and evolution? *Ecol. Evol.*, **11**, 5762–5776.
  77. Froy, O. and Chapnik, N. (2007) Circadian oscillation of innate immunity components in mouse small intestine. *Mol. Immunol.*, **44**, 1954–1960.
  78. Goo, R.H., Moore, J.G., Greenberg, E. *et al.* (1987) Circadian variation in gastric emptying of meals in humans. *Gastroenterology*, **93**, 515–518.
  79. Hoogerwerf, W.A. (2006) Biologic clocks and the gut. *Curr. Gastroenterol. Rep.*, **8**, 353–359.
  80. Konturek, P.C., Brzozowski, T. and Konturek, S.J. (2011) Gut clock: implication of circadian rhythms in the gastrointestinal tract. *J. Physiol. Pharmacol.*, **62**, 139–150.
  81. Peyric, E., Moore, H.A. and Whitmore, D. (2013) Circadian clock regulation of the cell cycle in the zebrafish intestine. *PLoS One*, **8**, e73209.
  82. Rosselot, A.E., Hong, C.I. and Moore, S.R. (2016) Rhythm and bugs: circadian clocks, gut microbiota, and enteric infections. *Curr. Opin. Gastroenterol.*, **32**, 7–11.
  83. Scheving, L.A. (2000) Biological clocks and the digestive system. *Gastroenterology*, **119**, 536–549.
  84. Stevenson, N.R., Sitren, H.S. and Furuya, S. (1980) Circadian rhythmicity in several small intestinal functions is independent of use of the intestine. *Am. J. Physiol.*, **238**, G203–G207.
  85. Neal, J.V. and Potten, C.S. (1981) Circadian rhythms in the epithelial cells and the pericryptal fibroblast sheath in three different sites in the murine intestinal tract. *Cell Prolif.*, **14**, 581–587.
  86. Bron, R. and Furness, J.B. (2009) Rhythm of digestion: keeping time in the gastrointestinal tract. *Clin. Exp. Pharmacol. Physiol.*, **36**, 1041–1048.
  87. Gómez, G.D. and Balcázar, J.L. (2008) A review on the interactions between gut microbiota and innate immunity of fish. *FEMS Immunol. Med. Microbiol.*, **52**, 145–154.
  88. Lauriano, E.R., Pergolizzi, S., Aragona, M. *et al.* (2019) Intestinal immunity of dogfish *Scyliorhinus canicula* spiral valve: a histochemical, immunohistochemical and confocal study. *Fish Shellfish Immunol.*, **87**, 490–498.
  89. Mitchell, C.D. and Criscitiello, M.F. (2020) Comparative study of cartilaginous fish divulges insights into the early evolution of primary, secondary and mucosal lymphoid tissue architecture. *Fish Shellfish Immunol.*, **107**, 435–443.
  90. Salter, S.J., Cox, M.J., Turek, E.M. *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.*, **12**, 87.
  91. Hajnsdorf, E. and Kaberdin, V.R. (2018) RNA polyadenylation and its consequences in prokaryotes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **373**, 20180166.