Tissue-specific transcriptomes reveal mechanisms of microbiome regulation in an ancient fish

Matt J. Thorstensen*[1], Alyssa M. Weinrauch[1], William S. Bugg[1], Ken M. Jeffries[1], and W. Gary Anderson[1]

[1] Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada.

*Corresponding author: Email: matt.thorstensen@gmail.com

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 aimed to provide 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 (±38.9 million std. dev.) 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, the gut represents a compartmentalized organ system with compartmentalized functions and the sequenced gut tissues were from each of these portions. Therefore, we focused our analysis on mRNA transcribed in different tissues of the gut and explored evidence of microbiome regulation. Gene set enrichment analyses were used to reveal the presence of photoperiod and circadian-related transcripts in the pyloric caecum, 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.

Keywords

RNA-seq, multi-tissue, lake sturgeon, gastrointestinal tract, bacteria, immune response, lake sturgeon

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 their 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 reduced representation sequencing for individual genotype information, and may miss patterns of admixture and hierarchical structure (6,7). Moreover, reference-free reduced representation sequencing is more vulnerable to stochasticity in results than reference-based approaches (i.e., with a reference genome or transcriptome) due to the bioinformatics pipelines used (8). 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 done using specific primers developed to assay mRNA abundance in the species (9–11), the lack of a publicly available lake sturgeon transcriptome and genome has hindered molecular physiology and environmental DNA work (12).

Furthermore, the early divergence of sturgeons (13,14) make representative species such as lake sturgeon useful for studying questions about vertebrate evolution. For example, the pyloric caecum was first studied by Aristotle, who hypothesized about storage, fermentation, and digestive functions and caeca in fish digestive tracts were then later determined to increase gut surface area for digestion and absorption (15,16). Sturgeons represent the first evolutionary appearance of fused caeca with increased surface area (17,18), 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 caecum should not be viewed as a basal state for Actinopterygii given that evolution in certain genes has been observed in other sturgeons and paddlefishes, and has presumably occurred in lake sturgeon as well (19–21). Nevertheless, the lake sturgeon pyloric caecum can be used as a representative tissue to study the evolution of vertebrate digestion (22).

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 (23–26). The lake sturgeon microbiome has been linked to its physiological state, providing evidence for host-microbe interactions (27–29). However, regulation of gut microbiota across different gut tissues has been well-characterized in only a few species, mostly humans and lab mice (26,30–32). With messenger RNA 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 Archaebacteria. Therefore, RNA sequencing in the lake sturgeon can be used to study gut microbiome heterogeneity and regulation, with implications for the evolution of gut microbiome regulation across vertebrates. While the community structure of the microbiome is heavily influenced by environmental factors (33), the hypothetical presence of heterogeneity in genera and genes in the microbial community across the lake sturgeon gut would suggest the presence of tissue-specific microbial regulatory mechanisms. Moreover, fish may have been the first group of microbiome hosts to evolve an innate capacity for microbiome regulation (33). Ancient bony fish such as the lake sturgeon are thus valuable for studying the dynamics between host and microbiome.

While assembling a genome for a polyploid fish involves extensive chromatin and long-read DNA sequencing (20), assembling a transcriptome is a more tractable task. Such transcriptomes enable in-depth analyses of molecular physiology, such as in population-specific thermal stress responses (34–36). While RNA-seq and transcriptome-based approaches are less commonly used for population genetics than DNA-based approaches, single nucleotide polymorphisms in RNA can be used to investigate population structure and signatures of selection (37–41). Transcriptomes would also enable investigations into evolution, both through descriptions of gene expression evolution (42,43) and by phylogenetic analyses of mutations (44,45). Therefore, transcriptome assembly, annotation, and dissemination enables broad research questions in physiology and genetics.

Multi-tissue and tissue-specific approaches to transcriptome assembly allow for more systematic and in-depth analysis than typical transcriptomics, which may use one tissue for more

91 focused investigations. For example, tissue-specific analyses revealed specialization in tissues
92 and stages in cell division, photosynthesis, auxin transport, stress responses, and secondary
93 metabolism in the tomato (*Solanum pimpinellifolium*) (46). In a livebearing fish, *Poeciliopsis*
94 *prolifica*, multiple tissues were used with RNA-seq data to investigate placental evolution, where
95 the abundance of clusters of transcripts was associated with different tissues (47). In the Atlantic
96 salmon (*Salmo salar*), a blood-specific transcriptome was compared to other tissue
97 transcriptomes to identify genes and gene ontology terms unique to blood (48). Thus
98 transcriptome assembly with multiple tissues would provide a stronger resource for molecular
99 research than single-tissue approaches.

100 One concern about transcriptomics in the lake sturgeon is that assembly may be affected
101 by the octoploid status of the species (1). For instance, in situations where a transcriptome must
102 be assembled without a reference genome, polyploid species are vulnerable to homeologs and
103 ambiguous but similar sequences that decrease accuracy in the final assembly (49). One solution
104 is to create transcriptomes specific to different conditions that may isolate different gene
105 isoforms (49), a strategy consistent with the benefits of a multi-tissue, tissue-specific approach.
106 In addition, different transcriptome assembly programs had varying success at accurately
107 assembling polyploid transcriptomes, and careful selection of an assembly program such as
108 Trinity can at least partly address challenges introduced by high ploidies (49–52).

109 In this study, we assembled, annotated, and analyzed the transcriptomes of 13 tissues in
110 the lake sturgeon, sequenced with short-read messenger RNA sequencing (i.e., Illumina). The
111 tissues were: brain, esophagus, gill, head kidney, heart, white muscle, liver, glandular stomach,
112 muscular stomach, anterior intestine, pyloric cecum, spiral valve, and rectum (Fig. 1). In
113 addition, all data used to assemble each of the 13 tissue-specific transcriptomes was 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). 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 compartmentalized functions throughout (18). Because the 7 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 gene ontology terms (53) 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 caecum because the tissue first appears in sturgeons in terms of increased surface area for digestion and absorption (16,17). In addition, 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 caecum, various types of innate immune regulation across gut tissues, and heterogeneity of bacteria and archaea associated transcripts in the lake sturgeon gut.

Methods

Sampling and Sequencing

Lake sturgeon of approximately 2 years old and unknown sex were sampled haphazardly from holding tanks and euthanized with an overdose of buffered tricaine methanesulfonate solution (MS-222; Sigma-Aldrich). Tissue from each of the gill, liver, brain, head kidney, white muscle, and heart were extracted from the sturgeon, stored in RNA*later* (Thermo Fisher), and the PureLink RNA Mini Kit (Thermo Fisher) was used for RNA extractions following manufacturer protocols. By contrast, gut samples (esophagus, glandular stomach, muscular stomach, anterior intestine, pyloric caecum, spiral valve, and rectum) were immediately placed in Trizol and extracted 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 (54). For all tissues, equal amounts of RNA were pooled from n=3 fish for sequencing.

Total RNA was sent to the Centre d'expertise et de services Génome Québec, Montreal, Quebec (http://gqinnovationcenter.com), where 250 nanograms of total RNA per tissue was used with the NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs). Stranded cDNA 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 reads on one lane of a NovaSeq 6000 (Illumina). A mean of 98.3 million (± 38.9 million standard deviation (s.d.)) reads were sequenced for each tissue (Table 1).

Transcriptome Assembly and Annotation

Trinity was used for transcriptome assembly (52,55), while Trinotate was used for transcriptome annotation (56–62). Both Trinity and Trinotate were used on mRNA sequencing

158 data from each tissue separately as tissue-specific transcriptomes, and with all data from the 13
159 tissues as an overall transcriptome. Trinity was used for its effectiveness at assembling polyploid
160 genomes (49). Transcriptome annotations were filtered for transcripts with E-values < $1x10^{-6}$ and
161 bit scores > 50 (63). BUSCO v5.1.2 was used to assess transcriptome completeness with respect
162 to the Actinopterygii odb10 dataset (64). The fishualize v0.2.3 package in the statistical
163 computing environment R v1.1.2 was used to visualize results (65,66). To assess divergence in
164 terms of mutation distance, Mash v1.1 was used to make pairwise comparisons between each
165 transcriptome (67). Because evolutionary distance is not expected among transcriptomes from
166 lake sturgeon sampled from a single population, the distances measured thus represent isoforms
167 and paralogs among the gene models in the assembled transcriptomes.

168

169 *Annotation Analyses*

170 The statistical computing environment R v1.1.2 and R package Tidyverse v1.3.1 were
171 used throughout functional analyses of lake sturgeon transcriptomes (66,68). A gene set
172 enrichment analyses was used to identify gene ontology terms for each transcriptome using
173 enrichR v3.0 with the Biological Process 2021, Molecular Function 2021, and Cellular
174 Component 2021 databases (53). Only gene ontology terms with a false discovery rate (q) < 0.05
175 were retained as significantly enriched. The R package UpSetR v1.4.0 was used to assess
176 uniqueness of gene ontology terms among tissues (69). This assessment of uniqueness among
177 tissues was repeated in an analysis excluding peripheral tissues and specific to the esophagus,
178 glandular stomach, muscular stomach, pyloric caecum, anterior intestine, spiral valve, and
179 rectum to identify patterns specific to lake sturgeon gut tissues.

Principal components analysis (PCA) was used to visualize differentiation among different tissues with respect to present genes and gene ontology terms. The overall transcriptome was excluded from PCA to explore variance among the tissue-specific transcriptomes. A PCA with *prcomp* in R was used on a table of 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 gene ontology databases. 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 the tissue-specific transcriptomes. However, the number of unique genes was assessed in the overall transcriptome to identify genes that were potentially missing or un-annotated from each tissue-specific transcriptome, but resolved with all data used in one assembly. The genes unique to the overall transcriptome were analyzed for gene ontology 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. Therefore, only transcripts annotated to bacteria or archaea remained from each transcriptome. A gene set enrichment analysis was performed on the microbe-annotated transcripts, but no significant gene ontology terms were identified. UpSet plots were used to identify 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 (± 24,405 s.d.), while the mean number of transcripts was 197,567 (± 51,481 s.d.) (Table 1). After filtering out transcripts with annotations with bit scores < 50 and E-values > $1x10^{-6}$, the number of transcripts remaining with annotations among the tissue-specific transcriptomes was a mean of 76,319 (± 19,737 s.d.) representing a mean 12,350 (± 1,217 s.d.) unique genes with annotations (Table 1). 3,065 genes were identified in analysis of uniqueness that included the overall and tissue-specific transcriptomes, although uniqueness from gene sets for the tissue-specific transcriptomes were 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% ± 15.7% s.d.) (Fig. 2). Divergence among tissue-specific transcriptomes, as assessed by 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 (Fig. 3A).

Annotation Analyses

A mean of 832 (± 123 s.d.) biological process gene ontology terms were identified among the 13 tissue-specific transcriptomes (Table 2). Among biological process gene ontology terms, 367 were shared across all tissues, but a substantial number were also unique to individual tissues such as 101 gene ontology terms in the liver and 71 in the heart (Fig. 4; Supplementary Tables S1-S13). Qualitatively similar patterns of shared gene ontology 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 & S3). A similar pattern of uniqueness was also present among biological process gene ontology 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 (± 19 s.d.) and 150 (± 21 s.d.) gene ontology terms were identified, respectively. Molecular function and cellular component gene ontology 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 gene ontology terms were identified from the overall transcriptome, possibly because the Fisher exact test used in enrichR was implemented for experimental designs, as opposed to surveys of gene presence (70). Five molecular function gene ontology terms were significant in a search of the 3,065 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 ontgoloy terms were: 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 gene ontology database (Fig. 3B; Supplementary Figure S5). Gut tissues tended to cluster together compared to

246 other tissues in each PCA, but because gut and peripheral tissues were extracted using separate
247 protocols, some differentiation between the two groups of tissues may be a technical artifact.

248 In the pyloric caecum, the gene ontology terms photoperiodism (GO:0009648) and
249 entrainment of circadian clock by photoperiod (GO:0043153) were uniquely present and related
250 to periodicity (Supplementary Table S11). Patterns of tissue-specific immune regulation were
251 uniquely present in several gut tissues. Rac protein signal transduction (GO: 0016601) was
252 uniquely present in the glandular stomach and may also represent a part of the innate immune
253 system with its role in neutrophil recruitment (Supplementary Table S8). Negative regulation of
254 immune response (GO:0045824) and autophagy of peroxisomes (GO:0030242) were unique to
255 the muscular stomach (Supplementary Table S9). Positive regulation of host by viral
256 transcription (GO:0043923) was uniquely present in the anterior intestine (Supplementary Table
257 S10). Toll-like receptor 9 signaling pathway (GO:0034162), toll-like receptor signaling pathway
258 (GO:0002224), and cellular response to interleukin-12 (GO:0071349) were each uniquely
259 present in the spiral valve, consistent with a role for the tissue in the innate immune system
260 (Supplementary Table S12). Positive regulation of viral life cycle (GO:1903902) was uniquely
261 present in the esophagus (Supplementary Table S2).

262

263 *Microbial Analyses*

264 A mean of 38 (± 19 s.d.) bacterial and archaeal genera were observed among 13
265 transcriptomes. A mean of 73 (± 52 s.d.) genes were annotated to bacteria or archaea among the
266 same 12 transcriptomes. Both microbial genera and annotated genes showed a pattern of high

267 uniqueness in each tissue, although 8 genera and 7 genes were present among all tissues (Fig. 5;
268 Supplementary Figure S6).

269

270 **Discussion**

271 *A Database of Transcriptomes*

272 The 13 tissue-specific transcriptomes and one overall transcriptome presented in this
273 study are a genomic resource publicly available for studying sturgeons. Transcriptomes are most
274 commonly used for molecular physiology, and the gill transcriptome presented here has already
275 been applied to study thermal stress between latitudinally separated populations of lake sturgeon
276 (36). These transcriptomic resources may be used to characterize physiological responses to
277 environmental conditions, which may in turn be used to inform conservation management (71).
278 As lake sturgeon represent a species that exhibits extensive phenotypic plasticity (9,72,73), these
279 transcriptomes also have potential for supporting fundamental research on the molecular basis of
280 resilience to environmental change. Gene ontology terms revealed tissue-specific patterns in each
281 of the transcriptomes presented here, such as 101 biological process terms unique to the liver and
282 71 unique to the heart. These terms thus represent transcriptional processes that would otherwise
283 have been missing from the transcriptome database if only a single tissue was considered.
284 Therefore, the 13 tissues we studied enable a broad range of analyses that would be otherwise
285 intractable, allowing for in-depth assessments of shared and tissue-specific processes, along with
286 genetic and physiological studies.

287 Among the 13 tissue-specific transcriptomes assembled, 7 were from the gut (esophagus,
288 glandular stomach, muscular stomach, pyloric caecum, anterior intestine, spiral valve, and

289 rectum), and 6 were from peripheral tissues (brain, gill, head kidney, heart, liver, and white
290 muscle). The gut and peripheral tissue transcriptomes separated into different groups using two
291 methods (PCA with present genes and metagenome distance estimation), although some
292 separation between the two groups of tissues may be attributed to different RNA extraction
293 methods used. Nevertheless, the distinction between the two groups is consistent with differences
294 in physiological function. The peripheral tissue transcriptomes enable a variety of research
295 questions on the lake sturgeon, such as liver and gill often used in work exploring the vertebrate
296 stress response (74,75). Tissues such as the brain, heart, head kidney, and white muscle can be
297 informative for developmental questions, with potential connections to nutrition and stress
298 among other biological processes (74,76–79). Meanwhile, the 7 gut transcriptomes represent the
299 major anatomically distinct regions of the gut. Because the gut encompasses an organ system
300 with distinct compartmentalization of function (18), the 7 gut transcriptomes provide an
301 opportunity to study distinct digestion-related mechanisms. Gut transcriptomics was predicted to
302 accelerate research on intestinal pathogen responses, dietary manipulations, and osmoregulatory
303 challenges (26), and the present data contribute to a longstanding body of work investigating
304 physiological mechanisms of the vertebrate gut (15,16). We thus provide all 13 tissue-specific
305 transcriptomes and one overall transcriptome as a scientific resource from this study, but focus
306 on discussing observations among the gut tissues.

307

308 *Circadian Rhythm Transcripts in the Pyloric Caecum*

309 The pyloric caecum is a tissue of interest because it is absent in Agnatha and
310 Chondrichthyes, but is present in Actinopterygii (17,18,22). Given sturgeon's status as an ancient
311 actinopterygian, sister to the rest of the clade (13), they represent an early evolutionary

312 appearance of the pyloric caecum. Analyses of tissue-specific transcriptome annotations revealed
313 notable patterns of transcript presence within pyloric caecum, which may have implications for
314 different mechanisms of lake sturgeon digestion. For instance, gene ontology terms
315 photoperiodism (GO:0009648) and entrainment of circadian clock by photoperiod (GO:
316 0043153) were unique to the pyloric caecum. In addition, among the core clock genes of *clock*,
317 *bmal1*, *per* (*1*, *2*, and *3*), and *cry* (*1* and *2*), all expressed in teleost pineal organs (80), *clock*,
318 *cry1*, and *cry2* were present in the lake sturgeon pyloric caecum. *Cry1* and *cry2*, which are
319 photoreceptors with important roles in circadian rhythms, were also transcribed in brain, liver,
320 heart, retina, muscle, spleen, gill, and intestine of European seabass (*Dicentrarchus labrax*),
321 where rhythmic expression was observed in the brain and liver (81). Notably, the clock-related
322 gene ontology terms observed in the present data were unique to the pyloric caecum, but
323 individual genes were present among other tissues such as the brain. Gene ontology terms used
324 in the present analyses were filtered for significance from a Fisher exact test (70). Therefore,
325 transcriptomes with annotated clock genes but without enriched gene ontology terms present
326 represent those with too few genes within the clock-related gene ontology terms to be significant.
327 The present results do not contradict prior work that identified clock genes in other tissues, but
328 do provide novel findings of core clock genes in the pyloric caecum that may be related to
329 feeding periodicity.

330 Feeding periodicity has been observed in numerous fish species (e.g., *Merlagius*
331 *merlangus*, (82); *Limanda limanda*, (83)), across herbivores, detritovores, insectivores,
332 zooplanktivores, and macrophyte feeders (84). Because feeding periodicity is phylogenetically
333 and ecologically widespread among fishes, we predict that physiological digestive mechanisms
334 may contribute to the phenomenon. A circadian rhythm in metabolic rate was observed in lake

335 sturgeon exposed to a 12-hour light-dark cycle, where metabolism was highest at sunrise (85).
336 The lake sturgeon used for the present study were also fed predictably, three times a day with a
337 12-hour light-dark cycle. Given the dual observations that feeding periodicity exists across fish
338 species (84) and the presence of several core clock genes in the lake sturgeon pyloric caecum, we
339 developed alternative hypotheses that may address underlying mechanisms of periodicity in the
340 lake sturgeon, that may be applicable in other fishes (86).

341 First, we hypothesized that physiological mechanisms of digestion periodicity may be
342 regulated by diel circadian clock rhythms in the pyloric caecum. Therefore, we predict diel
343 fluctuations in transcript abundance of *clock*, *cry1*, and *cry2* along with other circadian rhythm-
344 related genes only in the pyloric caecum of laboratory-held lake sturgeon consistent with feeding
345 times and a 12-hour light-dark cycle. Specific roles for physiological mechanisms of digestion
346 periodicity could include intestinal motility, intestinal function, innate immunity, microbiome
347 regulation, or cell proliferation (87–94). Alternatively, we hypothesized that the circadian rhythm
348 genes observed in the pyloric caecum may represent a part of a whole-gut circadian response
349 wave consistent with a phenomenon hypothesized in lab mice (95). That is, a whole-gut
350 circadian response may have been initiated at feeding and was observed by chance in the pyloric
351 caecum by sampling individuals approximately 17-18 hours after feeding. Therefore, from this
352 hypothesis we predict diel fluctuations of transcript abundance of *clock*, *cry1*, and *cry2* in the
353 pyloric caecum, and other gut tissues. More posterior gut tissues, such as the spiral valve, may
354 therefore show expression of the three predicted genes but chronologically later than the pyloric
355 caecum. By contrast, the glandular and muscular stomachs may show evidence of this circadian
356 response wave earlier in time than the pyloric caecum because of their position prior to the
357 caecum in the gut. While the first hypothesis about physiological functions of digestion as

358 regulated by the pyloric caecum focuses on one gut tissue, the hypotheses are not necessarily
359 mutually exclusive. A circadian response wave may pass through different gut tissues, but the
360 pyloric caecum may play key roles in downstream processes from the circadian wave. This
361 circadian response wave may be entrained by food intake times if it is consistent with
362 mammalian physiology (96). Therefore, a timepoint- and tissue-specific approach is needed to
363 test this hypothesis.

364

365 *Microbial Observations*

366 As lake sturgeon represent an early stage of Actinopterygian gut evolution, several
367 observations in the present database of tissue-specific transcriptomes were notable. One example
368 is the presence of transcripts related to innate immunity in the spiral valve. As gut tissues may be
369 in contact with food and potential associated pathogens from the external environment, innate
370 immunity and immune responses involved in digestion may help to protect the fish from food or
371 environment related pathogens (97). Cartilaginous fishes have gut-associated lymphoid tissues in
372 the spiral valve (98,99), consistent with the present observations of innate immune-related
373 transcripts in the lake sturgeon spiral valve. Gene ontology terms related to toll-like receptor
374 signaling pathways and cellular responses to interleukin were unique to the spiral valve in lake
375 sturgeon, while terms related to innate immune function were also present in the muscular
376 stomach, glandular stomach, and anterior intestine. The gene ontology term positive regulation
377 of viral life cycle is not an immune response in itself, but its unique presence in the esophagus
378 provides some evidence for the necessity of an innate immune response in other gut tissues.
379 These heterogeneous signals of host-microbiome interactions along the gut are consistent with

380 evidence of host-microbiome interactions from 16S rRNA and DNA sequencing from microbial
381 samples and the lake sturgeon spiral valve (27).

382 The tissue-specific transcriptome database enabled analyses of bacterial and archaeal
383 transcripts as well as genera across different gut tissues. Many of these microbial transcripts and
384 genera were unique to different gut sections, thus we concluded that the microbiome is likely
385 heterogenous across the lake sturgeon gut. A caveat is that the presence of microbial genera and
386 transcripts in certain tissues may be attributed to contamination, such as in the brain, white
387 muscle, and heart, along with transcripts or genera shared among all tissues (7 genes and 8
388 genera identified as shared among all 13 tissues). However, patterns of microbial presence were
389 consistent with microbiome regulation and tissue-specific function for gut tissues. For example,
390 among microbial genera unique to each tissue, the muscular stomach had the greatest number
391 present (15), followed by the spiral valve (12), and the anterior intestine (6). A qualitatively
392 similar pattern was found with transcripts of genes annotated to bacteria and archaea unique to
393 each tissue, with the muscular stomach (114), spiral valve (101), esophagus (50), glandular
394 stomach (23), rectum (18), and pyloric caecum (14) all supporting unique microbial
395 communities. These results demonstrate that the greatest number of unique microbial genera and
396 genes were identified in gut tissues as opposed to tissues outside of the gut. Therefore, the
397 present results are consistent with a heterogeneous microbial community with tissue-specific
398 mRNA transcription in the lake sturgeon gut.

399 Other work identified microbial community shifts in the lake sturgeon spiral valve in
400 response to a failure to transition diets and with feeding cessation (29). Similarly, spiral valve
401 microbiome community composition changed in response to exposure to common antibiotics,
402 drugs, and chemicals used in lake sturgeon aquaculture (28). Therefore, gut microbiome

403 community composition is dynamically connected to the physiological state of lake sturgeon
404 (27). Because unique patterns of microbial genera and genes were found in both the spiral valve
405 and other gut tissues in the present data, the analyses of gut transcriptomes demonstrate that host-
406 microbiome interactions may occur along much of the lake sturgeon gut, and that the interactions
407 may be spatially heterogeneous and specific to different gut tissues. Thus the transcriptomes used
408 here may support work in the lake sturgeon that resolves spatially distinct mechanisms of host-
409 microbiome interactions.

410

411 *Conclusions*

412 In the present study, 13 tissue-specific transcriptomes and one overall transcriptome were
413 presented as a resource for lake sturgeon research. Overlap of gene ontology terms was analyzed
414 among tissues. While shared patterns indicated consistent transcriptomic functions among
415 tissues, the presence of unique gene ontology terms showed that sequencing transcriptomes from
416 multiple tissues enabled research questions that would otherwise be intractable. Moreover, the
417 analysis of unique gene ontology terms among tissues revealed the presence of transcribed genes
418 related to photoperiodicity in the pyloric caecum, an observation consistent with a role for
419 periodicity in digestive physiology in an ancient fish. Transcripts involved in innate immune
420 function were found in the spiral valve and other gut tissues, which provide evidence in support
421 of a prior hypothesis about the emergence of innate immunity in the gut of cartilaginous fishes
422 and are consistent with specialization in immune function across gut tissues. An analysis of
423 genes annotated to bacteria and archaea indicated potentially heterogeneous microbiota and
424 microbial functions along different gut tissues, consistent with specialization in immune function
425 and microbiome regulation along the lake sturgeon gut. As lake sturgeon 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 differentiation of the gut as an organ system with distinct functions provided insights into digestive function, immunity, and microbiome regulation. These results are both a resource for lake sturgeon research and provide information about the mechanisms of compartmentalized function across gut tissues.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The lake sturgeon transcriptomes and annotations are available on Figshare https://figshare.com/projects/Lake_Sturgeon_Transcriptomes/133143. Code used in analyses of the data in the present study are available on GitHub (https://github.com/BioMatt/lakesturgeon_transcriptomes).

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.

Funding

The present work was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC)/Manitoba Hydro Industrial Research Chair awarded to W.G.A. Funding from NSERC Discovery Grants awarded to W.G.A. (#05348) and K.M.J. (#05479) was also used.

Conflicts of Interest

455 The authors declare no conflicts of interest.

456

457 **Authors' contributions**

458 M.J.T. performed bioinformatics analyses on the data and wrote the initial draft of the
459 manuscript, with assistance from A.M.W. and W.S.B. All authors edited the manuscript. In
460 addition, all authors designed and conceived of the study. W.G.A and K.M.J. provided funding
461 and supervision.

462

463 **References**

464 1. Fontana, F., Bruch, R.M., Binkowski, F.P., et al. (2004) Karyotype characterization of the lake
465 sturgeon, *Acipenser fulvescens* (Rafinesque 1817) by chromosome banding and fluorescent in
466 situ hybridization. *Genome*, **47**, 742–746.

467 2. Welsh, A. and May, B. (2006) Development and standardization of disomic microsatellite
468 markers for lake sturgeon genetic studies. *J Appl Ichthyol*, **22**, 337–344.

469 3. McQuown, E., Gall, G.A.E. and May, B. (2002) Characterization and Inheritance of Six
470 Microsatellite Loci in Lake Sturgeon. *Transactions of the American Fisheries Society*, **131**, 299–
471 307.

472 4. Welsh, A.B., Blumberg, M. and May, B. (2002) Identification of microsatellite loci in lake
473 sturgeon, Acipenser fulvescens, and their variability in green sturgeon, *A. medirostris*: PRIMER
474 NOTE. *Molecular Ecology Notes*, **3**, 47–55.

475 5. Whitaker, J.M., Price, L.E., Boase, J.C., et al. (2020) Detecting fine-scale population structure
476 in the age of genomics: a case study of lake sturgeon in the Great Lakes. *Fisheries Research*,
477 **230**, 105646.

478 6. Bohling, J., Small, M., Von Bargen, J., et al. (2019) Comparing inferences derived from
479 microsatellite and RADseq datasets: a case study involving threatened bull trout. *Conserv Genet*,
480 **20**, 329–342.

481 7. Lemopoulos, A., Prokkola, J.M., Uusi-Heikkilä, S., et al. (2019) Comparing RADseq and
482 microsatellites for estimating genetic diversity and relatedness — Implications for brown trout
483 conservation. *Ecol Evol*, **9**, 2106–2120.

484 8. Shafer, A.B.A., Peart, C.R., Tusso, S., et al. (2017) Bioinformatic processing of RAD-seq data
485 dramatically impacts downstream population genetic inference. *Methods Ecol Evol*, **8**, 907–917.

486 9. Bugg, W.S., Yoon, G.R., Schoen, A.N., et al. (2020) Effects of acclimation temperature on the
487 thermal physiology in two geographically distinct populations of lake sturgeon (*Acipenser*
488 *fulvescens*). *Conservation Physiology*, **8**, coaa087.

10. Bugg, W.S., Jeffries, K.M. and Gary Anderson, W. (2021) Survival and gene expression responses in immune challenged larval lake sturgeon. *Fish & Shellfish Immunology*, **112**, 1–7.

11. 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*). *Journal of Fish Biology*, **99**, 1912–1927.

12. 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. *Conservation Gesnet Resour*, **12**, 17–19.

13. Gardiner, B.G. (1984) Sturgeons as Living Fossils. In Eldredge, N., Stanley, S. M. (eds.), *Living Fossils*, Casebooks in Earth Sciences, Springer New York, New York, NY, pp. 148–152.

14. Choudhury, A. and Dick, T.A. (2003) The historical biogeography of sturgeons (Osteichthyes: Acipenseridae): a synthesis of phylogenetics, palaeontology and palaeogeography. *Journal of Biogeography*, **25**, 623–640.

15. Aristotle. Parts of Animals. Movement of Animals. Progression of Animals. *Parts of Animals. Movement of Animals. Progression of Animals*; Peck, A. L., Forster, E. S. (eds.); The Loeb classical library; Reprinted.; Harvard Univ. Press, Cambridge, Mass.

16. 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.

17. Buddington, R.K. and Diamond, J.M. (1987) Pyloric ceca of fish: a "new" absorptive organ. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, **252**, G65–G76.

18. Wilson, J.M. and Castro, L.F.C. (2010) Morphological diversity of the gastrointestinal tract in fishes. *Fish Physiology*, Elsevier, Vol. 30, pp. 1–55.

19. 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 Letters*, **584**, 3331–3339.

20. 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.

21. 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. *Molecular Biology and Evolution*, **38**, 1595–1607.

22. 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.

23. 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.

24. Holt, C.C., Bass, D., Stentiford, G.D., et al. (2021) Understanding the role of the shrimp gut microbiome in health and disease. *Journal of Invertebrate Pathology*, **186**, 107387.

525 25. Pratte, Z.A., Perry, C., Dove, A.D.M., et al. (2022) Microbiome structure in large pelagic
526 sharks with distinct feeding ecologies. *Anim Microbiome*, **4**, 17.

527 26. Martin, A.M., Sun, E.W., Rogers, G.B., et al. (2019) The Influence of the Gut Microbiome
528 on Host Metabolism Through the Regulation of Gut Hormone Release. *Front. Physiol.*, **10**, 428.

529 27. Abdul Razak, S. and Scribner, K.T. (2020) Ecological and Ontogenetic Components of
530 Larval Lake Sturgeon Gut Microbiota Assembly, Successional Dynamics, and Ecological
531 Evaluation of Neutral Community Processes. *Appl Environ Microbiol*, **86**, e02662-19.

532 28. Abdul Razak, S., Bauman, J.M., Marsh, T.L., et al. (2022) Changes in Lake Sturgeon Gut
533 Microbiomes Relative to Founding Origin and in Response to Chemotherapeutant Treatments.
534 *Microorganisms*, **10**, 1005.

535 29. Abdul Razak, S., Valentine, S., Marsh, T., et al. (2022) Compositional Dynamics of
536 Gastrointestinal Tract Microbiomes Associated with Dietary Transition and Feeding Cessation in
537 Lake Sturgeon Larvae. *Microorganisms*, **10**, 1872.

538 30. Zhang, Z., Geng, J., Tang, X., et al. (2014) Spatial heterogeneity and co-occurrence patterns
539 of human mucosal-associated intestinal microbiota. *ISME J*, **8**, 881–893.

540 31. Swidsinski, A. (2005) Spatial organization of bacterial flora in normal and inflamed intestine:
541 A fluorescence *in situ* hybridization study in mice. *WJG*, **11**, 1131.

542 32. Laukens, D., Brinkman, B.M., Raes, J., et al. (2016) Heterogeneity of the gut microbiome in
543 mice: guidelines for optimizing experimental design. *FEMS Microbiology Reviews*, **40**, 117–132.

544 33. Sullam, K.E., Essinger, S.D., Lozupone, C.A., et al. (2012) Environmental and ecological
545 factors that shape the gut bacterial communities of fish: a meta-analysis: FISH GUT
546 BACTERIAL COMMUNITIES. *Molecular Ecology*, **21**, 3363–3378.

547 34. Hale, M.C., McCormick, C.R., Jackson, J.R., et al. (2009) Next-generation pyrosequencing
548 of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits
549 of normalization and rarefaction in gene discovery. *BMC Genomics*, **10**, 203.

550 35. Doering, J.A., Tang, S., Peng, H., et al. (2016) High Conservation in Transcriptomic and
551 Proteomic Response of White Sturgeon to Equipotent Concentrations of 2,3,7,8-TCDD, PCB 77,
552 and Benzo[a]pyrene. *Environ. Sci. Technol.*, **50**, 4826–4835.

553 36. Bugg, W.S., Thorstensen, M.J., Marshall, K.E., et al. (2022) Transcriptome-wide patterns
554 reveal conserved and population-specific responses to increasing acclimation temperatures in
555 developing lake sturgeon (*Acipenser fulvescens*). *bioRxiv*, (2022) .

556 37. Harder, A.M. and Christie, M.R. (2022) Genomic signatures of adaptation to novel
557 environments: hatchery and life history-associated loci in landlocked and anadromous Atlantic
558 salmon (*Salmo salar*). *Can. J. Fish. Aquat. Sci.*, **79**, 761–770.

559 38. Thorstensen, M.J., Jeffrey, J.D., Treberg, J.R., et al. (2020) Genomic signals found using
560 RNA sequencing show signatures of selection and subtle population differentiation in walleye
561 (*Sander vitreus*) in a large freshwater ecosystem. *Ecol Evol*, **10**, 7173–7188.

562 39. Thorstensen, M.J., Baerwald, M.R. and Jeffries, K.M. (2021) RNA sequencing describes
563 both population structure and plasticity-selection dynamics in a non-model fish. *BMC Genomics*,
564 **22**, 273.

565 40. Yin, X., Martinez, A.S., Sepúlveda, M.S., et al. (2021) Rapid genetic adaptation to recently
566 colonized environments is driven by genes underlying life history traits. *BMC Genomics*, **22**,
567 269.

568 41. Yin, X., Martinez, A.S., Perkins, A., et al. (2021) Incipient resistance to an effective pesticide
569 results from genetic adaptation and the canalization of gene expression. *Evol Appl*, **14**, 847–859.

570 42. Harrison, P.W., Wright, A.E. and Mank, J.E. (2012) The evolution of gene expression and
571 the transcriptome–phenotype relationship. *Seminars in Cell & Developmental Biology*, **23**, 222–
572 229.

573 43. Yang, J., Ruan, H., Xu, W., et al. (2019) TreeExp2: An Integrated Framework for
574 Phylogenetic Transcriptome Analysis. *Genome Biology and Evolution*, **11**, 3276–3282.

575 44. Minh, B.Q., Schmidt, H.A., Chernomor, O., et al. (2020) IQ-TREE 2: New Models and
576 Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and*
577 *Evolution*, **37**, 1530–1534.

578 45. Pereira, R.J., Barreto, F.S., Pierce, N.T., et al. (2016) Transcriptome-wide patterns of
579 divergence during allopatric evolution. *Mol Ecol*, **25**, 1478–1493.

580 46. Pattison, R.J., Csukasi, F., Zheng, Y., et al. (2015) Comprehensive Tissue-Specific
581 Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit
582 Development. *Plant Physiol.*, **168**, 1684–1701.

583 47. Jue, N.K., Foley, R.J., Reznick, D.N., et al. (2018) Tissue-Specific Transcriptome for
584 *Poeciliopsis prolifica* Reveals Evidence for Genetic Adaptation Related to the Evolution of a
585 Placental Fish. *G3 Genes|Genomes|Genetics*, **8**, 2181–2192.

586 48. Andrew, S.C., Primmer, C.R., Debes, P.V., et al. (2021) The Atlantic salmon whole blood
587 transcriptome and how it relates to major locus maturation genotypes and other tissues. *Marine*
588 *Genomics*, **56**, 100809.

589 49. Voshall, A. and Moriyama, E.N. (2020) Next-generation transcriptome assembly and
590 analysis: Impact of ploidy. *Methods*, **176**, 14–24.

591 50. He, B., Zhao, S., Chen, Y., et al. (2015) Optimal assembly strategies of transcriptome related
592 to ploidies of eukaryotic organisms. *BMC Genomics*, **16**, 65.

593 51. Madritsch, S., Burg, A. and Sehr, E.M. (2021) Comparing de novo transcriptome assembly
594 tools in di- and autotetraploid non-model plant species. *BMC Bioinformatics*, **22**, 146.

595 52. Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly
596 from RNA-Seq data without a reference genome. *Nat Biotechnol*, **29**, 644–652.

597 53. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., et al. (2016) Enrichr: a comprehensive gene
598 set enrichment analysis web server 2016 update. *Nucleic Acids Research*, **44**, W90–W97.

599 54. Heumüller-Klug, S. (2015) Degradation of intestinal mRNA: A matter of treatment. *WJG*,
600 **21**, 3499.

601 55. Haas, B.J., Papanicolaou, A., Yassour, M., et al. (2013) De novo transcript sequence
602 reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.
603 *Nat Protoc*, **8**, 1494–1512.

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

606 57. Punta, M., Coggill, P.C., Eberhardt, R.Y., et al. (2012) The Pfam protein families database.
607 *Nucleic Acids Research*, **40**, D290–D301.

608 58. Petersen, T.N., Brunak, S., von Heijne, G., et al. (2011) SignalP 4.0: discriminating signal
609 peptides from transmembrane regions. *Nat Methods*, **8**, 785–786.

610 59. Krogh, A., Larsson, B., von Heijne, G., et al. (2001) Predicting transmembrane protein
611 topology with a hidden markov model: application to complete genomes. *Journal of Molecular*
612 *Biology*, **305**, 567–580.

613 60. Altschul, S.F., Gish, W., Miller, W., et al. (1990) Basic local alignment search tool. *Journal*
614 *of Molecular Biology*, **215**, 403–410.

615 61. Kanehisa, M., Goto, S., Sato, Y., et al. (2012) KEGG for integration and interpretation of
616 large-scale molecular data sets. *Nucleic Acids Research*, **40**, D109–D114.

617 62. Powell, S., Szklarczyk, D., Trachana, K., et al. (2012) eggNOG v3.0: orthologous groups
618 covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, **40**, D284–
619 D289.

620 63. Pearson, W.R. (2013) An Introduction to Sequence Similarity ("Homology") Searching.
621 *Current Protocols in Bioinformatics*, **42**.

622 64. Simão, F.A., Waterhouse, R.M., Ioannidis, P., et al. (2015) BUSCO: assessing genome
623 assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–
624 3212.

625 65. Schiettekatte, N., Brandl, S. and Casey, J. (2019) fishualize: Color Palettes Based on Fish
626 Species.

627 66. R Core Team (2022) R: A language and environment for statistical computing.

628 67. Ondov, B.D., Treangen, T.J., Melsted, P., et al. (2016) Mash: fast genome and metagenome
629 distance estimation using MinHash. *Genome Biol*, **17**, 132.

630 68. Wickham, H., Averick, M., Bryan, J., et al. (2019) Welcome to the Tidyverse. *JOSS*, **4**, 1686.

631 69. Conway, J.R., Lex, A. and Gehlenborg, N. (2017) UpSetR: an R package for the
632 visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.

633 70. Chen, E.Y., Tan, C.M., Kou, Y., et al. (2013) Enrichr: interactive and collaborative HTML5
634 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.

635 71. Connon, R.E., Jeffries, K.M., Komoroske, L.M., et al. (2018) The utility of transcriptomics in
636 fish conservation. *Journal of Experimental Biology*, **221**, jeb148833.

637 72. Dammerman, K.J., Steibel, J.P. and Scribner, K.T. (2016) Increases in the mean and
638 variability of thermal regimes result in differential phenotypic responses among genotypes
639 during early ontogenetic stages of lake sturgeon (*Acipenser fulvescens*). *Evol Appl*, **9**, 1258–
640 1270.

641 73. Loeppky, A.R., Belding, L.D., Quijada-Rodriguez, A.R., et al. (2021) Influence of
642 ontogenetic development, temperature, and pCO2 on otolith calcium carbonate polymorph
643 composition in sturgeons. *Sci Rep*, **11**, 13878.

644 74. Aluru, N. and Vijayan, M.M. (2009) Stress transcriptomics in fish: A role for genomic
645 cortisol signaling. *General and Comparative Endocrinology*, **164**, 142–150.

646 75. Jeffries, K.M., Teffer, A., Michaleski, S., et al. (2021) The use of non-lethal sampling for
647 transcriptomics to assess the physiological status of wild fishes. *Comparative Biochemistry and*
648 *Physiology Part B: Biochemistry and Molecular Biology*, **256**, 110629.

649 76. Ju, Z., Karsi, A., Kocabas, A., et al. (2000) Transcriptome analysis of channel catfish
650 (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene*, **261**, 373-382.

651 77. Carney, S.A., Chen, J., Burns, C.G., et al. (2006) Aryl Hydrocarbon Receptor Activation
652 Produces Heart-Specific Transcriptional and Toxic Responses in Developing Zebrafish. *Mol*
653 *Pharmacol*, **70**, 549–561.

654 78. Mazurais, D., Darias, M., Zambonino-Infante, J.L., et al. (2011) Transcriptomics for
655 understanding marine fish larval development. *Can. J. Zool.*, **89**, 599–611.

656 79. Sarropoulou, E., Kotoulas, G., Power, D.M., et al. (2005) Gene expression profiling of
657 gilthead sea bream during early development and detection of stress-related genes by the
658 application of cDNA microarray technology. *Physiological Genomics*, **23**, 182–191.

659 80. Saha, S., Singh, K.M. and Gupta, B.B.P. (2019) Melatonin synthesis and clock gene
660 regulation in the pineal organ of teleost fish compared to mammals: Similarities and differences.
661 *General and Comparative Endocrinology*, **279**, 27–34.

662 81. del Pozo, A., Vera, L.M., Sánchez, J.A., et al. (2012) Molecular cloning, tissue distribution
663 and daily expression of cry1 and cry2 clock genes in European seabass (*Dicentrarchus labrax*).
664 *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **163**,
665 364–371.

666 82. Pedersen, J. (2000) Food consumption and daily feeding periodicity: comparison between
667 pelagic and demersal whiting in the North Sea. *Journal of Fish Biology*, **57**, 402–416.

668 83. Gwyther, D. and Grove, D.J. (1981) Gastric emptying in *Limanda limanda* (L.) and the
669 return of appetite. *J Fish Biology*, **18**, 245–259.

670 84. Weliange, W.S., Amarasinghe, U.S., Moreau, J., et al. (2006) Diel feeding periodicity, daily
671 ration and relative food consumption in some fish populations in three reservoirs of Sri Lanka.
672 *Aquat. Living Resour.*, **19**, 229–237.

673 85. Svendsen, J.C., Genz, J., Anderson, W.G., et al. (2014) Evidence of Circadian Rhythm,
674 Oxygen Regulation Capacity, Metabolic Repeatability and Positive Correlations between Forced
675 and Spontaneous Maximal Metabolic Rates in Lake Sturgeon *Acipenser fulvescens*. *PLoS ONE*,
676 **9**, e94693.

677 86. Betts, M.G., Hadley, A.S., Frey, D.W., et al. (2021) When are hypotheses useful in ecology
678 and evolution? *Ecol. Evol.*, **11**, 5762–5776.

679 87. Froy, O. and Chapnik, N. (2007) Circadian oscillation of innate immunity components in
680 mouse small intestine. *Molecular Immunology*, **44**, 1954–1960.

681 88. Goo, R.H., Moore, J.G., Greenberg, E., et al. (1987) Circadian variation in gastric emptying
682 of meals in humans. *Gastroenterology*, **93**, 515–518.

683 89. Hoogerwerf, W.A. (2006) Biologic clocks and the gut. *Curr Gastroenterol Rep*, **8**, 353–359.

684 90. Konturek, P.C., Brzozowski, T. and Konturek, S.J. (2011) Gut clock: Implication of
685 circadian rhythms in the gastrointestinal tract. *J Physiol Pharmacol*, **62**, 139-150.

686 91. Peyric, E., Moore, H.A. and Whitmore, D. (2013) Circadian Clock Regulation of the Cell
687 Cycle in the Zebrafish Intestine. *PLoS ONE*, **8**, e73209.

688 92. Rosselot, A.E., Hong, C.I. and Moore, S.R. (2016) Rhythm and bugs: circadian clocks, gut
689 microbiota, and enteric infections. *Current Opinion in Gastroenterology*, **32**, 7–11.

690 93. Scheving, L.A. (2000) Biological clocks and the digestive system. *Gastroenterology*, **119**,
691 536–549.

692 94. Stevenson, N.R., Sitren, H.S. and Furuya, S. (1980) Circadian rhythmicity in several small
693 intestinal functions is independent of use of the intestine. *American Journal of Physiology-*
694 *Gastrointestinal and Liver Physiology*, **238**, G203–G207.

695 95. Neal, J.V. and Potten, C.S. (1981) Circadian Rhythms In the Epithelial Cells and the
696 Pericryptal Fibroblast Sheath In Three Different Sites In the Murine Intestinal Tract. *Cell Prolif*,
697 **14**, 581–587.

698 96. Bron, R. and Furness, J.B. (2009) Rhythm of digestion: keeping time in the gastrointestinal
699 tract. *Clinical and Experimental Pharmacology and Physiology*, **36**, 1041–1048.

700 97. Gómez, G.D. and Balcázar, J.L. (2008) A review on the interactions between gut microbiota
701 and innate immunity of fish. *FEMS Immunol Med Microbiol*, **52**, 145–154.

702 98. Lauriano, E.R., Pergolizzi, S., Aragona, M., et al. (2019) Intestinal immunity of dogfish
703 *Scyliorhinus canicula* spiral valve: A histochemical, immunohistochemical and confocal study.
704 *Fish & Shellfish Immunology*, **87**, 490–498.

705 99. Mitchell, C.D. and Criscitiello, M.F. (2020) Comparative study of cartilaginous fish divulges
706 insights into the early evolution of primary, secondary and mucosal lymphoid tissue architecture.
707 *Fish & Shellfish Immunology*, **107**, 435–443.

708

Tables and Figures

Table 1. Assembly and annotation statistics for each of 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*). Number of reads represents the number of short read mRNA sequences used in assembly, number of putative genes refers to the number of gene models assembled in Trinity, and number of transcripts represents the total number of transcripts identified. The numbers of transcripts with annotations and number of unique annotated genes refers to annotations performed with Trinotate and associated programs. 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.

Tissue	Number of Reads	Number of Putative Genes	Number of Transcripts	Number of Transcripts with Annotations	Number of Unique Annotated Genes
Brain	52583855	116196	185242	120970	13252
Gill	64243963	165299	250456	103501	12848
Head Kidney	51005951	104228	174924	147225	12363
Heart	61776431	115141	161496	58872	10709
Liver	52606238	68560	98998	47739	9817
White Muscle	65292771	64372	93273	50022	10143
Esophagus	133351646	136859	251276	78883	13319
Glandular Stomach	149907748	130979	243955	77889	13263
Muscular Stomach	138604614	120933	221499	69953	12807
Pyloric Caecum	146607316	118211	220417	71752	12836
Anterior Intestine	114334754	105674	197080	66478	12388
Spiral Valve	112481602	122567	225357	71406	13004
Rectum	134824388	127892	244401	80825	13580
Mean	98278559.77	115147.00	197567.23	80424.23	12333.00
Standard Deviation	38868780.36	25504.38	51480.67	27125.92	1215.21
Overall	1277621277	484570	770984	326367	18290

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

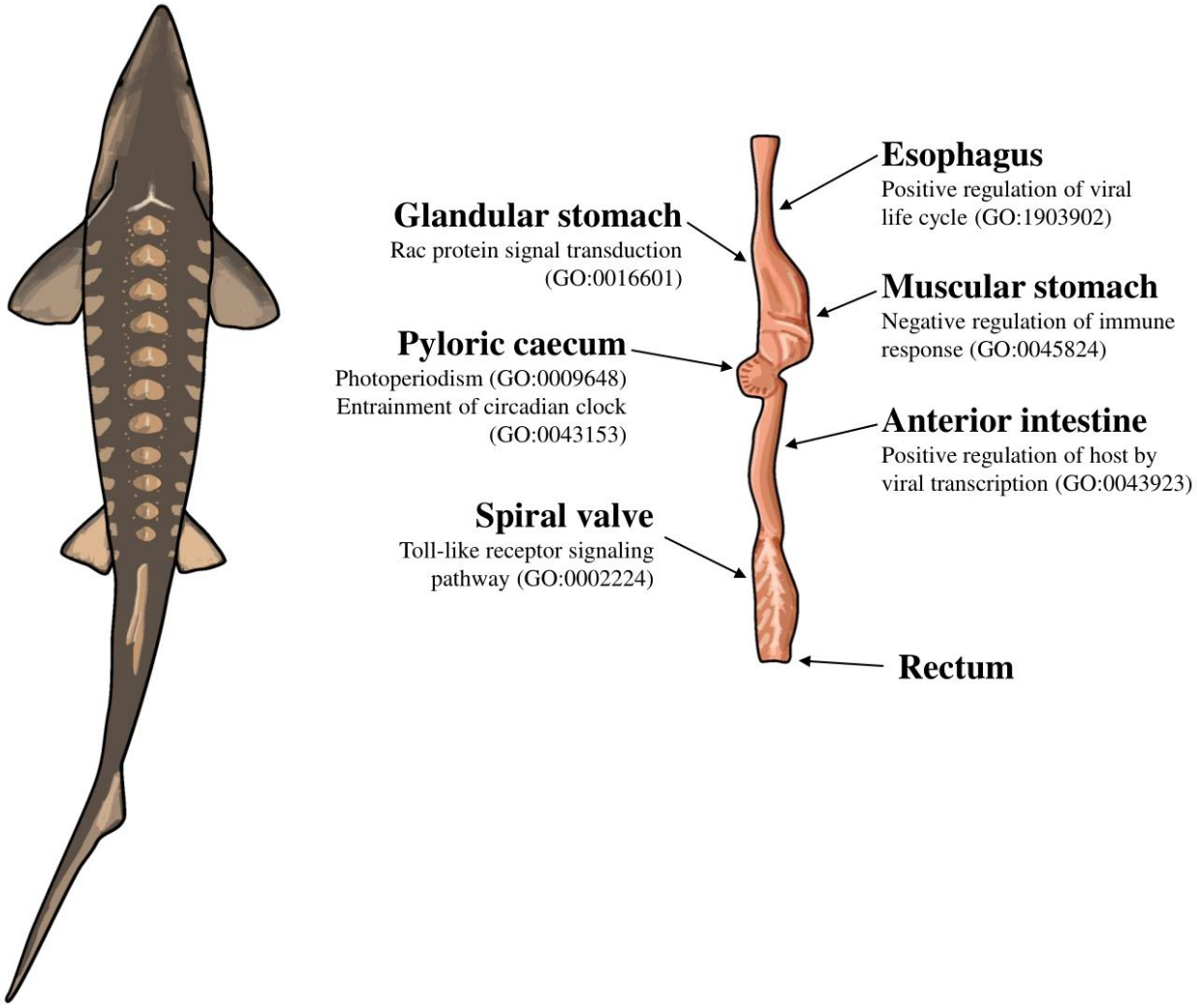
728

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	1017	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 Caecum	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

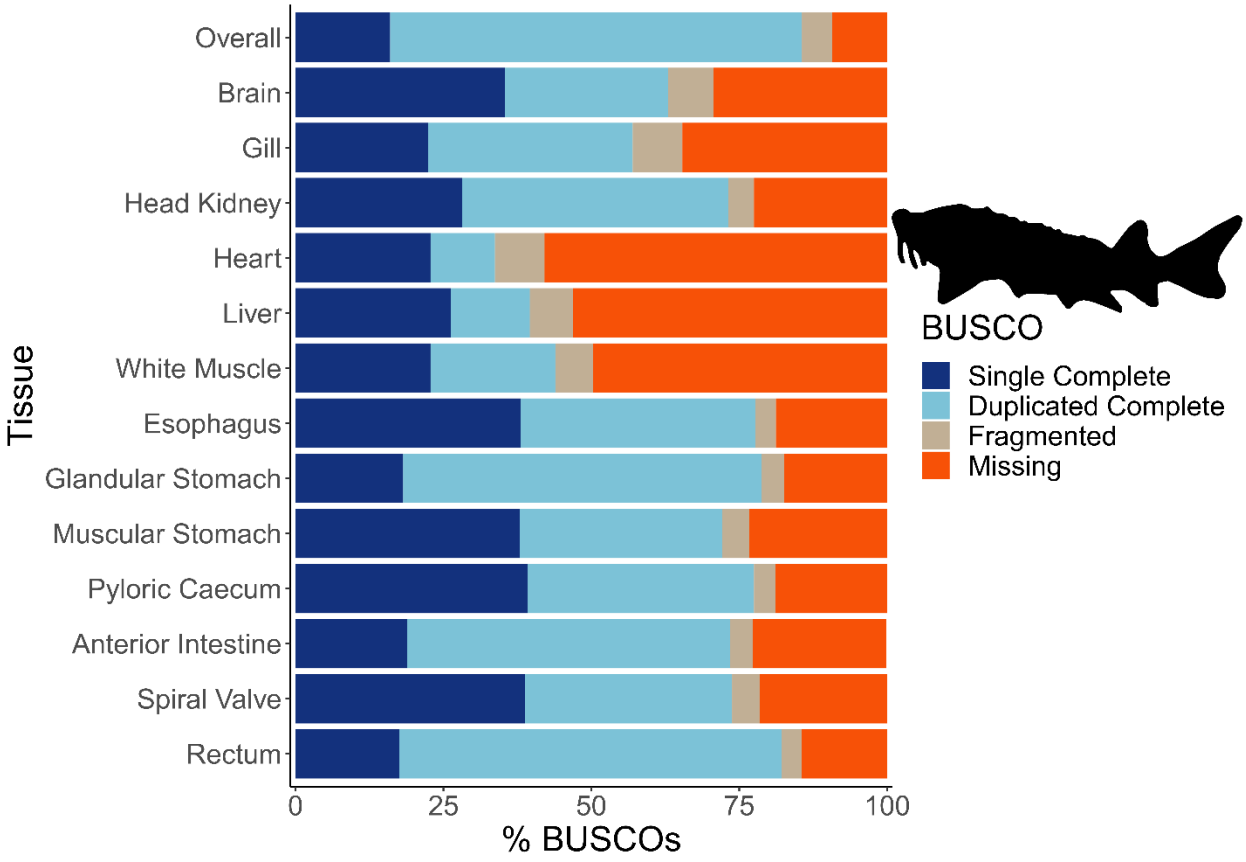
729

730

Figure 1. 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 ($q < 0.05$), gene ontology terms unique to the tissue identified with EnrichR. The gene ontology 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 gene ontology terms present in the pyloric caecum were processes related to circadian rhythms, unique to the tissue among the transcriptomes analyzed.



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

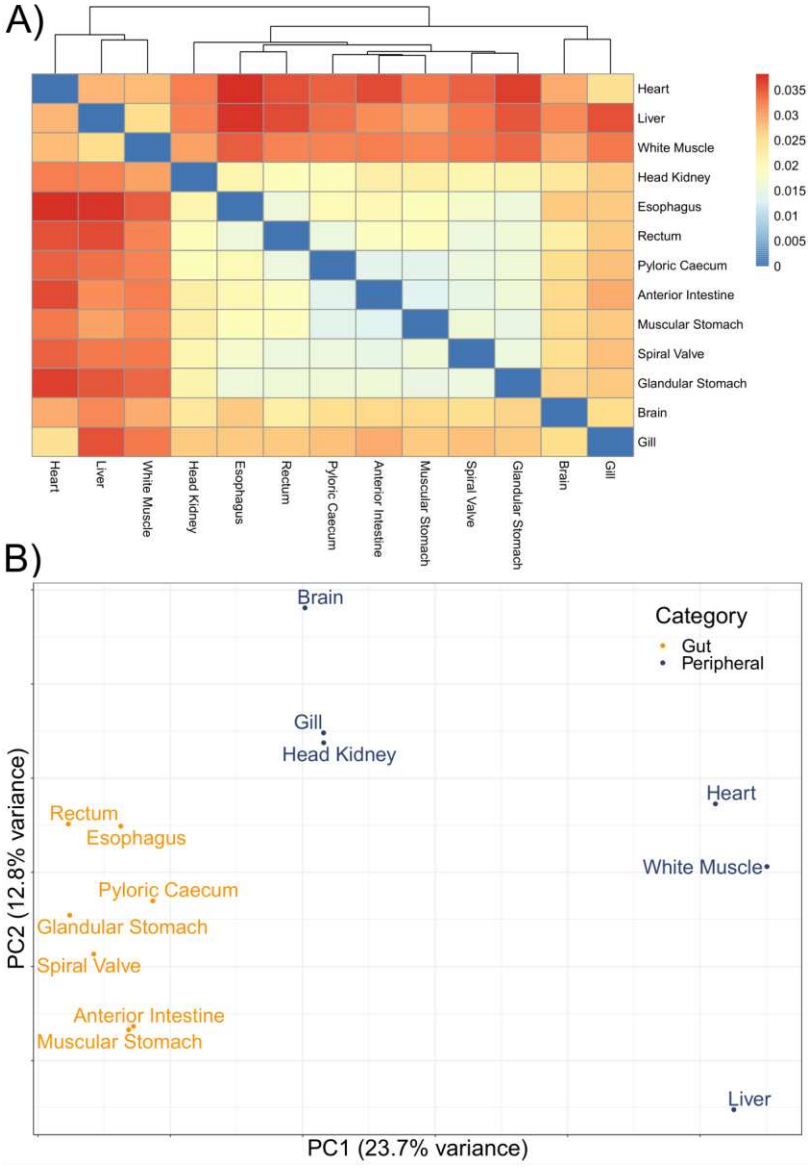


749

750

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

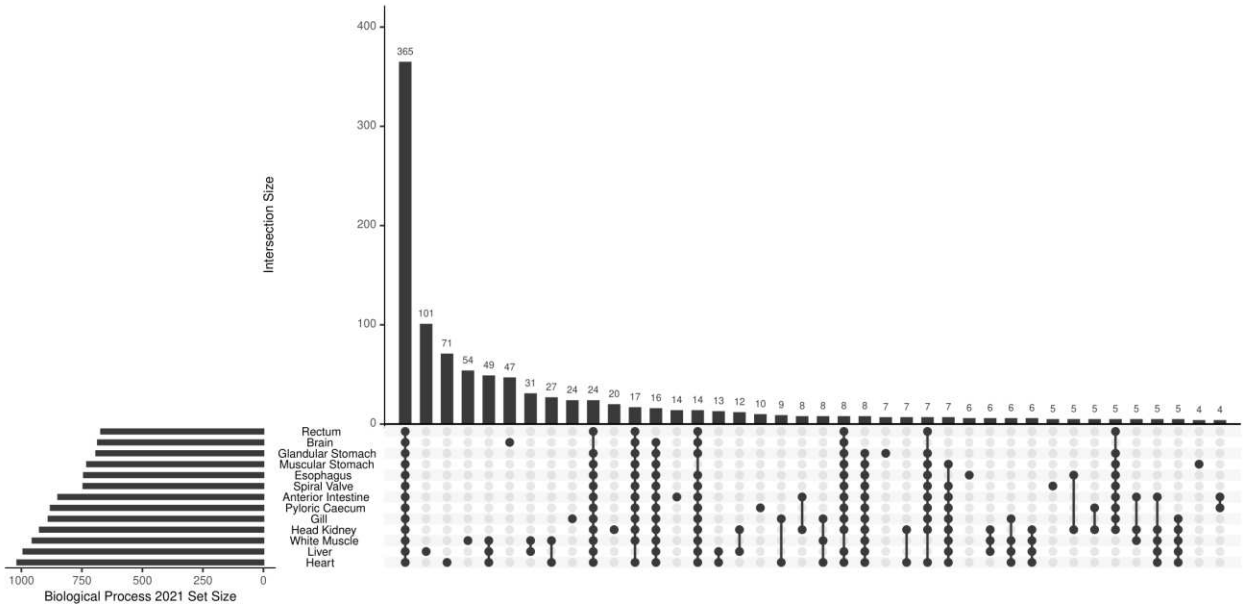
763



764

765

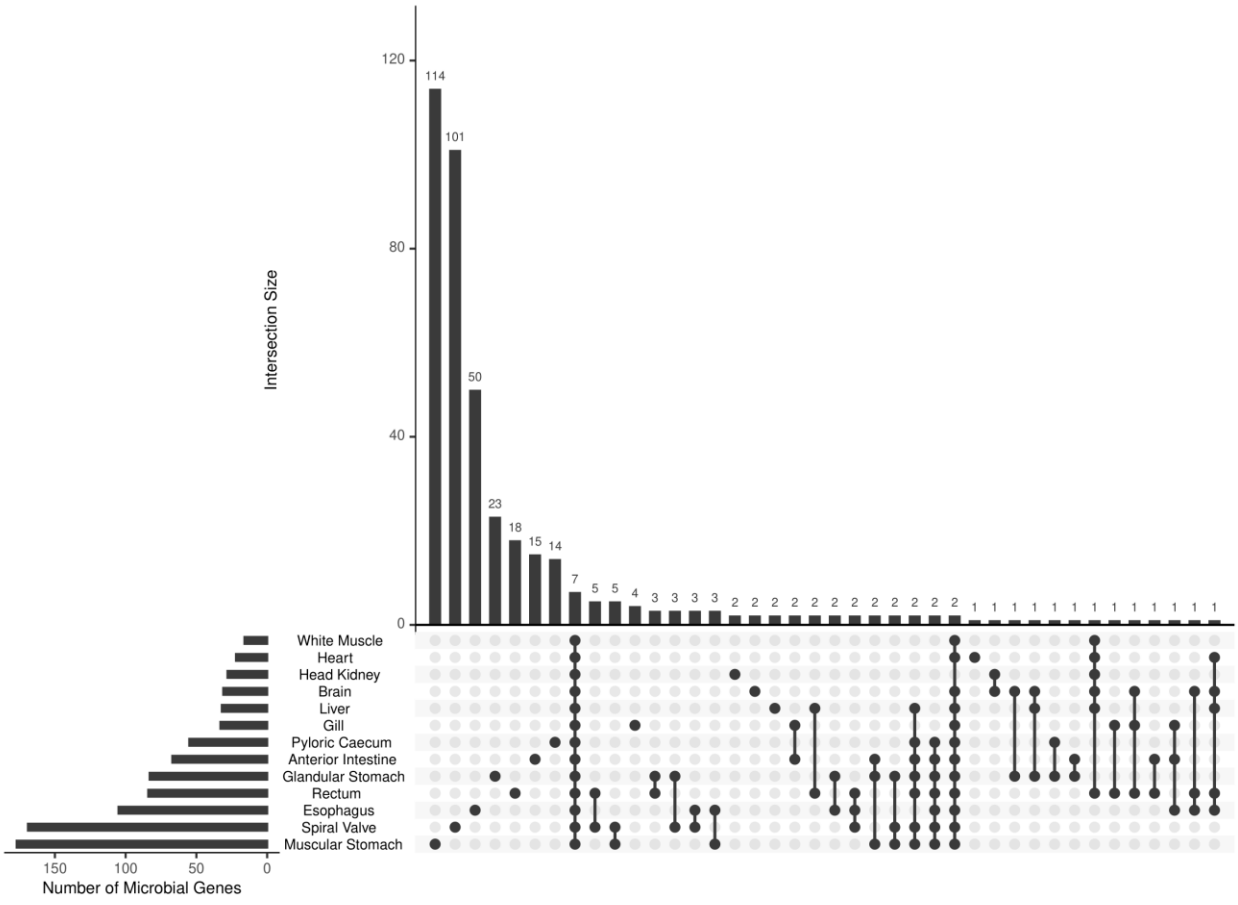
766 **Figure 4.** UpSet plot of shared and unique gene ontology (GO) terms from each of 13 tissue-
767 specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*). The GO terms presented here
768 are from the Biological Process 2021 database, significant at $q < 0.05$. The R package UpSetR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



770

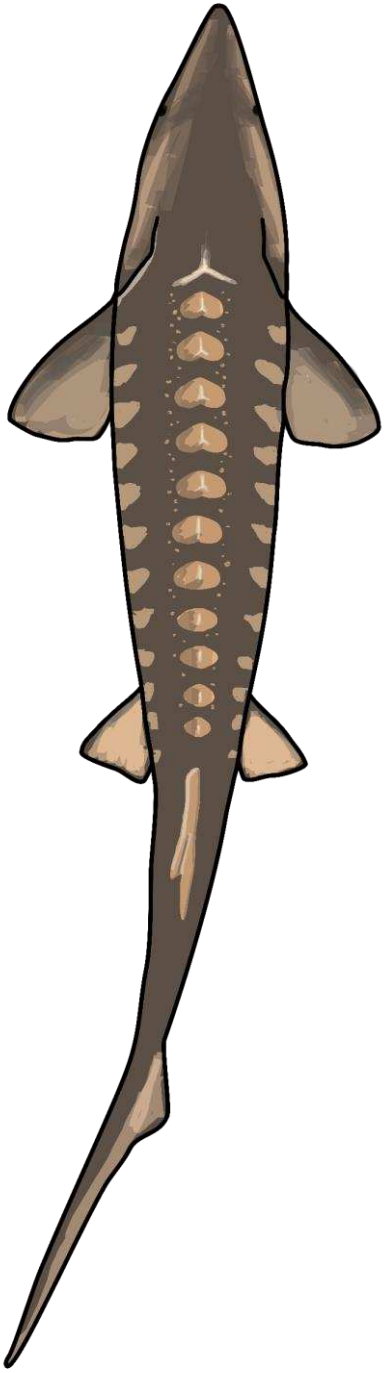
771

772 **Figure 5.** UpSet plot of shared and unique genes annotated to microbes (bacteria or archaea)
773 from each of 13 tissue-specific transcriptomes of the lake sturgeon (*Acipenser fulvescens*).
774 Annotations were performed with Trinotate, and bacterial or archaeal genes were identified by
775 filtering for those groups among filtered transcriptome annotation reports.



776

777



Esophagus
Positive regulation of viral life cycle (GO:1903902)

Glandular stomach
Rac protein signal transduction (GO:0016601)

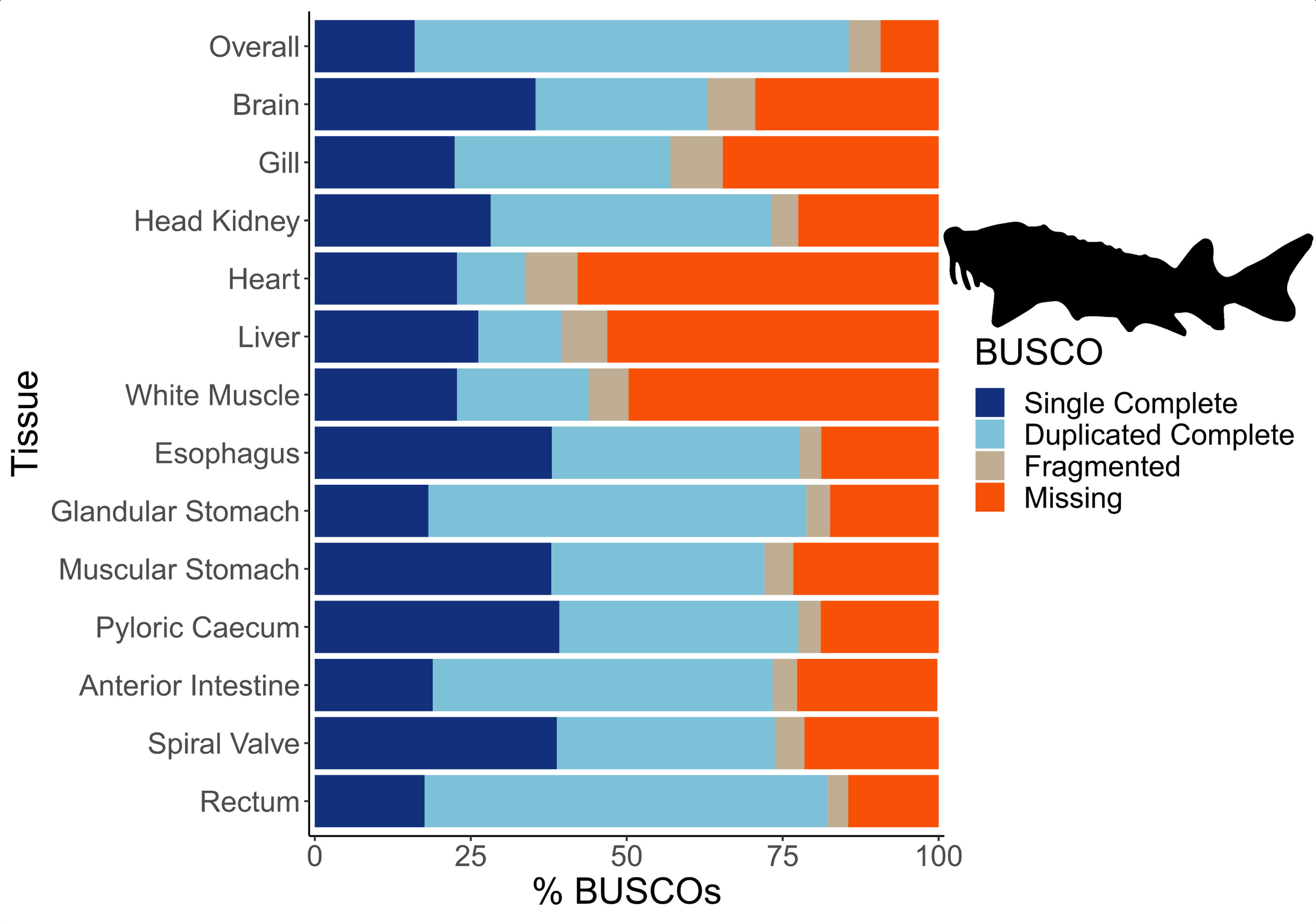
Muscular stomach
Negative regulation of immune response (GO:0045824)

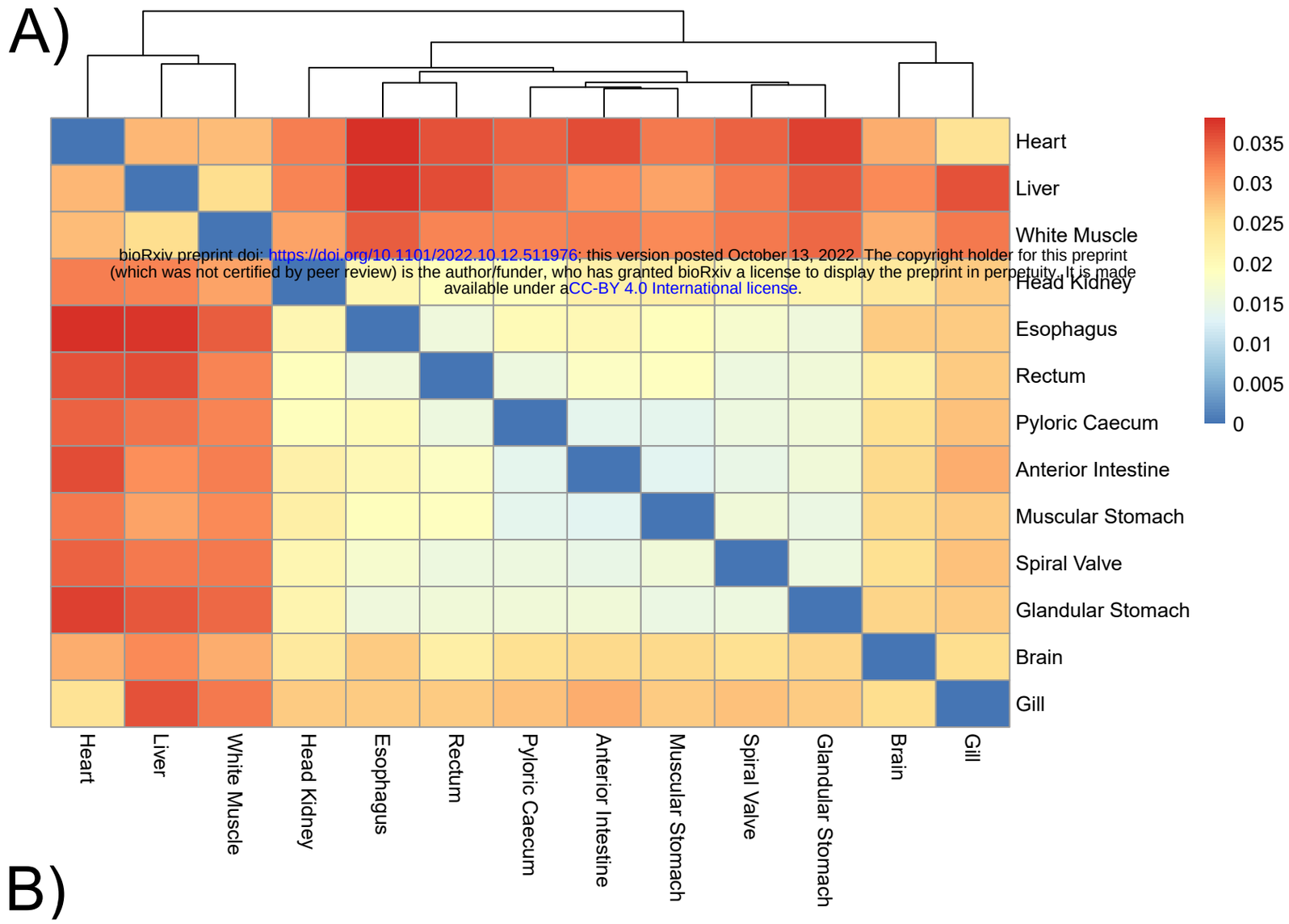
Pyloric caecum
Photoperiodism (GO:0009648)
Entrainment of circadian clock (GO:0043153)

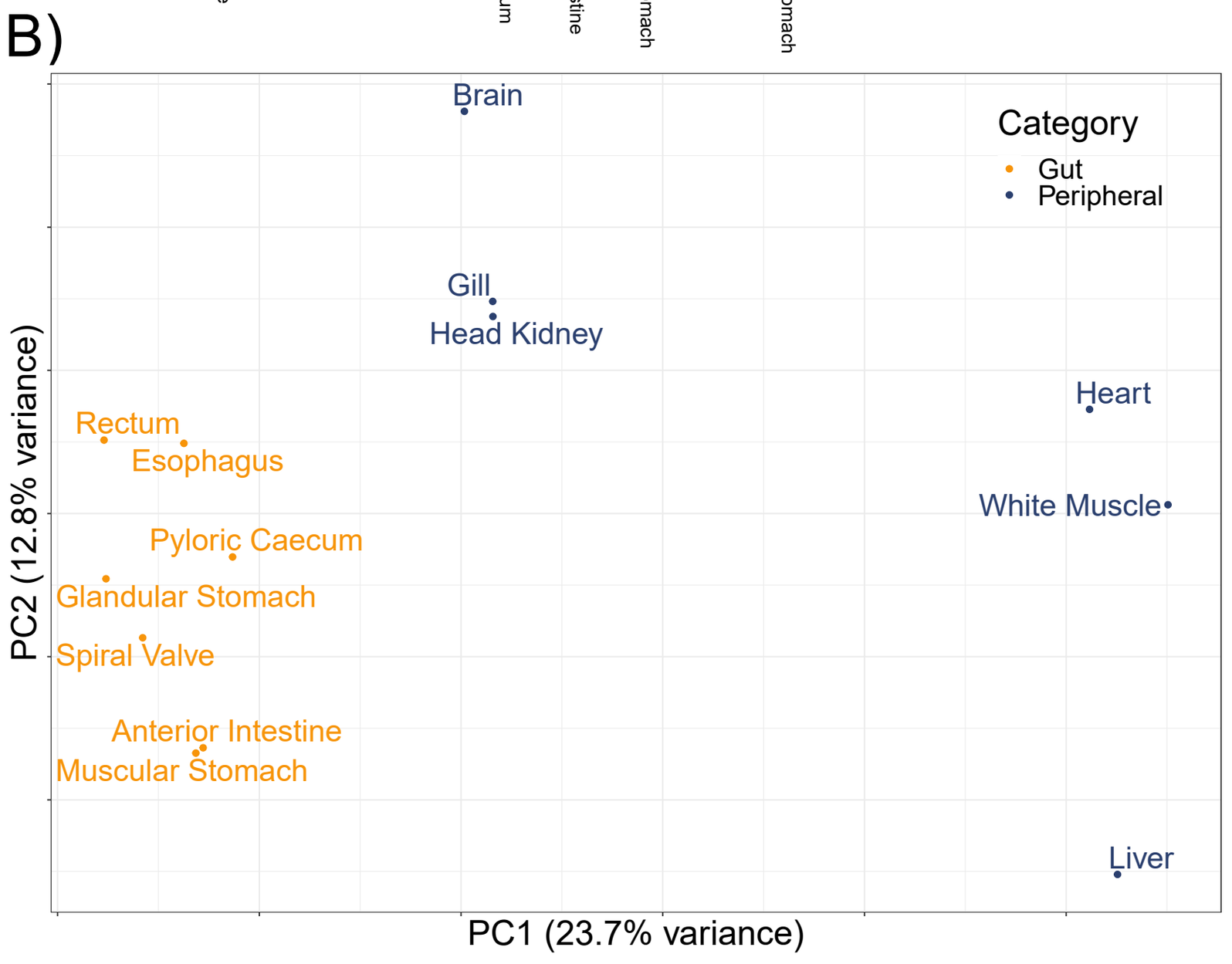
Anterior intestine
Positive regulation of host by viral transcription (GO:0043923)

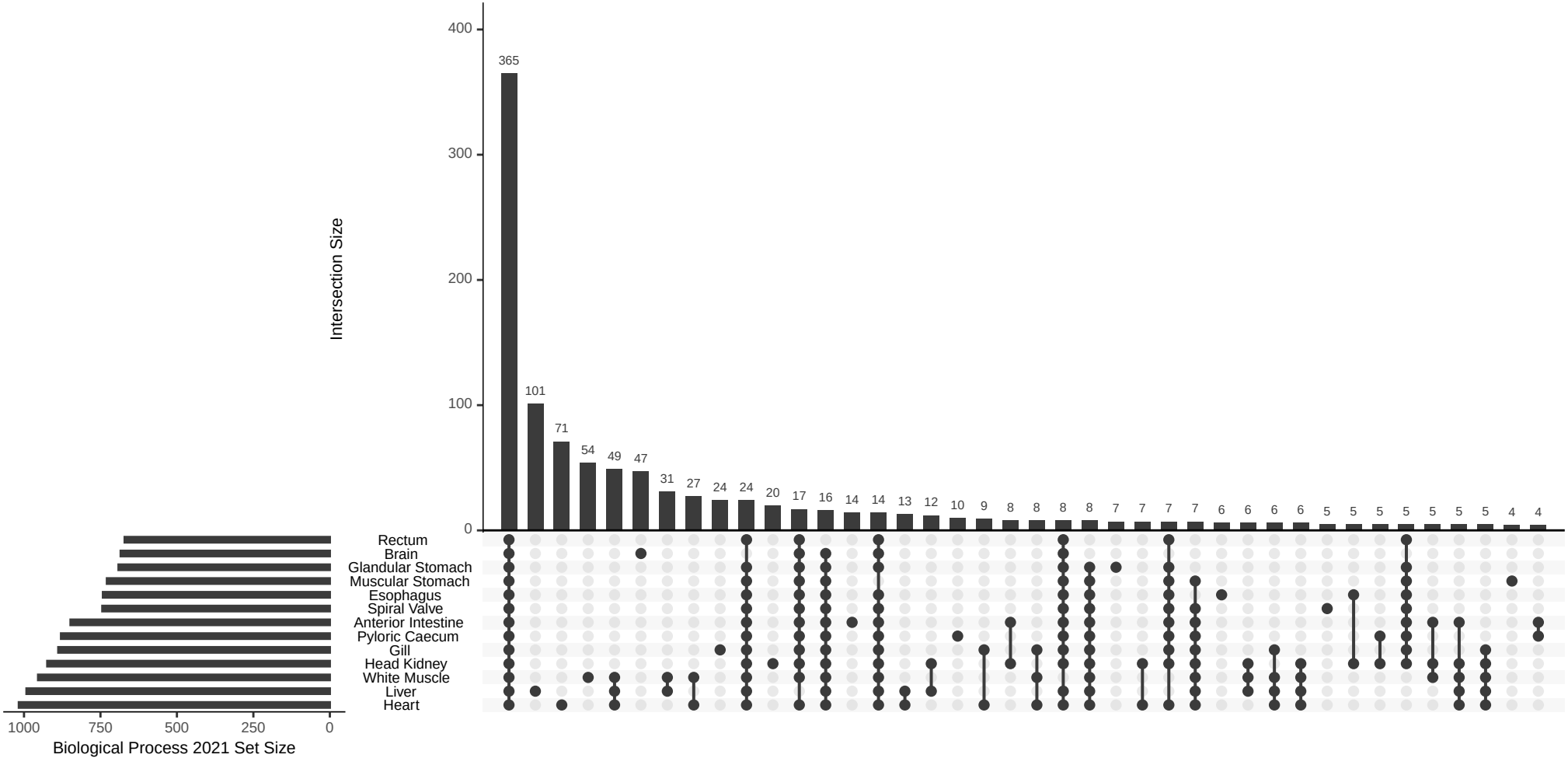
Spiral valve
Toll-like receptor signaling pathway (GO:0002224)

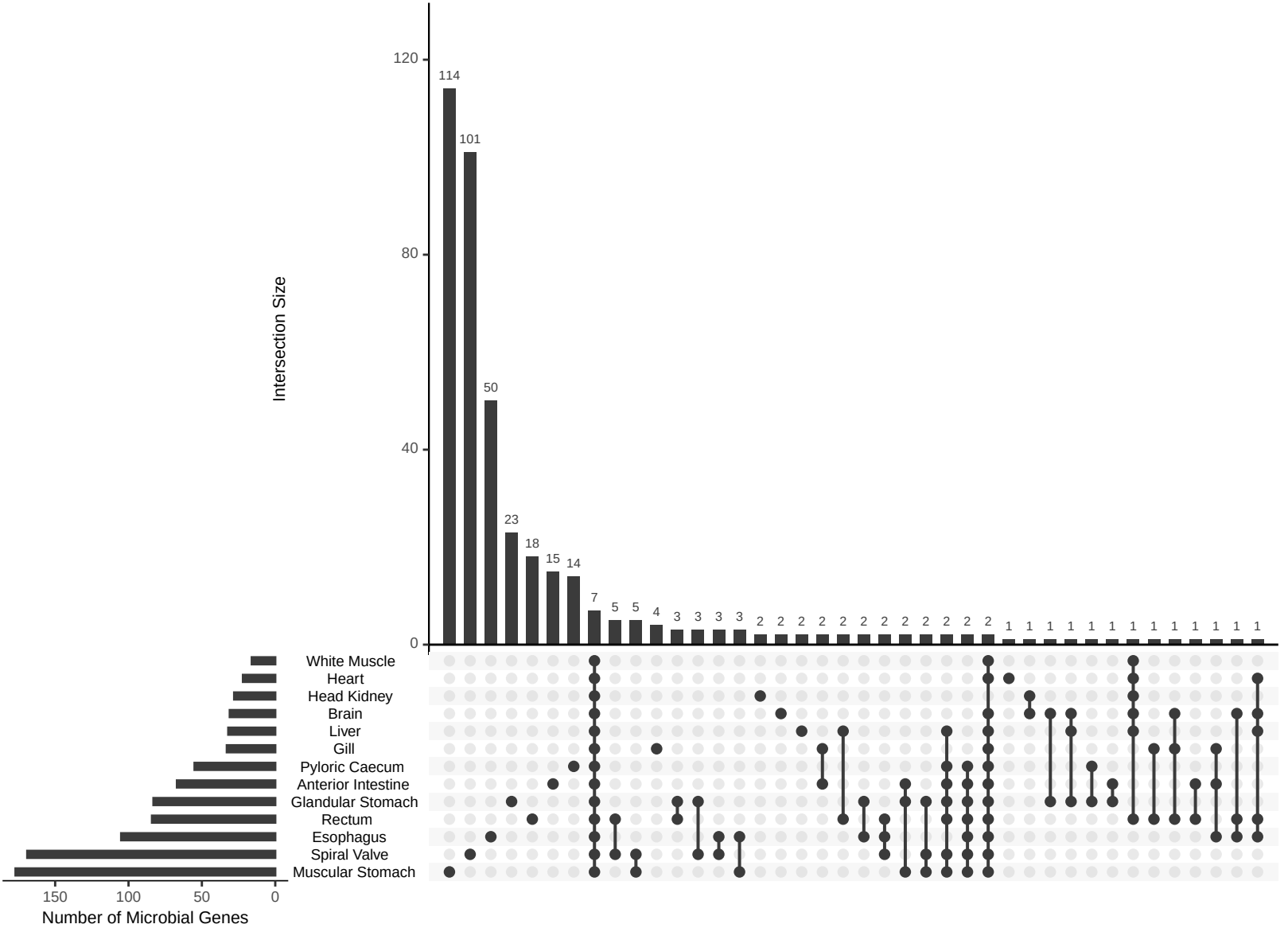
Rectum


Tissue
Overall
Brain
Gill
Head Kidney
Heart
Liver
White Muscle
Esophagus
Glandular Stomach
Muscular Stomach
Pyloric Caecum
Anterior Intestine
Spiral Valve
Rectum
0
25
50
75
100
% BUSCOs
BUSCO
Single Complete
Duplicated Complete
Fragmented
Missing






Intersection Size
400
300
200
100
0
365
101
71
54
49
47
31
27
24
24
20
17
16
14
14
13
12
10
9
8
8
8
8
7
7
7
7
6
6
6
6
5
5
5
5
5
5
5
4
4
Rectum
Brain
Glandular Stomach
Muscular Stomach
Esophagus
Spiral Valve
Anterior Intestine
Pyloric Caecum
Gill
Head Kidney
White Muscle
Liver
Heart
1000
750
500
250
0
Biological Process 2021 Set Size


Intersection Size
120
80
40
0
114
101
50
23
18
15
14
7
5
5
4
3
3
3
3
2
2
2
2
2
2
2
2
2
2
2
2
1
1
1
1
1
1
1
1
1
1
1
1
1
White Muscle
Heart
Head Kidney
Brain
Liver
Gill
Pyloric Caecum
Anterior Intestine
Glandular Stomach
Rectum
Esophagus
Spiral Valve
Muscular Stomach
150
100
50
0
Number of Microbial Genes