

Article

De Novo Hepatic Transcriptome Assembly and Systems Level Analysis of Three Species of Dietary Fish, *Sardinops sagax*, *Scomber japonicus*, and *Pleuronichthys verticalis*

Dylan J. Richards¹, Ludivine Renaud^{2,3}, Nisha Agarwal⁴, E. Starr Hazard^{3,5} , John Hyde⁶ and Gary Hardiman^{2,3,4,7,8,9,*} 

¹ Bioengineering Department, Clemson University, Charleston, SC 29425, USA; djricha@clemson.edu

² Department of Medicine, Medical University of South Carolina, Charleston, SC 29425, USA; renaudl@musc.edu

³ Center for Genomic Medicine, Bioinformatics, Medical University of South Carolina, Charleston, SC 29425, USA; hazards@musc.edu

⁴ Biomedical Informatics Research Center, San Diego State University, San Diego, CA 92182, USA; nisha2683@gmail.com

⁵ Academic Affairs Faculty & Computational Biology Resource Center, Medical University of South Carolina, Charleston, SC 29425, USA

⁶ NOAA Fisheries, Southwest Fisheries Science Center, La Jolla, CA 92037, USA; john.hyde@noaa.gov

⁷ Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC 29425, USA

⁸ Laboratory for Marine Systems Biology, Hollings Marine Laboratory, Charleston, SC 29412, USA

⁹ School of Biological Sciences & Institute for Global Food Security, Queens University Belfast, Stranmillis Road, Belfast BT9 5AG, UK

* Correspondence: G.Hardiman@qub.ac.uk

Received: 4 October 2018; Accepted: 17 October 2018; Published: 25 October 2018



Abstract: The monitoring of marine species as sentinels for ecosystem health has long been a valuable tool worldwide, providing insight into how both anthropogenic pollution and naturally occurring phenomena (i.e., harmful algal blooms) may lead to human and animal dietary concerns. The marine environments contain many contaminants of anthropogenic origin that have sufficient similarities to steroid and thyroid hormones, to potentially disrupt normal endocrine physiology in humans, fish, and other animals. An appropriate understanding of the effects of these endocrine disrupting chemicals (EDCs) on forage fish (e.g., sardine, anchovy, mackerel) can lead to significant insight into how these contaminants may affect local ecosystems in addition to their potential impacts on human health. With advancements in molecular tools (e.g., high-throughput sequencing, HTS), a genomics approach offers a robust toolkit to discover putative genetic biomarkers in fish exposed to these chemicals. However, the lack of available sequence information for non-model species has limited the development of these genomic toolkits. Using HTS and de novo assembly technology, the present study aimed to establish, for the first time for *Sardinops sagax* (Pacific sardine), *Scomber japonicas* (Pacific chub mackerel) and *Pleuronichthys verticalis* (hornyhead turbot), a de novo global transcriptome database of the liver, the primary organ involved in detoxification. The assembled transcriptomes provide a foundation for further downstream validation, comparative genomic analysis and biomarker development for future applications in ecotoxicogenomic studies, as well as environmental evaluation (e.g., climate change) and public health safety (e.g., dietary screening).

Keywords: sardine; mackerel; turbot; RNA sequencing; de novo transcriptome assembly; genomic biomarkers

1. Introduction

The monitoring of marine species as sentinels for ecosystem health has long been a valuable tool worldwide, giving insight into how both anthropogenic pollution and naturally occurring phenomena (i.e., harmful algal blooms) may lead to human and animal dietary concerns [1–8]. In addition to global influences of human activity on the environment (e.g., climate change), the coastal and estuarine environments contain many organic and inorganic contaminants from anthropogenic effluents [9]. Many of these contaminants have sufficient structural and chemical similarities to steroid and thyroid hormones, suggesting that they are potentially able to disrupt normal endocrine physiology in humans, fish, and other animals. An appropriate understanding of the effects of these endocrine disrupting chemicals (EDCs) on forage fish (e.g., sardine, anchovy, mackerel) can lead to significant insight into how these contaminants may affect local ecosystems in addition to their potential impacts on human health.

The San Diego South Bay region, which includes the cities of Coronado, Imperial Beach, and Tijuana, Mexico, has many diverse sources of pollution to the marine environment of both U.S. and Mexican origin leading to a long history of serious and ongoing water quality issues that have significant human health impacts [10–13]. This region is exposed to chemicals from domestic, agricultural, and industrial sources [14] which enter the coastal environment through the South Bay Ocean Outfall (SBOO), the Tijuana River Estuary, and other point and non-point sources. As the production of these compounds is likely to continue, there is increasing concern about the short- and long-term impacts of EDC exposure on human populations [15–17].

In this study, we focused on three common species of fish present in the marine environment of the San Diego South Bay region. Pacific sardine (*Sardinops sagax*) is a schooling coastal pelagic fish, often comprising the bulk of the diet for predatory vertebrates in this coastal ecosystem including several federally protected species (e.g., California least tern, brown pelican, bottlenose dolphin) [18]. This species is present along the west coast of North America, existing across gradients of temperature, salinity, and anthropogenic pollution [19]. As filter feeders, these coastal plankton feeders can readily accumulate toxins from the environment [19]. The Pacific chub mackerel (*Scomber japonicus*) overlaps in habitat with Pacific sardine, but feeds at a slightly higher trophic level on copepods, crustaceans, juvenile fishes, and squid. This species has been shown to be an effective sentinel species that helped to evaluate the effects of the Deepwater Horizon oil spill, which released crude oil and contaminated surface water habitats for pelagic fish for more than 12 weeks [20]. The hornyhead turbot (*Pleuronichthys verticalis*) is a demersal flatfish that resides on the mainland shelf from southern Baja California, Mexico to northern California, feeding primarily on sedentary tube-dwelling polychaetes that are present in sediment [21,22]. As sediment-associated and benthic feeders, hornyhead turbot are good sentinels for pollution monitoring due to their higher risk of exposure to chemicals that accumulate in sediments. Moreover, they inhabit a limited area, which allows one to localize the point source of chemical pollution [23]. In addition to their utility as sentinels for ecosystem health, these fish are also popular and important food sources for humans, owing to their rich omega-3 fatty acid content and flavor [24].

As a means to link bioavailability measurements for EDCs and fish response, a genomics approach offers a robust toolkit to associate the presence of these chemicals with genomic biomarkers. Advances in molecular methods, such as microarrays, quantitative PCR (q-PCR) and high-throughput sequencing (HTS) have substantially improved the sensitivity of analyses for assessing the biological effects of EDCs on animal physiology [23,25–30]. Recently developed genomics technologies have also facilitated assessment of the effects of xenobiotics in the environment on human health [31]. However, the paucity of available sequence information for non-model species is a challenge in developing these genomic toolkits.

The present study aimed to establish, for the first time for these three species, a de novo global transcriptome database for the liver, the primary organ involved in detoxification using HTS and de novo assembly technology. By combining the RNA-Seq output of liver samples from each of the

three species of fish, we assembled and annotated their hepatic transcriptomes. The assembled transcriptomes provide a foundation for further downstream validation, comparative genomic analysis and biomarker development for future applications in ecotoxicogenomic studies, as well as environmental evaluation (e.g., climate change) and public health safety (e.g., dietary screening).

2. Materials and Methods

2.1. Fish Sampling

Hornyhead turbot, sardine, and mackerel were chosen for study as they are common in the San Diego region and are species that have been used in California seafood contaminant monitoring programs. Fish were obtained either from a local commercial live bait provider (Everingham Brothers, La Mesa, CA, USA) or by bottom trawl (hornyhead turbot), all fish originated from the San Diego region. Fish (hornyhead turbot) collections were approved by and conducted under scientific collection permits issued by the California Department of Fish and Wildlife. This species is not threatened or endangered, and the collection sites were not located in ecological reserves or areas receiving special ecological protection. The sampling method (otter trawl) limits bycatch mortality by using a relatively small net, holding the catch in flowing seawater, and promptly returning non-target individuals to the ocean. These methods are the same as those used in regional monitoring programs, which have been approved by local and federal fish and wildlife and regulatory agencies [27,32]. Humane handling of the fish was assured by the use of a Standard Operating Procedure developed specifically for the study and approved by a Steering Committee composed of the study participants. Fish were transferred rapidly to holding tanks with flowing and aerated seawater at ambient temperature (~17 °C) for a minimum of 30 days prior to experiments to ensure that the fish show no signs of disease or injury and had recovered from initial handling and transport stress. Fish were humanely euthanized using either an overdose of tricaine methanesulfonate (250 mg/L) dissolved in seawater or by cervical dislocation prior to dissection. The livers were harvested and frozen in liquid nitrogen and stored at −80 °C.

2.2. RNA Extraction

Isolation of total RNA from liver samples was performed using TRIzol reagent (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA) and the extracted RNA were further purified using the RNeasy Mini kit (Qiagen, Valencia, CA, USA). All RNA samples were subjected to on-column digestion of DNA during RNA purification from cells, to ensure highly pure RNA-free from DNA contamination. The concentrations were determined by absorbance readings (OD) at 260 nm using an ND-1000 (Nanodrop, Wilmington, DE, USA). RNA was further assessed for integrity with the 6000 Nano LabChip assay from Agilent (Santa Clara, CA, USA). Only RNA samples with a RNA Integrity Number (i.e., RIN) score of >7.0 were used for RNA-Seq.

2.3. High-Throughput Sequencing

For the RNA-Seq experiments, five individual liver samples from each species were pooled. To prepare mRNA-Seq libraries, the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA) was utilized; 100–200 ng of total input RNA was used in accordance with the manufacturer's protocol. High-throughput sequencing was performed using Illumina instruments (GAIX or HiSeq2000) with each species' mRNA library sequenced to a minimum depth of ~8 million reads. A 100 bp single-end sequencing strategy was employed. Data were subjected to Illumina quality control procedures (>80% of the data yielded a Phred score of 30).

All sequencing data has been uploaded to the Sequence Read Archive (SRA) database under reference PRJNA493102.

2.4. De Novo Assembly and Transcriptome Annotation

Illumina adapter sequences and sequences that did not meet quality thresholds were removed using Trimmomatic [33]. All reads were pooled and assembled de novo using Trinity (version 2.2.0) with k -mer length set at 25 according to the original default strategy [34]. Trans-decoder (version 3.0.0) was used for coding sequence prediction [35], and Trinotate (2.0.2) for functional annotation [36]. Trinotate is a comprehensive annotation suite specifically designed for automatic functional annotation of de novo assembled transcriptomes of non-model organisms, including homology search to known sequence data (BLAST+/SwissProt) [37], protein domain identification (HMMER/PFAM) [38], transmembrane domain prediction (tmHMM) [39,40], and leveraging various annotation databases (eggNOG/GO/KEGG databases) [41–45], reporting the best hits in the databases (<http://trinotate.github.io/>). The eggNOG database is based on the original idea of COGs (clusters of orthologous groups), the KEGG (Kyoto Encyclopedia of Genes and Genomes) database is a collection of databases with genomes, biological pathways, diseases, drugs, and chemical substances and the Gene Ontology (GO) project provides functional information in the context of cellular component, (i.e., the parts of a cell or its extracellular environment), molecular function, (the activities of a gene product at the molecular level, e.g., binding or catalysis, and biological process (molecular events with a defined beginning and end). A summary data file of each fish annotation file (i.e., non-redundant Trinotate output) can be found in Supplementary Tables S1–S3. The Trinity fasta assemblies were analyzed for annotation and completeness with the BUSCO [46,47] program with the Actinopterygii_odb9 [48] dataset with zebrafish as reference. Bundled software with the BUSCO2 virtual machine included a UBUNTU image, NCBI BLAST, Augustus software version 3.2.2. [49] (<http://bioinf.uni-greifswald.de/augustus/references>), and hmmer version hmmer-3.1b2-linux-intel-ia32 (<http://www.hmmer.org>). BUSCO2 was run as a virtual machine via the Oracle VirtualBox software (<https://www.virtualbox.org/>). All analysis was carried out on a local high-performance compute cluster at the Medical University of South Carolina. The overall workflow is summarized graphically in Figure 1.

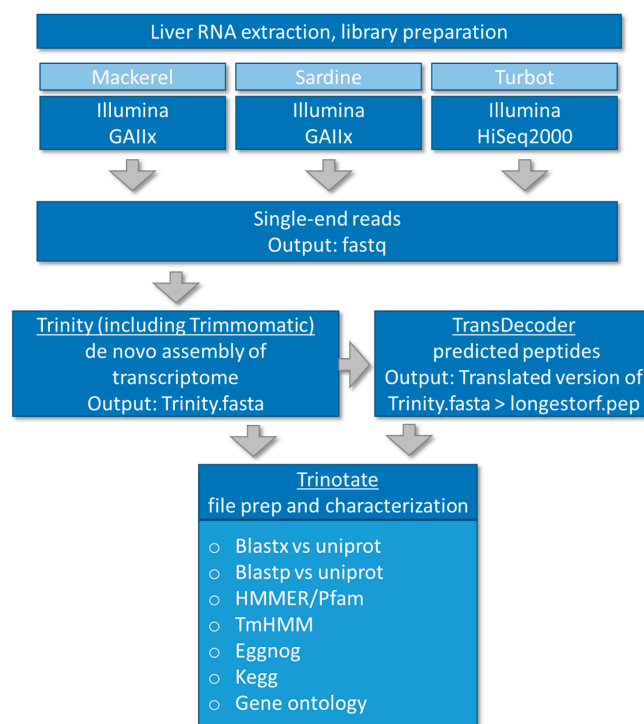


Figure 1. Workflow of de novo transcriptome assembly and annotation of three dietary fish species, mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*).

The annotation depth of zebrafish (*Danio rerio*) and humans (*Homo sapiens*) was collected from the Gene Ontology Consortium over four time points and the ratio of human to zebrafish was taken for each metric [45,50].

2.5. Systems Level Analysis

Blastp hits were compared for overlap using area-proportional Venn diagrams created using VENNY 2.1 [51]. To gain a functionalized understanding of the assembled transcriptomes, gene ontology (GO) for biological process were obtained from the list of blastp hits from each fish. After gathering a list of blastp hits per fish, species origin identifiers were removed (e.g., GAPDH_HUMAN -> GAPDH), and all duplicates were removed, so as to reduce overlapping redundancies in the functional analysis. When using zebrafish as the background annotation, the GO online tool g:Profiler was used with default settings with $p < 0.05$ and statistical domain size of “all known genes” [52]. Human annotated GO queries were performed using ToppFun from the ToppGene Suite with default settings [53]. GO terms with $p < 0.05$ were then visualized in semantic similarity-based scatterplots using REduce & Visualize Gene Ontology (REViGO) that combines redundant terms into a single, representative term based on a simple clustering algorithm relying on semantic similarity measures [54]. Selected GO terms were highlighted to indicate overall similarities and spatial organization.

3. Results

3.1. De Novo Assembly of Transcriptomes Using Trinity

By combining the RNA-Seq output of liver samples from each of the three fish species, we were able to assemble and annotate novel transcriptomes (Figure 1). As seen in Figure 2 and summarized in Supplementary Table S4, sequencing coverage (i.e., total input reads) was highest for turbot (T: 331,569,698), followed by sardine (S: 56,432,715) and mackerel (M: 8,868,248) (Figure 2A), which resulted in differing numbers of total assembled contigs and total TransDecoder contigs/peptides (Figure 2B,C). This also reflected the sequencing instrument utilized and the depth to which each sequencing library was sequenced. Despite differences in sequencing coverage, de novo assembly using Trinity produced similar sized contigs in each N -percentile category with the median contig length ~300–400 bp for each of the three species (Figure 2D). Increased sequencing coverage did, however, provide increased amounts of GO, evolutionary genealogy of genes and pathway annotated information, as seen in the Trinotate annotation hits output (Figure 2E, Supplementary Tables S1–S3). Based on the final annotation outcome, assembled contigs were re-characterized after selecting for Trinity contigs with at least one annotation hit derived from the Trinotate characterization for each fish (i.e., annotated contigs) (Figure 2F). This resulted in a higher N -percentile and median numbers for mackerel and sardine, while slightly decreasing numbers for turbot (Figure 2G). In support of increased annotation terms with increased sequencing coverage, Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis for the completeness of assembled transcriptomes showed that the lower sequence coverage (i.e., mackerel) had 18% partial/complete BUSCOs, whereas higher sequencing coverage (i.e., sardine, turbot) had ~30% partial/complete BUSCOs (Supplementary Table S5).

3.2. Functional Analysis of De Novo Hepatic Transcriptomes

The assembled transcriptomes provided a foundation for further downstream validation and comparative genomic analysis. As each of these samples is derived from the liver of several fish per species, the assembled transcriptomes represented a pooled transcriptomic view of each species and permitted a systems-level analysis on a per species basis. To focus on a functional view of each species, the protein-based annotation (i.e., TransDecoder/blastp) was used for continued analysis using, and a visual workflow can be seen in Figure 3A. The list of identified hits from the individual blastp files were used to assess hepatic gene expression profiles for each of the three fish species and

facilitate cross-species comparisons. In Figure 3B, comparison of the blastp hits from each species using a Venn diagram revealed a common set of 4886 genes accounting for ~19% of all genes across all three species (Supplementary Table S6). When using a GO term annotation library based on zebrafish, GO analysis showed an overlap of 795 GO terms representing ~20% of all GO terms (Figure 3C, Supplementary Table S7). As many of the GO terms are likely functionally redundant or closely related, REViGO analysis was used to reduce GO term redundancy, prioritize by GO term *p*-value, and visualize semantic similarity-based scatterplots for each species. By using the top-ranked GO terms by *p*-value (max. 350), REViGO plots revealed a strikingly similar spatial distribution for each species despite the differences in sequencing coverage (Figure 3D, Supplementary Table S8).

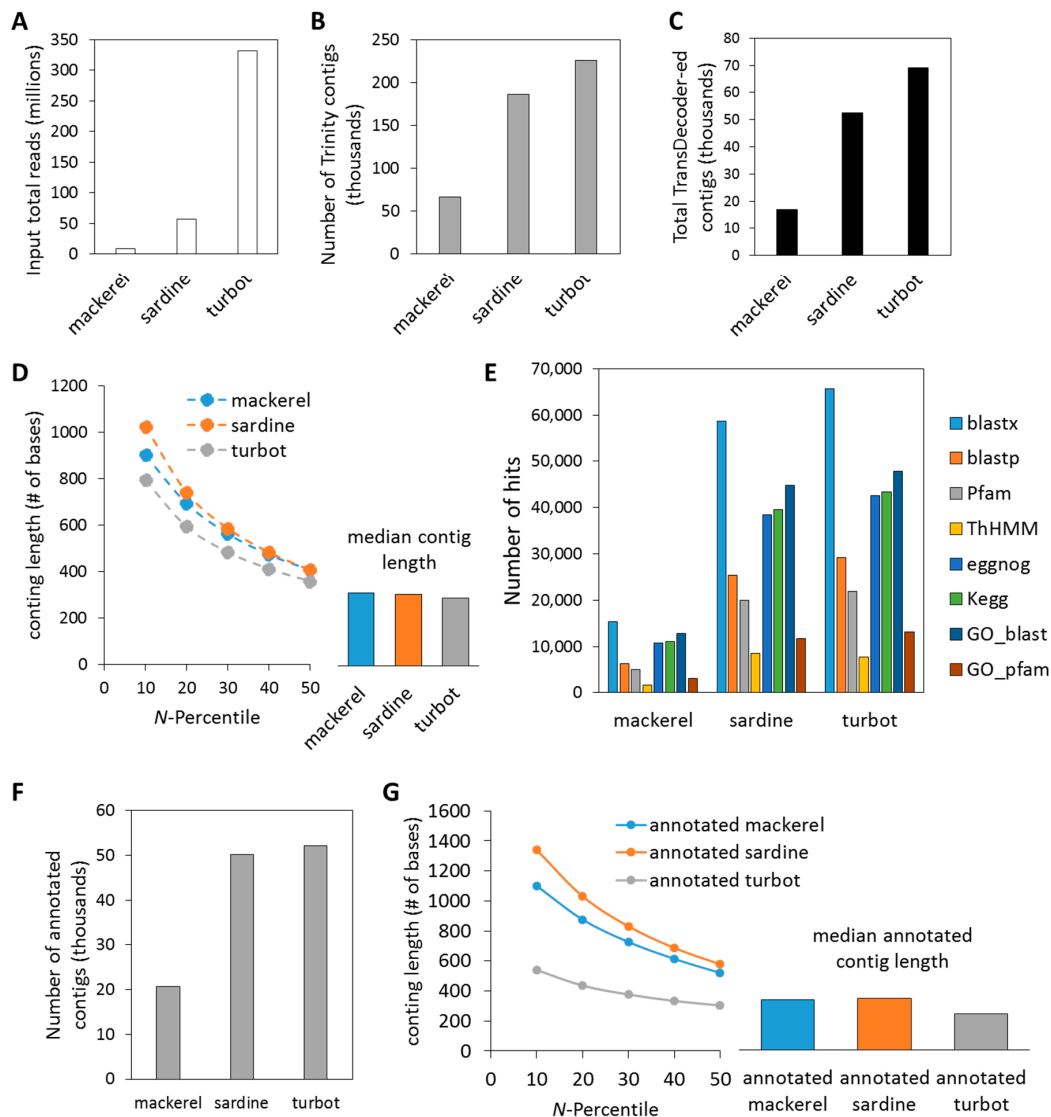


Figure 2. Quantitative Trinotate summary of Trinity transcriptome assembly output. (A) Each fish had unique number of input total number of reads that resulted in (B) differing number of resulting total assembled Trinity contigs and (C) total TransDecoder-ed contigs. (D) *N*-percentiles and median length of total contigs. (E) Sardine and turbot showed notable increases in non-redundant annotation hits over mackerel, attributed to differences in sequencing coverage. (F) Number of contigs with at least one annotation hit. (G) Re-evaluation of *N*-percentiles and median length of contigs with at least one annotation hit.

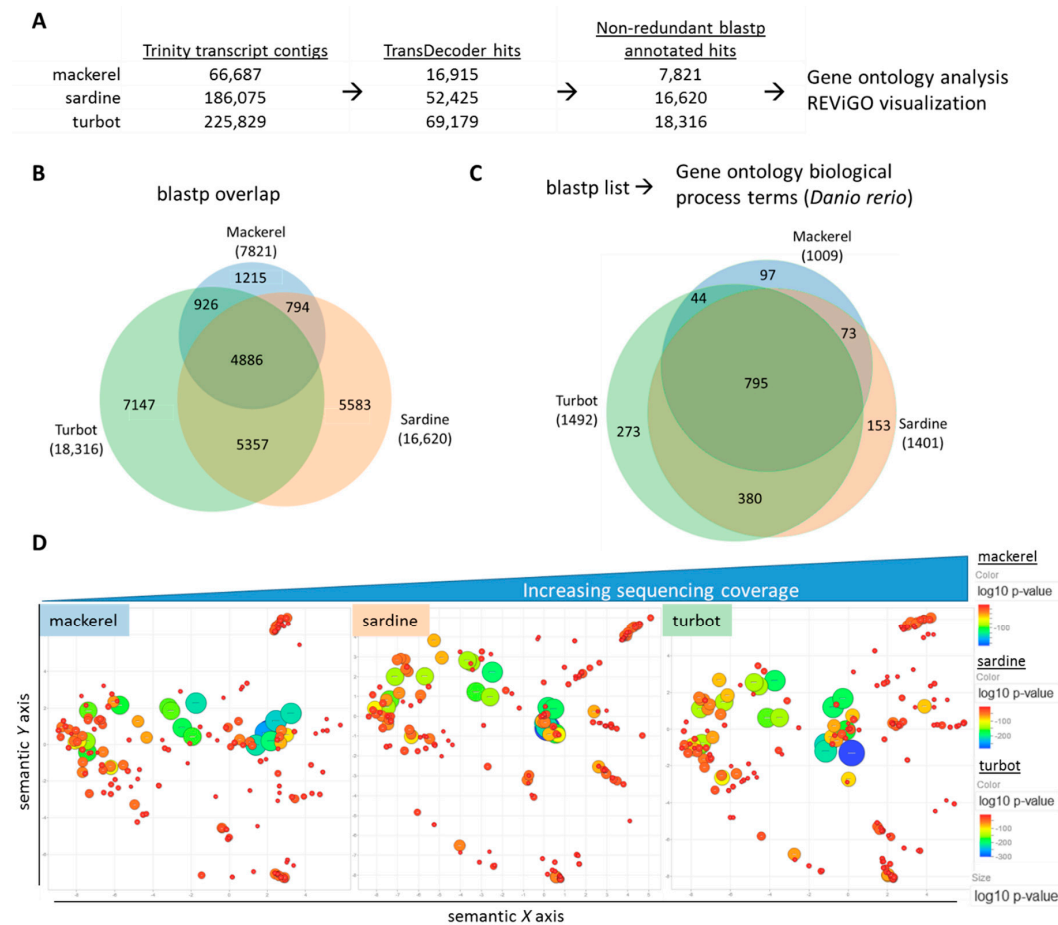


Figure 3. Comparative functional analysis of assembled transcriptomes. (A) Workflow of selecting non-redundant blastp hits from original Trinity contigs for comparative functional analysis. (B) A Venn diagram showing the overlap of non-redundant, identified hits from blastp query with each total number reflecting sequencing coverage. (C) A Venn diagram showing the overlap of gene ontology (GO) terms for biological processes derived from the blastp list for each fish using the zebrafish (*Danio rerio*) species as the annotation background. (D) REViGO plot of the top 350 significant GO terms after grouping of redundant terms (colored/sized by p -value of GO term category) reveals similar functional spatial distribution of top functional GO terms between fish.

In comparison to the depth of available annotation terms for zebrafish, the human (*Homo sapiens*) annotation background is notably larger than previously discussed (Figure 4A) [55]. When the fish genes were projected onto their human orthologs with a human GO term annotation library during GO term analysis, the fish GO terms overlapped by 909 terms (~28% of the total number) and the number of unshared terms was increased (Figure 4B, Supplementary Table S9). With a list based on “humanized” GO terms, REViGO plots of the top-ranked GO terms by p -value (max. 350) revealed unique patterning for each fish albeit with a similar overall distribution of related terms, as seen by the representative labeled terms (Figure 4C, Supplementary Table S10). Furthermore, a REViGO plot of the GO terms of the 4886 shared blastp hits from Figure 3A and Supplementary Table S4 showed a similar spatial distribution (Figure 4D), supporting the overall predicted functional similarity between fish liver samples across a range of sequencing coverage. Analysis of unique humanized GO terms showed an increasing number of unique terms according to sequencing coverage (Figure 5A). REViGO plots labeled with the top significant terms showed general physiological function-related terms that may reflect the added insight from increased sequencing coverage (Figure 5B, Supplementary Table S11).

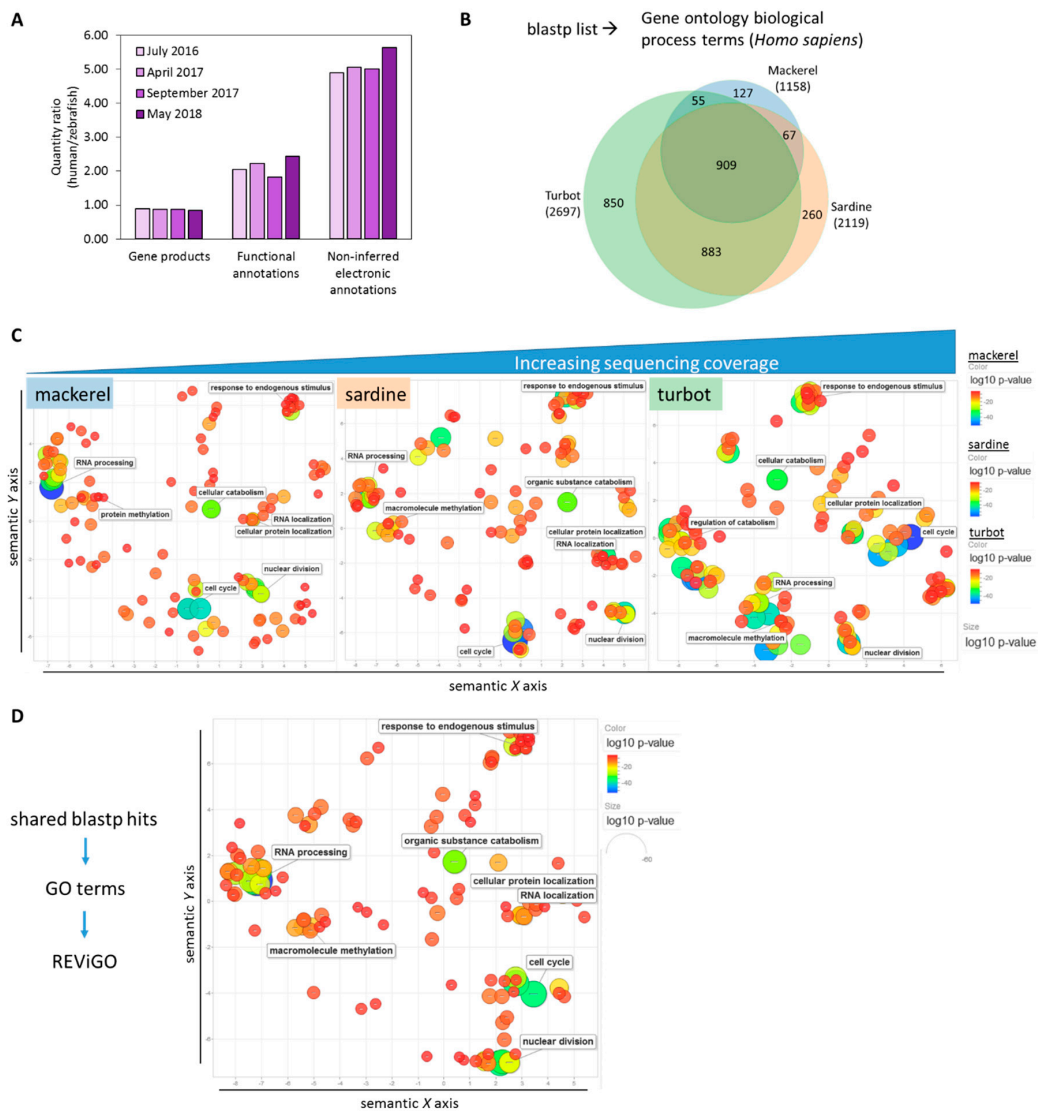


Figure 4. Comparative functional analysis of assembled transcriptomes using *Homo sapiens* annotation background. **(A)** Although gene product quantity is similar, human (*Homo sapiens*) annotation depth continues to be notably higher than zebrafish (*Danio rerio*) annotation based on the ratio of annotation metrics of human/zebrafish (as of May 2018). **(B)** A Venn diagram showing the overlap of gene ontology (GO) terms for biological processes derived from the blastp list for each fish using human (*Homo sapiens*) as the annotation background. **(C)** REViGO plot of the top 350 significant GO terms after grouping of redundant terms reveals similar functional spatial distribution of top functional GO terms between fish despite differences in sequencing coverage. Color (blue) and size (large) indicate low p -value of GO term category. **(D)** GO term identification and REViGO plot of overlapping blastp hits produce similar spatial distribution as shown in **(C)**, supporting the overall functional similarity across sequencing coverage.

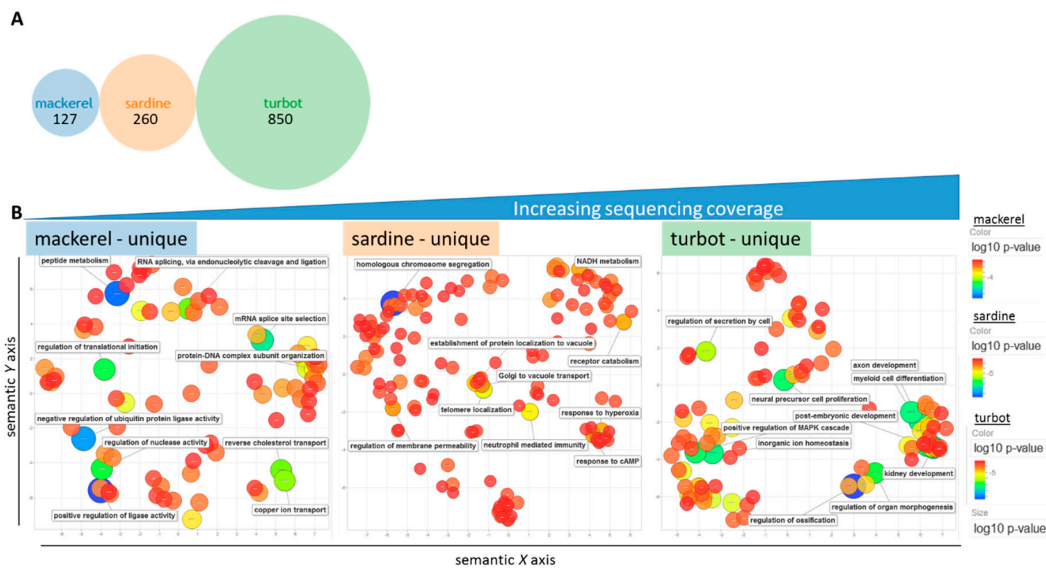


Figure 5. Unique GO terms for each fish. **(A)** Venn diagram showing the number of unique humanized GO terms from previous comparative analysis. **(B)** REViGO plot of the 127, 260, and top 350 significant unique GO terms from mackerel, sardine, and turbot, respectively, after grouping of redundant terms reveals extended information on general physiological terms despite differences in sequencing coverage. Color (blue) and size (large) indicate low p -value of GO term category.

4. Discussion

The assembly of these global transcriptomes for three sentinel coastal fish species provides a platform and template for future sequencing-based comparative studies. De novo assembly, as opposed to reference-genome based assembly, offers a direct evaluation of genetic information without interspecies discrepancies or biases. High-throughput screening creates large amounts of short-read data that poses a number of computational challenges in re-assembling. The development of the de Bruijn strategy (central to the Trinity method) to reconstruct transcriptional contigs offered a more accelerated, efficient computational path for next-generation sequencing to assemble large amounts of short-read data [56]. Given the current use of short-reads for sequencing, Trinity has been shown to be a robust and accurate method to construct de novo assemblies no matter the sequencing coverage [57]. Nevertheless, as sequencing strategies are further developed, the optimized method of assembly should take into account the sequencing parameters (e.g., depth, read length, genome size) and the overall purpose of the study (e.g., toxicity screening, splice variant exploration) [58].

In the context of toxicological studies, as systems-level/global molecular tools (e.g., RNA-Seq) become more available, genomic characterization will serve to provide an important view of the effects of contaminants, such as EDCs, and environmental shifts on fish health [59–62]. Here, by focusing on the shared genes and biological processes in the hepatic transcriptomes of three different fish species, we showed that major physiological functions can be observed and tracked at the genetic level despite differences in sequencing coverage and fish habitat. This contributes to the growing understanding of how sequencing coverage influences the downstream analysis. Notably, lower coverage, such as in the mackerel samples, could allow for high-throughput analysis in the context of transcriptomic screening, while retaining the ability to track changes in major biological processes. It is important to note that while overlapping genes may serve as the foundation to indicate relevant functions across species, trophic levels, and/or habitats, increased coverage provided extended details that may be useful in a targeted study, especially if the target environmental contaminants-affected transcript is not represented in low coverage data. Furthermore, additional optimization techniques of de novo assembly of non-model organisms may enhance transcriptomic information, supporting downstream analysis [63,64]. Future work will focus on evaluating the effect of different depths of sequencing

combined with de novo assembly parameters (e.g., *k*-mer length) to determine their effect on discovery of contaminant-affected transcripts.

Comparative analysis of shared hepatic functions across mackerel, sardine, and turbot demonstrated here also contributes to transcriptomic analysis across species that can be expanded to include other fish species and marine systems. The assembly of de novo transcriptomes bypasses the challenges of finding an appropriate reference genome for non-model fish while also expanding our knowledge of phylogenetic relationships [65–67]. Even with advances in transcriptome assembly of non-model organisms, functional annotation relies in part on the detailed research of model organisms. This study however, facilitates a “humanized” interpretation of the effect of contaminants on fish liver function, which may give insight into downstream concerns on human health [55]. Similarities in conserved functions and phylogenetic relationships can be combined to provide further insight into suitable biomarker selection across species and habitats [68,69].

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/9/11/521/s1>, Table S1: Mackerel (*Scomber japonicas*) Trinity annotation for contigs with at least one annotation hit, Table S2: Sardine (*Sardinops sagax*) Trinity annotation for contigs with at least one annotation hit, Table S3: Turbot (*Pleuronichthys verticalis*) Trinity annotation for contigs with at least one annotation hit, Table S4: Quantitative characterization of Trinity de novo assemblies of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*), Table S5: Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis with the Actinopterygii_odb9 [48] dataset with zebrafish as reference for the completeness of assembled transcriptomes for mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*), Table S6: List of blastp hits and their overlap from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*), Table S7: List of gene ontology biological process identifiers and their overlap from the blastp hits from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*) using a zebrafish gene ontology annotation background, Table S8: List of the top 350 significant gene ontology biological process identifiers and p-values from the blastp hits from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*) using a zebrafish gene ontology annotation background, Table S9: List of gene ontology biological process identifiers and their overlap from the blastp hits from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*) using a human gene ontology annotation background, Table S10: List of the top 350 significant gene ontology biological process identifiers and p-values from the blastp hits from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*) using a human gene ontology annotation background, Table S11: List of the gene ontology biological process identifiers (max. 350 terms) and p-values that were unique to each fish from the blastp hits from de novo hepatic assembled transcriptomes of mackerel (*Scomber japonicas*), sardine (*Sardinops sagax*), and turbot (*Pleuronichthys verticalis*) using a human gene ontology annotation background.

Author Contributions: Conceptualization, D.J.R., E.S.H., and G.H.; methodology, D.J.R., N.A., and E.S.H.; formal analysis, D.J.R. and E.S.H.; investigation, D.J.R. and E.S.H.; resources, G.H. and J.H.; data curation, D.J.R., N.A., and E.S.H.; writing—original draft preparation, D.J.R.; writing—review and editing, D.J.R., L.R., E.S.H., J.H. and G.H.; visualization, D.J.R. and E.S.H.; supervision, G.H.; project administration, L.R. and G.H.

Funding: This work was supported by the funding from SC EPSCoR (Gary Hardiman, 2017) and start-up funding from the College of Medicine at the Medical University of South Carolina.

Acknowledgments: We thank the staff and faculty at the MUSC Center for Genomic Medicine. We are grateful to Michael Baker and Alena Pribyl Rinke for many useful discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aloo, P.A. Health problems associated with consumption of fish and the role of aquatic environments in the transmission of human diseases. *Afr. J. Health Sci.* **2000**, *7*, 107–113. [PubMed]
2. Boltana, S.; Sanhueza, N.; Aguilar, A.; Gallardo-Escarate, C.; Arriagada, G.; Valdes, J.A.; Soto, D.; Quinones, R.A. Influences of thermal environment on fish growth. *Ecol. Evol.* **2017**, *7*, 6814–6825. [CrossRef] [PubMed]
3. Essington, T.E.; Beaudreau, A.H.; Wiedenmann, J. Fishing through marine food webs. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3171–3175. [CrossRef] [PubMed]
4. Glomset, J.A. Fish, fatty acids, and human health. *N. Engl. J. Med.* **1985**, *312*, 1253–1254. [CrossRef] [PubMed]

5. Greene, J.; Ashburn, S.M.; Razzouk, L.; Smith, D.A. Fish oils, coronary heart disease, and the environment. *Am. J. Public Health* **2013**, *103*, 1568–1576. [[CrossRef](#)] [[PubMed](#)]
6. Harley, C.D.; Randall Hughes, A.; Hultgren, K.M.; Miner, B.G.; Sorte, C.J.; Thornber, C.S.; Rodriguez, L.F.; Tomanek, L.; Williams, S.L. The impacts of climate change in coastal marine systems. *Ecol. Lett.* **2006**, *9*, 228–241. [[CrossRef](#)] [[PubMed](#)]
7. Petersen, A.; Andersen, J.S.; Kaewmak, T.; Somsiri, T.; Dalsgaard, A. Impact of integrated fish farming on antimicrobial resistance in a pond environment. *Appl. Environ. Microbiol.* **2002**, *68*, 6036–6042. [[CrossRef](#)] [[PubMed](#)]
8. Szuwalski, C.S.; Hilborn, R. Environment drives forage fish productivity. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3314–E3315. [[CrossRef](#)] [[PubMed](#)]
9. Chen, G.; White, P.A. The mutagenic hazards of aquatic sediments: A review. *Mutat. Res.* **2004**, *567*, 151–225. [[CrossRef](#)] [[PubMed](#)]
10. Ahn, J.H.; Grant, S.B.; Surbeck, C.Q.; DiGiacomo, P.M.; Nezlin, N.P.; Jiang, S. Coastal water quality impact of stormwater runoff from an urban watershed in southern California. *Environ. Sci. Technol.* **2005**, *39*, 5940–5953. [[CrossRef](#)] [[PubMed](#)]
11. Given, S.; Pendleton, L.H.; Boehm, A.B. Regional public health cost estimates of contaminated coastal waters: A case study of gastroenteritis at southern California beaches. *Environ. Sci. Technol.* **2006**, *40*, 4851–4858. [[CrossRef](#)] [[PubMed](#)]
12. Jiang, S.; Noble, R.; Chu, W. Human adenoviruses and coliphages in urban runoff-impacted coastal waters of southern California. *Appl. Environ. Microbiol.* **2001**, *67*, 179–184. [[CrossRef](#)] [[PubMed](#)]
13. Sercu, B.; Van De Werfhorst, L.C.; Murray, J.; Holden, P.A. Storm drains are sources of human fecal pollution during dry weather in three urban southern California watersheds. *Environ. Sci. Technol.* **2009**, *43*, 293–298. [[CrossRef](#)] [[PubMed](#)]
14. Parnell, P.E.; Groce, A.K.; Stebbins, T.D.; Dayton, P.K. Discriminating sources of PCB contamination in fish on the coastal shelf off San Diego, California (USA). *Mar. Pollut. Bull.* **2008**, *56*, 1992–2002. [[CrossRef](#)] [[PubMed](#)]
15. Soto, A.M.; Vandenberg, L.N.; Maffini, M.V.; Sonnenschein, C. Does breast cancer start in the womb? *Basic Clin. Pharmacol. Toxicol.* **2008**, *102*, 125–133. [[CrossRef](#)] [[PubMed](#)]
16. Diamanti-Kandarakis, E.; Bourguignon, J.P.; Giudice, L.C.; Hauser, R.; Prins, G.S.; Soto, A.M.; Zoeller, R.T.; Gore, A.C. Endocrine-disrupting chemicals: An Endocrine Society scientific statement. *Endocr. Rev.* **2009**, *30*, 293–342. [[CrossRef](#)] [[PubMed](#)]
17. Grun, F.; Blumberg, B. Minireview: The case for obesogens. *Mol. Endocrinol.* **2009**, *23*, 1127–1134. [[CrossRef](#)] [[PubMed](#)]
18. Francis, M.; Lowry, M.S.; Yochem, P.K.; Stewart, B.S.; Heath, C.B. Seasonal and annual variability in the diet of California sea lions, *Zalophus californianus*, at San Nicolas Island, California, 1981–1986. *Fish. Bull.* **1991**, *89*, 331–336.
19. LLeFebvre, K.; Silver, M.; Coale, S.; Tjeerdema, R. Domoic acid in planktivorous fish in relation to toxic *Pseudo-nitzschia* cell densities. *Mar. Biol.* **2002**, *140*, 625–631. [[CrossRef](#)]
20. Klinger, D.H.; Dale, J.J.; Machado, B.E.; Incardona, J.P.; Farwell, C.J.; Block, B.A. Exposure to Deepwater Horizon weathered crude oil increases routine metabolic demand in chub mackerel, *Scomber japonicus*. *Mar. Pollut. Bull.* **2015**, *98*, 259–266. [[CrossRef](#)] [[PubMed](#)]
21. Cross, J.; Roney, J.; Kleppel, G. Fish food-habits along a pollution gradient. *Calif. Fish Game* **1985**, *71*, 28–39.
22. Miller, D.J.; Lea, R.N. *Guide to the Coastal Marine Fishes of California*; UCANR Publications: Oakland, CA, USA, 1972.
23. Baker, M.E.; Ruggeri, B.; Sprague, L.J.; Eckhardt-Ludka, C.; Lapira, J.; Wick, I.; Soverchia, L.; Ubaldi, M.; Polzonetti-Magni, A.M.; Vidal-Dorsch, D.; et al. Analysis of endocrine disruption in southern California coastal fish using an aquatic multispecies microarray. *Environ. Health Perspect.* **2009**, *117*, 223–230. [[CrossRef](#)] [[PubMed](#)]
24. Kris-Etherton, P.M.; Harris, W.S.; Appel, L.J.; American Heart Association; Nutrition, C. Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. *Circulation* **2002**, *106*, 2747–2757. [[CrossRef](#)] [[PubMed](#)]
25. Baker, M.E.; Hardiman, G. Transcriptional analysis of endocrine disruption using zebrafish and massively parallel sequencing. *J. Mol. Endocrinol.* **2014**, *52*, R241–R256. [[CrossRef](#)] [[PubMed](#)]

26. Vidal-Dorsch, D.E.; Bay, S.M.; Ribocco, C.; Sprague, L.J.; Angert, M.; Ludka, C.; Ricciardelli, E.; Carnevali, O.; Greenstein, D.J.; Schlenk, D.; et al. Genomic and phenotypic response of hornyhead turbot exposed to municipal wastewater effluents. *Aquat. Toxicol.* **2013**, *140–141*, 174–184. [[CrossRef](#)] [[PubMed](#)]
27. Baker, M.E.; Sprague, L.J.; Ribocco, C.; Ruggeri, B.; Lekmine, N.; Ludka, C.; Wick, I.; Soverchia, L.; Ubaldi, M.; Šášik, R.; et al. Application of a targeted endocrine q-PCR panel to monitor the effects of pollution in southern California flatfish. *Endocr. Disrupt.* **2014**, *2*, e969598. [[CrossRef](#)]
28. Xu, E.G.; Mager, E.M.; Grosell, M.; Pasparakis, C.; Schlenker, L.S.; Stieglitz, J.D.; Benetti, D.; Hazard, E.S.; Courtney, S.M.; Diamante, G.; et al. Time- and oil-dependent transcriptomic and physiological responses to deepwater horizon oil in mahi-mahi (*Coryphaena hippurus*) embryos and larvae. *Environ. Sci. Technol.* **2016**, *50*, 7842–7851. [[CrossRef](#)] [[PubMed](#)]
29. Xu, E.G.; Khursigara, A.J.; Magnuson, J.; Hazard, E.S.; Hardiman, G.; Esbaugh, A.J.; Roberts, A.P.; Schlenk, D. Larval red drum (*Sciaenops ocellatus*) sublethal exposure to weathered deepwater horizon crude oil: Developmental and transcriptomic consequences. *Environ. Sci. Technol.* **2017**, *51*, 10162–10172. [[CrossRef](#)] [[PubMed](#)]
30. Xu, E.G.; Mager, E.M.; Grosell, M.; Stieglitz, J.D.; Hazard, E.S.; Hardiman, G.; Schlenk, D. Developmental transcriptomic analyses for mechanistic insights into critical pathways involved in embryogenesis of pelagic mahi-mahi (*Coryphaena hippurus*). *PLoS ONE* **2017**, *12*, e0180454. [[CrossRef](#)] [[PubMed](#)]
31. Miracle, A.L.; Ankley, G.T. Ecotoxicogenomics: Linkages between exposure and effects in assessing risks of aquatic contaminants to fish. *Reprod. Toxicol.* **2005**, *19*, 321–326. [[CrossRef](#)] [[PubMed](#)]
32. Baker, M.E.; Vidal-Dorsch, D.E.; Ribocco, C.; Sprague, L.J.; Angert, M.; Lekmine, N.; Ludka, C.; Martella, A.; Ricciardelli, E.; Bay, S.M.; et al. Molecular analysis of endocrine disruption in hornyhead turbot at wastewater outfalls in southern California using a second generation multi-species microarray. *PLoS ONE* **2013**, *8*, e75553. [[CrossRef](#)]
33. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
34. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
35. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)]
36. Bryant, D.M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M.B.; Payzin-Dogru, D.; Lee, T.J.; Leigh, N.D.; Kuo, T.H.; Davis, F.G.; et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **2017**, *18*, 762–776. [[CrossRef](#)]
37. The UniProt, C. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [[CrossRef](#)]
38. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)]
39. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)]
40. Sonnhammer, E.L.; von Heijne, G.; Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 175–182.
41. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)]
42. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
43. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)] [[PubMed](#)]
44. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [[CrossRef](#)]

45. The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **2017**, *45*, D331–D338. [[CrossRef](#)]
46. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
47. Waterhouse, R.M.; Seppey, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **2017**. [[CrossRef](#)]
48. Zdobnov, E.M.; Tegenfeldt, F.; Kuznetsov, D.; Waterhouse, R.M.; Simao, F.A.; Ioannidis, P.; Seppey, M.; Loetscher, A.; Kriventseva, E.V. OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **2017**, *45*, D744–D749. [[CrossRef](#)] [[PubMed](#)]
49. Keller, O.; Kollmar, M.; Stanke, M.; Waack, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **2011**, *27*, 757–763. [[CrossRef](#)] [[PubMed](#)]
50. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
51. Venny Oliveros, J. An interactive Tool for Comparing Lists with Venn's Diagrams. 2007. Available online: <http://bioinfo.cnb.csic.es/tools/venny/index.html> (accessed on 25 October 2018).
52. Reimand, J.; Arak, T.; Adler, P.; Kolberg, L.; Reisberg, S.; Peterson, H.; Vilo, J. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **2016**, *44*, W83–W89. [[CrossRef](#)]
53. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **2009**, *37*, W305–W311. [[CrossRef](#)] [[PubMed](#)]
54. Supek, F.; Bosnjak, M.; Skunca, N.; Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)] [[PubMed](#)]
55. Huff, M.; da Silveira, W.A.; Carnevali, O.; Renaud, L.; Hardiman, G. Systems analysis of the liver transcriptome in adult male zebrafish exposed to the plasticizer (2-ethylhexyl) phthalate (DEHP). *Sci. Rep.* **2018**, *8*, 2118. [[CrossRef](#)] [[PubMed](#)]
56. Compeau, P.E.; Pevzner, P.A.; Tesler, G. How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.* **2011**, *29*, 987–991. [[CrossRef](#)] [[PubMed](#)]
57. Zhao, Q.Y.; Wang, Y.; Kong, Y.M.; Luo, D.; Li, X.; Hao, P. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: A comparative study. *BMC Bioinform.* **2011**, *12* (Suppl. 14), S2. [[CrossRef](#)]
58. Li, Z.; Chen, Y.; Mu, D.; Yuan, J.; Shi, Y.; Zhang, H.; Gan, J.; Li, N.; Hu, X.; Liu, B.; et al. Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. *Brief Funct. Genom.* **2012**, *11*, 25–37. [[CrossRef](#)] [[PubMed](#)]
59. Sudhagar, A.; Kumar, G.; El-Matbouli, M. Transcriptome analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: A comprehensive review. *Int. J. Mol. Sci.* **2018**, *19*, 245. [[CrossRef](#)] [[PubMed](#)]
60. Uren Webster, T.M.; Shears, J.A.; Moore, K.; Santos, E.M. Identification of conserved hepatic transcriptomic responses to 17 β -estradiol using high-throughput sequencing in brown trout. *Physiol. Genom.* **2015**, *47*, 420–431. [[CrossRef](#)] [[PubMed](#)]
61. Oomen, R.A.; Hutchings, J.A. Variation in spawning time promotes genetic variability in population responses to environmental change in a marine fish. *Conserv. Physiol.* **2015**, *3*, cov027. [[CrossRef](#)] [[PubMed](#)]
62. Foo, S.A.; Byrne, M. Acclimatization and adaptive capacity of marine species in a changing ocean. *Adv. Mar. Biol.* **2016**, *74*, 69–116. [[CrossRef](#)] [[PubMed](#)]
63. Lu, B.; Zeng, Z.; Shi, T. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Sci. China Life Sci.* **2013**, *56*, 143–155. [[CrossRef](#)] [[PubMed](#)]
64. Ungaro, A.; Pech, N.; Martin, J.F.; McCairns, R.J.S.; Mevy, J.P.; Chappaz, R.; Gilles, A. Challenges and advances for transcriptome assembly in non-model species. *PLoS ONE* **2017**, *12*, e0185020. [[CrossRef](#)] [[PubMed](#)]
65. Malmstrom, M.; Matschiner, M.; Torresen, O.K.; Jakobsen, K.S.; Jentoft, S. Whole genome sequencing data and de novo draft assemblies for 66 teleost species. *Sci. Data* **2017**, *4*, 160132. [[CrossRef](#)] [[PubMed](#)]

66. Malmstrom, M.; Matschiner, M.; Torresen, O.K.; Star, B.; Snipen, L.G.; Hansen, T.F.; Baalsrud, H.T.; Nederbragt, A.J.; Hanel, R.; Salzburger, W.; et al. Evolution of the immune system influences speciation rates in teleost fishes. *Nat. Genet.* **2016**, *48*, 1204–1210. [[CrossRef](#)] [[PubMed](#)]
67. Carruthers, M.; Yurchenko, A.A.; Augley, J.J.; Adams, C.E.; Herzyk, P.; Elmer, K.R. De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genom.* **2018**, *19*, 32. [[CrossRef](#)]
68. Kroon, F.; Streten, C.; Harries, S. A protocol for identifying suitable biomarkers to assess fish health: A systematic review. *PLoS ONE* **2017**, *12*, e0174762. [[CrossRef](#)] [[PubMed](#)]
69. Yuan, Z.; Liu, S.; Zhou, T.; Tian, C.; Bao, L.; Dunham, R.; Liu, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genom.* **2018**, *19*, 141. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).