

De novo transcriptome assembly and functional annotation of cobia (*Rachycentron canadum*) liver

Bruno Cavalheiro Araújo, Alexandre Wagner Silva Hilsdorf, Renata G. Moreira, Luiz R. Nunes, Fabiano Bezerra Menegidio, David Aciole Barbosa

UNIVERSIDADE DE MOGI DAS CRUZES

Abstract

Rachycentron canadum (cobia) is considered the marine fish with the highest production potential in Brazil, mainly due to its fast growth, good feedstock conversion rate, and excellent fillet quality. However, the lack of genomic information hinders more comprehensive investigations in this species, a problem that could be minimized by de novo transcriptome assembly, providing data that could contribute to a variety of studies including conservation genetics, selective breeding, reproductive biology, and fish nutrition. In this report, we provide a de novo assembled transcriptome, using material from of *R. canadum* liver cells, highlighting the main functional annotations for this economically relevant fish. Briefly, 90 cobia juveniles (128.85 ± 18.43 g), obtained from a commercial hatchery, were randomly distributed in 3 tanks (2.000 L) under controlled temperature (23 ± 1.5 oC) and photoperiod. Fish were equally hand-fed (twice a day) with commercial marine fish diet, during 6 weeks, anesthetized (4 g benzocaine in 10 ml ethanol), placed into 40 L of seawater, and euthanized by spinal cord section. Liver samples were frozen in liquid nitrogen and stored at -80 oC, until total RNA was extracted and transcriptome sequencing performed, using 150 bp (2 X 75) paired-end strategy, in an Illumina Nextseq sequencer. This analysis resulted in 1, 761, 965, 794 sequences, which were further processed in a Galaxy server (usegalaxy.eu). Libraries were submitted to quality control (FastQC and MultiQC) and low-quality reads ($Q < 30$), adapters and other contaminant sequences were removed (Fastp), providing 1, 652, 319, 304 high-quality reads (93.8% of raw data), which were then used for de novo transcriptome assembly and metrics evaluation, using Trinity. This process resulted in the identification of 101, 789 unigenes and 163, 096 isoforms, with an average length of 1, 617.34 and 950.61 bp respectively. Median sizes (N50) were 7, 843 bp for unigenes and 2, 312 bp for isoforms. In total, 163, 096 transcripts were generated, with 95, 075 (58%) presenting more than 500 bp. Transcriptome completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO), identifying 81.7% of the universal complete actinopterygii genes (3746/4584, OrthoDB v9) supporting the high quality of our transcriptome assembly.. Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) functionally annotated 75, 554 transcripts (RefSeq: 25, 728 [34%]; Nr: 29, 155 [39%], Swiss-Prot: 15, 507 [21%]; EggNOG: 75, 068 [99%]), which were related to 75, 060 Gene Ontology (GO) terms (molecular function: 47, 766 [64%]; biological process: 48, 550 [65%]; cellular component: 34, 492 [46%]) and 24, 100 [32%] Kegg pathway assignments.

Funding: CAPES, CNPq, FAPESP

Link to Video: