Data in Brief 45 (2022) 108746

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Comprehensive data and workflow for mapping global proteome and post-translational modifications in Indian Major Carp, *Labeo rohita*

Mehar Un Nissa[a], Anwesha Banerjee[b], Mukunda Goswami[c], Sanjeeva Srivastava[a,*]

[a] *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India*
[b] *Presidency University, 86/1 College Street Kolkata 700073, West Bengal, India*
[c] *Central Institute of Fisheries Education, Indian Council of Agricultural Research, Versova, Mumbai, Maharashtra 400061, India*

ARTICLE INFO

Article history:
Received 8 August 2022
Revised 25 October 2022
Accepted 7 November 2022
Available online 13 November 2022

Dataset link: Fish Proteome Map (Reference Data)
Dataset link: Proteomic profiling of Labeo Rohita; a widely cultivated fish (Reference Data)
Dataset link: Organ wise proteomic profiling of Indian major carp, Labeo rohita (Reference Data)
Dataset link: Organ wise targeted mass spectrometric analysis of proteins in Labeo rohita using SRM/MRM approach (Reference Data)

ABSTRACT

We present the data for the global proteome and post-translational modification mapping of *Labeo rohita* (Rohu) which consists of mass-spectrometric (MS) data for 8498 proteins at 1% false discovery rate, which constitutes 26% of the total protein-coding sequences in Rohu. This data consists of deep proteomics of 17 normal tissues including eye, spinal cord, brain, male gonad, female gonad, gill, air bladder, gall bladder, gut, liver, heart, kidney, skin, scales, muscle, fin, spleen, as well as blood plasma and embryo of Rohu. The data from SRM-based targeted analysis to validate the presence of few key proteins is also included. Global post translational modification-based analysis (global PTM) was also performed in the studied tissues and its background data is also publicly accessible. This data and the web-based proteome map may aid applied and basic research endeavors in aquaculture to meet the food demands and nutritional security challenges of an increasing

* Corresponding author.
E-mail address: sanjeeva@iitb.ac.in (S. Srivastava).

https://doi.org/10.1016/j.dib.2022.108746

Keywords:
Aquaculture
Protein expression
Mass spectrometry
Labeo rohita
Post-translational modifications

world population. The data here is related to the research article "Organ-based proteome and post-translational modification profiling of a widely cultivated tropical water fish, *Labeo rohita*" in the Journal of Proteome Research [1].

Specifications Table

Subject	Omics: Proteomics (Biological sciences)
Specific subject area	Proteome map of *Labeo rohita*.
Type of data	Table, Figure
How the data were acquired	Data was acquired using liquid chromatography-tandem mass spectrometry (LC-MS/MS) through an Easy-nLC 1200 nano-flow liquid chromatography system coupled with Orbitrap Fusion Tribrid mass spectrometer. SRM/MRM data was acquired using an HPLC system (Thermo Vanquish) connected with Triple Quadrupole Mass spectrometer (TSQ Altis Thermo) Comparative protein expression data was obtained through Proteome Discoverer analysis of the mass spectrometry raw data. The expression data is presented in a web-based portal www.fishprot.org/ The Django framework was used in designing this portal. Currently it allows the data visualization as heatmap across the studied tissue. The data for PTMs (Phosphorylation, Methylation and Acetylation) was obtained using PTMProphet tool in the Trans-Proteomic Pipeline.
Data format	Analyzed
Description of data collection	For discovery based proteomic data, 19 normal tissues were taken for protein, peptide and PTM profiling where FDR of 1% was considered. For targeted proteomic comparison, peak intensities obtained from MRM analysis were compared across 9 tissues.
Data source location	Institution: Indian Institute of Technology, Bombay. City/Town/Region: Mumbai 400076. Country: India. Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: 19.1334° N, 72.9133° E
Data accessibility	Dataset 1: MS raw data and the protein database (.FASTA) is available at the Proteome-Xchange Consortium identifier PXD026377. All msf files (.msf) obtained from Proteome Discoverer analysis are available at PRIDE under the identifier PXD027141. Dataset 2: Datasets for selected/multiple reaction monitoring (S/MRM) experiment at Panorama public- https://panoramaweb.org/rohuorganwisesrm.url All the data is freely available at https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00759. Fish Proteome Map portal www.fishprot.org/
Related research article	Nissa, Mehar Un, Nevil Pinto, Arijit Mukherjee, Panga Jaipal Reddy, Biplab Ghosh, Zhi Sun, Saicharan Ghantasala, Chetanya Chetanya, Sanjyot Vinayak Shenoy, Robert L. Moritz, Mukunda Goswami, and Sanjeeva Srivastava. 2022. "Organ-Based Proteome and Post-Translational Modification Profiling of a Widely Cultivated Tropical Water Fish, *Labeo Rohita*." *Journal of Proteome Research* 21(2):420–37. doi:10.1021/acs.jproteome.1c00759.

Value of the Data

- This data is of great significance to the scientific community and researchers who focus on basic biological or industrial (fisheries) research.
- The provided data on organ-wise proteome, methylome, acetylome, and phosphoproteome will help in exploring the role of proteins and post-translationally modified proteins in this food fish.
- This dataset provides a map of proteins expressed by a tissue which gives clues to their function and organization in the cell.
- The provided data can be further analyzed and compared with other studies to identify the protein targets for accessing other aspects such as eco-toxicological monitoring.
- Our data extends a thorough understanding of commercially and ecologically important fish Rohu, and can benefit researchers involved in basic research, on a fish model as well as researchers associated with aquaculture and food industry.

1. Rationale and Objectives

Aquaculture is one of the food industries with the fastest growth rates. However, progress in aquaculture research has been significantly hampered by the scarcity of multi-omics data for the majority of the cultivated aquaculture species. One of the economically significant aquaculture species is Rohu. The recent release of Rohu's complete genome sequence [2] has increased the demand for creating an equivalent proteome map. In order to achieve this goal, this dataset was created to offer a thorough organ-based protein and PTM map of different tissue samples for this species.

2. Data Description

We describe two datasets acquired to develop a global proteome and PTM map for Rohu, a significant aquaculture species. Dataset 1 consists of discovery proteomic data obtained using Orbitrap Fusion based LC-MS/MS through data dependent acquisition mode. Dataset 2 consists of targeted proteomic data obtained using Multiple reaction monitoring approach of mass spectrometry. Before performing these LC-MS/MS analyses, protein was extracted from different tissue samples, followed by SDS-PAGE based fractionation. Protein was digested using in-gel method. The peptides were taken first to generate the discovery proteomic data which was acquired through 289 mass spectrometry runs by injecting one microgram of peptide each time. This raw data (dataset 1) obtained was analyzed in two different tools to perform the comparative protein expression and PTM identification in all the tissue samples taken (Fig. 1). MRM data consisting of 54 raw files was acquired for a set of 45 proteins targets. The data presented here consists of three figures and three tables (given in supplementary). Table 1 (.docx) has the details of all the reagents and equipment used to acquire or analyze this dataset. Fig. 2 and Table 2 (.xls) shows the distribution of proteins in each tissue based on number of unique peptide identifications. Number of peptide identifications with zero missed cleavage are represented in Fig. 3 and Table 3 (.xls).

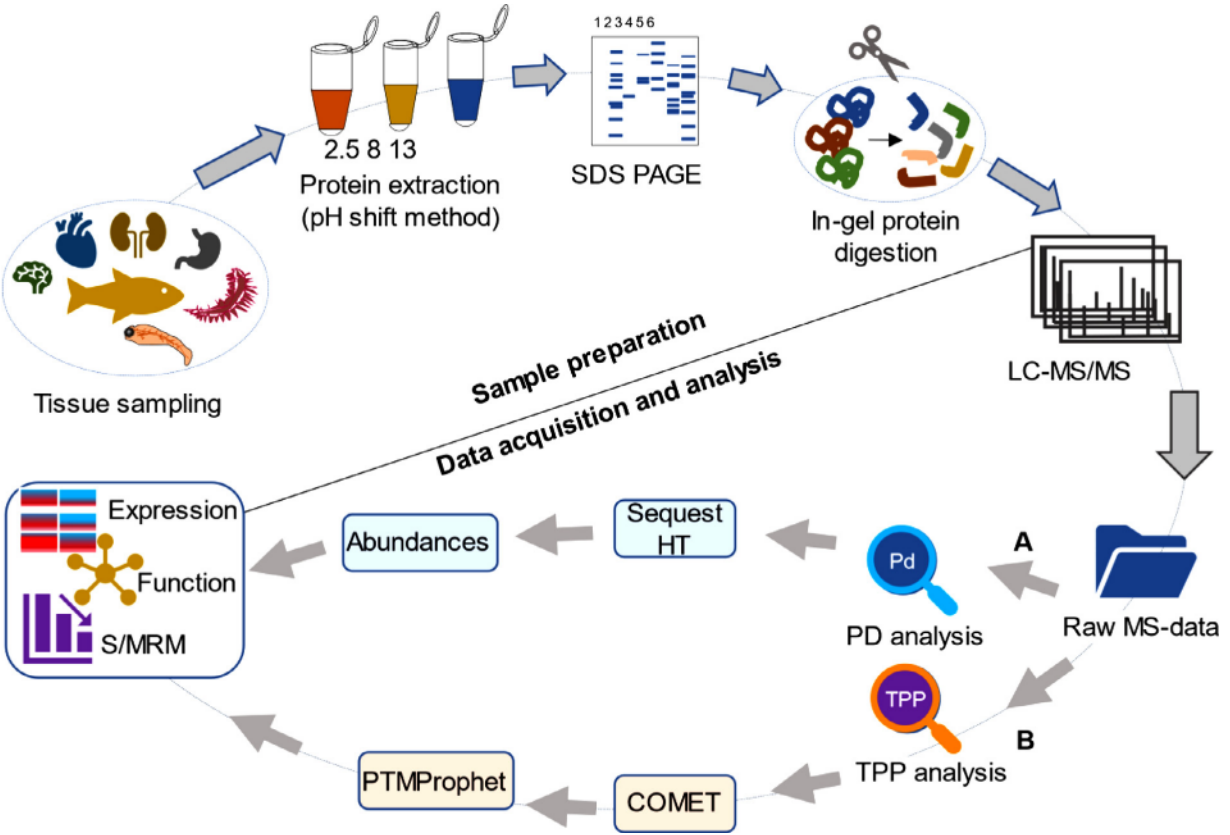

Fig. 1. Schematic representation of workflow for sample preparation and data analysis: Tissue samples were collected from fish and proteins were extracted using different methods as pH shift method. SDS-PAGE was performed to fractionate the sample. Gel pieces were processed for in-gel digestion to obtain peptides for Liquid chromatography tandem mass spectrometry (LC-MS/MS). Raw data obtained was analyzed using two different pipelines; **A**. Proteome Discoverer (PD) tool using Sequest HT for protein search to finally obtaining protein abundances and, **B**. Trans-Proteomic Pipeline (TPP) with COMET and PTM prophet tools to identify the post-translational modifications (PTMs) across the tissue samples. Data was further analysed to obtain comprehensive picture of protein expression and function. Validation experiment for trends of protein expression was performed using Selected/Multiple reaction monitoring (S/MRM).