Muscle Proteomics of the Indian Major Carp Catla (*Catla catla*, Hamilton)

Bimal Prasanna Mohanty1*, Sudeshna Banerjee1, Soma Bhattacharjee1, Tandrima Mitra1, Gopal Krishna Purohit1, Anil Prakash Sharma1, Dhanasekar Karunakaran1 and Sasmita Mohanty2

1ICAR- Central Inland Fisheries Research Institute, Biochemistry Laboratory, FREM Division, Barrackpore, Kolkata-700 120, India
2KIIT School of Biotechnology, KIIT University, Bhubaneswar- 752024, India

Abstract

**Background:** Skeletal muscle proteomics aims at global identification, cataloging and biochemical characterization of the entire protein complement of voluntary contractile tissues. In the lower vertebrates like fish, skeletal muscle contributes 34-48% of the total body weight and the muscle composition contributes strongly to the quality. Characterization of the muscle proteome is a key to many aspects of aquaculture, encompassing physiology, growth, food safety, seafood authentication and quality, traceability and shelf-life. *Catla catla* is a commercially important carp species contributing a major share to be freshwater aquaculture production in the Indian subcontinent; however, little omics information is available on this species.

**Methods:** In the present study, protein spots excised from 2-D gels of catla muscle proteome were identified through MALDI-TOF-MS. Based on the muscle proteomic information, transcript information on the identified proteins was generated.

**Results:** A reference muscle proteome map for *Catla catla* was generated and 70 protein spots from 2-D gels, representing 22 proteins, have been identified by MALDI-TOF MS. We have also generated the partial gene sequence information on the identified proteins which have been submitted at GenBank.

**Conclusion:** This is the first study on the muscle proteogenomics of the commercially important carp *Catla catla*. Besides adding to the existing knowledgebase on comparative muscle proteomics, the information generated would also serve as the baseline proteogenomic information on this Indian major carp.

**Keywords:** *Catla catla*; 2-D electrophoresis; Proteogenomics; MALDI-TOF-MS; Reference muscle proteome; Indian major carp

Introduction

Proteomics is an unbiased, technology driven approach for the comprehensive cataloging of entire protein complements and represent an ideal analytical tool for the high throughput discovery of protein alterations in health and disease [1]. Mass spectrometry-based proteomics is concerned with the global analysis of protein composition, posttranslational modifications and the dynamic nature of expression levels. The generation of large data sets on protein expression levels makes proteomics a preeminent hypothesis generating approach in modern biology [2,3]. Organisms with no available genome sequence data can be studied using comparative proteogenomic approach [3]. Aiming to better understand proteome alterations, it is vital to have a reference proteome map for a specific tissue and species. In this respect, proteogenomics is a thorough approach for the detailed biochemical analysis of heterogeneous and plastic types of tissue, such as muscles.

Muscle plays a central role in whole-body protein metabolism by serving as the principal reservoir for amino acids to maintain protein synthesis in vital tissues and organs [4]. Skeletal muscle fibers represent one of the most abundant cell types in the vertebrates [5] and contractile fibers of skeletal muscle tissues provide coordinated excitation-contraction-relaxation cycles for voluntary movements and postural control [6], besides playing a central physiological role in heat homeostasis and presenting itself as a crucial metabolic tissue that integrates various biochemical pathways [7]. Skeletal muscle proteomics aims at the global identification, detailed cataloguing and biochemical characterization of the entire protein complement of voluntary contractile tissues in normal and pathological specimens [8]. Muscle proteomics has been applied to the comprehensive biochemical profiling of developing, maturity and ageing muscle, as well as the analysis of contractile tissues undergoing physiological adaptations seen in disuse atrophy, physical exercise and chronic muscular transformations [2].

Among the higher vertebrates muscle proteomic information is available on human [9], rat [10], rabbit [11] and chicken [12] and in lower vertebrates such information has been generated on the piscines zebrafish [13], common carp [14], seabream [15], cod [16] and the snakehead [17]; however, little is known about the patterns, mechanisms of muscle growth and muscle proteome of the Indian major carps rohu (*Labeo rohita*), catla (*Catla catla*) and mrigal (*Cirrhinus mrigala*), the major aquacultured species in the Indian subcontinent. In fish, skeletal muscle is the largest organ system and represents the edible part. It constitutes 34-48% of the total body weight of fish [15]. Muscle composition contributes strongly to quality; in fact texture, elasticity, and water holding capacity, all features highly related to quality, are dependent on number and integrity of muscular fibers [18]. The number of muscle fiber recruited during growth is subjected to variation depending on several factors, such as the fish strain, diet, exercise training and temperature [15].
Catla ([**Catla catla**](http://www.uniprot.org/uniprot/) (n=12), weighing 800-1000 g (length 35-40 cm), were procured from a reputed fish farm and hatchery in Kolkata. The species status of the specimens was confirmed by analyzing for species-specific RAPD markers for catla [19].

### Preparation of muscle extracts and protein quantification

Axial white skeletal muscle from midway down the body, under the dorsal fin and above the lateral line, was swiftly dissected out from fishes euthanized with MS 222 (>100-200 mgL−1). For muscle protein extraction, white muscle tissues were pooled and mechanically homogenized in ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4·7H2O, 1.4 mM KH2PO4), pH 7.3 containing protease inhibitor cocktail (Sigma) [20]. To minimize protein modification or degradation, all dissection and sample processing was performed on ice. The homogenates were centrifuged in a high speed refrigerated centrifuge (Biofuge FRESCO, Heraeus) at 10,000 rpm at 4°C for 10 min and supernatants (representing the soluble protein extracts) were aspirated out. Protein concentration of the extracts was determined using Bradford method [21] using BSA (Sigma) as the standard. The samples were stored as aliquots at -40°C, until further use.

### Gel electrophoresis

Prior to 2-D GE analysis, the proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [22] to check the protein quality and also to ensure equal loading. For 2-D GE analysis, the first dimension run (isoelectric focusing) was performed using a Bio-Rad Protean XI cell (Bio-Rad). The gels were stained with Coomassie Brilliant blue or Coomassie-silver double stained and stacking gel on a PROTEAN XI cell (Bio-Rad). The gels were stained with Coomassie Brilliant Blue or Coomassie-silver double stained and the 2-D gel images were acquired by ImageScanner III LabScan 6.0 (GE Healthcare Biosciences).

### MALDI-TOF-TOF-MS

Protein spots of interest were cut from the 2-D polyacrylamide gel, destained in methanol and ammonium bicarbonate buffer, and digested overnight with trypsin. The resulting peptides were extracted following standard techniques [24] by two 20-min incubations with 10-20 µL ACN containing 1% TFA, depending on the size of the gel piece. The resulting tryptic peptide extract was dried by rotary evaporation and stored at -20°C for further analysis by MS. The peptides were analysed by MALDI-TOF-TOF mass spectrometer using a 5800 Proteomics Analyzer (AB Sciei).

The dry peptide samples were reconstituted in 10 µL standard diluent (30:70; ACN:water). The resulting solution was diluted 1:10 with matrix solution (CHCA, 10 mg/mL) and spotted on a 384-well Opti-TOF stainless steel plate. The spotted samples were analyzed using a first run of standard TOF-MS. The system was set to perform a second run of MS/MS focused on the 20 most intensive peaks of the first MS (excluding peaks known to be of trypsin). The laser was set to fire 400 times per spot in MS and 2000 times per spot in MS/MS mode. Laser intensity was 2800 J and 3900 J for MS and MS/MS, respectively. A mass range of 800-4000 amu with a focus mass of 2100 amu was used.

For protein identification, peptide masses from trypsin digests derived using the MALDI-TOF-MS were used to search against Ludwig NR Database using the MASCOT program ([www.matrixscience.com](http://www.matrixscience.com)) [25]. The MASCOT search parameters were as follows: peptide mass accuracy was 100 ppm, protein modifications: cysteine as S-carbamidomethyl-derivative and oxidation of methionine allowed. The default search parameters used were: enzyme, trypsin; maximum missed cleavages, 1; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M); peptide tolerance+0.4 Da; Fragment mass tolerance+0.4 Da; Protein mass, unregistered; instrument–Default.

### Data and network pathway

Proteins were identified by MS analyses and protein database (Mascot) searches, basic descriptions of identified protein spots including protein names, accession numbers, theoretical molecular mass and pl values were provided. In addition to these descriptions other fundamental information relevant to identified proteins were acquired through further search of databases available in the public domain. Database accession numbers for identified proteins were provided as inputs to the Uniprot database ([http://www.uniprot.org/uniprot/](http://www.uniprot.org/uniprot/)) to obtain gene ontology annotation and sequence length. The consensus lists of proteins obtained were then analyzed for their respective molecular function, biological processes involved protein class and associated metabolic pathway by using Panther classification system software ([www.pantherdb.org/pathway/](http://www.pantherdb.org/pathway/)).

### Tissue collection and RNA extraction for transcript analysis

White muscle tissues were collected from anaesthetized fishes and stored at (-80°C) in RNA later (Sigma) to avoid RNA degradation. Subsequently, in total RNA isolation was completed by using RNA Express Reagent (Himedia). Briefly, samples (~1 gm) were homogenized in 1 ml RNA Express reagent; chloroform was added to separate the mixture into protein, DNA and RNA. The RNA supernatant was precipitated with 2- propanol (Himedia). The pellets were resuspended in 25 µl of DEPC (Himedia) treated water and stored at -80°C until further use.
Citation: Mohanty BP, Banerjee S, Bhattacharjee S, Mitra T, Purohit GK, et al. (2013) Muscle Proteomics of the Indian Major Carp Catla (Catla catla, Hamilton). J Proteomics Bioinform 6: 252-263. doi:10.4172/jpb.1000288

**cDNA preparation by Reverse Transcriptase**

Following isolation of RNA, the concentration of each RNA sample was measured and DNase (Fermentas) treatment was done to remove genomic DNA carryover. RNA (1μg) was reversed transcribed using M-MuLV reverse transcriptase. RT reactions were conducted in 20 μl reactions containing 5 μg DNase I-treated RNA, 1 μl oligo dT primers (Fermentas), 4μl dNTPs (Himedia), 0.5 μl RNase inhibitor (Fermentas), 5× RT reaction buffer, and 1 μl of RevertAid RT (Fermentas). Detailed procedures were followed from the manufacturer’s instructions.

**Polymerase chain reaction**

To amplify the genes from the catla muscle cDNA library, primers designed against the nucleotide sequences of the zebrafish, common carp and other related species with the assistance of Primer3 software were used (http://frodo.wi.mit.edu/). The reaction mixture (25 μl) consisted of 10 ng cDNA, 10X buffer (Himedia), 10 mM dNTPs (Himedia), 10 pM gene specific primer and 5U of Taq Polymerase (Himedia). PCR reaction was carried out by using thermal cycler (Veriti, Applied Biosystem). The amplification conditions were as follows: initial predenaturation at 95°C for 3 min followed by 45 cycles of amplification (denturation at 95°C for 45 s, annealing for 45 s at temperatures optimized for specific genes (Table 1), 72°C for 1min and a final extension at 72°C for 10 min. 8 μl of the PCR product was analyzed in 1.6% agarose gel in ImageQuant LAS 4000 (GE Healthcare).

| Sl No. | Target Gene                          | Primer Sequence                                      | Annealing Temperature |
|--------|-------------------------------------|------------------------------------------------------|-----------------------|
| 1      | Pyruvate kinase                     | F:SCAGYTGTTTGGAGGAGCTAC                              | 52°C                  |
|        |                                     | R:CCTGTCACCAATCACACCAG                               |                       |
| 2      | Creatine Kinase                     | F:TCACCCCTGCCTCAACAA                                  | 42°C                  |
|        |                                     | R:TGGCCCTGGACTACATCC                                  |                       |
| 3      | Beta actin                          | F:AGAGAGAATTGTGGCTGACATC                             | 60°C                  |
|        |                                     | R:CCTCGACTGACAGAGATTGTTGG                            |                       |
| 4      | Phosphorylase kinase                | F:GACAGGAGCATGAGTACGACA                              | 50°C                  |
|        |                                     | R:AAACCCATCCCAACAGGACGAC                             |                       |
| 5      | Adenylate Kinase                    | F:CATTGAGGCAGTACAGTAC                                | 45°C                  |
|        |                                     | R:CACTGACATCCACAGGACGAC                               |                       |
| 6      | Enolase                             | F:ATGCAGCGATCAATACA                                  | 47°C                  |
|        |                                     | R:CCTGCTTCTCTATGACGCTC                                |                       |
| 7      | Phospho-hydroxide phosphate          | F:CCTCGAGGCTGCTGCTGCT                                 | 47°C                  |
|        |                                     | R:ACACCGCAGACGACGACGAC                                |                       |
| 8      | PDZ and LIM                         | F:GCAGAACATGCCACAGGACGAC                              | 47°C                  |
|        |                                     | R:TCATTTGCGAGGCTGCTG                                 |                       |
| 9      | CapZB                               | F:CATGCCTGCGATAACACCAAGA                              | 50°C                  |
|        |                                     | R:CGGGTGGAGGGGTGAAACTAC                              |                       |
| 10     | Aspartate aminotransferase          | F:CATGCCTGCGATAACACCAAC                               | 46°C                  |
|        |                                     | R:ACACGGTCAAGGATCTCTTCTT                              |                       |
| 11     | Glyceraldehyde 3 phosphate dehydrogenase | F:CTGGTGACCCGTTGGCTGCTT                                | 42°C                  |
|        |                                     | R:TTTGGCGGCTTCTGCCTTATAA                              |                       |
| 12     | Apolipoprotein                      | F:CTCTCCCTCATCCTGCTGCCCTTATAA                         | 45°C                  |
|        |                                     | R:CTCCAGCAGCTACTCTGACAGA                              |                       |
| 13     | Fructose 1,6 bisphosphate           | F:AAFMGKACCATYCAACGAMCGC                              | 50°C                  |
|        |                                     | R:ACACGCTGRTGCRCAAGATACAC                             |                       |
| 14     | Lactate Dehydrogenase               | F:AGTAGCAGCCCCAATCCACGTC                              | 47°C                  |
|        |                                     | R:CCACATTACACACAGCATC                                 |                       |
| 15     | Phosphoglycerate kinase             | F:AGGGGGGCAGGTGCAAAAGATA                              | 48°C                  |
|        |                                     | R:CCTGGTGGTGTGGCTTCTCAT                               |                       |
| 16     | Transferrin                         | F:GGACTACCAAGGCTTATGGCAT                              | 64°C                  |
|        |                                     | R:GGCCACCATCGACTCAAT                                  |                       |
| 17     | hap47                               | F:CACCTGAGGATGAGGATGCAAT                              | 59°C                  |
|        |                                     | R:AAGGGAATGAGGATGACGTC                                |                       |
| 19     | hap60                               | F:AAAGATGGGGTGCAAGAGTGGC                              | 54°C                  |
|        |                                     | R:TGTTAGGAGGACAGATGCTG                                |                       |
| 17     | hap70                               | F:GCATGGTGACACCATTTGTG                                | 53°C                  |
|        |                                     | R:CCTCGGCTTGGAGAAATCC                                 |                       |
| 18     | hsc71                               | F:TAAGGACATGGGCTTTTAAT                                 | 60°C                  |
|        |                                     | R:CCACAGCAGTGATGGACTC                                 |                       |
| 20     | hap90                               | F:GGGAAATTTTCTCTCCAGAGGAC                             | 52°C                  |
|        |                                     | R:CCGAGATGACCGATGACGAGA                               |                       |
| 21     | 18s RNA                             | F:GGAGGTTGCAAGGAGGATCACG                              | 62°C                  |
|        |                                     | R:AAACCCAGAAATCCAGGACCA                                |                       |

Table 1: PCR Primers used for transcript analysis.

Source: [28]
Sequencing of the transcripts

Bands of appropriate size were purified with Hipura Quick gel purification Kit (Himedia) and were further sequenced by using Sanger’s dideoxy sequencing protocol (ABI3730 XL) with specific primers. To confirm the identity of the gene the partial sequences were subjected to Blastn and Blastx [26] for comparing with the Genbank Nucleotide and Protein database.

Results

Gel electrophoresis

SDS-PAGE of *Catla catla* soluble muscle proteins separated the muscle extracts into 22 bands in the MW range of 14 to >205 kDa (Figure 1A). 2-D gel electrophoresis profiles of catla muscle proteins were generated; a representative 2-D gel profile of the catla muscle proteome is shown (Figure 1B). Coomassie-silver double staining enabled visualization of soluble white muscle proteins into 130 spots. Majority of the separated proteins fall in the range of 20-60 kDa and pI 6-8, except few with MW >90 kDa and pI 6.5-7.5.

MALDI-TOF-MS

Proteomic analysis led to identification of 70 protein spots (Table 2 and Figure 2), which have been cataloged in an in-house fish proteomic database FISHPROT ([http://www.cifri.ernet.in/fishprot.html](http://www.cifri.ernet.in/fishprot.html)) [27]. Out of the 70 protein spots identified, 43 were identified as metabolic enzymes involved in carbohydrate (32) (Supporting Information Figure S1), protein (1), lipid (5) and nucleotide (5) metabolism. Another 19 spots were found to be cytoskeletal proteins, of which 14 are CK (Supporting Information Figure S2). Three proteins were identified to be associated with signal transduction and five were categorized separately based on their functions (Table 2 and Figure 2).

Proteins identified in other fish muscle proteomes reported earlier via gel-based proteomics were compared and newly identified protein spots in catla, which were not shown in other species proteome maps, prior to this study, are marked (Supporting Information Table S1). Similarly, comparative muscle protein profiles in the muscle proteome of higher and lower vertebrates, as identified via gel-based methods vis-à-vis proteins identified in catla (Supporting Information, Table S2).

Pathway analysis

The major pathways associated with the identified proteins of catla based on MALDI-TOF-MS data are mostly related to different metabolic pathways (Table 2). Upon analyzing the muscle proteome dataset by Panther classification system software, the total biological pathway was divided into eight categories: viz. asparagines and aspartate biosynthesis pathway, de novo purine biosynthesis, fructose-galactose metabolism, glycolysis, G-protein signaling pathway, phenylalanine biosynthesis, pyruvate metabolism, tyrosine biosynthesis (Supporting Information Figure S3 A-i). Large groups were found to be involved in glycolysis pathway (Supporting Information Figure S1) and proteins associated with this pathway were subdivided as pyruvate kinase, aldolase, enolase (Supporting Information Figure S3A-ii). G-protein signaling pathway represents both phosphorylases- A and -B (Supporting Information Figure S3A-iii).

According to biological processes, proteins are divided into three
Figure 2: Classification of protein spots from the muscle proteome of *Catla catla*, identified by MALDI-TOF MS. The proteins are clustered into different categories based on the pathways involved.

| Sample no | MASCOT Score | Accession no | Sequence Coverage | Calculated pI; Nominal Mass | Protein identified                                      | Species              | Sequence length | Gene |
|-----------|--------------|--------------|-------------------|-----------------------------|---------------------------------------------------------|----------------------|----------------|------|
| **Carbohydrate metabolism** | | | | | | | | |
| C-23      | 619          | A3RH18       | 30%               | 6.90;26588                  | Triosephosphate isomerase                               | Poecilia reticulate  | 247 AA         | TPI  |
| C-24      | 672          | A3RH18       | 30%               | 6.90;26588                  | Triosephosphate isomerase                               | Poecilia reticulate  | 247 AA         | TPI  |
| C-25      | 672          | A3RH18       | 21%               | 6.96;26794                  | Triosephosphate isomerase                               | Poecilia reticulate  | 247 AA         | TPI  |
| C-112     | 527          | E3TC72       | 29%               | 6.90;26539                  | Triosephosphate isomerase                               | Ictalurus furcatus   | 248 AA         | TPISB |
| C-111     | 287          | C1J0M6       | 29%               | 5.89;25057                  | Triose phosphate isomerase                              | Gillichthys mirabilis| 234 AA         | TPI  |
| C-41      | 70           | B5DGU1       | 12%               | 8.20;58295                  | Pyruvate kinase                                          | *Salmo salar*        | 530 AA         | pk   |
| C-44      | 61           | Q6DG64       | 3%                | 6.88;58317                  | Pyruvate kinase                                          | *Danio rerio*        | 530 AA         | pkm2b|
| C-86      | 182          | B5DGU1       | 12%               | 8.20;58295                  | Pyruvate kinase                                          | *Salmo salar*        | 530 AA         | pk   |
| C-114     | 414          | Q8JH72       | 25%               | 8.27;39701                  | aldolase A                                               | *Danio rerio*        | 364 AA         | Aldoa|
| C-84      | 253          | Q6IQP5       | 21%               | 6.16;47030                  | Enolase 1, (Alpha)                                       | *Danio rerio*        | 432 AA         | eno1 |
| C-116     | 596          | Q6TH14       | 23%               | 6.25;47442                  | Enolase 1, (Alpha)                                       | *Danio rerio*        | 433 AA         | eno3 |
| CM-162    | 255          | F1QBW7       | 8%                | 7.46;51287                  | Enolase (Fragment)                                       | *Danio rerio*        | 467 AA         | eno3 |
| C-78      | 137          | Q3B7R7       | 17%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-95      | 887          | Q3B7R7       | 28%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-97      | 809          | Q3B7R7       | 26%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-98      | 493          | Q3B7R7       | 23%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-99      | 776          | Q3B7R7       | 28%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-101     | 115          | Q3B7R7       | 9%                | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-115     | 311          | Q3B7R7       | 20%               | 9.07;46662                  | Enolase 3 protein (Fragment)                             | *Danio rerio*        | 423 AA         | eno3 |
| C-39      | 211          | Q76BD4       | 10%               | 5.76;41622                  | Phosphoglycerate kinase                                  | Acipenser baerii     | 389 AA         | PGK  |
| C-82      | 260          | Q76BD4       | 10%               | 5.76;41622                  | Phosphoglycerate kinase                                  | Acipenser baerii     | 389 AA         | PGK  |
| C-120     | 271          | Q5XJ10       | 17%               | 8.20;35761                  | Glycerinaldehyde-3-phosphate dehydrogenase              | *Danio rerio*        | 333 AA         | gapdh|
| C-119     | 95           | Q9PVK5       | 15%               | 6.83;36223                  | L-lactate dehydrogenase                                  | *Danio rerio*        | 333 AA         | lidha|
Proteins related to Glycogenolysis

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-59| Q503C7    | Phosphorylase, glycogen (Muscle) | Danio rerio      | 8%      | pgyma            | 6.29;42825  |
| C-122| Q503C7    | Phosphorylase, glycogen (Muscle) | Danio rerio      | 18%     | pgyma            | 6.67;96870  |
| C-91 | Q7SXW7    | Phosphoglucomutase 1             | Danio rerio      | 27%     | pgm1             | 5.74;61090  |
| C-92 | Q7SXW7    | Phosphoglucomutase 1             | Danio rerio      | 28%     | pgm1             | 5.74;61090  |
| CM-90| F1QF00    | Phosphoglucomutase 1             | Danio rerio      | 23%     | pgm1             | 5.74;61150  |
| CM-124| F1QF00   | Phosphoglucomutase 1             | Danio rerio      | 13%     | pgm1             | 5.74;61150  |

Proteins related to Gluconeogenesis

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| CM-74| G3NP1     | Fructose-bis-phophatase class    | Gasterostes       | 4%      | FBP2             | 6.66;36679  |
| CM-123| G3NP1     | Fructose-bis-phophatase class    | Gasterostes       | 4%      | FBP2             | 6.66;36679  |
| C-104| A5WVL5    | Fructose-1,6-bisphosphatase 2    | Danio rerio      | 12%     | fbp2             | 6.15;36769  |

Proteins related to Amino Acid Metabolism

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-38| Q7ZUW8    | Aspartate amino transferase      | Danio rerio      | 7%      | got1             | 5.77;29842  |

Lipid Metabolism

Proteins related to Cholesterol Transport

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-111| A92OV6    | Apolipoprotein A-I               | Gobiocypris       | 4%      | APOA1            | 5.77;29842  |

Proteins related to Glycerol Phosphate Shuttle

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-73| QT1E0     | Glycerol-3-phosphate dehydrogenase 1b | Danio rerio | 20%     | gpd1b            | 6.38;38169  |
| C-75| QT1E0     | Glycerol-3-phosphate dehydrogenase 1b | Danio rerio | 12%     | gpd1b            | 6.38;38169  |
| C-77| QT1E0     | Glycerol-3-phosphate dehydrogenase 1b | Danio rerio | 10%     | gpd1b            | 6.38;38169  |
| C-76| QT1E0     | Glycerol-3-phosphate dehydrogenase 1b | Danio rerio | 16%     | gpd1b            | 6.38;38169  |

Proteins related to Nucleic Acid Metabolism

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-8 | E3TD55    | Adenylate kinase                 | Ictalurus         | 17%     | adk              | 6.00;21531  |

Proteins related to Musculo-skeletal system

| ID  | Accession | Description                      | Species          | Percent | Primary Accession | Mowse Score |
|-----|-----------|----------------------------------|------------------|---------|------------------|-------------|
| C-57| D0VBL9    | Muscle-type creatine kinase CKM1 | Platichthys       | 6%      | CKM1             | 6.22;42982  |
| C-58| Q7T306    | Creatine kinase CKM5             | Danio rerio      | 8%      |ckmb             | 6.29;42825  |
| C-65| Q9Y116    | Creatine kinase CKM1             | Cyprinus carpio  | 9%      |ckmb             | 6.21;42751  |
| C-86| Q9Y116    | Creatine kinase CKM1             | Cyprinus carpio  | 9%      |ckmb             | 6.21;42751  |
| C-87| Q7T306    | Creatine kinase CKM3             | Danio rerio      | 13%     |ckmb             | 6.29;42825  |
| C-68| Q7T306    | Creatine kinase CKM3             | Danio rerio      | 6%      |ckmb             | 6.29;42825  |
| C-69| Q9Y115    | Creatine kinase CKM2             | Danio rerio      | 21%     |ckmb             | 6.22;42901  |
| C-70| Q7T306    | Creatine kinase CKM3             | Danio rerio      | 23%     |ckmb             | 6.29;42825  |
| C-72| D1ME15    | Muscle-type creatine kinase CKM1 | Plagius major    | 13%     |ckmb             | 6.58;42920  |
| C-80| Q9Y115    | Creatine kinase CKM2             | Cyprinus carpio  | 19%     |ckmb             | 6.22;42901  |
| C-110| Q7T306    | Creatine kinase CKM3             | Danio rerio      | 21%     |ckmb             | 6.29;42825  |
| C-117| A2BHA3    | Creatine kinase, muscle a        | Danio rerio      | 29%     |ckmb             | 6.32;42788  |
| CM-142| Q7T306   | Ckmb protein                     | Danio rerio      | 14%     |ckmb             | 6.29;42825  |
| CM-143| Q9Y116    | Creatine kinase M1               | Cyprinus carpio  | 13%     |ckmb             | 6.21;42751  |
| CM-138| G3NTW0   | CAPZB gene product               | Gasterostes aculeatus | 23%     |CAPZB            | 5.40;31032  |
categories, viz. immune system, system process and metabolic process (Supporting Information, Figure S3B-i) of which metabolic process represented the largest group, which is subdivided as primary and ROS metabolic process (Supporting Information, Figure S3B-ii). Primary metabolic process is further divided in to carbohydrate, cellular amino acid, lipid and nucleotide metabolism (Supporting Information Figure S3B-iii).

On the basis of molecular function, proteins are divided into two classes: viz. catalytic activity and antioxidant activity (Supporting Information Figure S4A-i) where catalytic activities of the identified proteins are further divided into transferase, isomerase, lyase and oxidoreductase catalysis (Supporting Information, Figure S4A-ii). Transferase activity of the enzyme proteins are further divided into glycosyltransferase, transaminase and a large group associated with kinase activity (Supporting Information Figure S4A-iii), which are further divided into nucleotide, amino acid, carbohydrate kinase (Supporting Information Figure S4A-iv).

Dataset proteins were grouped into five protein classes’ viz. transferase, isomerase, kinase, lyase and oxidoreductase (Supporting Information Figure S4B-i). Proteins of transferase class are further sub grouped into transaminase, kinase and phosphorylase (Supporting Information Figure S4B-ii). The kinases (Supporting Information Figure S4B-iii) are divided as described in Supporting Information Figure S4A-iv and oxidoreductase activity is further divided as peroxidase and dehydrogenase (Supporting Information Figure S4B-iv).

Pathway coverage of the major enzymes involved in carbohydrate metabolism is illustrated (Figure 3). The associated gene names, and sequence coverage are presented in adjacent grey boxes of the identified proteins (Figure 3).

**Transcript information on the identified proteins**

Based on the muscle proteome data, transcript informations on the identified proteins were generated. Out of 22 proteins identified, partial gene sequences of 16 proteins were generated which included pyruvate kinase (pk), CK (ck), betaactin (ACTB), glycogen phosphorylase (pygma), adenylate kinase (adk), enolase (eno), phosphohistidine phosphatase (si:dkey-51e6.1), PDZ and LIM domain 7 (pdlim7), and phosphoglycerate kinase (PGK). The partial sequence information on transferrin gene (TF) exactly matched with existing sequence information available at GenBank (Accession No. AM690341 [28]. Additionally, sequence information on IBS RNA and heat shock protein genes, hsp47, hsp60, hsp70, hsc71, and hsp90 has also been generated. Sequence information on all these genes has been deposited at GenBank (Table 3).

### Discussion

Proteomics, the global analysis of protein synthesis, studies the end product of gene expression i.e. the protein, for which it is also termed as functional genomics. The nucleic acid based technologies for analyzing differential gene expression assay only mRNA expression, which is not always reflected in the levels of protein synthesis. Two-dimensional protein gels, combined with peptide mass mapping by MALDI-TOF MS for protein identification, are widely used for determining differential protein synthesis in biological systems. In the current study, we have generated muscle proteogenomic information on the commercially important carp *Catla catla*.
Table 3: Partial gene sequences with GenBank Accession Nos. of genes coding for the proteins identified in muscle proteome of *Catla catla*.

| Sl No. | Target Gene | Amplicon Length (bp) | GenBank Accession No. |
|-------|-------------|----------------------|-----------------------|
| 1     | Pyruvate kinase | 535                  | KC707842              |
| 2     | Creatine Kinase  | 204                  | KC788423              |
| 3     | Beta actin      | 372                  | KC788424              |
| 4     | Phosphorylase kinase | 202              | KC788754              |
| 5     | Adenylate Kinase | 240                  | KC816537              |
| 6     | Enolase         | 290                  | KC816538              |
| 7     | Phosphohistidine phosphatase | 210         | KC816539              |
| 8     | PDZ and LIM     | 390                  | KC816540              |
| 9     | CapZB           | 200                  | KC816541              |
| 10    | Aspartate aminotransferase | 280     | KC816542              |
| 11    | Glyceroldehyde 3 phosphate dehydrogenase | 206    | KC887542              |
| 12    | Apolipoprotein  | 327                  | KC887543              |
| 13    | Fructose 1,6 bisphosphate | 207        | KC887540              |
| 14    | Lactate Dehydrogenase | 234      | KC887544              |
| 15    | Phosphoglycerate kinase | 543         | KC887545              |
| 16    | Transferrin     | 633                  | AM690341              |
| 17    | hsp47           | 460                  | KC915024              |
| 18    | hsp70           | 373                  | KC599207              |
| 19    | hsc71           | 415                  | KC800800              |
| 20    | hsp90           | 244                  | KC800801              |
| 21    | 18s RNA         | 514                  | KC915025              |

Source: [28]

Figure 3: Proteomic coverage of major enzymes involved in carbohydrate metabolism in *Catla catla* white muscle. Positional variants of enzymes involved in glycolysis, gluconeogenesis and glycogenolysis are shown: proteins identified are shown in boxes and the associated gene names and sequence coverage are shown in adjacent grey boxes.
mass fingerprints (PMF) which represented 22 proteins; this indicated the presence of positional variants, isotypes and/or the extensive post-translational modifications occurring in a physiologically active cell (Table 2).

The majority of protein identifications matched those determined from 2-D PAGE were enzymes of carbohydrate metabolism, same in the case of the previous study [14]. In many cases, the same protein was identified in multiple gel spots of similar molecular mass (Figure 1 and Table 2); these spots differed only in charge and are positional variants.

In the non-redundant Swiss-Prot database approximately 2500 sequences correspond to ray-finned fish of which 89 are common carp sequences [14]. Therefore, to accurately determine the identity of fish muscle proteins we have combined cross-species matching with MALDI-TOF-MS data. As the full genome sequence for any of the Indian major carps rohu - Labeo rohita, catla- Catla catla and mirror-Cirrhinus mrigala is not available, the carp muscle proteins were identified by matching to other species, mainly zebrafish (Danio rerio), which has been sequenced and extensively annotated. This approach is facilitated by the close taxonomic relationship of carp to zebrafish; both are cyprinids. Whilst the availability of further sequence data for carp will enhance the identification of proteins in this type of study, the results indicate that it is possible to identify fish muscle proteins through cross-species matching to a taxonomic near neighbor.

Comparative proteomic data analysis across species

Proteins identified in this study are dominated by proteins which are mainly composed of enzymes such as enolase, GAPDH, pyruvate kinase, and creatine kinase (CK), which are associated with energy production pathways. Aldolase, enolase, pyruvate kinase, CK, and their fragments have been reported with detailed characterization in the sea bream [15], snakehead [16] and common carp [14] muscle proteome to provide a number of insights on the size and environment-related variability. Muscle proteome changes in association with development and exercise, by means of 2-DE and MALDI-TOF MS studies performed in zebrafish [13] earlier. The proteome map of catla further strengthens the knowledge base for comparative muscle proteomics among different fish species (Supporting Information, Table S1). The available datasets would act as a basis for studies related to physiological status assessment of Catla catla under different environmental conditions, screening for diseases and biomarker identification for assessment of fish quality. It has been found that twelve proteins viz. aspartate amino transferase, glycerol-3-phosphate dehydrogenase (GPDH), CK M3, uncharacterized actin binding protein (CAPZB gene product), PDZ and LIM domain 7 (Development GDNF family signaling), Zgc:165344 protein, Zgc:91930 (adenylate kinase family), proteosome subunit alpha type 6, a-1-antitrypsin homolog, peroxiredoxin, DJ-1 (Fragment), novel protein similar to phosphohistidine phosphatase 1 (PHPT1) identified in Catla catla muscle proteome map are new identifications in gel-based proteomes; they have not been observed in earlier studies on zebrafish, common carp, seabream, snakehead fish and cod fish (Supporting Information Table S1) [13-17]. In an earlier study, proteome cataloging using 1-D PAGE protein separation, nano LC peptide fractionation and linear trap quadrupole (LTQ) mass spectrometry of cod Gadus morhua [16] identified a total of 4804 peptides. Moreover, proteomic signature of muscle has also been established also via non-gel based methods in Rainbow trout, Oncorhynchus mykiss [28].

The proteins identified in catla muscle proteome have also been compared with higher vertebrates; rat [10], rabbit [11], chicken [12] and human [9]. Using a combination of one-dimensional gel electrophoresis and HPLC-ESI-MS/MS, 954 different proteins were identified in human muscle [9]. Proteome analysis of rat skeletal muscle led to identification of ~50 proteins [10]. A proteomic reference map for the gastrocnemius muscle of rabbit has also been generated and 45 proteins have been identified [11]. In the present study, 10 protein spots have been identified in catla muscle proteome which have not been shown in higher vertebrate muscle proteome by gel-based proteomics [9-12]. These are aspartate amino transferase, CK M2 and M3, Zgc:165344 protein, adenylate kinase D, proteosome subunit a type 6, a-1-antitrypsin homolog, peroxiredoxin, DJ-1 (fragment) and novel protein similar to phosphohistidine phosphatase 1 (PHPT1) (Supporting Information Table S2).

Protein dataset analysis

Proteomics technologies are under continuous improvements and new technologies are introduced. Nowadays high throughput acquisition of proteome data is possible. The young and rapidly emerging field of bioinformatics in proteomics is introducing new algorithms to handle large and heterogeneous data sets and to improve the knowledge discovery process. Local proteomics bioinformatics platforms viz. FISHPROT is a database management systems and is a knowledge base for fish proteomic data.

Although all the proteins identified in catla were grouped according to their biochemical properties (Table 2 and Figure 2), they were further verified by putting the respective gene names into ‘Panther classification system software’ http://www.pantherdb.org/pathway/ which is a commercial and freely available software. This system uses the gene names of the identified proteins and classifies them into different groups on basis of their similarity to specific organisms already available in its database; the zebrafish Danio rerio in case of fish. As evident from classification of identified protein spots of Catla catla by this software, on the basis of biological pathway (Supporting Information Figure S3a) and molecular function (Supporting Information Figure S4a), majority of identified proteins are housekeeping proteins such as those involved in metabolism of carbohydrates, proteins, lipids and nucleotides (Supporting Information Figure S3b) and the musculoskeletal proteins.

According to the classification of proteins on the basis of ‘Total Biological processes’, a large chunk of identified proteins were classified under ‘Primary metabolic process’, which is further divided into carbohydrate, cellular amino acid, lipid and nucleotide metabolism (Figure 2). 22 protein spots (Table 2) identified in the catla muscle proteome represent six glycolytic enzymes, triose phosphate isomerase, pyruvate kinase, aldolase A, enolase, phosphoglycerate kinase and glyceraldehydes-3- phosphate dehydrogenase (GAPDH) (Supporting Information Figure S1). However, the Panther classification system software is able to identify only three enzyme proteins viz. pyruvate kinase, aldolase, and enolase, out of the six identified (Supporting Information Figure S3a- ii); this may be so because this classification system is limited to the information available in its database for specific organisms (under the piscines generic information only on zebrafish is available in its database) and therefore, except for proteins showing homology with zebrafish, other identified proteins were not taken in to account.

Carbohydrate metabolism

Glycolysis is a primary pathway for energy generation in most
organisms. In the present study, out of the 10 glycolytic enzymes (6/10), namely, aldolase (C-114), triose phosphate isomerase (C-23, 24, 25, 112, 113), GAPDH (C-120), phosphoglycerate kinase (C-39, 82), enolase (C-78, 84, 95, 97, 98, 99, 101, 115, 116, CM-162) and pyruvate kinase (C-41, 44, 86) have been identified and they represent 22 number of spots on the catla muscle proteome (Table 2 and Figure 3; Supporting Information Figure S1). Lactate dehydrogenase (C-119), the enzyme involved in anaerobic glycolysis, converting pyruvate to lactate, has also been identified.

Two spots (C-59, 122) were identified as glycogen phosphorylase. Glycogen phosphorylase catalyzes and regulates the entry of glucose residues into glycolysis from glycogen via glycogenolysis pathway. This is a regulatory enzyme present in both liver and muscle. In skeletal muscles the enzyme occurs in two forms, a catalytically active phosphorylated form (phosphorylase a) and a much less active dephosphorylated form (phosphorylase b). In muscle, the rate of conversion of glycogen units into glucose 1-phosphate is regulated by the ratio of the active phosphorylase a to the less active phosphorylase b. Glycogen synthase, the rate-limiting enzyme in glycogen biosynthesis is also regulated by glycogen phosphorylase (Figure 3).

Four spots (C-91, 92; CM-90, 124) were identified as phosphoglucomutase. Phosphoglucomutase is a key enzyme in the metabolism of glycogen and protein glycosylation. It is responsible for the reversible interconversion of glucose 1-phosphate to glucose 6-phosphate, both of which are key intermediates in the synthesis and breakdown of glycogen and galactose metabolism. It is also important for the formation of UDP-glucose which is an important intermediary metabolite in protein glycosylation. Inhibition of phosphoglucomutase has drastic effects on carbohydrate metabolism which reduces the steady-state levels of UDP-glucose, resulting in a defect of glycogen and trehalose biosynthesis, while galactose metabolism is inhibited, leading to galactosemia, accumulation of galactose 1-phosphate and Glucose 1-phosphate i.e., poor glycogen turnover.

Three spots (C-104, CM-74, CM-123) were identified as Fructose-bis-phosphatase, which catalyzes the conversion of Fructose-1,6-bisphosphatase to Fructose-6-phosphate, an important reaction step in gluconeogenesis.

**Lipid metabolism**

Four spots (C-73, 75, 76, 77) were identified as Glycerol 3-phosphate dehydrogenase 1b. Glycerol 3-phosphate and fatty acyl-CoA's are the common precursors for triacylglycerols and glycerol phosphorylases. Glycerol phosphate is formed in two ways; it is formed from the dihydroxy acetone phosphate generated during glycolysis by the action of cytosolic NAD-linked GPDH. It is also be formed from glycerol by the action of glycerol kinase.

Spot number C-111 has been identified as apolipoprotein A1 (Apo A-1) (Supporting Information Table S1). This protein has earlier been identified in Cyprinus carpio [14]. Apo A-1 is the major protein component of high density lipoprotein (HDL) in plasma, which confers water solubility to the lipoprotein complex thus facilitating lipid transport and metabolism. It promotes cholesterol efflux from tissues to the liver for excretion. It is a cofactor for lecithin cholesterolacyltransferase (LCAT) which is responsible for the formation of plasma cholesterol esters. In lower vertebrate species Apo A-1 is also synthesized in a number of peripheral tissues, e.g., rainbow trout gill [29] and liver [30], carp optic nerve [31] and skin [32]. In the cod (Gadus morhua) Apo A-1 is closely associated with the C3 component of the complement system [33]. Apo A-1 has also been reported to have a restorative or protective role and plays significant role in maintaining epithelial integrity [29].

**Nucleotide metabolism**

Five spots have been grouped as proteins related to nucleic acid metabolism; out of these, one spot (C-8) has been identified as adenylate kinase and the other 4 spots (C-13, 15, 20, 109) have been identified as adenylate kinase D (Table 2). The adenylate kinase D has not been reported in the vertebrates by gel-based proteomics.

**CK and other contractile proteins**

CK catalyzes the transphosphorylation between phosphocreatine and ADP and is central to the regulation of muscle bioenergetics. Large number of protein spots (13) across a broad range of MW and pl were identified as positional variants of muscle CK in the present study (Supporting Information Figure S2). It has been reported earlier that creatine/phosphocreatine interconversion played an important role in ATP regeneration since depletion of glycolytic enzymes in carp resulted in anoxia. High tissue CK activity, whether constitutive, induced, or both, may rather directly enhance contractile responses by enhancing cellular energy and contractile reserve. This high CK activity may alter local ADP levels at the contractile proteins and contribute to the enhanced contractility and myosin ATPase activity.

Complex patterns of CK isoforms exist in the skeletal muscle of fish [34-36] and three isoforms of CK have previously been identified in carp skeletal muscle [37]. These are referred to as M1, M2 and M3 and have predicted masses of 43 kDa and pl 6.22-6.32 [37]. In the muscle proteome of catla, 14 spots have been identified as CK; out of these, six (C-57, 65, 66, 72, 117, CM-143) are CK muscle type -1 (CK M-1), three are CK M-2 (C-69, 80, 142) and five are CK M-3 (C-58, 67, 68, 70, 110) (Table 2; Supporting Information Figure S2). This observation is consistent with the earlier reports on presence of multiple forms of this enzyme in muscle tissue of other fish [35,37]. CK M-3 has not been reported earlier in any vertebrates by gel-based proteomics.

Four spots (C-2, CM-126, 137, 138) were identified as actin (Table 2). Actin, a 42-kDa globular protein found in all eukaryotic cells, participates in many important cellular processes including muscle contraction, cell motility, cell division and cytokinesis, vesicle and organelle movement, cell signaling, and the establishment and maintenance of cell junctions and cell shape.

CM-141, identified as PDZ and LIM domain -7 protein, is a muscle-specific protein [9]. It is representative of a family of proteins composed of conserved PDZ and LIM domains. PDZ is an acronym combining the first letters of three proteins - Post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and Zonula occludens-1 protein (zo-1) - which were first discovered to share the domain [38]. The PDZ domain is a common structural domain of 80-90 amino-acids found in the signaling proteins of bacteria, yeast, plants, viruses and animals. LIM domains are protein structural domains, composed of two contiguous zinc finger domains, separated by a two-amino acid residue hydrophobic linker; they are named after their initial discovery in the proteins Lin11, Isl-1 and Mec-3. LIM-domain containing proteins have been shown to play roles in cytoskeletal organization, organ development and oncogenesis. LIM-domains mediate protein: protein interactions that are critical to cellular processes [39]. LIM domains are proposed to function in...
protein-protein recognition in a variety of contexts including gene transcription, development and in cytoskeletal interaction. The LIM domains of this protein bind to protein kinases, whereas the PDZ domain binds to actin filaments. The gene product is involved in the assembly of an actin filament-associated complex essential for transmission of ret/ptc2 mitogenic signaling. The biological function is likely to be that of an adapter, with the PDZ domain localizing the LIM-binding proteins to actin filaments of both skeletal muscle and non muscle tissues [9]. This protein has not been reported earlier in any fish species, by gel-based proteomics.

**Signal transduction**

Protein phosphorylation is a key regulatory mechanism for signal transduction in both prokaryote and eukaryotic cells. Most of our understanding regarding the signaling events in eukaryotes comes from tyrosine, serine/threonine kinases, and phosphatases. Histidine phosphorylation-dependent signaling in eukaryotes is less well characterized. The first vertebrate protein histidine phosphatase, PHPT1 was identified in 2002 [40] and it was found that PHPT1 is ubiquitously expressed in eukaryotes, from *C. elegans* to *Homo sapiens*. In the present study, one protein spot, CM160, has been identified as PHPT (Table 2); this protein has not been reported in other vertebrates by gel-based proteomics earlier (Supporting Information Table S1).

Two spots, C14 and C 18, were identified as peroxiredoxin and DJ-1, respectively (Table 1, Supporting Information Table S1 and S2). Peroxiredoxin controls cytokine induced peroxide levels and thereby mediating signal transduction in mammals. DJ-1, a chaperone protein having antioxidant properties, is involved in the cellular response to stress and has been described to play a role in apoptosis.

**Transcript analysis**

The partial gene sequence for the identified proteins (16/22) and for some additional proteins that include hsp47, hsp60, hsp70, hsc71, hsp90 and 18S RNA have been generated. Thus, this is the first study on this commercially important species with a proteogenomics approach. Proteogenomics aims mainly to use proteomics data and technologies for identifying novel genomic features, such as novel un-annotated genes, and for improving, correcting or confirming the structural and functional genome annotation. This approach is ideal to harness the wealth of information available at the proteome level and apply it to the available genomic information of organisms [3,41]. This would help to focus on the functional genome, rather than the whole genome and could possibly help to identify protein variants that could cause diseases, to identify protein biomarkers, to study genome variation and to identify QTLs associated with production traits. Multi-prolonged approaches such as transcriptomics and proteomics in addition to genomics should be included in future studies, as proteogenomic analyses provide a more accurate catalog of protein-coding genes [3,41].

In fish, flesh quality is dependent on environmental factors, mainly water and food quality for product safety and food composition for flesh nutritional quality. Nevertheless, amongst sensory quality, flesh texture is mainly determined by biological factors such as muscle organization, protein content, and composition. In fish, the best quality is firm and cohesive flesh with good water holding capacity. These traits are mainly determined by proteins' nature and properties, so proteomic tools appear especially of interest to study fish flesh quality. However, very few studies were undertaken to identify flesh quality biomarkers [42-45] and is a research gap needing attention of researchers.

The primary objective of the study was to establish a reference muscle proteome map for *Catla catla* and to identify the muscle proteins, which has been achieved to a great extent. The identified proteins in this study have been cataloged in the database http://www.cifriernet.in/fishprotf.html and would act as baseline information on proteogenomics of *Catla catla* and other aquacultured species. The information generated could also be useful for biotechnological interventions in fish health and disease management; besides adding to the existing knowledge base on comparative muscle proteomics.

**Acknowledgements**

This work was supported by the CIFRI core-Project No. FHE/ER/07/06/005 (to BPM) and ICAR National Fund for Basic, Strategic and Frontier Application Research in Agriculture (NFBFSFARA) project MAS-2001 (to BPM and SM). Sudeshna Banerjee and Tandrima Mitra are thankful to ICAR for the Research Fellowship. Soma Bhattacharjee was DBT Post Doctoral Fellow and GKP is a ICAR/NFBFSFARA Senior Research Fellow. The authors are thankful to the unknown reviewers for the valuable suggestions for improving the manuscript.

**Conflict of Interest Statement**

The authors declare that there are no conflicts of interest.

**References**

1. Hochstrasser DF, Sanchez JC, Appel RD (2002) Proteomics and its trends facing nature's complexity. Proteomics 2: 807-812.

2. Cravatt BF, Simon GM, Yates JR 3rd (2007) The biological impact of mass-spectrometry-based proteomics. Nature 450: 991-1000.

3. Renuse S, Chaerkady R, Pandey A (2011) Proteogenomics. Proteomics 11: 620-630.

4. Wolfe RR (2006) The underappreciated role of muscle in health and disease. Am J Clin Nutr 84: 475-482.

5. Ohlendiick K (2011) Proteomic Profiling of Fast-To-Slow Muscle Transitions during Aging. Front Physiol 2: 105.

6. Gordon AM, Homsher E, Regnier M (2000) Regulation of contraction in striated muscle. Physiol Rev 80: 953-924.

7. Abdul-Ghani MA, DeFronzo RA (2010) Pathogenesis of insulin resistance in skeletal muscle. J Biomed Biotech 2010: 479279.

8. Doran P, Donoghue P, O’Connell K, Gannon J, Ohlendiick K (2007) Proteomic profiling of pathological and aged skeletal muscle fibres by peptide mass fingerprinting (Review). Int J Mol Med 19: 547-564.

9. Hejland K, Yi Z, Hwang H, Bowen B, Leftor N, et al. (2008) Characterization of the human skeleton muscle proteinome by one-dimensional gel electrophoresis and HPLC-ESI-MS/MS. Mol Cell Proteomics 7: 257-267.

10. Piec I, Listrat A, Alliot J, Chambon C, Taylor RG, et al. (2005) Differential proteome analysis of aging in rat skeletal muscle. FASEB J 19: 1143-1145.

11. Almeida AM, Campos A, van Harten S, Cardoso LA, Coelho AV (2009) Establishment of a proteomic reference map for the gastrocnemius muscle in the rabbit (Oryctolagus cuniculus). Res Vet Sci 87: 196-199.

12. Doherty MK, McLean L, Hayter JR, Pratt JM, Robertson DHL, et al. (2004) The proteome of chicken skeletal muscle: Changes in Soluble protein expression during a growth in a layer strain. Proteomics 4: 2082-2089.

13. Bosworth CA 4th, Chou CW, Cole RB, Rees BB (2005) Protein expression patterns in zebrafish skeletal muscle: initial characterization and the effects of hypoxic exposure. Proteomics 5: 1362-1371.

14. McLean L, Young IS, Doherty MK, Robertson DH, Cossins AR, et al. (2007) Global cooling: cold acclimation and the expression of soluble proteins in carp skeletal muscle. Proteomics 7: 2687-2691.

15. Addis MF, Cappuccinnelli R, Tedde V, Pagnozzi D, Porcu MC, et al. (2010) Proteomic analysis of muscle tissue from gilthead sea bream (Sparus aurata, L.) farmed in offshore floating cages. Aquaculture 309: 249-252.

16. Gebriel M, Uleberg KE, Larsen S, HjelleBjersnads A, Sivertsvik M, et al. (2010) Cod (Gadusmorhua) muscle proteome cataloguing using 1D-PAGE: protein separation, nano-liquid chromatography peptide fractionation, and Linear Trap Quadrupole (LTQ) Mass Spectrometry. J Agric Food Chem 58: 12307-12312.
17. Gam L, Leow CY, Baie S (2006) Proteomic analysis of Snakehead fish (Channa striatus) muscle tissue. Malaysian J Biochem Mol Biol 14: 25-32.
18. Johnston IA (1999) Muscle development and growth: potential implications for flesh quality in fish. Aquaculture 177: 99-115.
19. Islam MS, Ahmed ASI, Azam MS, Alam MS (2005) Genetic analysis of three river populations of Catla catla (Hamilton) using randomly amplified polymorphic DNA markers. Asian J Anim Sci 18: 453-457.
20. Barik SK, Banerjee S, Bhattacharjee S, Das Gupta SK, Mohanty S, et al. (2013) Proteomic analysis of sarcoplasmic peptides of two related fish species for food authentication. Appl Biochem Biotechnol 171: 1011-1021.
21. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Chem 72: 248-254.
22. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.
23. O’Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250: 4007-4021.
24. Bringans S, Eriksen S, Kendrick T, Gopalakrishnakone P, Livk A, et al. (2008) Proteomic analysis of the venom of Heterometrus longimanus (Asian black scorpion). Proteomics 8: 1081-1096.
25. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.
27. http://www.cifri.ernet.in/fishpro.html
28. Sahoo PK, Mohanty BR, Kumari J, Barat A, Sarangi N (2009) Cloning, nucleotide sequence and phylogenetic analyses and tissue-specific expression of the transferrin gene in Asian buffalo, Bubalus bubalis. J Mol Biol 215: 403-410.
29. Bhattacharjee S, Mohanty S, Sharma AP, Mohanty BP (2011) Effect of storage temperature as a preanalytical variable on the lens crystallins protein quality for genomic and nucleotide sequence of the muscle isoforms of creatine kinase from the Antarctic TELEOS. Comp Biochem Physiol B 42: 221-226.
30. Winnard P, Cashon RE, Sidel BD, Vayda ME (2003) Isolation, characterization and nucleotide sequence of the muscle isoforms of creatine kinase from the Antarctic teleost Chaoeophalus aceratus. Comp Biochem Physiol B Biochem Mol Biol 140: 629-634.
31. Sun HW, Hui CF, Wu JL (1998) Cloning, characterization, and expression in Escherichia coli of three creatine kinase isoenzymes from carp (Ctenopharyngodon idellus). J Biol Chem 273: 33737-33740.
32. Kennedy MB (1995) Origin of PDZ (DHR, GLGF) domains. Trends Biochem Sci 20: 350.
33. Bach I (2000) The LIM domain: regulation by association. MechDev 9: 5-17.
34. Busam RD, Thorsell AG, Flores A, Hammarström M, Persson C, et al. (2006) First structure of a eukaryotic phosphohistidine phosphatase. J Biol Chem 281: 33850-33854.
35. Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. Brief Funct Genomic Proteomic 7: 50-62.
36. Picard B, Lefèvre F, Lebret B (2010) Meat and fish flesh quality improvement with proteomic applications. Animal Front 2: 18-25.
37. Kjaersgård IV, Jessen F (2003) Proteome analysis elucidating post-mortem changes in cod (Gadus morhua L) muscle proteins. J Agric Food Chem 51: 3985-3991.