



unesp

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



Caracterização do transcriptoma do músculo estriado esquelético do pacu

(Piaractus mesopotamicus)

EDSON ASSUNÇÃO MARECO

Tese apresentada ao Instituto de Biociências,
Câmpus de Botucatu, UNESP, para obtenção
do título de Doutor no Programa de Pós-
Graduação em Biologia Geral e Aplicada,
Área de concentração *Biologia Celular
Estrutural e Funcional*

Orientadora: Profa. Dra. Maeli Dal Pai

**BOTUCATU – SP
2015**

Instituto de Biociências - Seção Técnica de Pós-Graduação
Distrito de Rubião Júnior s/n CEP 18618-970 Cx Postal 510 Botucatu-SP Brasil
Tel (14) 3880-0780 posgraduacao@ibb.unesp.br



unesp

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de Botucatu



UNIVERSIDADE ESTADUAL PAULISTA
“Julio de Mesquita Filho”
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU

Caracterização do transcriptoma do músculo estriado esquelético do
pacu (*Piaractus mesopotamicus*)

EDSON ASSUNÇÃO MARECO

MAELI DAL PAI

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Biologia Geral e Aplicada, Área de concentração Biologia Celular Estrutural e Funcional

Orientadora: Profa. Dra. Maeli Dal Pai

**BOTUCATU – SP
2015**

Instituto de Biociências - Seção Técnica de Pós-Graduação
Distrito de Rubião Júnior s/n CEP 18618-970 Cx Postal 510 Botucatu-SP Brasil
Tel (14) 3880-0780 posgraduacao@ibb.unesp.br

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TEC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO -CRB 8/7500

Mareco, Edson Assunção.

Caracterização do transcriptoma do músculo estriado esquelético do pacu (*Piaractus mesopotamius*) / Edson Assunção Mareco. - Botucatu, 2015

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Maeli Dal Pai
Capes: 20204000

1. Proteínas. 2. Ubiquitina. 3. Pacu (Peixe). 4. Expressão gênica. 5. Sistema musculoesquelético. 6. Jejum.

Palavras-chave: Deubiquitinas ; E3 ubiquitinas;
Jejum-realimentação; Parálogos; Proteases
ubiquitina-especifica.

Dedico este trabalho...

A minha querida esposa Raquel, por estar ao meu lado, sempre me apoiando e me incentivando durante todos os momentos da vida.

AGRADECIMENTOS

A orientadora, Dra. Maeli Dal Pai, pela competência, paciência e por ter confiado em mim.

I would like to thank Prof. Dr. Johnston for giving me the opportunity of doing part of my PhD in his lab, which was for me a great experience, that I will carry with me during all my life.

I would like to thank Dr. Daniel Garcia de la serrana and Dr. Clara Coll Lladó, for all their help and support throughout the Ph.D. process. In particular, not only his help and advice about science, but also during beer's friday.

Also, I would like to thank Dr. Vera Vieira-Johnston, for all their support during my period in St Andrews. Always getting me with a good Brazilian's heart.

Ao Dr. Robson Francisco Carvalho, pelas colaborações e pelo companheirismo.

Aos meus amigos do Laboratório de Biologia do Músculo Esquelético (LBME) Aninha, Carlos, Fer Carani, Grasi, Ivan, Jason, Juarez, Leo, Paula, Paulinha, Raquel, Rodrigo, Rondi, Shola, Tata, Warlen pelo companheirismo e pelos momentos alegres de nosso convívio.

Aos todos os amigos do APTA – Presidente Prudente (Polo alta sorocabana) em especial ao Dr. Vander Bruno dos Santos, pelo companheirismo e pelo auxílio nos diversos períodos de nosso convívio.

Aos meus amigos de todos os laboratórios do Departamento de Morfologia pelo excelente convívio.

Aos técnicos de laboratório, José Eduardo e Ricardo e Vívian pelo auxílio profissional e amizade.

Aos funcionários do Departamento de Morfologia-IBB-UNESP-Botucatu por serem prestativos.

A Luciana, secretária do Departamento de Morfologia-IBB-UNESP-Botucatu, pela atenção e por todo auxílio.

Aos funcionários da secretaria de pós-graduação pelo profissionalismo e pela disposição em sempre querer ajudar.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão da bolsa de Doutorado (processo nº 2012/02489-4).

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de Doutorado sanduíche BEX 1450/13-1

Muito obrigado!

RESUMO

O pacu (*Piaractus mesopotamicus*) apresenta excelentes características zootécnicas como rápido crescimento, parâmetro que está diretamente relacionado com o aumento da musculatura estriada esquelética. Pesquisas envolvendo a caracterização de genes musculares têm sido realizadas utilizando-se tecnologias moleculares, em espécies de peixes modelo como o Zebrafish que possui o genoma sequenciado. No pacu, espécie que não possui o genoma sequenciado, esta falta de informação genética dificulta a realização de pesquisas relacionadas com a identificação das vias de sinalização que regulam o desenvolvimento, o crescimento e a manutenção do fenótipo muscular. Além disso, a falta de informação promove uma limitação para o desenvolvimento de programas de melhoramento para esta espécie. Os principais objetivos do presente estudo foram: 1) aumentar os recursos genéticos disponíveis para a espécie, 2) comparar os transcriptomas dos tipos de músculo, vermelho e branco explorando as vias de sinalização que controlam os tipos de fibras musculares 3) investigar a expressão de deubiquitinas pertencente a família do sistema de Proteases ubiquitinas-específicas (USP) no músculo branco em resposta ao jejum e realimentação. Com o sequenciamento, obteve-se aproximadamente 0.6Tb de leituras a partir de bibliotecas do músculo esquelético vermelho e branco. Aproximadamente, 665 milhões de leituras foram utilizadas para a realização da montagem, resultando em 504.065 contigs com um comprimento médio de 1,334bp e N50 = 2,772bp; 47% do transcriptoma foi anotado com êxito, onde pode-se identificar cerca de 15.000 genes considerados únicos e aproximadamente 8000 sequências com região codificante completas. Dentre os genes identificados destacam-se 319 genes de vias metabólicas e 380 microssatélites. Novecentos e cinquenta e seis e 604 genes foram diferencialmente expressos entre o músculo vermelho e branco, respectivamente. Identificamos 442 pares de genes parálogos, resultantes da duplicação genômica na base de origem dos peixes teleósteos. Foi possível identificar uma expressão diferencial dos genes parálogos, entre os tipos de fibras (301 genes no músculo vermelho e 245 genes no músculo branco). Quarenta e cinco membros da família USP foram identificados no transcriptoma. Os níveis de expressão foram quantificados por qPCR em um experimento de jejum (5 dias) seguido de 24 horas de realimentação. Os níveis de expressão dos genes das USP, demonstraram um aumento transitório no músculo rápido durante o jejum, semelhante a expressão das principais E3 ubiquitina-ligases. Através deste trabalho foi possível gerar o transcriptoma do músculo vermelho e branco do pacu (*Piaractus mesopotamicus*) com uma cobertura 53 vezes, aumentando significativamente as informações genéticas disponíveis para esta espécie que é importante para aquicultura brasileira. Observamos diferenças significativas na expressão gênica em componentes relacionados ao metabolismo, via Akt/PI3K/mTOR e miogênese. Com este projeto, identificamos uma grande quantidade de genes parálogos, onde foi possível realizar uma análise detalhada de expressão destes genes. Nós também realizamos uma descrição detalhada e abrangente da expressão genes membros da família de USP entre os tipos de fibras musculares e durante alterações de estado nutricional. Os resultados obtidos com este projeto possibilitarão estabelecer uma base sólida para estudos futuros do músculo esquelético em espécies brasileiras de importância econômica na aquicultura, em diferentes condições experimentais.

ABSTRACT

Pacu (*Piaractus mesopotamicus*) has excellent husbandry features like rapid growth. This parameter is directly related to the increase of skeletal muscle. Researches involving the characterization of muscle genes have been performed using molecular technologies, in species of fish such as zebrafish model that has sequenced the genome. In pacu, a species that does not have the sequenced genome, this lack of genetic information makes it difficult to conduct research related to the identification of signaling pathways that regulate development, growth and maintenance of muscle phenotype. Moreover, the lack of information promotes a limitation for the development of breeding programs for this species. The main objectives of this study were: 1) to increase the available genetic resources for the species, 2) compare the transcriptomes of red and white muscle types, exploring the signaling pathways that control the types of muscle fibers 3) to investigate the expression of deubiquitins belonging to the family of ubiquitin-specific proteases system (USP) in muscle in response to fasting and feedback. With the sequencing, it was obtained approximately 0.6Tb readings from red skeletal muscle and white libraries. Approximately 665 million readings were used to perform the assembly, resulting in 504 065 contigs with an average length of N50 = 1,334bp and 2,772bp; 47% of the transcriptome was noted successfully, which can be identified about 15,000 genes considered as unique and approximately 8000 sequences coding complete region. Among the identified genes, we could observe 319 metabolic pathways genes and 380 microsatellite. Nine hundred fifty six and 604 genes were differentially expressed between red and white muscle, respectively. We identified 442 pairs of paralogs genes resulting from genomic duplication at the base source of teleost fish. It was possible to identify differential expression of paralogous genes include the types of fibers (genes 301 and 245 in red muscle genes in muscle). Forty-five members of the family have been identified in USP transcriptome. Expression levels were quantified by qPCR in a fasting experiment (5 days) followed by a 24-hour refeeding. The expression levels of the genes of USP demonstrated a transient increase in muscle fast during fasting, similar expression of key E3 ubiquitin ligases. Through this work, it was possible to generate the transcriptome of red and white muscle of pacu (*Piaractus mesopotamicus*) with a 53-fold coverage, significantly increasing the genetic information available for this species which is important for Brazilian aquaculture. We observed significant differences in gene expression in components related to metabolism, Akt / PI3K / mTOR and myogenesis. With this project, we identified a lot of paralogs genes, which made it possible to conduct a detailed analysis of expression of these genes. We also conducted a detailed and comprehensive description of the members of the USP gene expression of family among the types of muscle fibers and during nutritional status changes. The results obtained from this project will enable lay a solid foundation for future studies on skeletal muscle in Brazilian species with economic importance in aquaculture, in different experimental conditions.

LISTA DE ABREVIATURAS

MPA : Ministério da Pesca e da Agricultura

SHh: Sonic Hedgehog

MRFs: Fatores Regulatórios Miogênicos (*do inglês: Myogenic regulatory factors*)

MyoD: Myoblast determination protein

Myf5: Myogenic factor 5

BMP: Bone morphogenetic protein

Prdm1a: positive regulatory domain I-binding factor 1

Sox6: Fator Transcricional SOX-6

KEGG: Kyoto Encyclopedia of Genes and Genomes

KAAS: KEGG Automatic Annotation Server

SBH: Método unidirecional de melhor resultado (*do inglês: Single-direcional best Hit*)

UCH: C-terminal ubiquitina (*do inglês: Ubiquitin C-terminal*)

USP: Proteases ubiquitinas-específicas (*do inglês: Ubiquitin-specific proteases*)

OTU: Protease de tumor ovariano (*do inglês: Ovarian tumour proteases*)

WGD: Duplicação completa do genoma (*do inglês: Whole Genome Duplication*)

LISTA DE FIGURAS

Figura 1: Juvenil de pacu (<i>Piaractus mesopotamicus</i>).....	12
Figura 2: Separação anatômica da musculatura esquelética do pacu.	13
Figura 3: Modelo esquemático, demonstrando a miogênese embrionária em vertebrados.	14
Figura 4: Desenvolvimento embrionário musculo vermelho e branco.	15
Figura 5: Modelo esquemático, demonstrando a miogênese pós-natal em peixes.....	19
Figura 6: Local de amostragem da musculatura esquelética do pacu.	23

SUMÁRIO

RESUMO	VII
ABSTRACT	VIII
1 INTRODUÇÃO	12
1.1 PACU	12
1.2 TECIDO MUSCULAR	13
1.3 JEJUM E REALIMENTAÇÃO	16
1.4 DUPLICAÇÃO GÊNOMICA EM PEIXES TELEÓSTEOS	19
2 HIPÓTESE E OBJETIVOS	22
3 MATERIAL E MÉTODOS	23
3.1 PEIXES E COLETA DAS AMOSTRAS MUSCULARES	23
3.2 PCR QUANTITATIVO EM TEMPO REAL	24
3.3 EXTRAÇÃO DO RNA TOTAL E TRANSCRIÇÃO DO CDNA	24
3.4 CONSTRUÇÃO DA BIBLIOTECA E SEQUENCIAMENTO	25
3.5 ANÁLISE DE BIOINFORMÁTICA	26
3.5.1 ANÁLISE DOS DADOS BRUTOS	26
3.5.2 MONTAGEM DOS TRANSCRITOS E ANÁLISE DOS GENES DIFERENCIALMENTE EXPRESSOS	26
3.5.3 ANOTAÇÃO FUNCIONAL	27
3.5.4 IDENTIFICAÇÃO DE SEQUÊNCIAS COM REGIÃO CODIFICANTE COMPLETA	27
3.5.5 Identificação de sequências parálogas	28
4 CAPÍTULO 1 - Characterization of the transcriptome of fast and slow muscle myotomal fibres in the Pacu (<i>Piaractus mesopotamicus</i>) *	29
5 REFERÊNCIAS	62

1 INTRODUÇÃO

1.1 PACU

O pacu (*Piaractus mesopotamicus*) (Figura 1) é um caracídeo neotropical de grande interesse econômico na aquicultura brasileira, sendo a segunda espécie nativa mais cultivada (MPA, 2011). Sua carne é saborosa e apresenta alto teor de gordura o que determina grande aceitação no mercado brasileiro e em outras regiões do mundo (MILSTEIN et al., 2000). Características zootécnicas como rusticidade, fecundidade elevada, rápido crescimento, precocidade e ótima adaptação à alimentação artificial, contribuem para sua expansão na cadeia produtiva. Dentre as peixes cultivados no Brasil, o pacu, é uma das espécies que apresenta maior potencial econômico (JOMORI et al., 2003), devido especialmente ao seu rápido crescimento corporal, com o peso variando de 1,0 a 1,5 kg no primeiro ano de criação, posteriormente, podendo atingir até 20 kg (URBINATI; GONÇALVEZ, 2006). A taxa de crescimento ou o desempenho nos peixes é um parâmetro importante para o sucesso da aquicultura e está diretamente relacionada com o crescimento da musculatura estriada esquelética, a qual corresponde à maior parte da massa corporal, sendo uma importante fonte de proteínas.



Figura 1: Juvenil de pacu (*Piaractus mesopotamicus*). Fonte: O Autor.

1.2 TECIDO MUSCULAR

O tecido muscular estriado esquelético é o tecido mais abundante do corpo dos vertebrados (BIRESSI; MOLINARO; COSSU, 2007) e, nos peixes, pode contribuir com aproximadamente 60% do peso total do animal (BONE, 1978). Este tecido possui uma plasticidade notável, adaptando-se a uma variedade de estímulos externos e internos. Embora a plasticidade muscular seja um fenômeno comum nos vertebrados (SCHIAFFINO; REGGIANI, 1996), nos peixes teleósteos, o tecido muscular é considerado ainda mais especializado para atender uma maior demanda de produção de força necessária tanto para a natação sustentada como para a realização de movimentos rápidos. A solução evolutiva adotada pelas espécies de peixes para suprir estas necessidades conflitantes de movimentação, foi a de dividir o sistema locomotor em diferentes compartimentos, aos quais contêm tipos específicos de fibras musculares (BONE, 1978; SÄNGER; STOIBER, 2001).

Diferentemente da musculatura dos mamíferos, caracterizada por uma mistura de tipos de fibras, na musculatura dos peixes os tipos de fibras musculares estão organizados em três camadas ou compartimentos distintos: vermelho/lento (superficial), intermediário e branco/rápido (profundo) (GREEK-WALKER; PULL, 1975) (Figura 2).

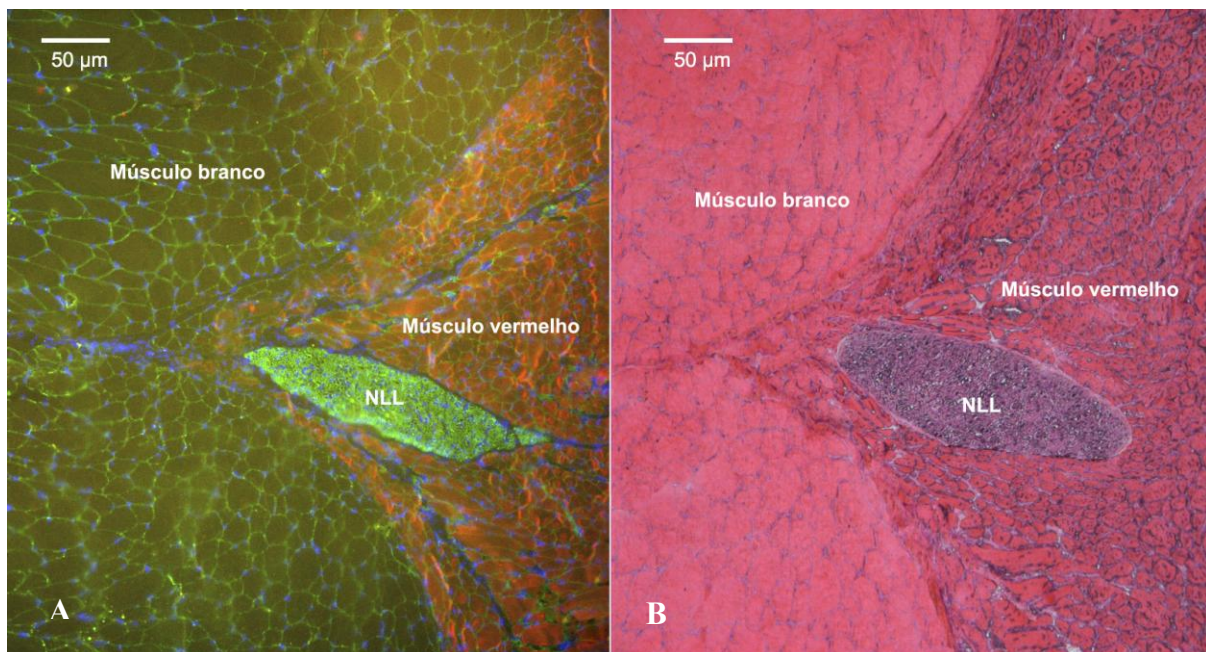
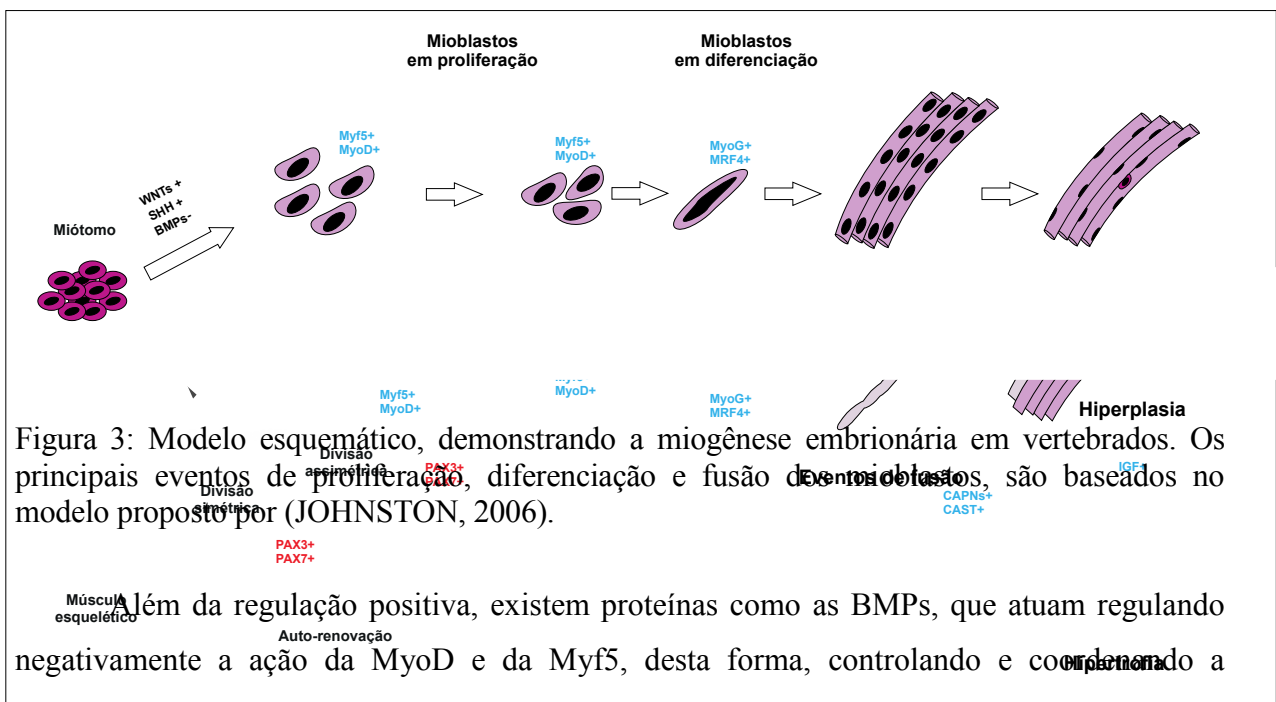


Figura 2: Separação anatômica da musculatura esquelética do pacu. (A) Secção transversal do músculo de pacu, incubada com um cocktail de anticorpo primário contra MHCI (BA-F8) (vermelho) e contra Lâminina, seguido por incubação com anticorpos secundários conjugados com fluorescência adequados. (B) Secção serial transversal do músculo de pacu corando com

hematoxilina e eosina (H&E). (NLL) Nervo da linha lateral. Barra representa 50 um. Fonte: O Autor.

A base para a diversificação das fibras musculares em rápidas e lentas, em peixes, tem sido atribuída diferentes linhagens celulares. A formação das fibras lentas, ocorre logo no início do desenvolvimento embrionário (DU, 2004). Onde, de modo geral, todos os tecidos que circundam os somitos do embrião, contribuem com sinais que induzem ou inibem vias de sinalização envolvidas com o processo de diferenciação das células somíticas em células precursoras miogênicas. Sabe-se que, além da notocorda, também participam deste processo o tubo neural, a superfície do ectoderma e a placa lateral do embrião. Uma grande variedade de glicoproteínas são secretadas por estas estruturas, que desencadeiam uma cascata de eventos, de comprometimento e diferenciação celular.

Neste sentido, as principais moléculas de sinalização extracelular que atuam regulando positivamente este processo são as SHh, Noggin e Wnt. Estas proteínas, exercem influência direta na ativação e expressão de Fatores Reguladores Miogênicos (MRFs), como o Myf5 e MyoD, conhecidos como MRFs primários. A MyoD e Myf5, fazem parte da família MyoD (Myf5, MyoD, Miogenina e MRF4) que atuam como fatores transcricionais de genes músculo específicos e são considerados os principais fatores que regulam o comprometimento das células somíticas em mioblastos.



MOLINARO; COSSU, 2007). Apenas uma fração de células progenitoras miogênicas se diferenciam logo no início da formação do miótomo primário. As células que se desenvolvem ao redor da notocorda (células adaxiais), consideradas como células precursoras lentas, são as primeiras a serem comprometidas em mioblastos e começam a se diferenciar em resposta a ação das SHh, que são secretadas pela notocorda (DU et al., 1997); desencadeando uma cascata de eventos promovendo a especificação das células adaxiais em fibras musculares lentas. Seguindo o desenvolvimento embrionário, as células adaxiais se alongam, e em seguida, migram radialmente através do somito, formando uma monocamada de fibras musculares lentas (BRYSON-RICHARDSON; CURRIE, 2008; DEVOTO et al., 1996). É neste momento que ocorre a formação do compartimento/Músculo vermelho (Figura 4).

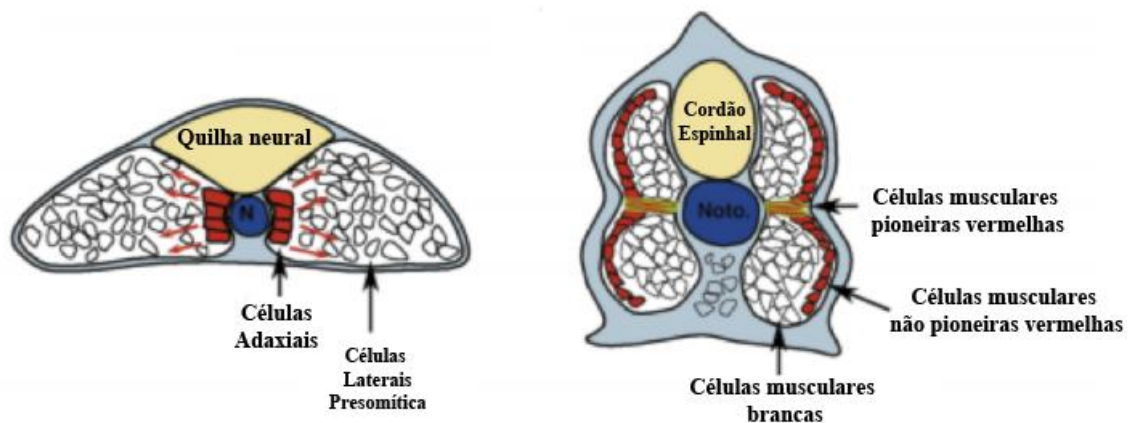


Figura 4: Desenvolvimento embrionário músculo vermelho e branco. Adaptado de (DU et al., 1997). N = Notocorda

Von Hofsten e colaboradores (2008) demonstraram que, além do SHh, o gene *Prdm1* é também um gene importante para a formação e manutenção do músculo vermelho, atuando principalmente como um repressor transcricional de genes responsáveis pela formação das fibras musculares brancas.

As fibras musculares que compõe o compartimento vermelho, são fibras exclusivamente de contração lenta e com metabolismo oxidativo. Elas apresentam um pequeno diâmetro (25-45 μm), alta concentração de mioglobina, muitas mitocôndrias e lipídios e um excelente suprimento sanguíneo. O recrutamento das fibras vermelhas ocorre durante a realização de movimentos lentos e de sustentação, como a migração (JOHNSTON, 1977; SÄNGER; STOIBER, 2001).

A maior contribuição para a massa muscular ocorre através da formação do compartimento branco (~70%) (SÄNGER; STOIBER, 2001), localizado em uma região mais profunda do músculo, sendo formado por fibras musculares brancas maiores (~200 µm de diâmetro). Estas fibras apresentam uma velocidade de contração rápida, metabolismo glicolítico, baixa concentração de mioglobina, poucas mitocôndrias e lipídios. Sua utilização ocorre principalmente durante os movimentos bruscos de natação, tais como na captura de alimento e fuga de predadores (DRIEDZIC; HOCHACHKA, 1976; JOHNSTON, 1980). A formação do compartimento branco ocorre através da diferenciação das células laterais pré-somáticas encontradas nos somitos do embrião (FIGURA 4). Esta diferenciação é desencadeada após a migração das células adaxiais e a ativação de diversos fatores transcricionais músculo branco-específico. Dentre eles, podemos destacar o fator de transcrição Sox6 o qual é importante no processo de diferenciação das fibras musculares brancas, por inibir a expressão de genes músculo vermelho-específico, durante a fase fetal (JACKSON et al., 2015).

Entre os compartimentos vermelho e branco encontra-se o compartimento intermediário (musculatura intermediária), com fibras que apresentam propriedades morfofisiológicas intermediárias entre as das fibras musculares brancas e vermelhas, como contração rápida e metabolismo oxidativo/glicolítico (JOHNSTON, 1977; SÄNGER; STOIBER, 2001).

1.3 JEJUM E REALIMENTAÇÃO

Os peixes ocupam nichos ecológicos extremamente diversificados e possuem hábitos alimentares definitivamente diferenciados. Durante o seu ciclo de vida normal, os peixes podem sofrer longos períodos de privação alimentar (FERREIRA; SILVA, 2012; GAHR; WEBER; REXROAD, 2006; GUDERLEY et al., 2003). Estes períodos normalmente ocorrem de modo natural, devido as flutuações de disponibilidade de alimento no ambiente aquático (SOUZA et al., 2003). Períodos como sazonalidade (diminuição da disponibilidade de alimento natural) e reprodução (diminuição do apetite), contribuem dramaticamente na diminuição das taxas de crescimento. Desta forma, cada espécie possui mecanismos específicos para lidar com períodos de escassez de alimento. Durante o jejum, os peixes mobilizam suas reservas energéticas para compensar a falta de alimento. Neste sentido, um dos principais tecidos afetado durante o jejum é o músculo esquelético. Isso ocorre

principalmente porque boa parte das proteínas armazenadas na musculatura é mobilizada para a manutenção da homeostase do animal (GAHR; WEBER; REXROAD, 2006).

De modo geral, em condições fisiológicas normais, o músculo esquelético possui um mecanismo de balanço natural de síntese (anabolismo) e degradação de proteínas (catabolismo). Inicialmente, em resposta ao jejum, ocorre uma mobilização de lipídios hepáticos e glicogênio muscular e hepático. No entanto, se a falta de alimento for extrema, ocorre o início da degradação das proteínas miofibrilares (GUDERLEY et al., 2003).

A degradação das proteínas musculares é um processo altamente complexo e dependente da ativação de diversos sistemas e enzimas proteolíticas (SALEM et al., 2007; SEILIEZ et al., 2008; VARSHAVSKY, 1998). Dentre esses sistemas destacam-se o complexo proteolítico multicatalítico/proteossomal (BOWER; TAYLOR; JOHNSTON, 2009; MARTIN et al., 2002; MATSAKAS; PATEL, 2009), o sistema lisossomal que inclui as catapsinas (CLEVELAND et al., 2009), caspases e as calpaínas (MACQUEEN et al., 2010; SALMERÓN et al., 2013), as aminopeptidases citoplasmáticas e proteases alcalinas, bem como, as elastases e collagenases (enzimas hidrolíticas), que atuam na degradação da matrix extracelular do músculo esquelético (DELBARRE-LADRAT et al., 2006).

A via ubiquitina-proteassoma (complexo proteolítico multicatalítico) é considerado o principal sistema responsável pela degradação de proteínas musculares. As ubiquitinas são os principais componentes desta via em estudos que visem a compreensão do processo de catabolismo em músculo de peixes. Diversos estudos (BOWER; JOHNSTON, 2010; FUENTES et al., 2012; TACCHI et al., 2010, 2012) têm demonstrado que a ativação desta via, é provavelmente mediada pela regulação da transcrição de genes E3 ubiquitina-ligases, Murf1 e Mafbx, os quais também participam de processos atroficos. As E3 ubiquitina-ligases são consideradas enzimas-chave no processo de degradação proteica ligando-se as proteínas alvo e promovendo a marcação (ubiquitinização), gerando uma especificidade ao sistema de degradação proteossomal (TACCHI et al., 2010). Após o período de jejum pode ocorrer a diminuição do diâmetro das fibras musculares e conseqüentemente a diminuição da massa muscular. Além disso, alguns autores sugerem que o jejum pode promover a diminuição da capacidade de diferenciação das fibras musculares imaturas (GAHR; WEBER; REXROAD, 2006) em função da diminuição da expressão de genes relacionados com crescimento muscular (GARCÍA DE LA SERRANA et al., 2013).

No entanto, para que ocorra um controle maior das proteínas a serem degradadas pelo sistema proteossomal, é necessário que ocorra a ativação de enzimas que controlem o

processo de ubiquitinação (marcação) das proteínas, realizado principalmente pelas E3 ubiquitinas (KOMANDER; CLAGUE; URBÉ, 2009). Este controle ocorre através da ação das deubiquitinas (DUBs), uma super-classe de famílias de enzimas (UCH, USP, OTU, Josephin e JAMM/MPN+) que exercem função contrária as ubiquitinas, promovendo a recuperação das proteínas marcadas pelas ubiquitinas (KOMANDER; CLAGUE; URBÉ, 2009; SACCO et al., 2010). A atividade das DUBs pode ocorrer de diferentes formas: inibição da formação dos complexos de ubiquitinas; remoção das cadeias de ubiquitinas ou ainda através da modificação das cadeias de ubiquitinas que se formam nas proteínas (SACCO et al., 2010).

A homeostase do músculo esquelético ainda depende do equilíbrio entre a proliferação e diferenciação de células específicas do músculo conhecidas como Células Satélites (CS). As CS são células miogênicas mononucleadas quiescentes localizado entre o sarcolema e membrana basal da fibra muscular (MAURO, 1961). Este grupo de células se originam a partir de mioblastos comprometidos que não se diferenciaram em fibras maduras, durante o desenvolvimento embrionário, permanecendo assim em estado quiescente através da ação contínua do Pax7, inibindo a sua diferenciação em fibras musculares (KUANG et al., 2007).

Durante o crescimento muscular pós-natal, ocorre a ativação e a proliferação das CS, processo que retoma os principais eventos da miogênese embrionária nas quais as CS sofrem a ação de diversos estímulos, dentre eles, a ação dos fatores regulatórios miogênicos (MyoD e Miogenina) que promovem a proliferação e diferenciação. Com a retomada da alimentação, ocorre a retomada do anabolismo e a recuperação do crescimento muscular. Nos peixes, o crescimento muscular pós-natal ocorre por hiperplasia (aumento do número de fibras musculares) e/ou hipertrofia (aumento do diâmetro das fibras musculares, através do acréscimo de mionúcleos) (KOUMANS; AKSTER, 1995; ROWLERSON; VEGGETTI, 2001), processos que envolvem a ativação, proliferação e diferenciação das células satélites (FAUCONNEAU; PABOEUF, 2001) (Figura 4). A manutenção da população de CS é igualmente importante para manutenção do fenótipo, uma vez que as CS contribuem para a formação do músculo adulto. A possível perda da capacidade da manutenção (auto renovação) da população de CS durante o crescimento muscular, tem por consequência a diminuição da capacidade de crescimento e regeneração do músculo (MOTOHASHI; ASAKURA, 2014).

A auto renovação das CS pode ocorrer através da divisão assimétrica no qual, para cada célula satélite “mãe”, origina-se um mioblasto e a outra CS. Processo, onde ocorre a

manutenção parcial da população das células satélites. No entanto, a divisão também pode ocorrer de forma simétrica, onde em algumas situações distintas pode ocorrer a formação de dois mioblastos, ou ainda a formação de duas células idênticas, ou seja, duas CS. (Kuang et al., 2007)

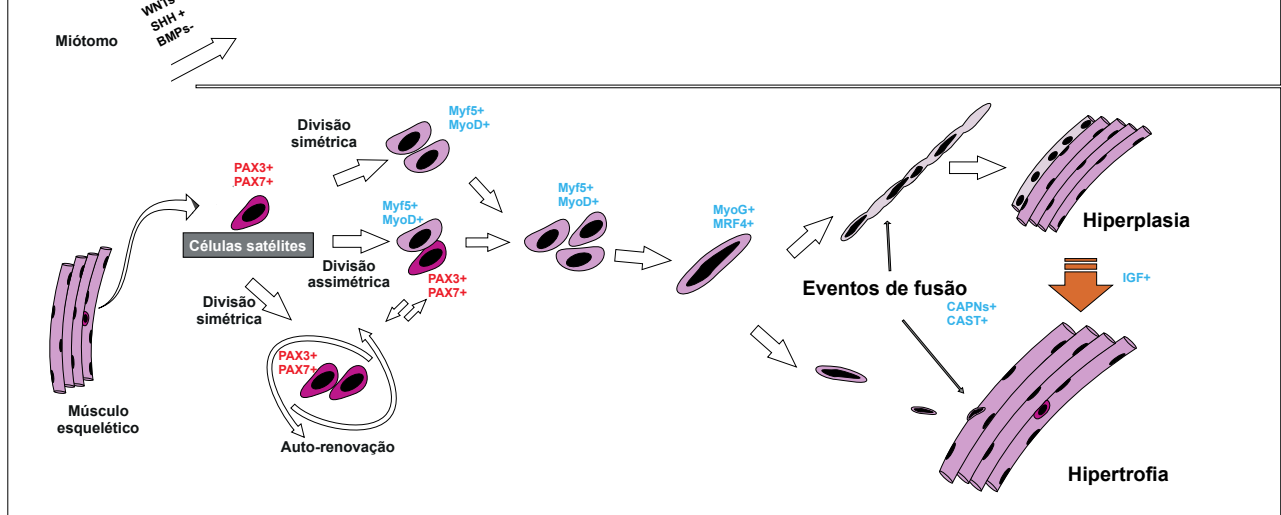


Figura 5: Modelo esquemático, demonstrando a miogênese pós-natal em peixes. Os principais eventos de proliferação, diferenciação e fusão dos mioblastos, são baseados no modelo proposto por (JOHNSTON, 2006). O modelo do processo de manutenção da população das células satélites (pax7+) foi baseado nos principais achados descritos por (KUANG et al., 2007).

Sabendo que o processo de jejum e realimentação pode ocorrer de forma natural, diversas estratégias têm sido utilizadas com o objetivo de maximizar o crescimento animal. A adoção de estratégias de alimentação com restrição alimentar seguida de realimentação tem sido uma forma de explorar a capacidade natural de recuperação metabólica e o crescimento dos peixes, possibilitando economia no fornecimento de alimento (SOUZA et al., 2003).

1.4 DUPLICAÇÃO GÊNOMICA EM PEIXES TELEÓSTEOS

Eventos de duplicação genômica (poliploidia) e mutações genéticas são considerados os principais fatores que contribuem para a diversificação das espécies (JAILLON et al., 2004; MEYER; SCHARTL, 1999). O aumento das informações genéticas provenientes de projetos de sequenciamento de genoma, vem fornecendo informações em larga escala dos eventos de duplicação completas e parciais de genomas dos mais diversos organismos (GARCIA DE LA SERRANA; MARECO; JOHNSTON, 2014). Neste sentido, análises filogenéticas dos genes duplicados, análises comparativas de suas localizações

cromossômicas e inferências evolutivas de cariótipo ancestral, contribuem para confirmar a hipótese inicialmente proposta por Ohno em 1970 (OHNO, 1970) de que todos os vertebrados passaram por dois estágios (2R) de duplicação genômica completa (WGD). Análises semelhantes, também trouxeram *insights* com relação a evolução de peixes teleósteos, demonstrando a existência de um terceiro evento de WGD específico desse grupo (GLASAUER; NEUHAUSS, 2014; SATO; NISHIDA, 2010). De modo geral, os eventos de WGD contribuíram para o aumento da complexidade das famílias gênicas, através da formação de novas cópias de um mesmo gene no genoma (GARCIA DE LA SERRANA; MARECO; JOHNSTON, 2014). As cópias parálogas sofrem a ação da pressão seletiva/evolutiva de forma diferente umas das outras. Mutações e rearranjos moleculares, podem promover mudanças significativas em uma das cópias destes genes e podendo levar a perda de sua função (perda da cópia) (GARCIA DE LA SERRANA; MARECO; JOHNSTON, 2014). No entanto, caso as alterações não sejam suficientes para a perda da funcionalidade, pode ocorrer a chamada sub-funcionalização do gene, sendo que neste caso a atividade dos parálogos podem ser semelhante a atividade do gene ancestral. Em geral a expressão dos parálogos pode ser condicionada ao tecido (DE SANTIS; JERRY, 2011), ao tempo (MACQUEEN; ROBB; JOHNSTON, 2007), ou ainda as condições intrínsecas e extrínsecas em que o indivíduo se encontra, evitando assim a redundância da função.

Diversos estudos tem demonstrado a presença e a expressão diferencial de genes parálogos em várias vias de regulação gênica, incluindo vias relacionadas com controle do metabolismo muscular (ERBAY et al., 2003; MACQUEEN; ROBB; JOHNSTON, 2007; TAN; JUN DU, 2002). MacQueen e colaboradores (2007) identificaram a expressão diferencial de parálogos de MyoD, expressos em diferentes períodos e em diferentes regiões do músculo esquelético de salmão durante a miogênese. A similaridade de função apresentada entre as cópias de MyoD nestes organismos demonstram um comportamento de sub-funcionalização condicionada ao tempo.

A compreensão da função da maioria dos parálogos ainda permanece desconhecida uma vez que manutenção e deleção de alguns genes parálogos ao longo da evolução pode ser linhagem-específica, fazendo com que a comparação entre as espécies seja ainda mais complexa (GARCIA DE LA SERRANA; MARECO; JOHNSTON, 2014). Como exemplo dessas diferenças, podemos citar o gene da miostatina (MSTN), inicialmente descrito em mamíferos por McPherron e colaboradores (1997) como um importante regulador negativo do crescimento e desenvolvimento muscular através da inibição da proliferação dos mioblastos.

No entanto, em peixes teleósteos é possível encontrar duas cópias de MSTN, descritas como *mstn 1/a* e *mstn 2/b*. Nos peixes a cópia de parálogo que mais se assemelha a função descrita em mamíferos é a isoforma conhecida como *mstn 2/b*, o qual é expresso exclusivamente no músculo esquelético e no cérebro. Já a isoforma *mstn 1/a* apresenta uma expressão mais variada em diversos tecidos, demonstrando um padrão de expressão diferente e com função ainda não caracterizada na maioria das espécies (MACCATROZZO et al., 2001).

2 HIPÓTESE E OBJETIVOS

A hipótese do nosso trabalho é que a análise do transcriptoma possibilitará avaliar genes e vias de sinalização diferencialmente expressas nos músculos vermelho e branco no pacu.

Os objetivos do estudo foram: 1) aumentar as informações de sequências genéticas disponíveis para a espécie, 2) comparar os transcriptomas dos músculos vermelho e branco para explorar e caracterizar genes envolvidos com a manutenção dos fenótipos musculares lento e rápido e 3) investigar a expressão de genes pertencentes à família de deubiquitinas, conhecida como Ubiquitina específicas Protease (USP), no músculo branco em resposta ao jejum e realimentação.

3 MATERIAL E MÉTODOS

3.1 PEIXES E COLETA DAS AMOSTRAS MUSCULARES

A caracterização do transcriptoma de músculo vermelho e branco, foi realizada utilizando amostras musculares de 5 pacu adultos (*Piaractus mesopotamicus*) ($1,50 \pm 0,61$ kg). Os peixes foram mantidos em tanques de 1000 litros no centro de piscicultura da Universidade do Oeste Paulista (Unoeste) Presidente Prudente, São Paulo, Brasil, sob fotoperíodo natural (12:12) e temperatura ($28^{\circ} \text{C} \pm 1^{\circ} \text{C}$, intervalo). O músculo esquelético branco foi removido a partir da região epaxial medial dorsal, enquanto que o músculo esquelético vermelho foi dissecada a partir da linha lateral (Figura 6). As amostras musculares foram imersas em solução RNAlater® para a estabilização do RNA (Ambion, Life Technologies, EUA) e mantidas a -20°C até o processamento.

Para a etapa do projeto que envolveu o jejum e a realimentação, foram utilizados juvenis de pacu com 15g ($n = 80$), mantidas em tanques de 500 litros e alimentados com uma dieta comercial até o início do experimento. Os peixes foram então mantidos em período de jejum durante 4 dias, seguido por um período de realimentação até a saciedade durante 24 h. Amostras da região da musculatura branca, foram removidas a partir da região epaxial medial dorsal, no início do jejum (-4d), e diariamente durante o período de privação de alimentar (-3d, -2d, -1d, 0d; $n = 8$) e 6, 12 e 24 horas ($n = 8$) após a re-alimentação.

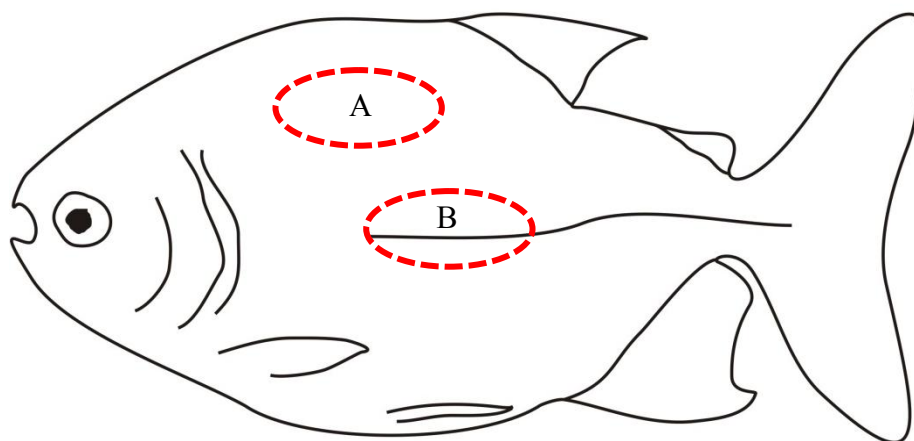


Figura 6: Local de amostragem da musculatura esquelética do pacu. (A) Músculo branco (B) Músculo vermelho. Fonte: O autor.

3.2 PCR QUANTITATIVO EM TEMPO REAL

A análise dos dados obtidos no qPCR seguiram o modelo mínimo de informação para publicação de experimentos quantitativos por PCR em Tempo Real (BUSTIN et al., 2009).

O desenho dos primers foi realizado utilizando o programa primer3, disponível em: <http://bioinfo.ut.ee/primer3-0.4.0/>. Sempre que possível, os primers foram desenhados para amplificar regiões de junções éxon-éxon. As junções éxon-éxon foram previstas e estimadas, alinhando a sequência pacu contra as informações genômicas do Zebrafish (*Danio rerio*) disponíveis no Ensembl (CUNNINGHAM et al., 2014) utilizando serviço online de alinhamento, Spidey (WHEELAN; CHURCH; OSTELL, 2001).

A reação de PCR quantitativa (qPCR) foi realizada utilizando o equipamento Mx3005P qPCR (Agilent, La Jolla, CA, EUA). Cada reação foi realizada com 7,5 qPCR µl de SensiFast (Bioline) Master Mix, 6 µl de cDNA (diluição de 80 vezes e diluição de 40 vezes para os genes de IGF) e 0,75 µl de cada primer a 500 nm até um volume final de 15 ul. As reações foram realizadas em placas de 96 poços (Agilent) com o seguinte protocolo: a ativação inicial de 95 °C durante 2 minutos seguido de 40 ciclos de 5s a 95 °C, 20s a 65 °C.

No final de cada reação foi adicionado um ciclo de 95 °C, referente a curva de dissociação, com o objetivo de confirmar a amplificação de um único produto. Para a análise e normalização das reações, foram avaliados 6 genes de referência: RPL13, RPL19, PPIAA, EF1A, GAPDH e HPRT1. Segundo o programa BestKeeper (PFAFFL et al., 2004), o gene HPRT1 foi o que demonstrou uma maior estabilidade durante as fases do experimento, por esta razão, foi o gene escolhido para normalização dos dados de expressão gênica. O cálculo da expressão relativa foi calculada utilizando o método Pfaffl (PFAFFL, 2001).

3.3 EXTRAÇÃO DO RNA TOTAL E TRANSCRIÇÃO DO CDNA

A extração do RNA dos animais do experimento de jejum e a realimentação foi realizada utilizando o reagente TRIreagent (Bioline, Londres, UK), seguindo as instruções do fabricante. A integridade do RNA total foi confirmada por electroforese em gel agarose com brometo de etídio. A concentração, e as razões 260/280 e 260/230 proporções foram estimadas usando um NanoDrop 1000 espectrofotômetro (Thermo Fischer Scientific, Waltman, MA). Todas as amostras de RNA utilizadas apresentaram razão 260/280 nm e 260/230 acima de 1,9 e 2,2, respectivamente. A transcrição reversa dos mRNAs obtidos do músculo esquelético será realizada utilizando o kit Quantitect (QIAGEN, Manchester, Reino

Unido) seguindo as orientações do fabricante. Para a reação foram utilizados aproximadamente 6 μ L de Master Mix para a transcrição reversa, ao qual foi adicionado o equivalente a 1 μ g de RNA, onde finalmente o volume completado para 20 μ L com água livre de nucleases. A mistura foi incubada nas seguintes condições: 42°C por 30 minutos seguido de 95°C por 3 minutos para a inativação da transcriptase reversa.

3.4 CONSTRUÇÃO DA BIBLIOTECA E SEQUENCIAMENTO

A preparação das bibliotecas e o sequenciamento de nova geração foram realizadas mediante uma cooperação entre a Universidade de St. Andrews e centro de genômica aplicada do Hospital *Sick Children (SickKids)*, Toronto, Canadá.

O RNA total foi extraído dos músculos vermelho e branco utilizando o reagente TRIzol, Life Technologies, conforme as instruções do fabricante (Life Technologies, Carlsbad, CA). Após a extração, amostras de cada animal foram purificadas em tampão, para atingirem de 0,1 a 4 μ g de RNA total. As bibliotecas foram então purificadas e amplificadas (200pb \pm 30 pb) por meio de PCR com primers específicos para os adaptadores (Illumina). Após o término desta etapa, avaliou-se a qualidade das bibliotecas, utilizando o equipamento Agilent 2100 Bioanalyser (Agilent Technologies) com o auxílio de um chip DNA 1000. As amostras de boa qualidade apresentaram fragmentos próximos ao tamanho de 260 pb.

O construção individual das bibliotecas de mRNA de 5 pacu adultos, dos músculos vermelho e branco, foram realizadas, utilizando o kit TruSeq RNA Sample Prep®. O preparo foi conduzido a partir das informações disponíveis no protocolo TruSeq RNA Sample Prep® Guide (Illumina – San Diego, EUA).

O processo de sequenciamento do mRNA foi realizado utilizando o aparelho o HiSeq-2000 (Illumina – San Diego, EUA), de acordo com as instruções do fabricante, utilizando o protocolo de pair-end reads. Com o kit TruSeq PE Cluster Kit v3-cBot-HS (Illumina – San Diego, EUA), as amostras foram ligadas a oligos complementares posicionados na superfície dos *lanes* (canaletas da lâmina) e posteriormente amplificadas, resultando em *clusters* (agrupamentos) de fragmentos iguais. A sequência dos agrupamentos foi determinada com uso do kit TruSeq SBS kit v3-HS (Illumina – San Diego, EUA) (200 cycles), conforme recomendações do fabricante. Em cada *lane* foi depositada seis amostras a partir das quais foi obtido em média 76 milhões de *reads* (leituras) (sequência de dados não tratados provenientes das máquinas de sequenciamento) por biblioteca.

O sequenciamento de RNA foi realizado através da plataforma Illumina utilizando a abordagem *sequencing-by-synthesis*, onde cada nucleotídeo é marcado com uma cor. Neste procedimento, cada biblioteca recebe a incorporação de adaptadores específicos nas duas extremidades de cada fragmento. Em seguida, o adaptador da extremidade livre do fragmento se liga ao seu oligonucleotídeo complementar formando uma estrutura em ponte, criando assim os *clusters*. Na etapa de desnaturação, as pontes de hidrogênio da molécula de DNA são desfeitas mediante elevação de temperatura, e após uma série de ciclos, foram obtidos clusters de moléculas idênticas ligadas à superfície de sequenciamento. Com a incorporação de nucleotídeos durante cada ciclo e a excitação a laser, é gerado um sinal o qual é captado e interpretado como um dos quatro possíveis nucleotídeos. O processo de incorporação dos nucleotídeos marcados, excitação e leitura é repetido para cada nucleotídeo. A leitura é então feita de forma sequencial permitindo a montagem da sequência completa de cada fragmento.

3.5 ANÁLISE DE BIOINFORMÁTICA

3.5.1 ANÁLISE DOS DADOS BRUTOS

Os dados brutos gerados pelo sequenciamento, foram processadas pelo departamento de Informática do Centro de Genômica Aplicada hospital *Sick Children (SickKids)*. Os dados iniciais, passaram por processo de identificação e em seguida de remoção dos adaptadores e reads de baixa qualidade. A identificação de reads com baixa qualidade e fragmentos de sequências superexpressas (possíveis adaptadores), foi realizada utilizando o programa *fastqc* (PATEL; JAIN, 2012). Leituras que apresentavam qualidade inferior a 28, foram descartadas, utilizando o programa *FASTQ Quality Filter* e para remoção dos possíveis adaptadores foi utilizado o *FASTA/Q Clipper*, (ambos os programas fazem parte do pacote de programas *fastx-toolkit*, disponível em: http://hannonlab.cshl.edu/fastx_toolkit), restando cerca de 86% dos reads iniciais.

3.5.2 MONTAGEM DOS TRANSCRITOS E ANÁLISE DOS GENES DIFERENCIALMENTE EXPRESSOS

A montagem dos transcritos foi realizada, utilizando o procedimento *de novo*, no qual os transcritos são montados sem um genôma ou transcriptoma de referência. O procedimento

de montagem foi realizado, utilizando o programa *Trinity* (HAAS; PAPANICOLAOU; YASSOUR, 2013). O programa *Trinity*, combina três módulos de software independentes: *Inchworm*, *Chrysalis*, e *Butterfly*. De modo geral, o programa particiona as sequências em fragmentos menores e constrói diversos gráficos individuais (gráficos Bruijn), cada um representando a complexidade de um determinado gene ou local, e em seguida, processa cada gráfico de forma independente. Desta forma, o programa é capaz de diferenciar possíveis isoformas de um mesmo gene e possíveis genes parálogos. Sendo cada parte do programa, responsável por determinar a melhor reconstrução dos transcritos. Ao final da execução, o programa gera, um arquivo do tipo fasta, contendo todos os transcritos identificados, chamados de contigs.

Após a montagem dos contigs, utilizou-se o algoritmo RSEM (LI; DEWEY, 2011), para identificar a abundância do transcritos, através da estimativa do número de leituras mapeada por contig. Desta forma, pode-se obter o valor de contagem de reads por contig em cada biblioteca. Os resultados das leituras mapeadas entre músculo vermelho e branco, foram utilizados para identificar os transcritos diferencialmente expressos, através do uso do algoritmo DEseq, disponível entre os pacotes Bioconductor/R (ANDERS et al., 2013).

3.5.3 ANOTAÇÃO FUNCIONAL

Os contigs foram anotados utilizando o software Blast2GO (GÖTZ et al., 2008). Todas as sequências foram confrontadas contra o banco de dados não redundante (NR) do NCBI, utilizando o algoritmo BLASTx com um valor de corte de 10^{-3} , seguido por anotação funcional, utilizando os parâmetros padrão de software.

Os contigs também foram mapeados contra vias moleculares e metabólica conhecidas de vertebrado, utilizando a função de anotação automática do banco de dados KEGG. A Anotação das vias foi realizada utilizando o método de SBH da ferramenta KAAS contra *Homo sapiens*, *Pan troglodites*, *Mus musculus*, *Rattus norvegicus*, *Sus scrofa*, *Gallus gallus*, *Meleagris gallopavo*, *Danio rerio* e *Xenopus laevis*.

3.5.4 IDENTIFICAÇÃO DE SEQUÊNCIAS COM REGIÃO CODIFICANTE COMPLETA

Para a identificação das sequências com região completa, utilizou-se as informações do proteoma completo do *Danio rerio*, como base para um alinhamento utilizando o

algoritmo tBLASTn. O alinhamento permitiu explorar e avaliar o percentual de cobertura da região codificante de cada sequência. Sequências que apresentaram uma cobertura maior que 90%, foram consideradas como sequência de codificação completa, ou seja, sequências completas.

3.5.5 Identificação de sequências parálogas

Os Contigs anotados foram alinhados contra o proteoma do Zebrafish, utilizando o algoritmo BLASTx (valor de corte de e^{-40}) presente no software BioEdit (HALL, 1999). Contigs que apresentaram similaridade com mais de uma sequência de Zebrafish, foram considerados como sequências candidatas parálogas. Em seguida, para confirmação das sequências parálogas foi realizado o alinhamento das sequências candidatas contra o proteoma de outras espécies de peixes (*Gasterosteus aculeatus*, *Takifugu rubripes*, *Oryzias latipes*, *Tetraodon nigroviridis*) e contra o proteoma humano (*Homo sapiens*), que posteriormente foi utilizado como grupo externo para a análise filogenética. As árvores filogenéticas foram construídas através de uma análise de máxima verossimilhança, utilizando o servidor on-line *PhyML* combinada com a opção G-Blocks (algoritmo que exclui regiões com pouco confiança de alinhamento) (DEREEPER et al., 2008). Para cada caso, o melhor modelo evolutivo foi estimado utilizando o programa MEGA5 (TAMURA et al., 2011).

4 CAPITULO 1 - Characterization of the transcriptome of fast and slow muscle myotomal fibres in the Pacu (*Piaractus mesopotamicus*) *

* Artigo publicado na Revista BMC Genomics.

Mareco E a, Garcia de la Serrana D, Johnston I a, Dal-Pai-Silva M: Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). BMC Genomics 2015, 16:1–13. DOI 10.1186/s12864-015-1423-6

Characterization of the transcriptome of fast and slow muscle myotomal fibres in the Pacu (*Piaractus mesopotamicus*)

Edson A Mareco^{1,2}, Daniel Garcia de la serrana², Ian A Johnston², Maeli Dal-Pai-Silva¹

¹Institute of Biosciences of Botucatu, São Paulo State University - UNESP, Botucatu, São Paulo, Brazil, 18618-970.

²School of Biology, Scottish Oceans Institute, University of St Andrews, Scotland, KY16 8LB

ABSTRACT

Background: The Pacu (*Piaractus mesopotamicus*) is a member of the Characiform family native to the Prata Basin (South America) and a target for the aquaculture industry. A limitation for the development of a selective breeding program for this species is a lack of available genetic information. The primary objectives of the present study were 1) to increase the genetic resources available for the species, 2) to exploit the anatomical separation of myotomal fibres types to compare the transcriptomes of slow and fast muscle phenotypes and 3) to systematically investigate the expression of Ubiquitin Specific Protease (USP) family members in fast and slow muscle in response to fasting and refeeding.

Results: We generated 0.6Tb of pair-end reads from slow and fast skeletal muscle libraries. Over 665 million reads were assembled into 504,065 contigs with an average length of 1,334bp and N50=2,772bp. We successfully annotated nearly 47% of the transcriptome and identified around 15,000 unique genes and over 8000 complete coding sequences. 319 KEGG metabolic pathways were also annotated and 380 putative microsatellites were identified. 956 and 604 genes were differentially expressed between slow and fast skeletal muscle, respectively. 442 paralogues pairs arising from the teleost-specific whole genome duplication were identified, with the majority showing different expression patterns between fibres types (301 in slow and 245 in fast skeletal muscle). 45 members of the USP family were identified in the transcriptome. Transcript levels were quantified by qPCR in a separate fasting and refeeding experiment. USP genes in fast muscle showed a similar transient increase in expression with fasting as the better characterized E3 ubiquitin ligases.

Conclusion: We have generated a 53-fold coverage transcriptome for fast and slow myotomal muscle in the pacu (*Piaractus mesopotamicus*) significantly increasing the genetic resources

available for this important aquaculture species. We describe significant differences in gene expression between muscle fibre types for fundamental components of the general metabolism, the Pi3k/Akt/mTor network and myogenesis, including detailed analysis of paralogue expression. We also provide a comprehensive description of USP family member expression between muscle fibre types and with changing nutritional status.

KEYWORDS

Paralogues, Ubiquitin-specific proteases (USP), aquaculture genomics

BACKGROUND

Fish myotomes are composed of anatomically segregated muscle fibre types each with distinct contractile and metabolic phenotypes [1]. Based on their contractile speed skeletal muscle fibres are classified as slow, intermediate or fast [2, 3]. Fast twitch (white) muscle fibres comprise the bulk of the myotome and are recruited for energetic movements associated with prey capture and escape behaviour [4]. Fast fibres have elevated densities of myofibrils, reduced myoglobin content and higher a capillary density than slow fibres and utilise phosphagen breakdown and anaerobic glycolysis to power contraction [5]. Sustained swimming activity is supported by superficial layers of slow (red) and intermediate (pink) twitch fibres which are recruited at slow and high cruising speeds respectively [3]. Slow fibres have extensive lipid and glycogen stores, abundant mitochondria and high capillary densities reflecting their reliance on aerobic metabolism [5]. Intermediate (pink) muscle fibre types are found between the slow and fast fibre layers, but express distinct isoforms of myosin heavy chains [6] and have intermediate contraction speeds and metabolic characteristics [3, 7].

Muscle growth reflects the balance between protein synthesis and degradation. These two processes are influenced by numerous biotic and abiotic factors including food availability, growth factors, age, sex, diet composition, swimming activity, oxygen saturation, light and temperature [8, 9]. The Insulin-like growth factor (Igf) network, composed of Igfs, binding proteins (Igfbp) and receptors (Igf1r and Igf2r), plays a pivotal role in integrating internal and external inputs to regulate muscle mass [8]. Igf1 regulates several signalling pathways including the Pi3k/Akt/mTor network that controls protein synthesis [8, 9]. Typically, fibre production continues until 45% of the maximum body length of fish, and subsequent growth is entirely by fibre elongation and hypertrophy [10-12]. Myogenesis

involves the activation, proliferation and fusion of a resident myoblast population involving hundreds of structural and regulatory genes [10, 11].

Muscle protein degradation occurs through three major pathways [13, 14] namely: membrane-bound lysosomal enzymes, calpain proteases [14, 15], and the Ubiquitin Proteasome (Ub) Pathway (UPP) [16]. UPP represents the most important system for degradation of unnecessary or damaged proteins. Targeted proteins are linked to ubiquitin, which acts as a recognition tag for the proteasome. Ubiquitin tagging of proteins requires the coordinated activity of three classes of enzymes known as E1, E2 and E3. It is the E3 enzymes, also known as E3-ubiquitin ligases, which conjugate ubiquitin to the target protein [17]. Ubiquitin mediated degradation can be reversed through action of deubiquitinating enzymes (DUB), a large group of proteases that cleave ubiquitin-protein conjugates removing the UPP signal, and play an essential role in the regulation of protein degradation [18]. DUBs are subdivided into four families: ubiquitin C-terminal hydrolases (UCHs), ovarian tumour proteases (OTUs), JAB1/MPN/MOV34 metalloenzymes and ubiquitin-specific proteases (USPs) [18]. USP is the largest family of DUB and regulates a wide variety of cellular processes. Although the essential role of USPs in protein degradation is well established, less is known about the function and regulation of specific family members: for example Usp7 has been associated with p53 and Akt turnover, Usp8 with receptor tyrosine endocytosis, Usp33 with the Von Hippel–Lindau disease (VHL) pathway and Usp19 is thought to have a role in muscle development [19]. There have been several studies of the expression of protein degradation related genes in fish, mostly in response to varying nutrition and focused on E3-ubiquitin ligases (particularly Fbox32 and Murf1) [20, 21]. In contrast, nothing is known about the transcriptional regulation of USP family members.

The pacu (*Piaractus mesopotamicus*) is a member of the characiform family (SuperOrder Ostariophysi) native to the Prata Basin (South America) and is a target for finfish aquaculture in Brazil. One of the main limitations for the development of a selective breeding program for this species is a lack of genetic information as well as limited knowledge about its physiology. Next Generation Sequencing (NGS) technologies have dramatically increased the amount of sequence data for teleosts and helped to overcome the lack of annotated genomes (so far only 12 fish annotated genomes are publicly available) [22].

The primary objectives of the present study were 1) to increase the genetic resources available for pacu 2) to exploit the anatomical separation of fibre types to characterise the expression signatures of fast and slow muscles and 3) to systematically investigate the

expression of USPs in fast and slow muscle in response to fasting and refeeding. Teleost fish underwent a teleost-specific genome duplication (TSGD) event around 450 million years ago (Mya) which was followed by diploidisation and gene loss [23, 24]. It is thought that around 15 to 20% of TSGD paralogues have been retained in the diploid genome of extant species [23–26]. Several studies have demonstrated that paralogues from the TSGD or the salmonid-specific whole-genome duplication (WGD), which occurred 88Mya [27], can display different patterns of expression during myogenesis and muscle growth [27–29]. A secondary objective was therefore to search for evidence of differential expression between teleost specific paralogues within and between muscle fibre types.

METHODS

FISH

For the generation of de novo transcriptome fast and slow skeletal muscles were dissected from 5 adult pacu (*Piaractus mesopotamicus*) ($1.50 \pm 0.61\text{kg}$; mean \pm SD body mass). Fish were maintained in 1000 litres fibreglass tanks at the Aquaculture Centre of the University of West Paulista (Unoeste) Presidente Prudente, São Paulo, Brazil, under natural photoperiod (12L: 12D) and temperature ($28^\circ\text{C} \pm 1^\circ\text{C}$, range). Fast skeletal muscle was dissected from the dorsal epaxial region at 0.5 fork length (FL) (FL, tip of snout to fork in the tail) whereas slow skeletal muscle was dissected from the lateral line and any remains of fast skeletal muscle carefully removed under a dissection microscope to obtain pure slow muscle. Tissues were preserve in RNAlater (Ambion/Applied Biosystems, Oslo, Norway) and frozen at -20°C until further analysis.

For the fasting-re-feeding experiment, 15g pacu ($n=80$) were maintained in duplicate fibreglass 500 litre tanks as described above and fed with a commercial diet until the start the experiment. Fish were fasted for 4 days followed by a period of satiation feeding for 24h. Fast skeletal muscle was sampled before fasting (-4d), daily during food deprivation (-3d, -2d, -1d, 0d; $n=8$) and 6, 12 and 24h ($n=8$) after re-feeding. Extracted fast skeletal muscle was preserved in RNAlater at -20°C until further analysis. All fish were sacrificed according to the Ethical Principles In Animal Research adopted by Brazilian College of Animal Experimentation (COBEA) and was approved by the Ethics Committee on Use of Animals/ Bioscience Institute/Unesp (CEUA=506)

SAMPLES SEQUENCING AND DE NOVO ASSEMBLY

Total RNA from adult pacu fast and slow skeletal muscle was used to prepare 10 individual Illumina libraries. Libraries preparation and sequencing was performed at the Centre for Applied Genomics of the Hospital for Sick Children (SickKids), Toronto, Canada. The resulting libraries were paired end sequenced using in an Illumina HiSeq2000.

Raw paired end reads generated were processed by the Department of Informatics of the Centre for Applied Genomics of SickKids hospital. After removing low quality reads, 86% of the paired end reads were de novo assembled using Trinity software [51]. RSEM application was used to identify transcript abundance by estimating the number of reads mapped per contig. The DEseq algorithm from the Bioconductor/R packages was used to identify differentially expressed transcripts [52].

FUNCTIONAL ANNOTATION

Contigs were annotated using Blast2GO software [53]. Sequences were blasted against the NCBI non-redundant (nr) database using BLASTx with an e-value cut-off of 10⁻³ followed by functional annotation using software default parameters [53]. Contigs were mapped against the known vertebrate metabolic and molecular pathways using the online KEGG Automatic Annotation Server (KAAS) [54]. KAAS annotation was performed using the single-directional Best Hit (SBH) method against *Homo sapiens*, *Pan troglodites*, *Mus musculus*, *Rattus norvegicus*, *Sus scrofa*, *Gallus gallus*, *Meleagris gallopavo*, *Danio rerio* and *Xenopus laevis*.

IDENTIFICATION OF COMPLETE CODING SEQUENCE

Annotated contigs were blasted against the Zebrafish complete proteome [22] using tBLASTn algorithm in BioEdit software [55]. BLAST alignments were explored to evaluate the percentage of coding sequence cover by the contig compared with its zebrafish orthologue. Sequences with more than 90% of coverage were considered as complete coding sequence (CDS). The CDS amino acids sequence was predicted using the Virtual Ribosome server [56].

MICROSATELLITE IDENTIFICATION

Sequences successfully annotated covering >90% of the CDS were investigated for SSR using msatcommander-1.0.2-alpha [57].

RNA EXTRACTION AND CDNA SYNTHESIS

Total RNA was extracted using 1ml TRIsure (Bioline, London, UK) following the manufacturer's recommendations. Integrity was confirmed by ethidium bromide gel electrophoresis of 1µg of total RNA. Concentration, 260/280 and 260/230 ratios were estimated using a NanoDrop 1000 spectrophotometer (Thermo Fischer Scientific, Waltman, MA). All RNA samples used had a 260/280 nm and 260/230 ratios above 1.9 and 2.2 respectively. 1µg of total RNA was reverse transcribed into cDNA for 30 min at 42°C using a Quantitect (QIAGEN, Manchester, UK) reverse transcription kit following manufacture's recommendations including a genomic DNA wipe-out step. To ensure that no genomic DNA was present in the samples a RT- control without the reverse transcriptase enzyme was performed simultaneously.

QUANTITATIVE REAL-TIME PCR

The following procedures were compliant with the minimal information requirements for publication of quantitative PCR guidelines [58]. Primers were designed to have a Tm of 60oC using Net primer online server (Premier BioSoft). Where possible primers were designed to cross exon-exon junctions. Exon-exon junctions were predicted by aligning the pacu contig against their zebrafish orthologue complete gene sequence retrieved from Ensembl [22] using Spidey online server [59]. Primers pairs, amplicon size, melting temperature and efficiency are listed in Supplementary 12.

Supplementary file 12 – qPCR primer sequences

Gene	Sequence		amplicon size	efficiency of reaction
	Forward (5'-3')	Reverse (5'-3')		
Mafbx	tctttggtgctcccctgtg	taaaaccgaggacggctgg	231	1.90
Fbox25	cgtaggtgaagtagccgcca	ggagaagtccagcctgttga	166	1.90
Mufla	ccattgatgccttggtgc	ccttcctgctctgctctttg	198	1.85
Muflb	ccccctgaaagacaaagaca	gctcagccttctgacttc	179	1.89
Siah1	cacaaatccatcaccacgc	gccatcatacttttctgc	150	1.89
Syvn1	cgacctacagagcgagaacc	cagccttcttaaattgcctca	185	1.88
Trip12	cgcagaagcagctcgtgataa	tgctgggggagactcactac	222	1.87

Huwe1	gagacctaactctccgccag	caccactgctctccacaatc	228	1.87
Ufd2	agtcgtcacagggagatgct	ttgggaacctcaggaatcag	206	1.89
Usp2a	agcctgcgagactactgt	actctgagcgtctctgtg	214	1.89
Usp2b	gcaccgacctcaacaacaac	acgaaacggggagcatactt	151	1.94
Usp4	gccttcttctgctggatggact	tcaggacagactagcgtgga	194	1.92
Usp5a	ccttcagcacttccagcaga	tgcgatttccaactccgta	199	1.90
Usp5b	cccagaccactttgagattcc	atttcgacttggcacactt	222	1.94
Usp8	agagggcacacacttcaggt	gcgagagcagaggagagaga	111	1.89
Usp9	cacaaacctcggaccaaaat	gcacagtcaaaaccctaacca	179	1.88
Usp11	caactaacctcggcaaacct	agaaactgggacgcaaaatg	249	1.90
Usp12a	tcctaatgacagtttcaa	gggtgagcaggttctctttgc	239	1.80
Usp12b	atactgatgacagtccgaaag	aggtcagcaggtctctttg	241	1.92
Usp14	ctctgaggaggtattcgggt	cagcaaacctgtgggaaggc	168	1.93
Usp16	ccgtctcactctccacctg	ctgtgctccacgattccata	181	1.89
Usp19	gccccacaagaaaccagtaa	gcacctcaaagcagaagagc	215	1.89
Usp21	tgtgctggatggagatgaga	cttctgtcggaggcaaacctc	210	1.81
Usp22	aactcctgcctcgtgtgtg	ttcttcccgttgcgtcttt	206	1.87
Usp24	aatcagaatcgtcggaggag	agtggcaaaagaggtgagga	173	1.90
Usp28	acagtttgtgtcaggtcccc	ctctcccacgtcgattct	155	1.91
Usp30	ctctctgtccattctctgtgt	gtcgtcctctggctctcaac	168	1.88
Usp36	gatgaagactgggtgccttgga	ccttggggcggagagagttt	118	1.90
Usp46	agcggttccgttgttctca	ctccacagcagcgcacaca	217	1.87
Igf1ra	gcttctcaggtcctgctcc	attcaccgtctaccgcatcg	248	1.87
Igf1rb	aggctggatgtgctctctt	gggtggctatcaaaacggt	223	1.88
Igf1	atttcagcaagccaacaggt	cgcacaatacatctcaagtcg	116	1.87
Igf2a	cagcaagcagaagacaacga	gcagcattctccactatcc	217	1.90
Igf2b	gttttgtcggtttggagga	gaaggtgctggaacaggaat	199	1.91
Igf3	ggctcactgaaggacaca	cgctctccacacaaaactc	183	1.90
Efla	attccactgagcccccttac	ttacgtcaaccttccatcc	190	1.94
Polr2a	atcaaccaacgaactccaac	cctctgtgctaaggcgaaac	205	1.92
Gapdh	acacacgacgacaagaccaa	gtcctctcgtgaaaactg	267	1.87
Hprt1	tgggctaaaggaaggatgag	tacaaagcgtgaggtggcta	202	1.89
ppiaa	attgtggttcgtgaagtcgc	ccgtgggcagagtgattat	170	1.87
sdha	acctgatgctgaatgctgtg	agtgtgcttctccagtgc	170	1.90
Rpl13	atcaacaggaagtagccc	aggatgagtttggagcggta	122	1.89
Rpl19	gcaaacgtggaaggatggt	cttgactccctgtaacgcc	215	1.85

Quantitative PCR (qPCR) was performed using a MX3005P qPCR machine (Agilent, La Jolla, CA, USA). Each qPCR reaction contained 7.5 µl of SensiFast (Bioline) Master Mix, 6 µl cDNA (80-fold dilution and 40-fold dilution for igf genes) and 0.75 µl of each primer at

500 nM to a final volume of 15 µl. Duplicate reactions were performed in 96-well plates (Agilent) with the following protocol: initial activation 95°C for 2min followed by 40 cycles of 5s at 95°C, 20s at 65°C. The qPCR was followed by a dissociation-melting curve from 60 to 95°C to confirm that a single product was amplified. Control reactions included no-template and RT- were simultaneously amplified.

Ribosomal protein 13 and 19 (rpl13, rpl19), Peptidylprolyl isomerase Aa (ppiaa), elongation factor 1 alpha, Glyceraldehyde-3-phosphate dehydrogenase (gapdh) and hypoxanthine phosphoribosyltransferase 1 (hprt1) were tested as reference genes. BestKeeper [60] analysis showed that hprt1 was the most stable reference gene and was used for data normalisation. Relative expression was calculated using the Pfaffl method [61].

TELEOST SPECIFIC PARALOGUES IDENTIFICATION

Contigs were blasted (BLASTx) against the Zebrafish proteome using BioEdit software with an e-value threshold of e^{-40} [55].

To confirm that contigs found were truly paralogues amino acids sequences from potential pacu paralogues were blasted (BLASTp) against the zebrafish (*Danio rerio*), stickleback (*Gasterosteus aculeatus*), takifugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), green pufferfish (*Tetraodon nigroviridis*), chicken (*Gallus gallus*), frog (*Xenopus laevis*) and human (*Homo sapiens*) proteomes using Ensembl BLAST server [22]. Best hits amino acids sequences from each proteome were retrieved. Sequences were aligned using the MAFFT online server [62]. Phylogenetic trees were constructed using a Maximum Likelihood analysis using PhyML online server combined with the G-Blocks option to cure unreliable aligned sections [63]. For each case the best evolutionary model was estimated using MEGA5 [64].

STATISTICAL ANALYSIS

Global DGE statistic analysis was performed using DEseq package from R-Bioconductor [52, 65]. For testing specific hypothesis involving differential mapping of specific pathways gene expression significance was tested using t-test, or Mann-Whitney U test when parametric were not fulfilled, followed by a Benjamin-Hochberg correction (False Discovery Rate, FDR). Differences between time-points in qPCR expression during fasting re-feeding were tested using Kruskal-Wallis test. Differences were considered significant when FDR, for differential mapping, or p-value, for qPCR expression, were <0.05 . Gene

expression data was clustered using an unsupervised hierarchical clustering algorithm using PermutMatrix [66].

RESULTS AND DISCUSSION

DE NOVO ASSEMBLY

Individual barcoded libraries from fast and slow skeletal muscle tissues were generated from 5 adult pacu ($1.5 \pm 0.61\text{kg}$; mean \pm SD) and sequenced using HiSeq 2000 platform yielding a total of 374,952,267 and 390,706,230 pair-end reads respectively ($\sim 0.6\text{Tb}$) (raw reads were deposited in the European Nucleotide Archive accession number PRJEB6656). 665,042,497 trimmed paired-end reads (86% of the total) were successfully assembled into 504,065 contigs with an average length of 1,334bp and N50=2,772bp (Table 1). The present study significantly improves upon previous teleost transcriptomes [30]. A total of 237,637 contigs (47%) were successfully annotated, representing over 15,000 unique genes with an average coverage of 53 and over 8,000 genes with $>90\%$ of the coding sequence (CDS) represented (Table 1; annotation results can be found in supplementary file 1). The number of genes found represent between 56-65% of all protein-coding genes currently identified in *Danio rerio* (Cypriniformes) and *Astyanax mexicanus* (Characiformes) the only two other Ostariophysi genomes available [22]. Pacu contigs were also annotated and classified into 319 different vertebrate signalling and metabolic pathways (supplementary file 2). In some cases, such as for the Pi3k/Akt/mTor network we found that over 90% of the genes were represented.

Table 1: Pacu *de novo* transcriptome metrics

Parameters	Slow skeletal muscle	Fast skeletal muscle
Reads	704,550,985	765,658,497
Reads assembled	521,568,496	665,042,497
Contig	504,065	
Contig mean length (bp)	1,334	
N50 (bp)	2,772	
Contig Annotated	232,637	
Unique genes	15,000	
Paralogues	406	
Differentially expressed genes	3,473	2,458
Annotated Pathway	314	

N50: The value was computed by sorting all contigs from largest to smallest and by determining the minimum set of contigs whose sizes total 50% of the entire transcriptome.

The current transcriptome dramatically increases the genetic resources available for the future development of genetic-based breeding programmes in pacu. Simple sequence repeats (SSRs) or microsatellites are widely used for parentage identification and stock management in family selection programs [31]. Screening for SSRs was focused on those contigs for which over 90% of the CDS was present to assure that repeated sequences were correctly identified in the UTR regions. A total of 380 SSRs were identified in the transcriptome, and the great majority of them (54%) were repeats of a dinucleotide motif (Table 2; supplementary file 3) increasing the number of potential microsatellites for this species [26]. We also screened the annotated transcriptome to identify TSGD-paralogues. It has been estimated that around 15 to 20% of teleost-specific paralogues have been retained in the diploid genomes of extant species [23-28] with a slightly higher proportion retained in Ostariophysi genomes [26]. We identified a total of 442 paralogues pairs in the pacu when compare with zebrafish (884 genes; Figure 1A, supplementary file 4). Although this is a significant improvement when compared with previous studies [22] the number of paralogues identified was still lower than what we could expect from a 16,000 gene transcriptome.

Table 2: Microsatellites identification

Motif	Count	%
AAC	1	0,3
AC	113	36,7
ACAG	1	0,3
ACCT	1	0,3
AG	107	34,7
AGCG	1	0,3
AGGC	1	0,3
AT	28	9,1
ATC	1	0,3
C	52	16,9
CG	2	0,6

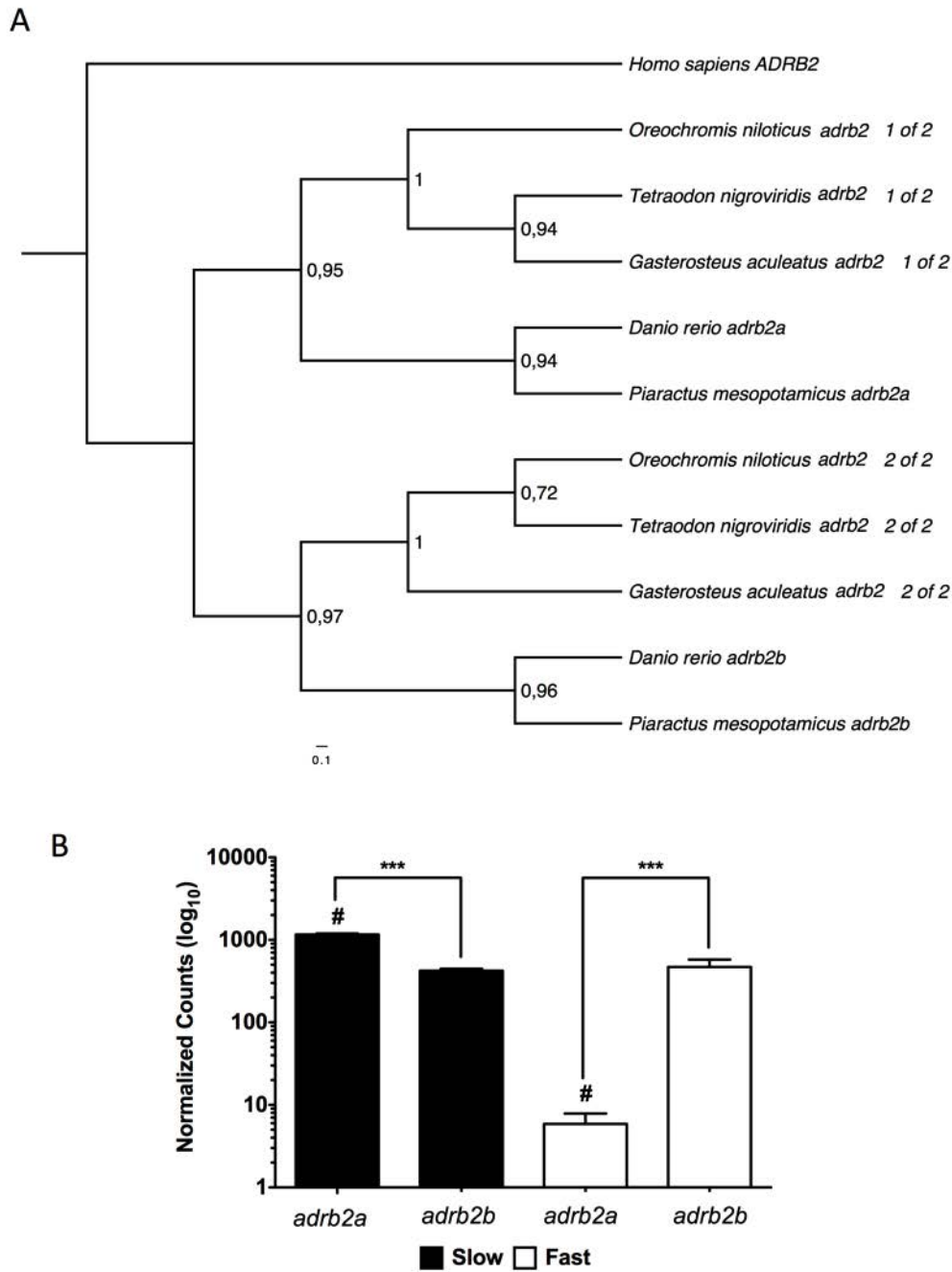


Figure 1: Example of a teleost-specific paralogue phylogenetic and digital expression analysis in pacu fast and slow skeletal muscle.

(A) Phylogenetic analysis of the *adr2* teleost-specific paralogs. Human *ADRB2* was used as an out-group. (B) *adr2a* and *adr2b* digital gene expression. Digital gene expression is represented using a logarithmic scale for slow (full bars) and fast (empty bars) skeletal muscle. Significant differences for paralogs within (***) and between (#; FDR<0.05) fibre types are indicated. Values represent mean \pm SE (n=5). *Adrb2*: beta-2 adrenergic receptor.

DIGITAL GENE EXPRESSION ANALYSIS

Individual fast and slow muscle libraries were mapped to the complete transcriptome. Mapped reads were normalized by contig length, library size and only those with an average of more than 10 mapped were used to estimate digital gene expression (DGE). Whereas there are several transcriptomic studies in teleost [29, 32, 33] to our knowledge this is the first one to compare global patterns of DGE in muscle different fibre types. Global DGE analysis revealed 956 and 604 genes differentially expressed between slow and fast skeletal muscle respectively (supplementary file 5).

Gene ontology (GO) enrichment analysis of the differently expressed genes (Table 3; supplementary file 6) revealed a significant enrichment in genes involved in oxidative phosphorylation and lipid metabolism (GO:0005739; 0006629; 0009055; 0005811; 0018925; Table 3) in slow skeletal muscle and carbohydrate metabolism and kinase activity (GO:0005975; 0019752; 0016301) in fast skeletal muscle. These results are in agreement with the main differences described between slow and fast muscle with respect to their predominant means of energy supply [7]. To further validate GO analysis results, reads mapped were individually explored in all components related to general metabolism pathways present in our transcriptome: glycolysis, β -oxidation, Krebs cycle and the electron transport chain (supplementary file 7). The majority of glycolysis-related enzymes found, such as glucose-6-phosphate isomerase (gpi), 6-phosphofructokinase (pfkm) or enolase (eno), were significantly more highly expressed in fast skeletal muscle (FDR<0.05) (Supplementary file 7). As expected slow muscle had a significantly higher DGE in almost all components of the Krebs cycle, β -oxidation pathway and the electron transport chain (supplementary file 8). Therefore DGE results captured fundamental physiological and metabolic differences between tissues. We also studied DGE in relation to the Pi3k/mTor pathway, myogenesis-related genes and TSGD-paralogues between fibre types. Previous studies have reported that TSGD-paralogues can display different expression profiles in muscle during myogenesis and in response to varying nutrition, however those studies were always restricted to fast skeletal muscle and to a limited number of paralogues [34, 35]. By using DGE we found that 301 and 245 TSGD-paralogues pairs were differentially expressed for slow and fast skeletal muscle respectively (FDR<0.05; example in Figure 1B; supplementary file 9). Similarly, we found 124 paralogues with significant differences in DGE between muscle types (FDR<0.05; Figure 1B). Various mechanisms have been postulated to explain TSGD paralogue retention during evolution including mutations leading to differential regulation in expression (sub-

functionalization) and/or the evolution of some novel function (neofunctionalization) [36]. Whilst we cannot distinguish between these possibilities in the present study it is apparent that TSGD paralogue retention has significantly contributed to phenotypic patterning of gene expression between fast and slow muscle fibre types.

Table 3: Gene ontology (GO) enrichment analysis

	GO Term	Name	Type	FDR	Genes in category	
	0003779	actin binding	F	2,30E-31	174	
	0005739	mitochondrion	C	1,20E-30	305	
	0006811	ion transport	P	6,20E-28	209	
	0006091	generation of precursor metabolites and energy	P	4,00E-21	125	
	0003774	motor activity	F	9,10E-21	79	
	0030154	cell differentiation	P	2,70E-14	398	
	0009653	anatomical structure morphogenesis	P	3,50E-14	397	
	0005198	structural molecule activity	F	6,10E-11	126	
	0007275	multicellular organismal development	P	9,30E-09	561	
Slow muscle	0005509	calcium ion binding	F	9,30E-09	107	
	0005216	ion channel activity	F	9,60E-09	63	
	0006629	lipid metabolic process	P	2,50E-08	153	
	0009055	electron carrier activity	F	8,40E-08	39	
	0007010	cytoskeleton organization	P	4,50E-07	181	
	0005886	plasma membrane	C	8,80E-07	420	
	0005783	endoplasmic reticulum	C	4,80E-06	149	
	0005615	extracellular space	C	2,00E-04	61	
	0000166	nucleotide binding	F	2,50E-04	397	
	0019725	cellular homeostasis	P	3,60E-04	93	
	0007267	cell-cell signaling	P	4,00E-04	105	
	0005811	lipid particle	C	2,10E-03	13	
	0019825	oxygen binding	F	3,20E-03	5	
	0005975	carbohydrate metabolic process	P	3,50E-03	99	
		0006091	generation of precursor metabolites and energy	P	7,50E-81	194
		0005975	carbohydrate metabolic process	P	1,00E-71	225
		0005783	endoplasmic reticulum	C	4,90E-23	174
Fast muscle	1901564	organonitrogen compound metabolic process	P	4,90E-23	115	
	0019752	carboxylic acid metabolic process	P	4,90E-23	115	
	0005216	ion channel activity	F	4,90E-23	80	
	0005509	calcium ion binding	F	7,20E-23	122	
	0006811	ion transport	P	4,90E-15	140	

0016301	kinase activity	F	8,60E-15	188
0019725	cellular homeostasis	P	3,50E-14	107
0003779	actin binding	F	5,50E-12	101
0005829	cytosol	C	1,20E-08	263
0005886	plasma membrane	C	1,60E-07	331
0005856	cytoskeleton	C	7,60E-06	223
0043234	protein complex	C	1,20E-04	321
0009056	catabolic process	P	3,10E-04	244

C = cellular component, P = biological process and F = molecular function

The Pi3k/Akt/mTor network is involved in several cellular processes: muscle growth, cell cycle, muscle differentiation and myoblast proliferation [8, 37, 38]. The majority of the pathway's components were identified in the transcriptome, including several paralogues (Figure 2; Supplementary file 10). We found differences in DGE between muscle types for 28 components (FDR<0.05; Figure 2). Only 6 genes were more abundant in fast skeletal muscle including insulin-like growth factor receptor 1a (*igfr1a*) and tuberous sclerosis 2 (*tsc2*) (Figure 2, empty circles). A total of 22 components of the pathway were found to be more highly expressed in slow skeletal muscle including insulin-like growth factor 2b (*igf2b*), several insulin-like growth factors binding proteins (*igfbp1a*, *igfbp2a*, *igfbp5a* and *igfbp7*), protein kinase B gamma (*akt3*), ribosomal protein S6 kinase beta 1a (*rps6kb1a*), eukaryotic translation initiation factor 4E-binding protein 1 (*eif4ebp1*) and mitogen-activated protein kinases (*mapk14a* and *mapk14b*) (Figure 2, red circles). Those genes significantly up regulated in slow skeletal muscle are directly involved in the stimulation of protein synthesis, suggesting a higher protein synthesis potential in this fibre type [39; 40]. Some of the genes significantly up regulated in fast skeletal muscle are associated with the regulation of protein synthesis during fasting. For instance, *lkb1* is expressed when energy levels are low, *mlst8* binds mTor to stabilize the complex when amino acids are not optimal and *tsc2* is related with protein synthesis inhibition [40].

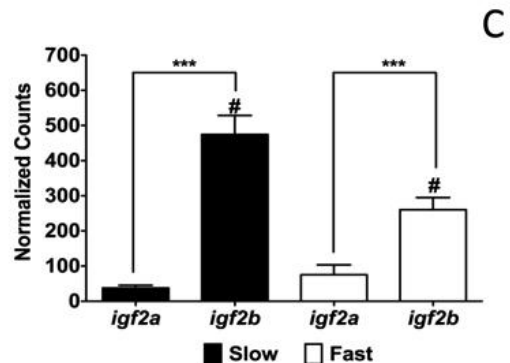
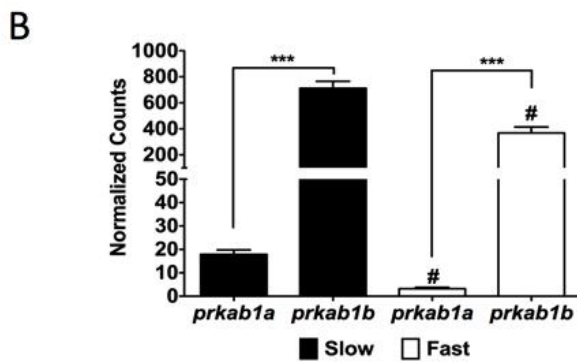
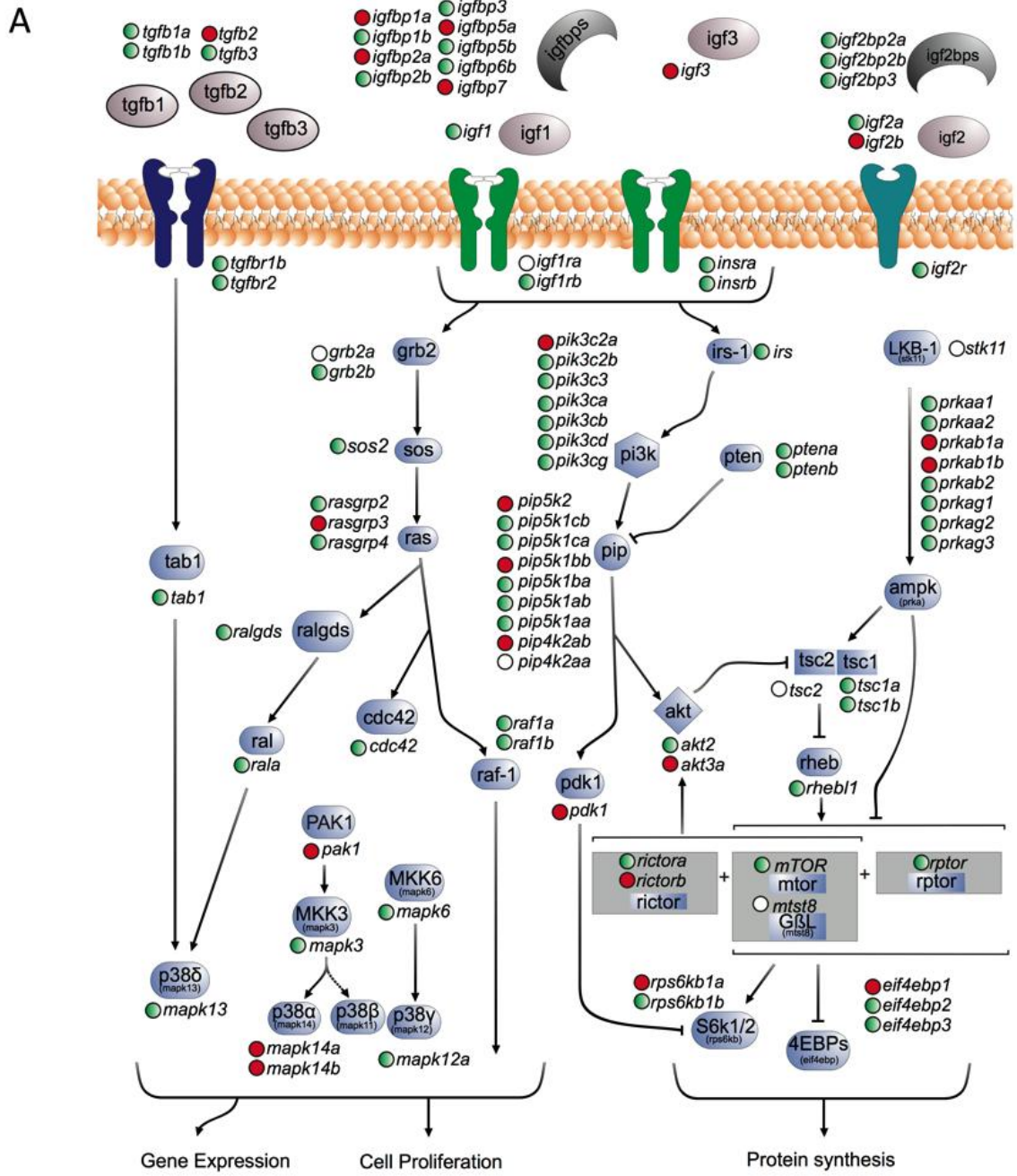


Figure 2: Digital gene expression of the Pi3k/mTor pathway components in pacu fast and slow skeletal muscle.

(A) Pi3k/mTor components represented in the transcriptome mapped into a reconstruction of the same pathway. Red circles (●) and empty circles (○) indicate components significantly higher in slow and fast skeletal muscle respectively muscle (FDR<0.05). Green circles (●) indicate components with no significant differences between fibre types. (B) *igf2a* and *igf2b* digital gene expression. (C) *prkab1a* and *prkab1b* digital gene expression. Digital gene expression is represented using a logarithmic scale for slow (full bars) and fast (empty bars) skeletal muscle. Significant differences for paralogues within (***) (FDR<0.001) and between (#; FDR<0.05) fibre types are indicated. Values represent mean ± SE (n=5). *Igf2*: insulin-like growth factor 2; *Prkab1*: 5' AMP-activated protein kinase subunit beta 1

DGE was also analysed for the genes associated with myoblast activation, proliferation, differentiation and migration (Figure. 3A). *Stac3* is a multifunctional signal adaptor protein which shows a strong correlation with myogenin expression both in cell culture and in vivo and is required for normal myotube formation and sarcomere assembly [41]. *Stac3* localises to T-tubules and also functions in the mediation of voltage-induced Ca²⁺ release and contractility [42]. The higher expression of *stac3*, in correlation with myogenin, in fast than slow muscle (Figure 3E) may therefore reflect the more extensive development of T-tubules and sarcoplasmic reticulum in this fibre type which are required to achieve shorter contraction cycles during high speed swimming [43]. In contrast, the other myogenic regulatory factors (Figure 3B) and members of the myocyte enhancer gene family (except *mef2ca*) (Figure 3C) had similar expression patterns in fast and slow muscles. Differences in expression for myostatin paralogues has been previously reported in other fish species including rainbow trout (*Oncorhynchus mykiss*) and tilapia (*Oreochromis niloticus*) with *mstna* expression (also known as *mstn2*) confined to brain, testes and spleen and *mstnb* (also known as *mstn1*) more expressed in heart and muscle [45].

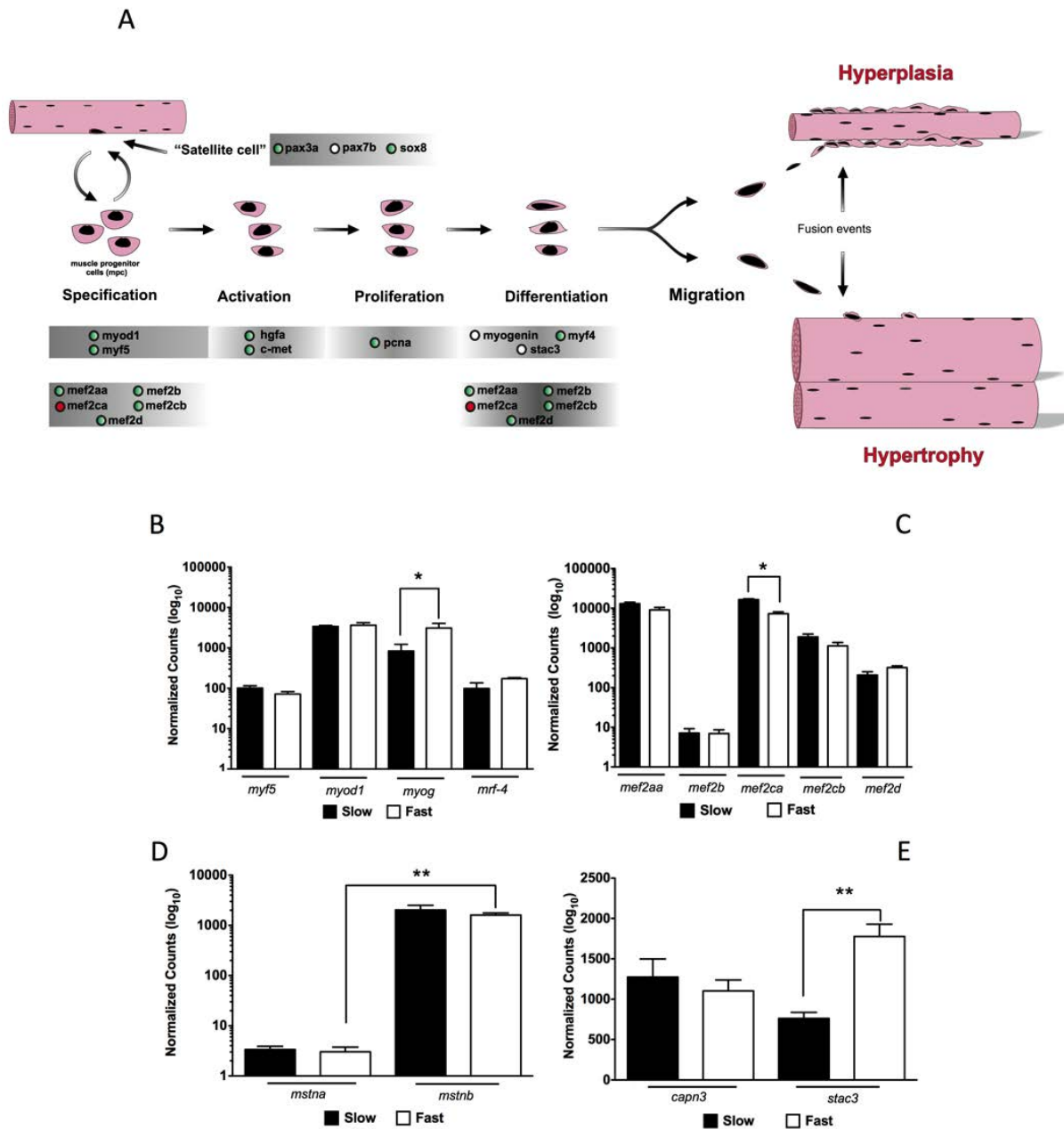


Figure 3: **Digital gene expression of myogenic related genes in pacu fast and slow skeletal muscle.**

(A) Myogenic related genes represented in the transcriptome mapped into a schematic reconstruction of the myogenesis process. Red circles (●) and empty circles (○) indicate components significantly higher in slow and fast skeletal muscle respectively muscle (FDR<0.05). Green circles (◉) indicate components with no significant differences between fibre types. (B) Myogenic regulatory factors (MRFs) digital gene expression. (C) Myocyte-specific enhancer factor 2 genes digital gene expression. (D) Myostatin digital gene expression. (E) Calpain 3 and cysteine rich domain 3 digital gene expression. Digital gene expression is represented using a logarithmic scale for slow (full bars) and fast (empty bars) skeletal muscle. Significant differences for paralogues within (***) and between

(#; FDR<0.05) fibre types are indicated. Values represent mean \pm SE (n=5). *Myod1*: myoblast determination protein 1; *myf5*: myogenic factor 5; *mrf4*: myogenic regulatory factor 4; *myog*: myogenin; *mef2*: myocyte specific enhancer factor 2; *capn3*: calpain-3; *stac3*: cysteine rich domain 3; *mstn*: myostatin.

TRANSCRIPTIONAL REGULATION OF USP AND E3-UBIQUITIN LIGASES

The quality of the present transcriptome allowed us to study the regulation of the Ubiquitin Specific Proteases (USP) family, a group of DUBs for which very little is known in skeletal muscle [18]. USPs are a highly diverse family with a common Ubiquitin carboxyl-terminal hydrolase (UCH 2_3) functional domain. Several USPs members contain a variety of other functional domains including dual specific phosphatases (DUSP), zinc-finger in ubiquitin carboxyl-terminal hydrolase (ZF-UBP), ubiquitin associated domains (UBA) or even an Hsp90 binding motif (MEEVD) (Supplementary file 11). A total of 45 USPs, representing almost the complete repertoire in identified zebrafish, were found in the pacu transcriptome. Again, the normalized number of reads mapped was used to estimate their abundance in fast and slow skeletal muscle (Figure 4). We found that 4 USPs were significantly more abundant in slow (*usp2b*, *usp10*, *usp43a* *usp48*) (Figure 4A) and 9 USPs were more highly expressed in fast skeletal muscle (*usp5b*, *usp9*, *usp14*, *usp19*, *usp21*, *usp24*, *usp28*, *usp45*, *usp47*) (Figure 4B). The lack of studies concerning USP function in fish makes it difficult to explain these differences in a physiological context. With the objective of gaining knowledge about the role of USPs in muscle growth we performed a nutritional challenge experiment to analyse their expression during the transition from a catabolic to an anabolic state. To this end pacu juveniles were fasted for 4 days followed by a short period of satiation feeding (24h) to stimulate protein synthesis and a subset of muscle expressed USP genes were measured by real-time PCR (GE) (Figure 4). The sub-set was chosen to reflect their diversity with respect to functional domains present (*usp2a*, *usp2b*, *usp4*, *usp5a*, *usp5b*, *usp8*, *usp9*, *usp11*, *usp12a*, *usp12b*, *usp14*, *usp16*, *usp19*, *usp21*, *usp28*, *usp30*, *usp36* and *usp46*). The expression of several Pi3k/Akt pathway components (*igf1*, *igf2a*, *igf2b*, *igf1ra*, *igf1rb* and *igf3*) and E3-ubiquitinating ligases (*mafbox*, *murf1a*, *murf1b*, *fbox-25*, *huwe*, *ufd2*, *trip12*, *syyna* and *syah*) were also measured to gain a better understanding of the metabolic context in which USP expression was occurring (Figure 5A). Transcript abundance from the transcriptome was highly correlated with that determined by qPCR ($R^2=0.76$; n=21; $P<0.0001$; data not show).

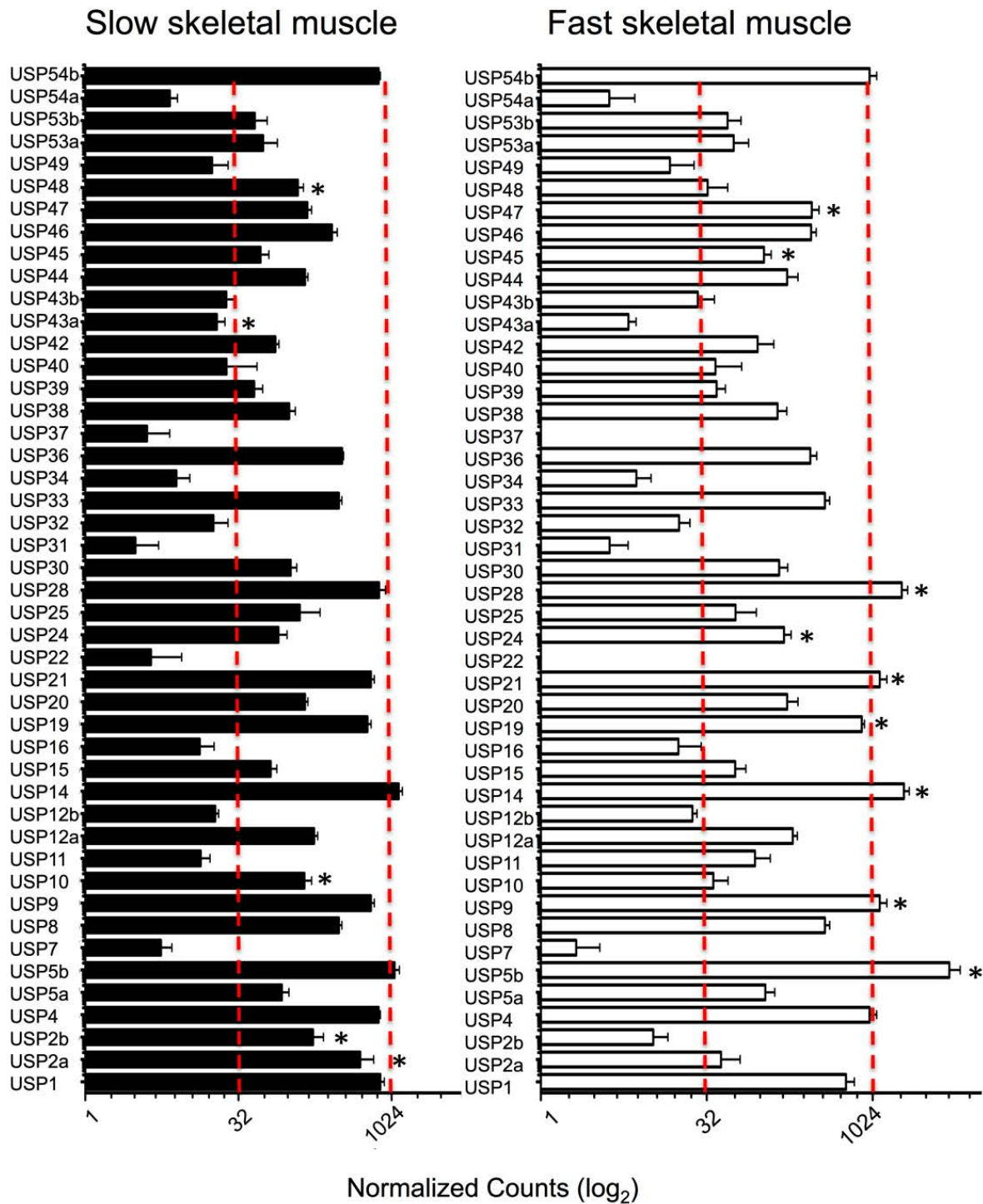


Figure 4: USP digital gene expression in fast and slow skeletal muscle.

Digital gene expression for all USPs is represented using a 2-logarithmic scale for slow (full bars) and fast (empty bars) skeletal muscle. Significant differences and between fibre types (FDR<0.05) are indicated (*). Values represent mean ± SE (n=5).

The E3-ubiquitin ligases are an essential part of the proteasome system, directly involved in protein degradation [45, 46]. All E3-ubiquitin ligases increased their relative abundance during fasting, for example, *fbox-32* a muscle specific E3-ubiquitin ligase increased 100-fold (Figure 5B) and *fbox25* increased 10-fold, in line with results from other similar studies [21, 47]. The majority of the USPs increased their expression around 2-fold with fasting and recovered pre-fasting levels 24h after re-feeding (Figure 5A). In contrast, *Usp12a*, showed a transient increase 6h after re-feeding, and *usp46* and *usp5b* did not change in expression (Figure 5D and E). The overlapping expression profiles of USPs and E3-ubiquitin ligases suggests that USPs play an important role during muscle atrophy. USPs can cleave ubiquitin from proteins, effectively removing the proteasome signal [15]. It is possible that the increase in their abundance in fasted fish is related to fine tuning of the regulation of protein degradation. For example, many USPs targets are essential to maintain cell homeostasis including *mdm4*, *p53*, *h2a*, *h2b*, *fbw7*, *fancd2* or *brca2* [17, 48-50], and it is possible that these proteins may be relatively spared during fasting.

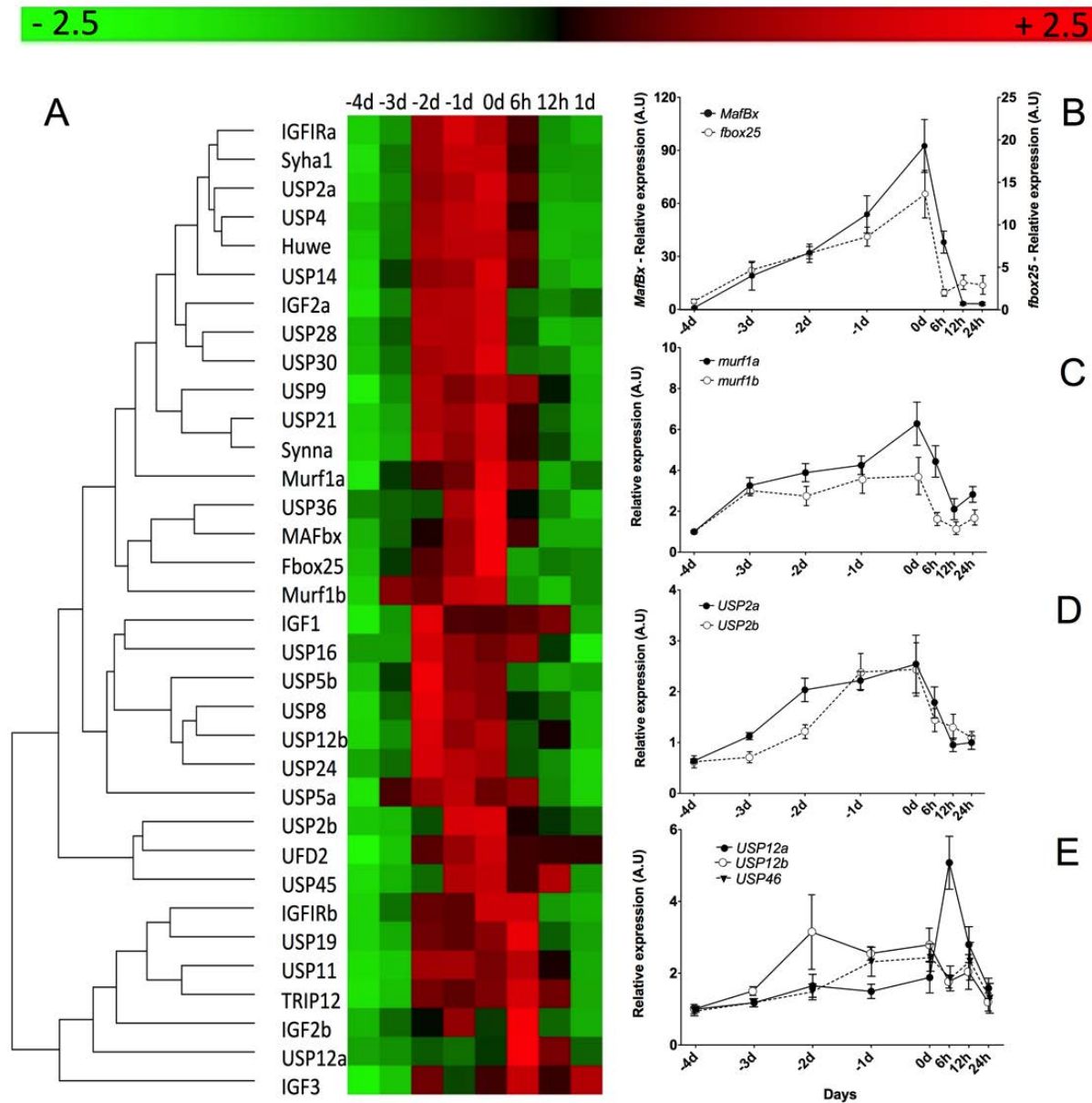


Figure 5: **Relative gene expression in response to fasting and satiation feeding in pacu juveniles in fast skeletal muscle.**

(A) Heat map summary and hierarchical cluster for all genes analysed in fast skeletal muscle during transition from 4 days of fasting (-4, -3, -2 and -1) to 24h of satiation feeding (0, 6, 12 and 24 hours). Rows are standardized to have a mean of 0 and standard deviation of 1. Red indicates high and green low expression values. Relative gene expression graphs for *mafbx* (full circles) and *fbox25* (empty circles) (B) *murf1a* (full circles) and *murf1b* (empty circles) (C) *usp2a* (full circles) and *usp2b* (empty circles) (D) *usp12a* (full circles), *usp12b* (empty circles) and *usp46* (inverted triangles) (E). Values represents mean \pm SE (n=8). *Usp*: ubiquitin specific peptidase; *fbox*: f-box only protein; *mafbx*: f-box only protein 32; *murf1*: E3-ubiquitin ligase TRIM63.

CONCLUSIONS

We have produced an in depth transcriptome for fast and slow myotomal muscle for the pacu (*Piaractus mesopotamicus*), an important South American aquaculture species. This resource allowed us to characterise the expression signatures of the main myotomal muscle fibre types and identify candidate microsatellite sequences that could be used in breeding programs. The availability of the transcriptome allowed us to identify and analyse the expression of teleost-specific paralogues retained in the Ostariophysi lineage. The transcriptome also enabled a comprehensive study of E3 ubiquitin ligase and USP gene expression in the context of the transition between the fasting (catabolic state) and satiation feeding (anabolic state). We identified differences in expression within gene family members thereby identifying candidates for further investigation.

REFERENCES

1. Sanger A, Stoiber W: Muscle fiber diversity and plasticity. *Fish Physiol* 2001, 18:187–250.
2. Altringham JD, Johnston IA: Modelling muscle power output in swimming fish. *J Exp Biol* 1990, 148:395-402.
3. Coughlin DJ, Rome LC: The Roles of Pink and Red Muscle in Powering Steady Swimming in Scup, *stentomus chrysops*. *Am Zool* 1996, 36:666–677.
4. Johnston IA, Calvo J, Guderley H, Fernandez D, Palmer L: Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. *J Exp Biol* 1998, 201:1-12.
5. Wakeling JM, Johnston IA: Muscle power output limits fast-start performance in fish. *J Exp Biol* 1998, 201 (Pt 10): 1505-26.
6. Mascarello F, Rowleron A, Radaelli G, Scapolo PA, Veggetti A: Differentiation and growth of muscle in the fish *Sparus aurata* (L): I. Myosin expression and organization of fibre types in lateral muscle from hatching to adult. *J Muscle Res Cell Motil* 1995, 16:213–22.

7. Johnston IA, Dvison W, Goldspink G: Energy metabolism of carp swimming muscle. *J Comp Physiol* 1977, 144:203-216.
8. Johnston IA, Bower NI, Macqueen DJ: Growth and the regulation of myotomal muscle mass in teleost fish. *J Exp Biol* 2011, 214:1617–1628.
9. Cleveland BM, Weber GM, Blemings KP, Silverstein JT: Insulin-like growth factor-I and genetic effects on indexes of protein degradation in response to feed deprivation in rainbow trout (*Oncorhynchus mykiss*). *Am J Physiol Regul Integr Comp Physiol* 2009, 297:R1332–42.
10. Rescan PY, Montfort J, Fautrel A, Rallièrre C, Lebret V: Gene expression profiling of the hyperplastic growth zones of the late trout embryo myotome using laser capture microdissection and microarray analysis. *BMC Genomics* 2013, 14:173.
11. Johnston IA: Environment and plasticity of myogenesis in teleost fish. *J Exp Biol* 2006, 209:2249–2264.
12. Johnston IA: Muscle development and growth: potential implications for flesh quality in fish. *Aquaculture* 1999, 177:99–115.
13. Tacchi L, Bickerdike R, Secombes CJ, Martin SA : Muscle-specific RING finger (MuRF) cDNAs in Atlantic salmon (*Salmo salar*) and their role as regulators of muscle protein degradation. *Mar Biotechnol* 2012, 14:35–45.
14. Salmerón C, García de la Serrana D, Jiménez-Amilburu V, Fontanillas R, Navarro I, Johnston IA, Gutiérrez J, Capilla E: Characterisation and Expression of Calpain Family Members in Relation to Nutritional Status, Diet Composition and Flesh Texture in Gilthead Sea Bream (*Sparus aurata*). *PLoS One* 2013, 8:e75349.
15. Argilés JM, López-Soriano FJ, Busquets S: Apoptosis signalling is essential and precedes protein degradation in wasting skeletal muscle during catabolic conditions. *Int J Biochem Cell Biol* 2008, 40:1674–8.
16. Lecker SH, Goldberg AL, Mitch WE: Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J Am Soc Nephrol* 2006, 17:1807

17. David Y, Ternette N, Edelmann MJ, Ziv T, Gayer B, Sertchook R, Dadon Y, Kessler BM, Navon A: E3 ligases determine ubiquitination site and conjugate type by enforcing specificity on E2 enzymes. *J Biol Chem* 2011, 286:44104–15.
18. Komander D, Clague MJ, Urbé S: Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* 2009, 10:550–63.
19. Daviet L, Colland F: Targeting ubiquitin specific proteases for drug discovery. *Biochimie* 2008, 90:270–83.
20. Amaral IPG, Johnston IA: Insulin-like growth factor (IGF) signalling and genome-wide transcriptional regulation in fast muscle of zebrafish following a single-satiating meal. *J Exp Biol* 2011, 214:2125–2139.
21. Bower N, Johnston IA: Characterisation and differential regulation of MAFbx/Atrogin-1 a and b transcripts in skeletal muscle of Atlantic salmon (*Salmo salar*). *Physiol Genomics* 2010, 396:114–130.
22. www.ensembl.org
23. Garcia de la Serrana D, Mareco E, Johnston IA: Systematic variation in the pattern of gene paralogue retention between the teleost super-orders Ostariophysi and Acanthopterygii. *Genome Biol Evol* 2014, 6:981-987.
24. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Cattolico L, Poulain J, et al.: Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 2004, 431:946–57.
25. Sato Y, Nishida M: Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environ Biol Fishes* 2010, 88:169–188.

26. Braasch I, Postlethwait J: Polyploidy in fish and the teleost genome duplication. In Polyploidy genome Evol. Edited by Soltis PS, Soltis DE. Springer Berlin Heidelberg; 2012:341–383.
27. Macqueen DJ, Johnston IA: A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci 2014, 281:20132881.
28. Amaral IPG, Johnston IA: Experimental selection for body size at age modifies early life-history traits and muscle gene expression in adult zebrafish. J Exp Biol 2012, 215:3895–3904.
29. Magnanou E, Klopp C, Noirot C, Besseau L, Falcón J: Generation and characterization of the sea bass *Dicentrarchus labrax* brain and liver transcriptomes. Gene 2014, 544:56–66.
30. Garcia de la serrana Castillo D, Estévez A, Andree K, Johnston I: Fast skeletal muscle transcriptome of the Gilthead sea bream (*Sparus aurata*) determined by next generation sequencing. BMC Genomics 2012, 13:181.
31. Mojekwu TO, Anumudu CI: Microsatellite markers in Aquaculture: Application in Fish population genetics. IOSR J Environ Sci Toxicol Food Technol 2013, 5:43–48.
32. Calduch-Giner JA, Bermejo-Nogales A, Benedito-Palos L, Estensoro I, Ballester-Lozano G, Sitjà-Bobadilla A, Pérez-Sánchez J: Deep sequencing for de novo construction of a marine fish (*Sparus aurata*) transcriptome database with a large coverage of protein-coding transcripts. BMC Genomics 2013, 14:178.
33. Palstra AP, Beltran S, Burgerhout E, Brittijin SA, Magnoni LJ, Henkel CV, Jansen HJ, van den Thillart GE, Spaink HP, Planas JV: Deep RNA sequencing of the skeletal muscle transcriptome in swimming fish. PLoS One 2013, 8:e53171.
34. Ganassi M, Badodi S, Polacchini a, Baruffaldi F, Battini R, Hughes SM, Hinitis Y, Molinari S: Distinct functions of alternatively spliced isoforms encoded by zebrafish *mef2ca* and *mef2cb*. Biochim Biophys Acta 2014, 1839:559-70.
35. Garcia de la Serrana D, Johnston IA: Expression of heat shock protein (Hsp90) paralogues is regulated by amino acids in skeletal muscle of Atlantic salmon. PLoS One 2013, 8:e74295.

36. Maere S, Peer YV.: Duplicate Retention After Small and Large-Scale Duplications. In *Evol after Gene Duplic.* 1st edition. Edited by Dittmar K, Liberles D. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010:31–56.
37. Rommel C, Bodine SC, Clarke BA, Rossman R, Nunez L, Stitt TN, Yancopoulos GD, Glass DJ: Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. *Nat Cell Biol* 2001, 3:1009–13.
38. Glass DJ: Molecular mechanisms modulating muscle mass. *Trends Mol Med* 2003, 9:344–350.
39. Lang CH, Vary TC, Frost RA: Acute in vivo elevation of insulin-like growth factor (IGF) binding protein-1 decreases plasma free IGF-I and muscle protein synthesis. *Endocrinology* 2003, 144:3922–33.
40. Huang K, Fingar DC: Growing knowledge of the mTOR signaling network. *Sem Cell Dev Biol* 2014.
- 41 Bower NI, Garcia de la serrana D, Cole NJ, Hollway GE, Lee HT, Assinder S, Johnston IA: STAC3 is required for myotube formation and myogenic differentiation in vertebrate skeletal muscle. *J Biol Chem* 2012, 287:43936-43949.
42. Nelson BR, Wu F, Liu Y, Anderson DM, McAnally J, Lin W, Cannon SC, Bassel-Duby R, Olson EN: Skeletal muscle-specific T-tubule protein STAC3 mediates voltage-induced Ca²⁺ release and contractility. *Proc Natl Acad Sci USA* 2013, 110:11881-11886.
43. Fleming JR, Crockford T, Altringham JD, Johnston IA: Effects of temperature acclimation on muscle relaxation in the carp: A mechanical, biochemical and ultrastructural study. *J Exp Zool* 1990, 255:286-295.
44. Gabillard JC, Biga PR, Rescan PY, Seilliez I: Revisiting the paradigm of myostatin in vertebrates: insights from fishes. *Gen Comp Endocrinol* 2013, 194:45-54.
45. Deshaies RJ, Joazeiro CA: RING domain E3 ubiquitin ligases. *Annu Rev Biochem* 2009, 78:399–434.

46. Berndsen CE, Wolberger C: New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* 2014, 21:301–7.
47. Bower NI, Taylor RG, Johnston IA: Phasing of muscle gene expression with fasting-induced recovery growth in Atlantic salmon. *Front Zool* 2009, 6:18.
48. Gomes LC, Di Benedetto G, Scorrano L: During autophagy mitochondria elongate are spared from degradation and sustain cell viability. *Nat Cell Biol* 2011, 13:589–598.
49. Cataldo F, Peche LY, Klaric E, Brancolini C, Myers MP, Demarchi F, Schneider C: CAPNS1 regulates USP1 stability and maintenance of genome integrity. *Mol Cell Biol* 2013, 33:2485–96.
50. Holowaty MN, Sheng Y, Nguyen T, Arrowsmith C, Frappier L: Protein interaction domains of the ubiquitin-specific protease, USP7/HAUSP. *J Biol Chem* 2003, 278:47753–61.
51. Haas B, Papanicolaou A, Yassour M: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013, 8:1494–512.
52. <http://www.bioconductor.org>
53. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 2008, 36:3420–35.
54. Moriya Y, Itoh M, Okuda S: KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids Res* 2007, 35 (Web Server issue):W182.
55. <http://www.mbio.ncsu.edu/bioedit/bioedit.html>
56. Wernersson R: Virtual Ribosome: a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res* 2006, 34(Web Server issue):W385–8.
57. Faircloth BC: Msatcommander: Detection of Microsatellite Repeat Arrays and Automated, Locus-Specific Primer Design. *Mol Ecol Resour* 2008, 8:92–4.

58. Bustin S A, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT: The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009, 55:611–22.
59. <http://www.ncbi.nlm.nih.gov/spidey/>
60. Pfaffl MW: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001, 29:e45.
61. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol Lett* 2004, 26:509–15.
62. Katoh K, Standley D: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013, 30:772–80.
63. Dereeper A, Guignon V, Blanc G: Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids Res* 2008, 36(Web Server issue):W465–9.
64. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28:2731–9.
65. Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biol* 2010, 11:R106.
66. Caraux G, Pinloche S: PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2005, 21:1280–1.

SUPPLEMENTARY INFORMATION

Supplementary file 1: **Annotated transcriptome**

Annotated nucleotide using blast2go from the pacu slow and fast muscle transcriptome.

Supplementary file 2: **KEEG Automatic Annotation Sever results**

Hierarchical information of the sequences from annotated transcriptome mapped to the KEGG pathways maps using the automatic annotation tool KAAS [54].

Supplementary file 3: **Simple Sequence Repeats Identification.**

Simple Sequence Repeats identified in contigs containing over 90% of the CDS using Msatcommander software [57].

Supplementary file 4: **Teleost-specific paralogues sequences and phylogenetic trees.**

Supplementary file contains the amino acids sequences for the pacu paralogues and the corresponding orthologues for zebrafish, stickleback, tilapia, green puffer fish and human, their MAFFT alignments and individual phylogenetic trees.

Supplementary file 5: **Fast and slow skeletal muscle global digital gene expression analysis.**

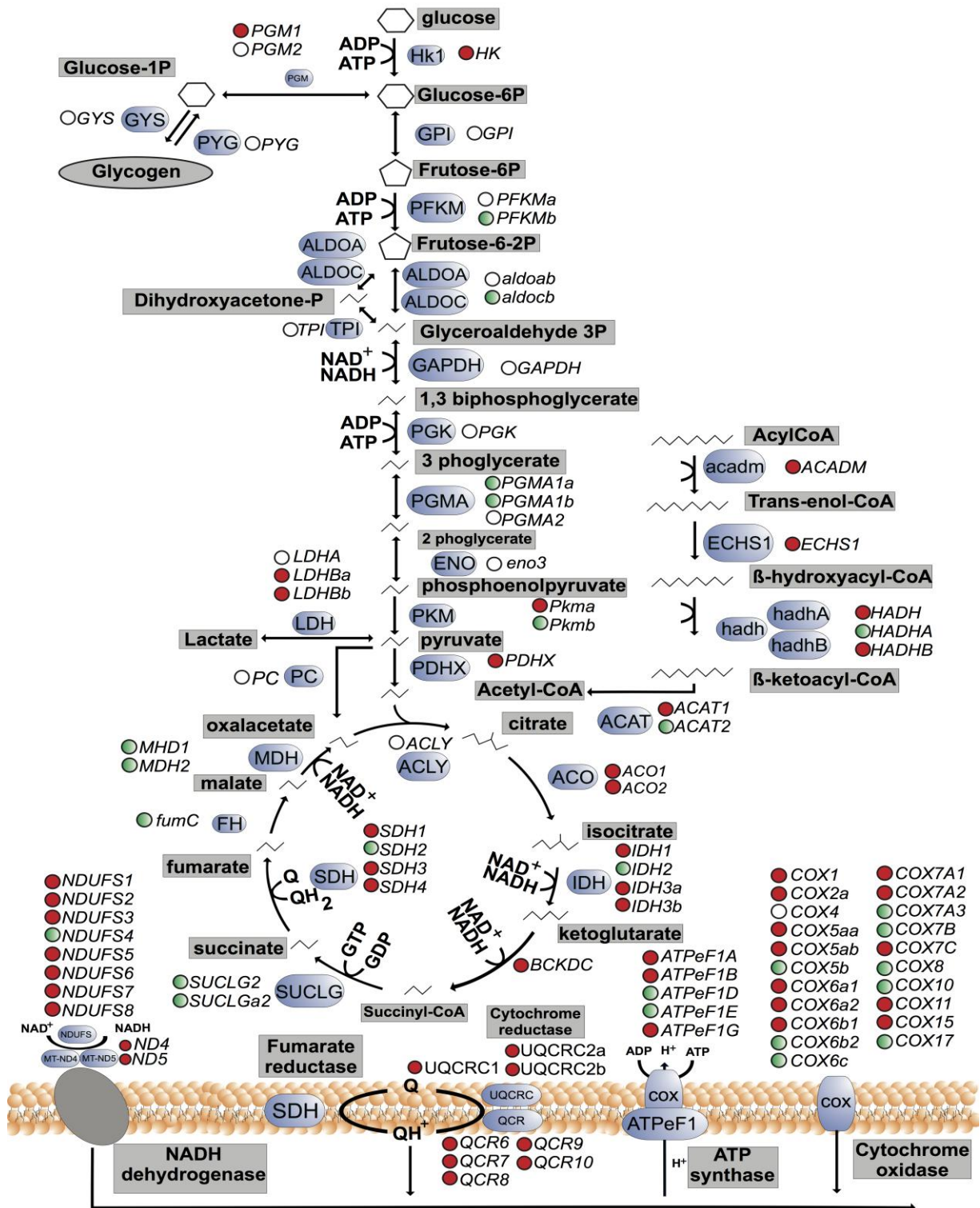
Global digital gene expression analysis between fast and slow skeletal muscle was performed using the number of reads mapped normalized by length and library size was investigated using R-Bioconductor. Normalized counts for individual animals per contig are shown. Differences between tissues were analysed using the DESEQ package. Significant differences were considered when $FDR < 0.05$.

Supplementary file 6: **Fast and slow skeletal muscle Gene Ontology (GO) analysis.**

GO analysis was performed for differently expressed contigs originating from the fast and slow skeletal muscle comparison. Significantly enriched categories were investigated using Fisher Exact Test with the whole annotated transcriptome as reference and considered significant when $FDR < 0.05$.

Supplementary file 7: **Digital gene expression of metabolic pathways in fast and slow skeletal muscle.**

Components from the different metabolic pathways were extracted from KAAS annotation results. Analysis between fast and slow skeletal muscle was performed using the number of reads mapped normalized by length and library size. Normalized counts for individual animals per contig are shown. Differences between tissues were analysed by t-test followed by a False Discovery Rate correction. Significant differences were considered when $FDR < 0.05$.



Supplementary file 8: Schematic representation of fast and slow skeletal muscle main metabolic pathways.

Metabolic genes represented in the transcriptome mapped into a reconstruction of the main metabolic pathways. Red circles (●) and empty circles (○) indicate components significantly

higher in slow and fast skeletal muscle respectively muscle (FDR<0.05). Green circles (●) indicate components with no significant differences between fibre types.

Supplementary file 9: Teleost-specific paralogues digital gene expression.

Teleost-specific paralogues digital gene expression in slow and fast skeletal muscle. Comparisons were made between paralogue pairs and fibres types. Analysis between fast and slow skeletal muscle was performed using the number of reads mapped normalized by length and library size. Normalized counts for individual animals per contig are shown. Differences between tissues were analysed by t-test followed by a False Discovery Rate correction. Significant differences were considered when FDR<0.05.

Supplementary file 10: Pi3k/mTor pathway components digital gene expression.

Pi3k/mTor components digital gene expression for slow and fast skeletal muscle. Analysis between fast and slow skeletal muscle was performed using the number of reads mapped normalized by length and library size. Normalized counts for individual animals per contig are shown. Differences between tissues were analysed by t-test followed by a False Discovery Rate correction. Significant differences were considered when FDR<0.05.

Supplementary file 11: Pacu ubiquitin specific proteases functional domains and digital gene expression.

Ubiquitin specific proteases (USP) digital gene expression analysis between slow and fast skeletal muscle. Analysis between fast and slow skeletal muscle was performed using the number of reads mapped normalized by length and library size. Normalized counts for individual animals per contig are shown. Differences between tissues were analysed by t-test followed by a False Discovery Rate correction. Significant differences were considered when FDR<0.05. Pacu, zebrafish and human protein USP functional domains were determined using InterProScan webserver. Relative gene expression was analysed for *fbox25*, *huwe*, *mafbox*, *murfla*, *murflb*, *syah1*, *syyna*, *trip12*, *ufd2*, *igf1*, *igf2a*, *igf2b*, *igf3*, *igf1ra*, *igf1rb*, *usp2a*, *usp2b*, *usp4*, *usp5a*, *usp5b*, *usp8*, *usp9*, *usp11*, *usp12b*, *usp12b*, *usp14*, *usp16*, *usp19*, *usp21*, *usp24*, *usp28*, *usp30*, *usp36* and *usp46*, Values represents mean ± SE (n=8 fish).

5 REFERÊNCIAS

- ANDERS, S. et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. **Nature Protocols**, v. 8, n. 9, p. 1765–1786, 2013.
- BIRESSI, S.; MOLINARO, M.; COSSU, G. Cellular heterogeneity during vertebrate skeletal muscle development. **Developmental Biology**, v. 308, n. 2, p. 281–293, 2007.
- BONE, Q. Locomotor Muscle. In: HOAR, W. S.; RANDALL, D. J. (Eds.). . **Fish Physiology-Locomotion**. 1. ed. New York: ACADEMIC PRESS INC., 1978. p. 361–424.
- BOWER, N. I.; TAYLOR, R. G.; JOHNSTON, I. A. Phasing of muscle gene expression with fasting-induced recovery growth in Atlantic salmon. **Frontiers in zoology**, v. 6, p. 18, jan. 2009.
- BOWER, N.; JOHNSTON, I. Characterisation and differential regulation of MAFbx / Atrogin-1 a and b transcripts in skeletal muscle of Atlantic salmon (*Salmo salar*). **Biochemical and biophysical research ...**, v. 396, n. 2, p. 265–271, 2010.
- BRYSON-RICHARDSON, R. J.; CURRIE, P. D. The genetics of vertebrate myogenesis. **Nature reviews. Genetics**, v. 9, n. 8, p. 632–646, 2008.
- CLEVELAND, B. M. et al. Insulin-like growth factor-I and genetic effects on indexes of protein degradation in response to feed deprivation in rainbow trout (*Oncorhynchus mykiss*). **American journal of physiology. Regulatory, integrative and comparative physiology**, v. 297, n. 5, p. R1332–42, nov. 2009.
- DE SANTIS, C.; JERRY, D. R. Differential tissue-regulation of myostatin genes in the teleost fish *Lates calcarifer* in response to fasting. Evidence for functional differentiation. **Molecular and Cellular Endocrinology**, v. 335, p. 158–165, 2011.
- DELBARRE-LADRAT, C. et al. Trends in postmortem aging in fish: understanding of proteolysis and disorganization of the myofibrillar structure. **Critical reviews in food science and nutrition**, v. 46, n. 5, p. 409–421, 2006.
- DEREEPER, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. **Nucleic acids research**, v. 36, n. Web Server issue, p. 465–469, 2008.
- DEVOTO, S. H. et al. Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. **Development (Cambridge, England)**, v. 122, p. 3371–3380, 1996.
- DRIEDZIC, W. R.; HOCHACHKA, P. W. Control of energy metabolism in fish white muscle. **The American journal of physiology**, v. 230, n. 3, p. 579–82, mar. 1976.
- DU, S. J. et al. Positive and negative regulation of muscle cell identity by members of the hedgehog and TGF-beta gene families. **The Journal of cell biology**, v. 139, n. 1, p. 145–156, 1997.

DU, S. J. Molecular regulation of fish muscle development and growth. In: GONG, Z.; KORZH, V. (Eds.). **Fish development and genetics: the zebrafish and medaka models**. Molecular ed. New Jersey: World Scientific Pub Co Inc, 2004. p. 339–391.

ERBAY, E. et al. IGF-II transcription in skeletal myogenesis is controlled by mTOR and nutrients. **Journal of Cell Biology**, v. 163, p. 931–936, 2003.

FAUCONNEAU, B.; PABOEUF, G. Muscle Satellite Cells in Fish. **Fish physiology, Muscle Development and Growth**, v. 18, p. 73–101, 2001.

FERREIRA, F. J.; SILVA, C. A. Atividade alimentar do pacu *Piaractus mesopotamicus* Holmberg (1887) criados em tanques rede. **Bioikos**, v. 26, n. 1887, p. 35–42, 2012.

FUENTES, E. N. et al. Catabolic Signaling Pathways, Atrogenes, and Ubiquitinated Proteins Are Regulated by the Nutritional Status in the Muscle of the Fine Flounder. **PLoS ONE**, v. 7, n. 9, p. e44256, 14 set. 2012.

GAHR, S. A.; WEBER, G. M.; REXROAD, C. E. Fasting and refeeding affect the expression of the Inhibitor of DNA Binding (ID) genes in rainbow trout (*Oncorhynchus mykiss*) muscle. **Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology**, v. 144, p. 472–477, 2006.

GARCÍA DE LA SERRANA, D. et al. Characterisation and expression of myogenesis regulatory factors during in vitro myoblast development and in vivo fasting in the gilthead sea bream (*Sparus aurata*). **Comparative biochemistry and physiology. Part A, Molecular & integrative physiology**, v. 167C, p. 90–99, 21 out. 2013.

GARCIA DE LA SERRANA, D.; MARECO, E. A.; JOHNSTON, I. Systematic variation in the pattern of gene paralogue retention between the teleost super-orders Ostariophysi and Acanthopterygii. **Genome biology and evolution**, 14 abr. 2014.

GLASAUER, S. M. K.; NEUHAUSS, S. C. F. Whole-genome duplication in teleost fishes and its evolutionary consequences. **Molecular Genetics and Genomics**, p. 1045–1060, 2014.

GÖTZ, S. et al. High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic acids research**, v. 36, n. 10, p. 3420–35, jun. 2008.

GREEK-WALKER, M.; PULL, G. A. A survey of red and white muscle in marine fish. **Journal of Fish Biology**, v. 7, n. 3, p. 295–300, 1975.

GUDERLEY, H. et al. Metabolic priorities during starvation: Enzyme sparing in liver and white muscle of Atlantic cod, *Gadus morhua* L. **Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology**, v. 135, n. 2, p. 347–356, 2003.

HAAS, B.; PAPANICOLAOU, A.; YASSOUR, M. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature protocols**, v. 8, n. 8, p. 1494–512, ago. 2013.

HALL, T. A. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT** Nucleic Acids Symposium Series, 1999.

JACKSON, H. E. et al. The role of Sox6 in zebrafish muscle fiber type specification. **Skeletal Muscle**, v. 5, n. 1, p. 1–17, 2015.

JAILLON, O. et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. **Nature**, v. 431, n. October, p. 946–957, 2004.

JOHNSTON, I. A. A comparative study of glycolysis in red and white muscles of the trout (*Salmo gairdneri*) and mirror carp (*Cyprinus carpio*). **Journal of Fish Biology**, v. 11, n. 6, p. 575–588, dez. 1977.

JOHNSTON, I. A. Contractile properties of fish fast muscle fibres. **Marine Biology Letters**, v. 1, p. 323–328, 1980.

JOHNSTON, I. A. Environment and plasticity of myogenesis in teleost fish. **The Journal of experimental biology**, v. 209, n. Pt 12, p. 2249–64, jun. 2006.

JOMORI, R. K. et al. Growth and survival of pacu *Piaractus mesopotamicus* (Holmberg, 1887) juveniles reared in ponds or at different initial larviculture periods indoors. **Aquaculture**, v. 221, n. 1-4, p. 277–287, 2003.

KOMANDER, D.; CLAGUE, M. J.; URBÉ, S. Breaking the chains: structure and function of the deubiquitinases. **Nature reviews. Molecular cell biology**, v. 10, n. 8, p. 550–63, ago. 2009.

KOUMANS, J. T. M.; AKSTER, H. Myogenic cells in development and growth of fish. **Comparative Biochemistry and Physiology Part A: Physiology**, v. 110, n. 1, p. 3–20, 1995.

KUANG, S. et al. Asymmetric Self-Renewal and Commitment of Satellite Stem Cells in Muscle. **Cell**, v. 129, n. 5, p. 999–1010, 2007.

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC bioinformatics**, v. 12, p. 323, jan. 2011.

MACCATROZZO, L. et al. A novel second myostatin gene is present in teleost fish. **FEBS Letters**, v. 509, n. 1, p. 36–40, 2001.

MACQUEEN, D. J. et al. A newly classified vertebrate calpain protease, directly ancestral to CAPN1 and 2, episodically evolved a restricted physiological function in placental mammals. **Molecular Biology and Evolution**, v. 27, n. 8, p. 1886–1902, 2010.

MACQUEEN, D. J.; ROBB, D.; JOHNSTON, I. A. Temperature influences the coordinated expression of myogenic regulatory factors during embryonic myogenesis in Atlantic salmon (*Salmo salar* L.). **Journal of Experimental**, n. 210, p. 2781–2794, 2007.

- MARTIN, S. A. M. et al. Ubiquitin-proteasome-dependent proteolysis in rainbow trout (*Oncorhynchus mykiss*): Effect of food deprivation. **Pflugers Archiv European Journal of Physiology**, v. 445, n. 2, p. 257–266, 2002.
- MATSAKAS, A.; PATEL, K. Intracellular signalling pathways regulating the adaptation of skeletal muscle to exercise and nutritional changes. p. 209–222, 2009.
- MAURO, A. Satellite Cell of Skeletal Muscle Fibers. **The Journal of biophysical and biochemical cytology**, v. 9, p. 493–495, 1961.
- MEYER, A; SCHARTL, M. Gene and genome duplication in vertebrates: the one-to-four (-to eight in fish) rule and the evolution of novel gene functions. v. 11, p. 699–704, 1999.
- MILSTEIN, A. et al. Low temperature tolerance of pacu , *Piaractus mesopotamicus*. **Environmental Biology of Fishes**, v. 58, p. 455–460, 2000.
- MOTOHASHI, N.; ASAKURA, A. Muscle satellite cell heterogeneity and self-renewal. **Front Cell Dev Biol.**, v. 2, n. 1, p. 1–21, 2014.
- MPA. **BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA 2011**, 2011.
- OHNO, S. **Evolution by gene duplication**. 1. ed. New York: Springer Science+Business Media, LLC 1970, 1970.
- PATEL, R. K.; JAIN, M. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. **PLoS ONE**, v. 7, n. 2, 2012.
- ROWLERSON, A.; VEGGETTI, A. Cellular Mechanisms of Post-Embryonic Muscle Growth in Aquaculture Species. In: JOHNSTON (Ed.). **Fish physiology, Muscle Development and Growth**. [s.l.] ACADEMIC PRESS INC., 2001. p. 103–140.
- SACCO, J. J. et al. Emerging roles of deubiquitinases in cancer-associated pathways. **IUBMB life**, v. 62, n. 2, p. 140–57, fev. 2010.
- SALEM, M. et al. Effect of starvation on global gene expression and proteolysis in rainbow trout (*Oncorhynchus mykiss*). **BMC genomics**, v. 8, p. 328, jan. 2007.
- SALMERÓN, C. et al. Characterisation and Expression of Calpain Family Members in Relation to Nutritional Status, Diet Composition and Flesh Texture in Gilthead Sea Bream (*Sparus aurata*). **PloS one**, v. 8, n. 9, p. e75349, jan. 2013.
- SÄNGER, A.; STOIBER, W. Muscle Fiber Diversity and Plasticity.pdf. **Fish physiology**, v. 18, p. 187–250, 2001.
- SATO, Y.; NISHIDA, M. Teleost fish with specific genome duplication as unique models of vertebrate evolution. **Environmental Biology of Fishes**, v. 88, p. 169–188, 2010.
- SCHIAFFINO, S.; REGGIANI, C. Molecular diversity of myofibrillar proteins: gene regulation and functional significance. **Physiol Rev**, v. 76, n. 2, p. 371–423, 1996.

SEILIEZ, I. et al. Feeding status regulates the polyubiquitination step of the ubiquitin-proteasome-dependent proteolysis in rainbow trout (*Oncorhynchus mykiss*) muscle. **The Journal of nutrition**, v. 138, n. 3, p. 487–91, mar. 2008.

SOUZA, V. L. et al. Avaliação do Crescimento e do Custo da Alimentação do Pacu (*Piaractus mesopotamicus* Holmberg, 1887) Submetido a Ciclos Alternados de Restrição Alimentar e Realimentação. **Revista Brasileira de Zootecnia**, v. 32, n. 1, p. 19–28, 2003.

TACCHI, L. et al. Ubiquitin E3 ligase atrogin-1 (Fbox-32) in Atlantic salmon (*Salmo salar*): sequence analysis, genomic structure and modulation of expression. **Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology**, v. 157, n. 4, p. 364–73, dez. 2010.

TACCHI, L. et al. Muscle-specific RING finger (MuRF) cDNAs in Atlantic salmon (*Salmo salar*) and their role as regulators of muscle protein degradation. **Marine biotechnology (New York, N.Y.)**, v. 14, n. 1, p. 35–45, fev. 2012.

TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular biology and evolution**, v. 28, n. 10, p. 2731–9, out. 2011.

TAN, X.; JUN DU, S. Differential expression of two MyoD genes in fast and slow muscles of gilthead seabream (*Sparus aurata*). **Development Genes and Evolution**, v. 212, p. 207–217, 2002.

URBINATI, E. C.; GONÇALVEZ, F. D. Pacu (*Piaractus mesopotamicus*). In: BALDISSEROTO, B.; GOMES, L. DE C. (Eds.). . **Espécies nativas para a piscicultura no Brasil**. Santa Maria – RS: Editora Universidade Federal Santa Maria, 2006. v. 1p. 225–256.

VARSHAVSKY, A. The ubiquitin system. **Annual review of biochemistry**, v. 22, n. 10, p. 383–387, out. 1998.

VON HOFSTEN, J. et al. Prdm1- and Sox6-mediated transcriptional repression specifies muscle fibre type in the zebrafish embryo. **EMBO reports**, v. 9, n. 7, p. 683–689, 2008.