



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE AGRONOMIA E MEDICINA VETERINÁRIA

CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (*Pseudoplatystoma reticulatum*) E HÍBRIDOS COM PINTADO (*P. corruscans*) COM TECNOLOGIAS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

LUCIANA CRISTINE VASQUES VILLELA

TESE DE DOUTORADO
EM CIÊNCIAS ANIMAIS

BRASÍLIA/DF
SETEMBRO DE 2015



**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE AGRONOMIA E MEDICINA VETERINÁRIA**

CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (*Pseudoplatystoma reticulatum*) E HÍBRIDOS COM PINTADO (*P. corruscans*) COM TECNOLOGIAS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

Luciana Cristine Vasques Villela

ORIENTADOR: Alexandre Rodrigues Caetano

**TESE DE DOUTORADO
EM CIÊNCIAS ANIMAIS**

PUBLICAÇÃO: 139D/2015

**BRASÍLIA/DF
SETEMBRO DE 2015**

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE AGRONOMIA E MEDICINA VETERINÁRIA

CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (*Pseudoplatystoma reticulatum*) E HÍBRIDOS COM PINTADO (*P. corruscans*) COM TECNOLOGIAS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

LUCIANA CRISTINE VASQUES VILLELA

TESE DE DOUTORADO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS ANIMAIS, COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS ANIMAIS.

APROVADA POR:

ALEXANDRE RODRIGUES CAETANO, Livre Docência (Embrapa Recursos Genéticos e Biotecnologia - CENARGEN) - (ORIENTADOR)

MARCIO BOTELHO DE CASTRO, Doutorado (Universidade de Brasília - UnB) - (EXAMINADOR INTERNO)

RODRIGO DIANA NAVARRO, Doutorado (Universidade de Brasília - UnB) - (EXAMINADOR INTERNO)

ANTONIO AUGUSTO MENDES MAIA, Livre Docência (Faculdade de Zootecnia e Engenharia de Alimentos - USP) - (EXAMINADOR EXTERNO)

CONCEPTA MARGARET MCMANUS PIMENTEL, Pós-doutorado (Universidade de Brasília - UnB) - (EXAMINADOR EXTERNO)

BRASÍLIA/DF, 14 DE SETEMBRO DE 2015

AGRADECIMENTOS

Um agradecimento mais que especial ao meu orientador, Alexandre (Dr. Alexandre Rodrigues Caetano), pela paciência, dedicação, apoio, amizade e ensinamentos ao longo destes anos de doutorado. Obrigada por tudo, Alexandre.

Outro agradecimento especial à amiga Naiara, sem sua ajuda eu não teria conseguido realizar todas as atividades no laboratório. Não posso deixar de agradecer ao Renato e ao Gleisson, que também sempre me ajudaram com as atividades e dúvidas no laboratório.

Aos amigos Anderson, Eduardo e Gigio, da Embrapa Pesca e Aquicultura, pela realização das coletas dos animais para a realização deste trabalho. Aos colegas da Embrapa CNPTIA pela realização das análises para o desenvolvimento desta tese. À Embrapa, de forma geral e à UnB, pela possibilidade de realização deste trabalho.

Aos amigos Bete, Cali, Lilith, e Rony pelos momentos passados no laboratório e fora dele também. Às amigas Fá e Ana Paula, por me “salvarem” nas horas em que precisei ficar até mais tarde no laboratório e pela amizade. Às amigas Renata e Lícia, pelas palavras de apoio, compreensão e incentivo nos momentos difíceis que encontrei pelo caminho, neste último ano.

À minha querida família: irmãs Carol e Gal e em especial meu pai Jorge e Carmem, por terem me proporcionado a base que permitiu que eu chegasse até aqui. Obrigada por estarem presente, me apoiando sempre, independente das decisões acertadas ou não que tomei.

Aos meus queridos marido (Gá) e filho Samuel (Sam, a mamãe conseguiu terminar o livro!), obrigada por estarem sempre ao meu lado e por compartilharem comigo mais esta etapa alcançada.

E à Deus por ter permitido que tudo isso acontecesse.

ÍNDICE

RESUMO	vii
ABSTRACT	ix
PARECER DA CEUA/Cenargen	xi
CAPÍTULO 1	1
1 INTRODUÇÃO	1
1.1 Caracterização do Problema	3
2 OBJETIVOS	3
2.1 Objetivo Geral	3
2.2 Objetivos Específicos	3
3 REVISÃO DE LITERATURA	5
3.1 A Aquicultura no Brasil e no Mundo	5
3.2 Considerações sobre a Biologia do Pintado e da Cachara	7
3.3 Potencial Aquícola do Pintado e da Cachara	9
3.4 Criação de Híbridos Inter-Específicos: Vantagens, Desvantagens e Riscos Ambientais	11
3.5 Metodologias de Identificação de Híbridos	12
3.6 Marcadores SNP – Bases Moleculares, Prospecção e Utilização	13
3.7 Tecnologias de Sequenciamento de Nova Geração	14
3.8 Sequenciamento do Transcriptoma Completo	15
3.9 Sequenciamento de ESTs/cDNAs	16
3.10 Genes tecido específico e <i>housekeeping</i>	17
3.11 Músculo Branco e Músculo vermelho em peixes	17
4 REFERÊNCIAS BIBLIOGRÁFICAS	19
CAPÍTULO 2 - DESENVOLVIMENTO DE UM ENSAIO DE MINISEQUENCIAMENTO PARA IDENTIFICAÇÃO DE HÍBRIDOS ENTRE CACHARA (<i>Pseudoplatystoma reticulatum</i>) E PINTADO (<i>Pseudoplatystoma corruscans</i>)	22
1 Resumo	22
2 Abstract	25
3 Introdução	27
4 Material e Métodos	29
5 Resultados e Discussão	36
6 Conclusões	58
7 Referências Bibliográficas	59
CAPÍTULO 3 - PROSPECÇÃO DE POLIMORFISMOS DE BASE ÚNICA (SNP) ESPÉCIE-ESPECÍFICOS NOS TRANSCRIPTOMAS DA CACHARA (<i>Pseudoplatystoma reticulatum</i>) E DE HÍBRIDOS DE PINTADO (<i>Pseudoplatystoma corruscans</i>)	66
1 Resumo	66
2 Abstract	68
3 Introdução	69
4 Material e Métodos	71
5 Resultados e Discussão	74
6 Conclusões	87
7 Referências Bibliográficas	88
CAPÍTULO 4 - CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (<i>Pseudoplatystoma reticulatum</i>): SEQUENCIAMENTO, MONTAGEM E ANÁLISE DE EXPRESSÃO TECIDO-ESPECÍFICA	90
1 Resumo	90
2 Abstract	92
3 Introdução	94
4 Material e Métodos	96
5 Resultados e Discussão	99
6 Conclusões	108
7 Referências Bibliográficas	109
CAPÍTULO 5 - SEQUENCIAMENTO, MONTAGEM DE NOVO E ANÁLISE DESCRITIVA DO GENOMA MITOCONDRIAL COMPLETO DA CACHARA (<i>Pseudoplatystoma reticulatum</i>)	111
1 Resumo	111
2 Abstract	113
3 Introdução	114

4	Material e Métodos	117
5	Resultados e Discussão	119
6	Conclusões	135
7	Referências Bibliográficas	134
CAPÍTULO 6 – CONSIDERAÇÕES FINAIS		146

RESUMO

CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (*Pseudoplatystoma reticulatum*) E DE HÍBRIDOS COM PINTADO (*P. corruscans*) COM TECNOLOGIAS DE SEQUENCIAMENTO DE NOVA GERAÇÃO

Autor: Luciana Cristine Vasques Villela, Mestre, Brasília/DF

Orientador: Alexandre Rodrigues Caetano, Livre Docência, Brasília/DF

A cachara do pantanal (*Pseudoplatystoma reticulatum*) é uma espécie de bagre neotropical nativa do Brasil historicamente importante para a pesca comercial e esportiva. A produção dos bagres e de seus híbridos em cativeiro apresentou uma taxa de crescimento de mais de 100% entre 2010 e 2011, chegando a 17.619 toneladas. A geração de híbridos com o pintado (*P. corruscans*) vem sendo amplamente utilizada pelo setor produtivo para obter peixes superiores para o cultivo, que apresentam vigor híbrido, além das características superiores de cada uma das espécies puras. Contudo, essa prática apresenta uma ameaça para o desenvolvimento sustentável da piscicultura dos surubins, já que os híbridos gerados são férteis e podem retrocruzar com populações selvagens e estoques puros de reprodutores mantidos em cativeiro. Ferramentas de vanguarda são necessárias para alavancar o desenvolvimento de tecnologias avançadas para a criação da cachara em cativeiro, como a detecção de introgressões interespecíficas, e métodos aprimorados de reprodução e de monitoramento e melhoramento genético de estoques de reprodutores. Inicialmente, um ensaio de minisequenciamento (SNaPshot®) foi desenvolvido e validado com base em marcadores SNP (*Single Nucleotide Polymorphism*) espécie-específicos disponíveis na literatura, derivados de dois genes mitocondrais (16S e COI) e quatro genes nucleares (RAG2, GLOB, 18S e EF1 α). Adicionalmente, análises do transcriptoma de sete tecidos (músculo branco, músculo vermelho, hipófise, gônada, rim, fígado e brânquia) de cacharas e híbridos de

cachara/pintado foram realizadas para caracterizar o transcriptoma da espécie e prospectar SNPs espécie-específicos. RNA extraído de cada um dos tecidos de 17 cacharas e nove híbridos foi combinado em quantidades equimolares e utilizado para gerar 16 bibliotecas de cDNA: uma biblioteca para cada tecido de cada espécie (n=14) e uma biblioteca contendo todos os tecidos para cada espécie (n=2); as quais foram sequenciadas com tecnologia Illumina com protocolos de sequenciamento de 100 e 300 pb de comprimento. Um total de 997.043.038 fragmentos sequenciados foram processados com os programas Trinity e BWA para montagem do transcriptoma das duas espécies (cachara e híbridos de pintado) e identificação de SNPs, respectivamente. A análise do transcriptoma da cachara revelou um total de 93.674 transcritos únicos, além de 37.917 transcritos tecido-específicos e 7.456 transcritos presentes em todos os tecidos analisados. Um total de 647.954 SNPs foram identificados nas sequências analisadas, sendo que 64 SNPs espécie-específicos foram identificados, satisfazendo critérios de espaçamento e posição nos transcritos analisados, os quais serão adequados para identificar introgressões da ordem de 1,65% entre genomas da cachara e do pintado. Adicionalmente, uma montagem do DNA mitocondrial da cachara foi produzida com base nas sequências de cDNA analisadas e sequências da região controle D-loop geradas com sequenciamento Sanger. O mitogenoma completo da cachara apresentou 16.576 pb de comprimento, e a mesma estrutura básica, ordem e organização gênica observada nos mitogenomas de outras espécies de vertebrados. As informações e conhecimentos gerados neste estudo serão disponibilizadas publicamente, e poderão ser empregadas no desenvolvimento de novos estudos em áreas como filogeografia, sistemática filogenética, genética de populações, conservação e melhoramento da cachara e de outros bagres tropicais.

Palavras-chave: Sequenciamento de nova geração; RNA-Seq; MiSeq; HiSeq; Genes *Housekeeping*

ABSTRACT

TRANSCRIPTOME CHARACTERIZATION OF CACHARA (*Pseudoplatystoma reticulatum*) AND PINTADO (*P. corruscans*) HYBRIDS USING NEW GENERATION SEQUENCING TECHNOLOGIES

Author: Luciana Cristine Vasques Villela, Mestre, Brasília/DF

Advisor: Alexandre Rodrigues Caetano, Livre Docência, Brasília/DF

Pantanal (Brazilian wetlands) cachara (*Pseudoplatystoma reticulatum*) is a Brazilian native neotropical catfish historically important for commercial and sport fisheries. Catfish and hybrids aquaculture production showed a growth rate over 100% between 2010 and 2011, reaching 17,619 tonnes. The hybrids generation with pintado (*P. corruscans*) has been widely used by growers to obtain superior fishes, with hybrid vigor, in addition to pure species superior characteristics. However, this practice is a menace to sustainable catfish aquaculture development hence the hybrids are fertile and can backcross with wild populations and pure broodstocks maintained in captivity. The use of cutting-edge tools is necessary to boost the advanced technologies development for cachara farming, like interspecific introgression detection, and improved bloodstock's reproduction, monitoring and breeding methods. Initially a minisequencing assay (SNaPshot©) was developed and validated based on specie-specific SNP (Single Nucleotide Polymorphism) tracers available at literature, derived from two mitochondrial genes (16S and COI) and four nuclear genes (RAG2, GLOB, 18 S and EF1 α). Additionally, transcriptome analysis of seven tissues (white muscle, red muscle, hypophysis, gonad, kidney, liver and gill) of cachara and hybrids of cachara/pintado were carried to characterize the transcriptome of the species and search for specie-specific SNPs. The RNA extracted from each tissue of 17 cachara and nine hybrids were combined in equimolar quantities and used to generate 16 cDNA libraries: one library,

for each tissue of each specie (n=14) and a library containing all tissues of each specie (n=2), which were sequenced with Illumina technology with sequencing protocols of 100 and 300 pb in length. A total of 997,043,038 sequenced fragments were processed with Trinity and BWA programs to transcriptome assembly of two species (cachara and pintado hybrids) and SNPs identification, respectively. The transcriptome analysis of cachara revealed a total of 93,674 unique transcripts, in addition to 37,917 tissue-specific transcripts and 7,456 transcripts present in all analyzed tissues. A total of 647,954 SNPs were identified in the analyzed sequences, amongst them 64 specie-specific SNPs, satisfying spacing and position criteria in the studied transcripts which were suitable for introgression identification in the order of 1.65% between cachara and pintado genome. Besides, a mitochondrial DNA assembly were produced based on the analyzed cDNA and D-loop control region sequences using Sanger sequencing. The cachara complete metagenome showed 16,576 bp in length, and the same basic structure, order and gene organization observed in the metagenomes of other vertebrate species. The novel information and knowledge generated by this study will be public available and can be used in the development of new researches in filo geography, filogenetic systematics, population genetics, cachara and other tropical catfish conservation and breeding.

Keywords: Next-generation sequencing; RNA-Seq; MiSeq; HiSeq; Housekeeping genes



Recursos Genéticos e Biotecnologia

M. CEUA N° 06/2014

Brasília, 08 de abril de 2014.

Ao Pesquisador Alexandre Rodrigues Caetano

Assunto: Parecer da CEUA/Cenargen sobre solicitação de autorização para uso de animais em experimento e/ou ensino

Prezado Alexandre,

Em reunião do dia 08 de abril de 2014, a Comissão de Ética no Uso de Animais da Embrapa Recursos Genéticos e Biotecnologia (CEUA/Cenargen) analisou e aprovou sua solicitação para uso de animais no projeto intitulado "**Prospecção de polimorfismos de base única (SNP) e caracterização do transcriptoma do pintado (*Pseudoplatystoma corruscans*) e da cachara (*Pseudoplatystoma reticulatum*) com tecnologias de seqüenciamento de nova geração para desenvolvimento de ferramentas genômicas de manejo e melhoramento genético**", protocolo CEUA 002-2014.

Lembramos que conforme determina a Resolução Normativa N°. 01, de 9 de julho de 2010, do CONCEA, Art. 6º, Inciso IX, o relatório final do projeto deverá ser encaminhado à CEUA.

Atenciosamente,

A handwritten signature in black ink that reads "Silvia Tereza Ribeiro Castro".

Silvia Tereza Ribeiro Castro
Coordenadora da CEUA/Cenargen
Embrapa Recursos Genéticos e Biotecnologia

CAPÍTULO 1

1 INTRODUÇÃO

1.1 Caracterização do Problema

A aquicultura diz respeito ao cultivo confinado e controlado de organismos (peixes, crustáceos, moluscos, anfíbios e plantas aquáticas) que vivem em ambiente predominantemente aquático (Camargo & Pouey, 2005). E tem apresentado expansivo crescimento nas últimas décadas, no Brasil e no mundo (FAO, 2014). No entanto, o Brasil ainda se encontra em uma posição modesta em relação aos maiores produtores e exportadores de animais aquáticos do mundo (Boletim Estatístico da Pesca e Aquicultura. Brasil, 2011).

Segundo a FAO (Organização das Nações Unidas para Alimentação e Agricultura), o País possui grande potencial para ser um dos maiores atores no mercado internacional num futuro próximo, especialmente, quando se leva em consideração seu clima e geografia favoráveis, sua extensão de águas costeiras e continentais, rica biodiversidade, tanto no mar quanto nos rios e lagoas, características importantes para o desenvolvimento e crescimento da aquicultura (Carlson, 2014).

Um dos ramos da aquicultura, a piscicultura (produção de peixes em cativeiro) vem demonstrando grande crescimento, proporcionando a produção de proteína animal de alta qualidade e baixo custo. Ainda mais quando se leva em consideração o número significativo de espécies nativas, aproximadamente 3000 espécies (Carlson, 2014), com potencial para a piscicultura que o Brasil possui. No entanto, a domesticação e o desenvolvimento de tecnologias necessárias para o cultivo (nutrição, reprodução, sistemas e manejo, genética e

melhoramento, etc.) das espécies nativas ainda estão por ser realizados, fazendo com que sua participação na piscicultura nacional fique abaixo dos 50% (Ostrensky et al., 2008; Carlson, 2014).

A cachara (*Pseudoplatystoma reticulatum*) e o pintado (*Pseudoplatystoma corruscans*) são espécies nativas de surubins com potencial para a piscicultura. A geração de híbridos através do cruzamento dessas duas espécies é uma prática comum entre os piscicultores, realizada para a obtenção de peixes com características mais adequadas para o cultivo em cativeiro, como diminuição de canibalismo, aumento da taxa de sobrevivência na larvicultura e após a eclosão, maior consumo de ração (Prado et al., 2011; Prado et al., 2012a).

No entanto, a hibridação também apresenta pontos negativos para o desenvolvimento da piscicultura dos surubins, uma vez que os híbridos gerados são férteis. É elevado o risco de contaminação dos estoques “puros” de reprodutores de pintado e cachara nas pisciculturas, pois o comércio de animais vivos normalmente é feito na fase juvenil de crescimento, quando não é possível diferenciar visualmente híbridos F1 de indivíduos puros com precisão. Assim como existe o risco de perda das populações selvagens devido à introgressão por cruzamentos com os escapes híbridos das pisciculturas (Prado et al., 2011; Prado et al. 2012a; Porto-Foresti et al., 2008).

O uso de ferramentas que identifiquem animais “puros” dessas espécies, ou seja, que não apresentem introgressão de material genético da outra espécie, é necessário para a formação de estoques de reprodutores puros. Métodos que permitem a identificação de híbridos interespecíficos F1 foram desenvolvidos (Hashimoto et al., 2013; Prado et al., 2011; Prado et al. 2012a; Prado et al. 2012b), porém ferramentas mais avançadas, que permitam a identificação de introgressões em menor escala, como por exemplo, através do cruzamento entre animais híbridos e de híbridos com animais puros, precisam ser desenvolvidas.

O tamanho dos genomas do pintado e cachara não é conhecido. Smith & Gregory (2009) determinaram que o valor C (quantidade de DNA contido em um núcleo haplóide) dos peixes da ordem Siluriforme, família Pimelodidae (a qual engloba o pintado e a cachara), varia entre 0,88 e 1,36 pg. Portanto, podemos inferir que se o tamanho dos genomas do pintado e da cachara for igual ao tamanho médio do genoma na família Pimelodidae, teremos genomas com 1,12 pg, o que equivale à aproximadamente 1 Gpb (1 milhão de pares de base). Portanto, para podermos detectar introgressões da ordem de 1,56% (equivalente a 5 gerações de retrocruzamento de um híbrido F1), serão necessários pelo menos 64 SNPs espécie-específicos (=100%/1,56%), ou um SNP a cada 17 Mpb (=1Gpb/64).

Portanto, a identificação de animais puros, sem qualquer nível de introgressão, e de eventuais contaminações/introgressões genéticas, é necessária para que tecnologias possam ser desenvolvidas de forma sustentável para a criação do pintado e da cachara. O desenvolvimento de um painel de marcadores SNPs espécie-específicos que possibilite a identificação de diferentes níveis de introgressão trará nova luz para este cenário atual.

Adicionalmente, uma das formas de contribuir com o desenvolvimento do cultivo dessas espécies no País é aumentando o conhecimento biológico das espécies. Neste sentido, a geração de informações genômicas e moleculares sobre uma das espécies em estudo, a cachara (*Pseudoplatystoma reticulatum*), inexistentes até o momento, possui significativa importância. Adicionalmente, a caracterização do genoma mitocondrial e do transcriptoma da cachara, além de gerar informações moleculares para a espécie, poderá estimular o desenvolvimento de trabalhos semelhantes para outras espécies de peixes.

2 OBJETIVOS

2.1 Objetivo Geral

Gerar e analisar informações úteis para realização de estudos moleculares avançados e para o desenvolvimento de ferramentas de manejo, conservação e melhoramento genético de estoques de reprodutores de cachara (*Pseudoplatystoma reticulatum*) e pintado (*Pseudoplatystoma corruscans*).

2.2 Objetivos Específicos

- a) Desenvolver um ensaio com a tecnologia SNaPshot para a análise de SNPs espécie-específicos com base em informações disponíveis na literatura;
- b) Gerar bancos de seqüências ESTs (*Expressed Sequence Tags*) de diferentes tecidos de pintado e cachara com a utilização de métodos de seqüenciamento de nova geração;

- c) Identificar marcadores SNP espécie-específicos nos bancos de EST, que sejam úteis para o desenvolvimento de painéis de marcadores que identifiquem introgressões genéticas em diferentes níveis;
- d) Identificar marcadores SNP com bom nível de informatividade que possam ser utilizados para compor painéis de marcadores com diferentes aplicações para as duas espécies em questão: testes de confirmação de paternidade, identificação e rastreabilidade de estoques/linhagens de reprodutores, estudos de variabilidade e estruturação genética de populações selvagens e em cativeiro (estoques de reprodutores), construção de mapas genéticos, etc.;
- e) Analisar os bancos de EST para caracterização do transcriptoma da cachara (*Pseudoplatystoma reticulatum*) e geração de um conjunto de informações que sejam úteis para estudos futuros de expressão gênica;
- f) Caracterizar o genoma mitocondrial da cachara (*Pseudoplatystoma reticulatum*) a partir da montagem do seu transcriptoma e gerar informações úteis para futuros estudos de filogenia, filogeografia, conservação, melhoramento genético, etc.

3 REVISÃO DE LITERATURA

3.1 A Aquicultura no Brasil e no Mundo

A aquicultura pode ser definida como o cultivo confinado e controlado de organismos (peixes, crustáceos, moluscos, anfíbios e plantas aquáticas) que têm na água o seu principal ou mais frequente ambiente de vida. É uma atividade que vem crescendo sensivelmente em relação à pesca extrativista, tornando-se uma importante alternativa para a produção de pescado (alimento de alto valor protéico), tanto em área continental como marinha (Camargo & Pouey, 2005).

Em 2010, foram produzidas 168 milhões de toneladas de pescado no mundo (provenientes tanto da pesca extrativa quanto da aquicultura). Os maiores produtores foram a China (63,5 milhões de toneladas), a Indonésia (11,7 milhões de toneladas), a Índia (9,3 milhões de toneladas) e o Japão (5,2 milhões de toneladas). O Brasil contribuiu com 1,26 milhões de toneladas, ocupando a 19ª colocação no *ranking* mundial (Tabela 1.1). De 2010 para 2011, houve um incremento de 13% na produção de pescado nacional (produção aproximada de 1,4 milhões de toneladas). A pesca extrativa marinha foi a maior fonte de produção de pescado nacional (553.670,0 t, correspondente à 38,7% do total de pescado produzido), seguida pela aquicultura continental (544.490,0 t, 38%), pesca extrativa continental (249.600,2 t, 17,4%) e aquicultura marinha (84.214,3 t, 6%) (Boletim Estatístico da Pesca e Aquicultura. Brasil 2011).

Tabela 1.1. Produção total de pescado (t) e classificação no *ranking* de produtores mundiais entre os anos de 2007 e 2010

País	2007	2008	2009	2010
China	56.160.587 (1°)	57.827.007 (1°)	60.474.939 (1°)	63.495.197 (1°)
Indonésia	8.176.567 (2°)	8.815.120 (2°)	9.820.818 (2°)	11.662.343 (2°)
Índia	6.971.536 (3°)	7.583.569 (3°)	7.865.598 (3°)	9.348.063 (3°)
Japão	5.687.766 (5°)	5.542.475 (5°)	5.465.155 (4°)	5.292.392 (4°)
Brasil	1.072.227 (21°)	1.156.423 (21°)	1.240.813 (19°)	1.264.765 (19°)

Adaptado de Boletim Estatístico da Pesca e Aquicultura. Brasil. 2011; Boletim Estatístico da Pesca e Aquicultura. Brasil. 2008 e 2009.

Levando-se em consideração a produção aquícola mundial (tanto marinha quanto continental), China, Indonésia e Índia produziram em 2010, 47,8 milhões, 6,3 milhões, 4,6 milhões de toneladas de pescado, respectivamente, enquanto o Brasil produziu 479.399 toneladas. Apesar de ocupar apenas a 17ª posição no *ranking* mundial de produção de pescado proveniente da aquicultura, a evolução deste setor no Brasil tem sido expressiva. Em 2011, houve um incremento de 31,1% em relação à 2010, quando a produção aquícola nacional atingiu 628.704,3 toneladas. Apenas a aquicultura continental foi responsável pela maior parte desta produção (86,6%, que corresponde à 544.490 t); o restante foi proveniente da aquicultura marinha (Figura 1.1) (Boletim Estatístico da Pesca e Aquicultura. Brasil 2011).

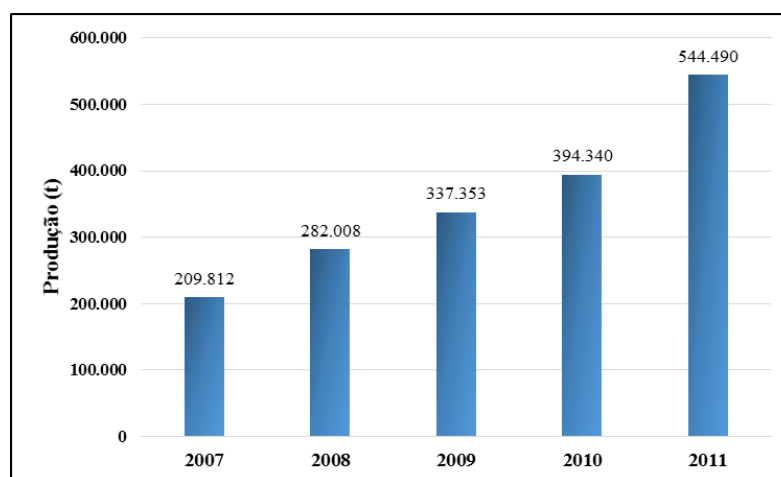


Figura 1.1. Evolução da produção de pescado proveniente da aquicultura continental no Brasil entre 2007 e 2011 (Adaptado de: Boletim Estatístico da Pesca e Aquicultura. Brasil 2008-2009; Boletim Estatístico da Pesca e Aquicultura. Brasil 2010; Boletim Estatístico da Pesca e Aquicultura. Brasil 2011)

A produção aquícola continental brasileira concentra-se nas espécies exóticas, como as tilápias, carpas e trutas, que juntas representaram aproximadamente 55,0% da produção em 2011. Em relação aos peixes nativos, apesar de o Brasil apresentar grande diversidade de espécies, a produção mais expressiva ficou por conta dos peixes redondos (tambaqui, tambacu e pacu): 182.591,4 toneladas. Apesar de ainda pouco expressiva, a produção das diferentes espécies de bagres (*Pseudoplatystoma reticulatum*, *P. corruscans*, *P.*

tigrinum, *Leiarius marmoratus*, etc.), juntamente com seus híbridos foi da ordem de 17.619,70 toneladas no ano de 2011 e apresentou um aumento de aproximadamente 300% em relação à produção no ano de 2007 (Figura 1.2) (Boletim Estatístico da Pesca e Aquicultura. Brasil 2008-2009; Boletim Estatístico da Pesca e Aquicultura. Brasil 2010; Boletim Estatístico da Pesca e Aquicultura. Brasil 2011).

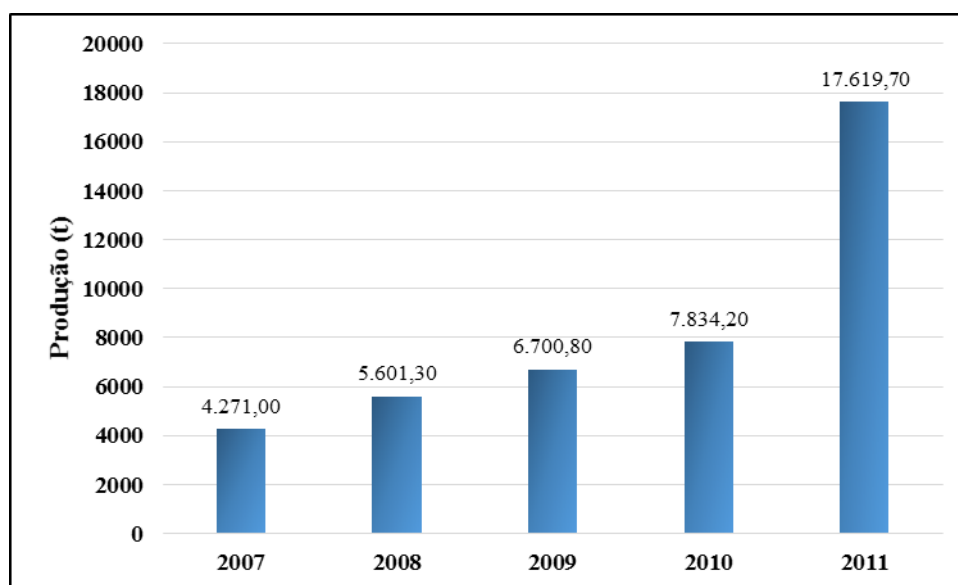


Figura 1.2. Evolução da produção de bagres e seus híbridos na aquicultura continental no Brasil entre os anos de 2007 e 2011 (Adaptado de: Boletim Estatístico da Pesca e Aquicultura. Brasil 2008-2009; Boletim Estatístico da Pesca e Aquicultura. Brasil 2010. Boletim Estatístico da Pesca e Aquicultura. Brasil 2011)

A participação das espécies nativas na piscicultura nacional, de forma geral, ainda é bastante modesta. O maior entrave para aumentar a produtividade dessas espécies é a falta de conhecimentos básicos e de tecnologias aplicadas que permitam que seu cultivo se torne zootécnica e economicamente viável (Ostrensky et al., 2008).

No entanto, a FAO O Brasil possui, portanto, um potencial natural para a produção piscícola, decorrente da sua disponibilidade hídrica (8,4 mil km de costa litorânea; 5,5 milhões de hectares de reservatórios de água doce); do clima preponderantemente tropical; da auto-suficiência na produção de grãos para fabricação de ração e da diversidade de espécies nativas de peixes com potencial para exploração aquícola (Ostrensky et al., 2008; Boletim Estatístico da Pesca e Aquicultura. Brasil 2011; Carlson. 2014).

3.2 Considerações sobre a Biologia da Cachara e do Pintado

A ordem Siluriforme é a segunda maior ordem de peixes neotropicais de água doce, conhecidos como peixes de couro. A família Pimelodidae é uma das mais especiosas, compreendendo entre 50 e 60 gêneros e 300 espécies (Reis et al., 2004; Torrico et al., 2009). Os gêneros de maior destaque dessa ordem são *Pseudoplatystoma*, *Brachyplatystoma*, *Pharactocephalus* e *Zungaro* (Ferreira et al., 1998).

O gênero *Pseudoplatystoma* constitui um grupo monofilético de bagres amplamente distribuídos pelas principais Bacias da América do Sul. São espécies migratórias, de grande importância para a aquicultura nas Bacias hidrográficas onde ocorrem (Buitrago-Suárez & Burr, 2007; Torrico et al., 2009; Carvalho-Costa et al., 2011), cuja classificação taxonômica ainda precisa ser melhor definida. Até recentemente, eram reconhecidas apenas três espécies do gênero: *P. fasciatum* (Linnaeus 1776), amplamente distribuída nas Bacias do Paraná, Amazônica, Orinoco e Magdalena, e rios da Guiana; *P. tigrinum* (Valenciennes 1840), nas Bacias do Orinoco e Amazônica; e *P. corruscans* (Spix e Agassiz 1829), restrito às Bacias do Paraná e São Francisco (Lundberg & Littmann, 2003; Burgess, 1989).

No entanto, a diversidade do gênero, aliada à incerteza em relação ao número de espécies nele existentes levaram Buitrago-Suárez & Burr (2007) a realizar uma revisão taxonômica do gênero *Pseudoplatystoma*. Com base em análises morfológicas detalhadas, os autores reconheceram oito espécies diferentes dentro do gênero *Pseudoplatystoma*: o amplamente distribuído surubim *P. fasciatum* foi dividido em cinco espécies: 1) *P. fasciatum* (*sensu stricto*), restrito às Bacias da Guiana e do rio Suriname; 2) *P. punctifer* (Casteunau 1855), presente na bacia do Rio Amazonas; 3) *P. reticulatum* (Eigenmann e Eigenmann 1889), amplamente distribuído nas bacias dos rios Paraná-Paraguai e presumivelmente também restrito a uma região central do rio Amazonas; 4) *P. orinocoense*, presente na bacia do rio Orinoco e 5) *P. magdaleniatum*, presente na bacia do Rio Magdalena; e outras três espécies: 6) *P. tigrinum* (Valenciennes 1840), presente na bacia do rio Amazonas; 7) *P. metaense*, presente na bacia do rio Orinoco; 8) *P. corruscans* (Spix e Agassiz 1829), presente nas bacias dos rios Paraná e São Francisco. Estas espécies estariam divididas em dois clados, suportados por características anatômicas: clado *P. fasciatum*, composto pelas espécies *P. fasciatum*, *P. punctifer*, *P. orinocoense*, *P. magdaleniatum*, *P. reticulatum*, e *P. corruscans* e clado *P. tigrinum*, composto pelas espécies *P. metaense* e *P. tigrinum*.

Por sua vez, baseados em análises filogenéticas moleculares envolvendo sequências dos genes Citocromo b e Região Controle do DNA mitocondrial, Torrico et al. (2009) identificaram cinco clados principais dentro do gênero *Pseudoplatystoma*: um clado composto por *P. metaense* e *P. orinocoense*; e outros quatro clados compostos individualmente por *P. reticulatum*, *P. tigrinum*, *P. corruscans* e *P. magdaleniatum*. Os autores não conseguiram diferenciar *P. fasciatum* de *P. punctifer* com base nos dados moleculares, indicando a necessidade de maior investigação em relação a estas duas espécies que haviam sido morfologicamente diferenciadas e colocando em dúvida a classificação morfológica realizada por Buitrago-Suárez & Burr (2007).

Já Carvalho-Costa et al. (2011), analisando filogenias baseadas em um gene mitocondrial (Citocromo b) e dois genes nucleares (gene de ativação da recombinação – RAG1 e o gene ribossomal S7), reconheceram apenas quatro espécies dentro do gênero *Pseudoplatystoma*: *P. magdaleniatum*, *P. corruscans*, *P. tigrinum* e *P. fasciatum*.

Dentre essas espécies, o *Pseudoplatystoma corruscans* (Spix e Agassiz 1829), conhecido como pintado, e o *Pseudoplatystoma reticulatum* (Eigenmann e Eigenmann 1889), conhecido como cachara, estão entre os maiores peixes das principais bacias hidrográficas da América do Sul (Figura 1.3). Na natureza, as duas espécies coexistem (vivem em simpatria) apenas na porção média da bacia do rio Paraná, incluindo sub-bacias do rio Paraguai (Buitrago-Suárez & Burr, 2007). A proximidade genética entre *P. corruscans* e *P. reticulatum* tem sido inferida principalmente com base na sua conhecida habilidade de gerar híbridos férteis.



Figura 1.3. *Pseudoplatystoma reticulatum* ou cachara (A) e *Pseudoplatystoma corruscans* ou pintado (B). Fotos: Anderson Luis Alves

De acordo com Benites (2008), a posição taxonômica do pintado é a seguinte:

Classe: Osteichthyes, **Subclasse:** Actinopterygii, **Infraclasse:** Teleostei, **Divisão:** Euteleostei, **Super Ordem:** Ostariophysii, **Ordem:** Siluriformes, **Família:** Pimelodidae,

Espécie: *Pseudoplatystoma corruscans*. A cachara, *Pseudoplatystoma reticulatum*, apresenta a mesma posição taxonômica do pintado.

No ambiente natural, o pintado e a cachara são espécies reofílicas (nadam contra a correnteza dos rios para amadurecer sexualmente), piscívoras, demersais (vivem a maior parte do tempo em contato com o fundo dos rios ou dos ambientes de criação) e de hábitos noturnos. As fêmeas apresentam maior porte e maior ganho de peso quando comparadas aos machos em condições de criação em viveiros, porém os machos atingem a maturação sexual antes das fêmeas. A diferenciação externa entre os sexos do pintado só é possível durante o período reprodutivo, que ocorre no verão, induzido pelo aumento da temperatura da água, maior fotoperíodo, chuvas e aumento do nível dos rios. É nessa época que eles realizam migrações rio acima até a cabeceira dos rios para desovarem (piracema). O período reprodutivo do pintado concentra-se nos meses de dezembro e janeiro, enquanto a cachara está apta para se reproduzir de novembro a março. Esses períodos podem variar em função da região, da bacia hidrográfica e das condições climáticas de cada ano. (Barbosa et al., 2011; Campos, 2010).

3.3 Potencial Aquícola do Pintado e da Cachara

O cultivo dessas espécies em larga escala esteve limitado inicialmente pela viabilização da produção de juvenis e pela ausência de tecnologias adequadas para a engorda dos peixes até o abate/despesca. Com o desenvolvimento de tecnologias para estes fins, o cultivo dessas espécies vem, gradualmente, se expandindo (Campos, 2010).

O pintado e a cachara são considerados importantes espécies nativas para a piscicultura continental devido à sua rusticidade, carne com baixo teor de gordura e ausência de espinhas intramusculares, porte avantajado, alto valor de mercado e importância histórica na produção da pesca das regiões onde ocorrem. O cultivo dessas espécies ainda envolve custos elevados. Apesar de apresentarem características zootécnicas favoráveis, a viabilidade econômica do cultivo dessas espécies depende do condicionamento alimentar dos alevinos ao arraçoamento, na superfície da água, durante o dia (Barbosa et al., 2011; Campos, 2010; Ostrensky et al., 2008).

Conforme observado na Figura 1.2, a produção aquícola dos bagres juntamente com seus híbridos apresentou um crescimento expressivo entre os anos de 2007 e 2011 (mais

de 300%), passando de 4 mil toneladas para mais de 17 mil toneladas. O pintado, a cachara e seus híbridos apresentam uma participação significativa em vários setores do mercado nacional de pescado (Campos, 2010). A produção dessas espécies é realizada em escala industrial em vários estados da região Centro-Oeste, porém o desenvolvimento e refinamento de novas tecnologias, principalmente nas áreas de reprodução, larvicultura, nutrição, identificação e controle de doenças, melhoramento e manejo genético dos plantéis de reprodutores, é necessário para o crescimento sustentável do setor.

Dada a velocidade da expansão da criação de híbridos de pintado/cachara, muitas vezes desordenada e sem os devidos controles, é essencial que sejam desenvolvidas ferramentas adequadas para o manejo e melhoramento genético de estoques de reprodutores de pintado e cachara, e para o monitoramento dos estoques selvagens dessas espécies. O desenvolvimento sustentável da criação dessas espécies em cativeiro depende da implantação de programas de domesticação e de melhoramento genético de linhagens mais adequadas aos sistemas de produção em cativeiro, assim como o monitoramento da dinâmica populacional dos estoques selvagens (Campos, 2010).

3.4 Criação de Híbridos Inter-Específicos: Vantagens, Desvantagens e Riscos Ambientais

A hibridação é um fenômeno bastante comum entre os peixes, podendo ocorrer de maneira natural no ambiente ou pela intervenção humana. A geração de híbridos interespecíficos é uma prática comum em várias partes do mundo. No caso de peixes criados para a produção de carne, a hibridação busca gerar vantagens produtivas, combinando características de duas espécies parentais em um único animal e vem sendo utilizada para aumentar a velocidade de crescimento, melhorar a qualidade da carne, aumentar a resistência às doenças, melhorar a tolerância às variações ambientais, maior taxa de sobrevivência na larvicultura e após a eclosão, menor canibalismo, e uma série de outras características (Bartley et al., 2010; Scribner et al., 2001). Os benefícios citados resultaram na disseminação em larga escala das práticas de hibridação nas pisciculturas do Brasil, onde são produzidos híbridos de várias espécies nativas (Tabela 1.1).

Tabela 1.1 - Híbridos interespecíficos mais comumente utilizados na piscicultura continental brasileira (adaptado de Porto-Foresti et al., 2010)

GERAÇÃO PARENTAL		Produto Híbrido
Parental Fêmea	Parental Macho	
Tambaqui (<i>Colossoma macropomum</i>)	Pacu (<i>Piaractus mesopotamicus</i>)	Tambacu
Pacu (<i>Piaractus mesopotamicus</i>)	Tambaqui (<i>Colossoma macropomum</i>)	Paqui
Tambaqui (<i>Colossoma macropomum</i>)	Pirapitinga (<i>Piaractus brachypomus</i>)	Tambatinga
Pirapitinga (<i>Piaractus brachypomus</i>)	Tambaqui (<i>Colossoma macropomum</i>)	Pirambaqui
Pacu (<i>Piaractus mesopotamicus</i>)	Pirapitinga (<i>Piaractus brachypomus</i>)	Patinga ou Papi
Pirapitinga (<i>Piaractus brachypomus</i>)	Pacu (<i>Piaractus mesopotamicus</i>)	Pirapicu
Piauçu (<i>Leporinus macrocephalus</i>)	Piapara (<i>Leporinus elongatus</i>)	Piaupara
Piapara (<i>Leporinus elongatus</i>)	Piauçu (<i>Leporinus macrocephalus</i>)	Piapaçu
Pintado (<i>Pseudoplatystoma corruscans</i>)	Cachara (<i>Pseudoplatystoma reticulatum</i>)	Pintachara
Cachara (<i>Pseudoplatystoma reticulatum</i>)	Pintado (<i>Pseudoplatystoma corruscans</i>)	Cachapinta
Pintado (<i>Pseudoplatystoma corruscans</i>)	Jurupoca (<i>Hemiosorubim platyrhynchus</i>)	Pintajuru
Pintado (<i>Pseudoplatystoma corruscans</i>)	Pirarara (<i>Phractocephalus hemioliopus</i>)	Pintapira
Cachara (<i>Pseudoplatystoma reticulatum</i>)	Pirarara (<i>Phractocephalus hemioliopus</i>)	Cachapira
Pintado (<i>Pseudoplatystoma corruscans</i>)	Jundiá Amazônico (<i>Leiarius marmoratus</i>)	Pintado Amazônico
Jundiá Amazônico (<i>Leiarius marmoratus</i>)	Pintado (<i>Pseudoplatystoma corruscans</i>)	Jundiara
Matrinxã (<i>Brycon amazonicus</i>)	Piracanjuba (<i>Brycon orbignyanus</i>)	Matrinjuba
Piracanjuba (<i>Brycon orbignyanus</i>)	Matrinxã (<i>Brycon amazonicus</i>)	Piracaxã

Por outro lado, a hibridação apresenta desvantagens e elevados riscos ambientais. O principal problema é a fertilidade dos híbridos interespecíficos, o que representa elevados riscos tanto para as populações selvagens, quanto para os estoques “puros” de reprodutores nas pisciculturas comerciais (essenciais para a produção sustentável de híbridos F1). Os escapes de peixes dos tanques de criação são frequentes (Orsi & Agostinho, 1999) e também representam um sério risco ambiental. Peixes híbridos férteis liberados no ambiente onde existem populações selvagens (cuja barreira reprodutiva é apenas comportamental) podem reproduzir entre si e com indivíduos das espécies puras utilizadas no cruzamento (Toledo-Filho et al., 1998).

A eventual contaminação genética acidental dos estoques de reprodutores “puros” nas pisciculturas (Padhi & Mandal, 1997) também é um risco eminente, já que o comércio de animais vivos normalmente é feito na fase juvenil de crescimento, quando não é possível diferenciar, visualmente, e com precisão, híbridos F1 de animais puros (Prado et al., 2011). Essas situações podem gerar desequilíbrios ecológicos (Einum & Fleming, 1997) e até mesmo levar à extinção das espécies nativas (Allendorf et al., 2001), restringindo, assim, os limites que poderão ser almejados por programas de melhoramento genético para produção das espécies puras em cativeiro (Prado et al., 2011).

3.5 Metodologias de Identificação de Híbridos

Algumas metodologias já foram descritas na literatura para a identificação de peixes híbridos. Em uma revisão, Scribner et al. (2001) verificaram que 45% dos trabalhos de identificação de híbridos de peixes empregaram métodos morfológicos; 35% utilizaram isoenzimas para essa identificação, 12% DNA mitocondrial, 4% DNA nuclear e 2% utilizaram cariótipos.

Hashimoto et al. (2013) e Prado et al. (2011), utilizando metodologias como PCR-RFLP (Reação em Cadeia da Polimerase – Polimorfismo de Comprimento de Fragmento de Restrição) e PCR-*multiplex* (ou o termo mais adequado: PCR espécie-específica), desenvolveram cinco marcadores moleculares para a identificação de híbridos entre *P. corruscans* e *P. reticulatum*. Quatro marcadores são baseados em genes nucleares: RAG2 (Gene de Ativação da Recombinase 2), EF1 α (Fator de Elongação 1 α), GLOB (β -Globina) e 18S (RNA Ribossomal 18S) e um dos marcadores é baseado no gene mitocondrial 16S (Gene Ribossomal 16S). Segundo Hashimoto et al. (2014), marcadores diagnósticos nucleares são importantes para diferenciar híbridos entre espécies. Os marcadores mitocondriais, embora haplóides, são necessários para identificar a origem materna do híbrido (Hashimoto et al., 2014).

PCR-RFLP e PCR-*multiplex* têm sido caracterizados como métodos eficientes para uma rápida e pouco dispendiosa identificação de híbridos (Hashimoto et al., 2013; Prado et al., 2011). Essas metodologias facilitam o diagnóstico baseadas na combinação de variantes de nucleotídeos simples em genes mitocondriais (16S) e nucleares (RAG2, EF1 α , GLOB e 18S) (Hashimoto et al., 2014). No entanto, segundo Boecklen & Howard (1997), quatro a cinco marcadores moleculares são suficientes para discriminar, grosseiramente, espécies puras de híbridos F1 ou indivíduos retrocruzados simples (como RC1). Porém, para a diferenciação entre indivíduos puros e indivíduos com níveis mais avançados de retrocruzamento (RC4, RC 5, etc.), uma quantidade substancialmente maior de marcadores (em torno de 70 marcadores) é necessária.

Os marcadores desenvolvidos por Hashimoto et al. (2013) e Prado et al. (2011) representam um avanço, porém não são adequados para a identificação de menores graus de introgressão genética. Considerando que os híbridos são férteis, a possibilidade de animais serem produzidos naturalmente ou artificialmente, através de cruzamentos entre híbridos e de híbridos com animais puros, é elevada. Esses cruzamentos (F1 x F1, F1 x cachara, F1 x

pintado, etc.) podem produzir peixes com diferentes níveis de introgressão, cuja detecção somente será possível com um painel de marcadores espécie-específicos que identifique blocos fracionados do genoma.

O desenvolvimento de novas tecnologias de seqüenciamento de DNA e de genotipagem de marcadores moleculares do tipo SNP (do inglês *Single Nucleotide Polymorphism*) abriu novos horizontes para o desenvolvimento de ensaios e métodos diagnósticos que poderão permitir a detecção de híbridos/eventos de introgressão em diferentes níveis, com custos baixos e métodos acessíveis ao sistema produtivo.

3.6 Marcadores SNP – Bases Moleculares, Prospecção e Utilização

Os marcadores SNP (Polimorfismos de Base Única, do inglês *Single Nucleotide Polimorphisms*) têm como base as alterações mais simples da molécula de DNA: mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). As mutações mais comuns são as transições, onde ocorrem trocas de uma purina por outra purina ($A \leftrightarrow G$) ou de uma pirimidina por outra pirimidina ($C \leftrightarrow T$). Menos frequentes, as transversões ocorrem quando há troca de uma purina por uma pirimidina, ou vice-versa ($C/T \leftrightarrow A/G$). Normalmente, os marcadores SNP são bi-alélicos, ou seja, são encontradas apenas duas variantes em uma espécie. Os SNPs podem ocorrer em regiões codificadoras ou com função regulatória, porém, na maior parte das vezes são encontrados em espaços intergênicos, sem função determinada (Caetano, 2009; Vignal et al., 2002). Distribuem-se de forma homogênea pelo genoma e são extremamente abundantes nos genomas de espécies não endogâmicas (Bovine Genome Sequencing and Analysis Consortium, 2009; Li et al., 2009).

A utilização de marcadores moleculares SNPs em estudos de associação e mapeamento genético, assim como em ensaios diagnósticos para confirmação de paternidade, identificação individual (rastreadabilidade), detecção de doenças genéticas e/ou polimorfismos associados a características de produção, esteve limitada por muito tempo devido às restrições tecnológicas.

O método padrão de prospecção de marcadores SNP era baseado no método de sequenciamento Sanger, no qual a detecção de SNPs distribuídos aleatoriamente pelo genoma se dava através do alinhamento de uma sequência de um fragmento aleatório do genoma com

uma sequência consenso. Esse processo era limitado pela capacidade de geração e análise de sequências de um laboratório, e pela necessidade da existência de uma sequência referência – o genoma da espécie precisava estar sequenciado. Programas de bioinformática como o PHRED, PHRAP e o CONSED (Ewin et al, 1998; Gordon et al., 1998) foram desenvolvidos para automatizar a fase de análise de dados, permitindo grandes avanços. Apesar das limitações, grandes conjuntos de SNPs foram detectados em todas as espécies que tiveram seus genomas sequenciados utilizando esta metodologia (Bovine Genome Sequencing and Analysis Consortium, 2009; Li et al., 2009).

Porém, as tecnologias de sequenciamento de nova geração alteraram este cenário e facilitaram a descoberta e a genotipagem em massa dos SNPs. Diferentes tecnologias estão amplamente disponíveis para a realização de ensaios de genotipagem de baixa densidade (de dezenas até centenas de marcadores), média densidade (de milhares a dezenas de milhares) e alta densidade (de centenas de milhares a milhões) (Vignal et al., 2002, Caetano, 2009). Dessa forma, as tendências metodológicas atuais para a geração de dados moleculares de genotipagem têm convergido para o uso de Polimorfismos de Base Única (SNPs).

Exemplos práticos da aplicação dessas tecnologias para prospecção de SNPs estão amplamente disponíveis. Um estudo com bovinos (Van Tassell et al., 2008) prospectou um total de 62.042 SNPs utilizando o Illumina/Solexa Genome Analyzer. Alternativas econômicas para prospecção de SNPs em espécies que não possuem genoma sequenciado, como algumas espécies de peixes, também já foram estabelecidas. Uma delas é a prospecção de SNPs no transcriptoma, através de RNA-Seq (Yu et al., 2014; Helyar et al., 2012; Xu et al., 2012; Hubert et al., 2010). Essa metodologia tem a vantagem de também gerar informações descritivas sobre o transcriptoma da espécie em questão, as quais podem ser utilizadas em estudos de genômica funcional, expressão gênica descritiva e aplicada.

3.7 Tecnologias de Sequenciamento de Nova Geração

Avanços nas tecnologias de sequenciamento do DNA (conhecidas como tecnologias de sequenciamento de nova geração), tornaram possível a geração de grandes quantidades de sequências de forma muito rápida e a um custo substancialmente baixo, permitindo o sequenciamento completo do transcriptoma e do genoma das espécies, assim

como o ressequenciamento. As diferentes plataformas existentes apresentam características específicas (Metzker, 2010). Em geral, os métodos de sequenciamento que geram sequências curtas são mais úteis em experimentos de ressequenciamento (para espécies que já possuem genoma referência sequenciado) para descoberta de SNPs (Van Tassel et al., 2008), estudos do transcriptoma (Wang et al., 2009) e estudos de padrões globais de metilação do genoma (Laird, 2010). Adicionalmente, metodologias para sequenciamento e montagem *de novo* de genomas complexos com sequências curtas também já foram desenvolvidas, como no caso do sequenciamento do genoma do panda gigante (Li et al., 2010). A grande restrição para o sequenciamento de genomas complexos com tecnologias que geram sequências mais curtas está nos métodos computacionais necessários para alinhamento e montagem das sequências. Porém, a evolução das ferramentas de bioinformática permitiu a superação destas limitações (Liu et al., 2011). As novas tecnologias de sequenciamento que produzem sequências longas ainda apresentam custos elevados, o que limita o sequenciamento a um número restrito de genomas de uma espécie. Enquanto é possível ter uma montagem do genoma humano com 95% de cobertura a partir de um total de 8X de cobertura de sequenciamento com sequências com comprimento médio de 400pb, são necessários 35X de cobertura de sequenciamento para obter uma montagem da mesma cobertura com sequências de 75pb (Metzker, 2010).

3.8 Sequenciamento do Transcriptoma Completo

O transcriptoma é um conjunto completo de moléculas de RNA (transcritos). Não é estável como o genoma, varia de acordo com o ambiente externo, condição fisiológica ou estágio de desenvolvimento do organismo (Wang et al., 2009). A análise do transcriptoma permite entender melhor como funciona um organismo. A análise do transcriptoma é uma importante ferramenta para compreender a relação entre o genótipo e o fenótipo, permitindo conhecer melhor as vias metabólicas e os mecanismos envolvidos nos processos biológicos de um organismo.

As tecnologias de sequenciamento de nova geração têm possibilitado essa análise, que ficou conhecida como sequenciamento do transcriptoma completo, sequenciamento de RNA ou simplesmente, RNA-Seq. No RNA-Seq, DNAs complementares (cDNAs) gerados a partir do RNA de interesse são diretamente sequenciados usando NGS. Utilizando algoritmos de bioinformática, os fragmentos obtidos podem ser alinhados a um

genoma referência ou montados *de novo*, de forma a permitir a caracterização do transcriptoma completo de uma espécie (Zmienko et al., 2011; Shen et al., 2012).

Como ele não necessariamente requer conhecimento prévio do genoma, pode ser aplicado à espécies que ainda não possuem genoma sequenciado. E permite não apenas a análise ou caracterização do transcriptoma, mas também a descoberta de polimorfismos SNPs e de novos transcritos, genes, estudos de expressão gênica diferencial; identificação da complexidade funcional dos transcriptomas; *splicing* alternativo; RNAs não codificadores, etc. (Zmienko et al.; 2011, Shen et al., 2012; Salem et al., 2012; Wang et al., 2009; Nagalakshmi et al., 2010).

O sequenciamento das extremidades de moléculas de cDNA produzidas a partir de RNA mensageiro extraído de diferentes tecidos de uma espécie para a geração de ESTs (do inglês *Expressed Sequence Tags*) foi uma estratégia desenvolvida no final dos anos 90 para a caracterização do transcriptoma (veja Gerhold & Caskey, 1996). O sequenciamento das extremidades 5' e 3' dos cDNAs permite a identificação parcial dos genes transcritos de origem e também a identificação de polimorfismos úteis para estudos genômicos (Novaes et al., 2008).

RNA-Seq vem sendo aplicado em diversos estudos com diferentes espécies de peixes: a) mapeamento do transcriptoma e anotação gênica (*Poecilia reticulata*, Fraser et al., 2011; *Danio rerio*, Collins et al., 2012; *Ictalurus punctatus*, Liu et al., 2012; *Onchorhynchus mykiss*, Salem et al., 2015); b) descoberta de novos transcritos (*Danio rerio*, Pauli et al., 2012; *Onchorhynchus mykiss*, Palstra et al., 2013; *Salmo salar*, Kure et al., 2013); c) descoberta de SNPs (*Onchorhynchus mykiss*, Salem et al., 2012; *Cyprinus carpio*, Xu et al., 2012); d) quantificação de expressão gênica (*Danio rerio*, Vesterlund et al., 2011; *Salmo trutta*, Uren Webster et al., 2013) etc.

3.9 Genes tecido específico e *housekeeping*

Uma variedade de genes está ativa dentro do núcleo de nossas células. Alguns são necessários para a manutenção diária das funções celulares, enquanto outros têm papéis mais específicos para certos tecidos ou tipos celulares. Conseqüentemente, cada tecido possui seu próprio perfil de atividade gênica. Como genes ativos produzem RNA, as diferenças de tecido na atividade dos genes podem ser pesquisadas pela caracterização do RNA que eles

contêm. O conjunto completo de RNAs ou transcriptoma é sequenciado de vários tecidos e esses dados são usados para comparar o grau de especialização de diferentes tecidos e investigar o conjunto de genes ativos em cada tecido.

A expressão de grande parte dos genes varia de acordo com os tipos celulares e de tecidos e de acordo com os estágios fisiológicos e de desenvolvimento de um organismo. No entanto, alguns genes são constitutivamente expressos em todos os tecidos e seus níveis de expressão são comparativamente constantes entre os diferentes tipos celulares. Estes genes são chamados de genes *housekeeping* e a hipótese é que eles constituem um pequeno conjunto de genes necessários para manter as funções básicas celulares (Butte et al., 2001; She et al., 2009). A expressão desses genes (tecido-específicos) é fundamental para manter a especificidade e determinar a complexidade dos organismos multicelulares, a forma como eles afetam o desenvolvimento, a função e a manutenção dos diferentes tipos celulares de um organismo (Salem et al., 2015).

Genes tecido-específico são genes predominantemente expressos em um único tecido biologicamente relevante (Chang et al., 2011). A expressão dos genes tecido-específicos é fundamental para manter a especificidade dos tipos celulares/tecido em um organismo. O estudo da expressão de genes *housekeeping* x genes tecido-específicos permite entender melhor o funcionamento de um organismo, sua complexidade e evolução. Além disso, a identificação de genes *housekeeping* e genes tecido-específicos fornecem informações moleculares importantes que poderão ser aplicadas no melhoramento genético dos peixes (Salem et al., 2015; Xiao et al., 2010).

3.10 Músculo Branco e Músculo vermelho em Peixes

Os peixes constituem um modelo interessante para o estudo dos processos biológicos, regulatórios e metabólicos envolvidos na locomoção e no crescimento, devido à características peculiares observadas no seu tecido muscular esquelético. De acordo com Carani et al. (2008); Santos et al. (2007), a maior parte da massa corporal dos peixes é constituída pelo tecido muscular (aproximadamente 60%), sendo a proporção deste tecido bem maior nos peixes do que em outros vertebrados, devido em parte à grande demanda proveniente do sistema locomotor, por causa da densidade do meio onde vivem. Diferente dos

mamíferos, cujo tecido muscular apresenta-se em forma de mosaico de fibras, nos peixes, o tecido muscular é compartimentalizado em diferentes tipos de fibras musculares:

- a) a musculatura branca compreende a maior parte da massa muscular (60%), sendo formada por fibras musculares maiores, de contração rápida e metabolismo glicolítico. Possui baixa concentração de mioglobina, poucas mitocôndrias e poucos lipídeos. Essas fibras são utilizadas durante a realização de movimentos mais bruscos de natação;
- b) a musculatura vermelha (compreende no máximo 30% do volume da massa muscular total e está restrita à região da linha lateral), localizada na região superficial adjacente ao tegumento, é formada por fibras menores, de contração lenta e metabolismo oxidativo. Apresenta alta concentração de mioglobina, mitocôndrias e lipídeos e excelente suprimento sanguíneo, sendo utilizada na realização de movimentos lentos e de sustentação;
- c) fibras rosas ou intermediárias.

O crescimento pós-natal do tecido muscular esquelético dos peixes envolve dois mecanismos:

- Hipertrofia – células mio-satélites se fundem a fibras musculares existentes, aumentando o número de núcleos para maior síntese de miofibrilas.
- Hiperplasia – ocorre a formação de novas fibras musculares (mosaico de fibras com diâmetros diferentes). Esse padrão de organização é melhor observado na musculatura branca.

Nos mamíferos a hiperplasia cessa em um curto período de tempo após o nascimento. Nos peixes de tamanho pequeno, o crescimento pós-natal ocorre principalmente por hipertrofia; a hiperplasia ocorre nas fases iniciais do desenvolvimento. Em peixes de crescimento indeterminado com tamanho maior, a hiperplasia se estende por um período maior. Normalmente, ocorre maior esforço nas fibras vermelhas do que nas brancas. Esse tipo de esforço específico parece influenciar os tipos e as quantidades de proteínas sintetizadas localmente. (Carani et al. 2008; Santos et al., 2007).

4 REFERÊNCIAS BIBLIOGRÁFICAS

- ALLENDORF, F.W.; LEARY, R.F.; SPRUELL, P. et al. The problem with hybrids: setting conservation guidelines. **Trends in Ecology and Evolution**, v. 16, p.613-622, 2001.
- BARBOSA, O.N.; RAIZER, J.; GONDA, M.F. et al. Desempenho e coeficiente intestinal de alevinos puros e híbridos de pintados em condicionamento alimentar. **Revista Brasileira de Zootecnia**, v. 40, n. 12, p.2621-2627, 2011.
- BARTLEY, D.M.; RANA, K.; IMMINK, A.J. The use of inter-specific hybrids in aquaculture fisheries. **Reviews in Fish Biology and Fisheries**, v. 10, p.325-337, 2001.
- BENITES, C. **Caracterização genética do pintado *Pseudoplatystoma corruscans* (Spix & Agassiz, 1829) (Siluriformes: Pimelodidae) da Bacia Paraná-Paraguai, por marcadores do tipo microssatélite**. 2008. 88 p. Tese (Doutorado) – Universidade Estadual Paulista, Centro de Aquicultura da UNESP, 2008.
- BOECKLEN, W.J.; HOWARD, D.J. Genetic analysis of hybrid zones: numbers of markers and power of resolution. **Ecology**, v. 78, n. 8, p.2611-2616, 1997.
- BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA. Brasil 2008 e 2009. Ministério da Pesca e Aquicultura. 99p., 2009.
- BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA. Brasil 2010. Ministério da Pesca e Aquicultura. 128p., 2010.
- BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA. Brasil 2011. Ministério da Pesca e Aquicultura. 60p., 2011.
- BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v. 324 n. 5926, p.522-528, 2009.
- BUITRAGO-SUÁREZ, U.A.; BURR, B.M. Taxonomy of the catfish genus *Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. **Zootaxa**, v. 1512, p.1-38, 2007.

- BURGESS, W.E. An atlas of freshwater and marine catfishes: a preliminary survey of the Siluriformes. **T.F.H. Publication**, Neptune City, 784p., 1989.
- BUTTE, A.J.; DZAU, V.J.; GLUECK, S.B. Further defining housekeeping, or “maintenance,” genes Focus on “A compendium of gene expression in normal human tissues”. **Physiol Genomics**, v. 7, p.95-96, 2001.
- CAETANO, A.R. SNP markers: basic concepts, applications in animal breeding and management and perspectives for the future. **Revista Brasileira de Zootecnia**, v. 38, p.64-71, 2009.
- CAMARGO, S.G.O. de; POUHEY, J.L.O.F. Aquicultura – Um mercado em expansão. **Revista Brasileira de Agrociência**, v. 11, n. 4, p.393-396, 2005.
- CAMPOS, J.L. O cultivo do pintado (*Pseudoplatystoma corruscans*, Spix; Agassiz, 1829), outras espécies do gênero *Pseudoplatystoma* e seus híbridos. p. 335-361. In: BALDISSEROTTO, B.; GOMES, L. de C. (Orgs.). **Espécies nativas para piscicultura no Brasil**. 2ª edição. Santa Maria, Rio Grande do Sul: Ed. da UFSM. 2010. 606p.
- CARANI, F.R.; AGUIAR, D.H.; ALMEIDA, F.L.A. de. et al. Morfologia e crescimento do músculo estriado esquelético no pirarucu *Arapaima gigas* Cuvier, 1817 (Teleostei, Arapaimidae). **Acta Scientiarum. Biological Sciences**, v. 30, n. 2, p.205-211, 2008.
- CARLSON, V. (Coord.). 1º Anuário Brasileiro da Pesca e Aquicultura, 2014. 133p., 2014.
- CARVALHO-COSTA, L.F.; PIORSKI, N.M.; WILLIS, S.C. et al. Molecular systematics of the neotropical shovelnose catfish genus *Pseudoplatystoma* Bleeker 1862 based on nuclear and mtDNA markers. **Molecular Phylogenetics and Evolution**, v. 59, n. 1, p.177-194, 2011.
- CHANG, C.W.; CHENG, W.C.; CHEN, C.R. et al. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. **Plos One**, v. 6, n. 7, p.1-10, 2011.
- COLLINS, J.E.; WHITE, S.; SEARLE, S.M.J. et al. Incorporating RNA-Seq data into zebrafish Ensembl genebuild. **Genome Research**, v. 22, p.2067-2078, 2012.
- EINUM, S.; FLEMING, I.A. Genetic divergence and interactions in the wild among native farmed and hybrid Atlantic salmon. **Journal of Fish Biology**, v. 50, p.634-651, 1997.
- EWING, B.; HILLIER, L.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v. 8, n. 3, p.175-85, 1998.
- FAO. The State of World Fisheries and Aquaculture 2014. Rome, 2014. 233p.
- FERREIRA, E.J.G.; ZUANON, J.A.S.; SANTOS, G.M. **Peixes comerciais do médio Amazonas: Região de Santarém, Pará**. IBAMA, Brasília. 1998.

- FRASER, B.A.; WEADICK, C.J.; JANOWITZ, I. et al. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. **BMC Genomics**, v. 12, n. 202, p.1-14, 2011.
- GERHOLD, D.; CASKEY, C.T. It's the genes! EST access to human genome content. **Bioessays**, v. 18, n. 12, p.973-81, 1996.
- HASHIMOTO, D.T.; PRADO, F.D. DO; SENHORINI, J.A.; FORESTI, F.; PORTO-FORESTI, F. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. **Aquaculture Research**, .v. 44, p.876-884, 2013.
- HASHIMOTO, D.T.; SENHORINI, J.A.; FORESTI, F. et al. Genetic Identification of F1 and Post-F1 Serrasalmid juvenile hybrids in Brazilian Aquaculture. **Plos One**, v.9, n. 3, p.1-8, 2014.
- HELYAR, S.J.; LIMBORG, M.T.; BEKKEVOLD, B. et al. SNP discovery using next generation transcriptomic sequencing in Atlantic Herring (*Clupea harengus*). **Plos One**, v. 7, n. 8, p.1-11, 2012.
- HUBERT, S.; HIGGINS, B.; BORZA, T.; BOWMAN, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). **BMC Genomics**, v. 11, n. 191, p.1-14, 2010.
- KURE, E.H.; SAEBO, M.; STANGELAND, A.M. et al. Molecular responses to toxicological stressor: Profiling microRNAs in wild Atlantic salmon (*Salmo salar*) exposed to acidic aluminum-rich water. **Aquatic Toxicology**, v. 138-139, p.98-104, 2013.
- LAIRD, P.W. Principles and challenges of genome-wide DNA methylation analysis. **Nature Review Genetics**, v. 11, n. 3, p.191-203, 2010.
- LI, R.; FAN, W.; TIAN, G.; et al. The sequence and de novo assembly of the giant panda genome. **Nature**, v. 463, n. 7279, p.311-7, 2010.
- LI, R.; YU, C.; LI, Y. et al. SOAP2: an improved ultrafast tool for short read alignment. **Bioinformatics**. v. 25, n. 15, p.1966-1967, 2009.
- LIU, S.; ZHOU, Z., LU, J. et al. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. **BMC Genomics**. v. 12, n. 53, p.1-13, 2011.
- LIU, S.; ZHANG, Y.; ZHOU, Z. et al. Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. **BMC Genomics**, v. 14, n. 1, p.1-18, 2012.
- LUNDBERG, J.G.; LITTMANN, M.W. Family Pimelodidae (Long-whiskered catfishes). In: REIS, R.; KULLANDER, S.O.; FERRARIS, C.J.Jr. (Eds.). **Check list of the freshwater fishes of South and Central America**. Edipucrs: Porto Alegre, 2003. p.432-455.

- METZKER M.L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, p.31-46, 2010.
- NAGALAKSHMI, U.; WAERN, K.; SNYDER, M. et al. RNA-Seq: A method for comprehensive transcriptome analysis. **Current Protocols in Molecular Biology**, 2010.
- NOVAES, E.; DROST, D.R.; FARMERIE, W.G.; et al. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. **BMC Genomics**, v. 9, n. 312, p.1-14, 2008.
- ORSI, L.M.; AGOSTINHO, A.A. Introdução de espécies de peixes por escapes acidentais de tanques de cultivo em rios da Bacia do Rio Paraná, Brasil. **Revista Brasileira de Zoologia**, v. 16, p.557-560, 1999.
- OSTRENSKY, A.; BORGHETTI, J.R.; SOTO, D. **Aquicultura no Brasil. O desafio é crescer**. Secretaria Especial de Aquicultura e Pesca/FAO, Brasília, DF, 2008. 276 p.
- PADHI, B.K.; MANDAL, R.K. Inadvertent hybridization in a carp hatchery detected by nuclear DNA RFLP. **Journal of Fish Biology**, v. 50, p.906-909, 1997.
- PALSTRA, A.P.; BELTRAN, S.; BURGERHOUT, E. et al. Deep RNA sequencing of the skeletal muscle transcriptome in swimming fish. **Plos One**, v. 8, n. 1, p. 1-14, 2013.
- PAULI, A.; VALEN, E.; LIN, M.F. et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. **Genome Research**, v. 22, p.577-591, 2012.
- PORTO-FORESTI, F.; HASHIMOTO, D.T.; ALVES, A.L. et al. Cytogenetic markers as diagnoses in the identification of the hybrid between Piauçu (*Leporinus microcephalus*) and Piapara (*Leporinus elongatus*). **Genetics and Molecular Biology**, v. 31, n. 1, p.195-202, 2008.
- PORTO-FORESTI, F.; HASHIMOTO, D.T.; SENHORINI, J.A. et al. Híbridaç o em piscicultura: Monitoramento e perspectivas. p. 589-606. In: BALDISSEROTTO, B.; GOMES, L. de C. (Orgs.). **Esp cies nativas para piscicultura no Brasil**. 2^a ediç o. Santa Maria, Rio Grande do Sul: Ed. da UFSM, 2010. 606p.
- PRADO, F.D.; HASHIMOTO, D.T.; MENDONÇA, F.F. et al. Molecular identification of hybrids between Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum*. **Aquaculture Research**, p.1-5, 2011.
- PRADO, F.D. do; HASHIMOTO, D;T.; SENHORINI, J.A. et al. Detection of hybrids and genetic introgression in wild stocks of two catfish species (Siluriformes: Pimelodidae): The impact of hatcheries in Brazil. **Fisheries Research**, v. 125-126, p.300-305, 2012a.
- PRADO, F.D. do; NUNES, T.L.; SENHORINI, J.A. et al. Cytogenetic characterization of F1, F2 and backcross hybrids of the Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Pimelodidae, Siluriformes). **Genetics and Molecular Biology**, v. 35, n. 1, p. 57-64, 2012b.

- REIS, R.E.; KULLANDER, S.O.; FERRARIS, C.J. **Check list of freshwater fishes of South and Central America**. Porto Alegre, RS: Edipucrs. 2004. 729p.
- SALEM, M.; VALLEJO, R.L.; LEEDS, T.D. et al. RNA-Seq identifies snp markers for growth traits in rainbow trout. **Plos One**. v. 7, n. 5, p 1-13, 2012.
- SALEM, M.; PANERU, B.; AL-TOBASEI, R. et al. Transcriptome assembly, geneannotation and tissue gene expression atlas of the rainbow trout. **Plos One**. p.1-27, 2015.
- SANTOS, V.B. dos. Aspectos morfológicos da musculatura lateral dos peixes. **Boletim do Instituto de Pesca**, v. 33, n. 1, p.127-135, 2007.
- SCRIBNER, K.T.; PAGE, K.S.; BARTRON, M.L. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. **Reviews in Fish Biology and Fisheries**, v. 10, p.293-323, 2001.
- SHE, X.; ROHL, C.A.; CASTLE, J.C. et al. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. **BMC Genomics**, v. 10, n. 269, p.1-12, 2009.
- SHEN, Y.; CATCHEN, J.; GARCIA, T.; AMORES, A.; BELDORTH, I.; WAGNER, J.; ZHANG, Z.; POSTLETHWAIT, J.; WARREN, W.; SCHARTL, M.; WALTER, R.B. Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F₁ interspecies hybrids. **Comparative Biochemistry Physiology, Part C**. v. 155, p.102-108, 2012.
- SMITH, E.M.; GREGORY, T.R. Patterns of genome size diversity in the ray-finned fishes. **Hydrobiologia**, v. 625, p.1-25, 2009.
- TOLEDO-FILHO, S.A.; ALMEIDA-TOLEDO, L.F.; FORESTI, F. et al. Programas genéticos de seleção, hibridação e endocruzamento aplicados à Piscicultura. Cadernos de Ictiogenética n° 4. Departamento de Biologia- CCS/USP, 56 P.Z. 1998,
- TORRICO, J. P.; HUBERT, N.; DESMARAIS, E. et al. Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): biogeographic and evolutionary implications. **Molecular Phylogenetics and Evolution**, v. 51, n. 3, p.588-594, 2009.
- UREN WEBSTER, T.M.; BURY, N.R.; VAN AERLE, R. et al. Global transcriptome profiling reveals molecular mechanisms of metal tolerance in a chronically exposed wild population of brown trout. **Environmental Science & Technology**, v. 47, p.8869-8877, 2013.
- VAN TASSELL, C.P.; SMITH, T.P.; MATUKUMALLI, L.K.; et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature Methods**. v.5, n.3, 2008. p.247-52.
- VESTERLUND, L.; JIAO, H.; UNNEBERG, P. et al. The zebrafish transcriptome during early development. **BMC Developmental Biology**, v.11, n. 30, 2011.

- VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M. et al. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 34, p. 275-305, 2002.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. The Review provides a comprehensive overview of recent advances and challenges in techniques that are used in transcriptome profiling methods that use NGS technologies (RNA-seq). **Nature Reviews Genetic**, v. 10, p.57–63, 2009.
- XIAO, S.J.; ZHANG, C.; ZOU, Q. et al. TiSGeD: a database for tissue-specific genes. **Bioinformatics applications note**, v. 26, n. 9, p. 1273-1275, 2010.
- XU, J.; JI, P.; ZHAO, Z. et al. Genome-wide SNP discovery from transcriptome of four common carp strains. **PloS One**, v. 7, n. 10, p.1-9, 2012.
- YU, Y.; WEI, J.; ZHANG, X. et al. SNP discovery in the transcriptome of white pacific shrimp *Litopenaeus vannamei* by next generation sequencing. **PloS One**. v. 9, n.1, p.1-9, 2014.
- ZMIENKO, A.; JACKOWIAK, P.; FIGLEROWICZ, M. Transcriptome sequencing: next generation approach to RNA functional analysis. **Journal of Biotechnology, Computational Biology and Bionanotechnology**, v. 92, n. 4,0 p.311-319, 2011.

CAPÍTULO 2

DESENVOLVIMENTO DE UM ENSAIO DE MINISEQUENCIAMENTO PARA IDENTIFICAÇÃO DE HÍBRIDOS ENTRE CACHARA (*Pseudoplatystoma reticulatum*) E PINTADO (*Pseudoplatystoma corruscans*)

1 RESUMO

Cachara (*P. reticulatum*) e pintado (*P. corruscans*) são bagres conhecidos pela capacidade de gerar híbridos férteis, arriscando a manutenção de estoques puros de reprodutores nas pisciculturas e a existência dos estoques selvagens. Um ensaio SNaPshot® foi desenvolvido para interrogar SNPs espécie-específicos de *P. reticulatum* e *P. corruscans*, localizados em quatro genes nucleares (RAG2, GLOB, 18S e EF1 α) e em dois genes mitocondriais (16S e COI). O ensaio foi desenvolvido com base em informações previamente publicadas sobre marcadores utilizados para diferenciar indivíduos puros de híbridos das espécies *P. reticulatum* e *P. corruscans*. Onze posições polimórficas, consideradas espécie-específicas com base no alinhamento das sequências dos genes disponibilizadas no *GenBank* foram inicialmente selecionadas para comporem o ensaio, que investigou 26 exemplares morfologicamente identificados como 17 cacharas e nove pintados: posições 271 (C/T) e 356 (T/C) do gene 16S; 243 (A/C) e 409 (G/A) do COI; 243 (A/G) e 254 (C/A) do RAG2; 484 (T/A) e 587 (C/T) do EF1 α ; 300 (A/G) e 374 (A/T) do GLOB e posição 185 (T/C) do gene 18S. Resultados obtidos após a otimização do ensaio permitiram concluir que dois dos genes indicados na literatura como marcadores espécie-específicos não são adequados para a diferenciação dos híbridos: o gene mitocondrial 16S e o gene nuclear GLOB. Ambos foram

retirados do ensaio. O método SNaPshot© adaptado às informações disponíveis no *GenBank* permitiu a identificação de 13 híbridos entre os 26 indivíduos coletados. Todos os pintados coletados foram identificados como híbridos. Os resultados foram confirmados por sequenciamento dos seis fragmentos investigados. O ensaio SNaPshot© foi adequado para a diferenciação de híbridos F1 e pós-F1 das espécies *P. reticulatum* e *P. corruscans*, porém existe a necessidade de desenvolvimento de ferramentas que permitam a identificação de níveis menores de introgressão entre estas espécies.

Palavras-chave: Minisequenciamento; Hibridação interespecífica; Sanger.

**DEVELOPMENT OF A MINISEQUENCING ASSAY TO IDENTIFY HYBRIDS
BETWEEN CACHARA (*Pseudoplatystoma reticulatum*) AND PINTADO
(*Pseudoplatystoma corruscans*)**

2 ABSTRACT (CORRIGIR)

We carried out a SNaPshot assay to interrogate nine specie-specific SNPs of *P. reticulatum* and *P. corruscans* localized in four nuclear genes (RAG₂, GLOB, 18S e EF1 α) and in one mitochondrial gene (16S). This assay was performed based on previously published information about the used tracers for differentiate pure and hybrid individuals of the species *P. reticulatum* and *P. corruscans* (Hashimoto et al., 2013; Prado et al., 2011). Nine polymorphic positions considered specie-specific based on the gene sequence alignment available in the *GenBank* were picked to compose this assay: gene 16S position 271 (C/T) and 356 (T/C); RAG₂ position 243 (A/G) and 254 (C/A); EF1 α position 484 (T/A) and 587 (C/T); GLOB position 300 (A/G) and 374 (A/T) and 18S position 185 (T/C). The results obtained after the assay optimization allowed us to find that two of the indicated genes as specie-specific tracers in the review, were not suitable for hybrid differentiation: mitochondrial gene 16S and nuclear gene GLOB. In the SNaPshot minisequencing carried out in this thesis, the gene 16S were replaced with mitochondrial gene COI. We selected to interrogate the polymorphic positions 243 (A/C) and 409 (G/A) in the COI's sequence. The results of the GLOB genotyping were discarded and the gene were removed from the assay. The SNaPshot method adapted to *GenBank* available information allowed us to identify 13 hybrid amongst the 26 captured individuals (morphologically indicated as 17 cacharas and new pintados) that would be used to develop the other experiments idealized for this PhD thesis. All of the captured pintados were identified as hybrids. This result changed the initial idea of the

experiments. In face of the SNaPshot minisequencing results, the only alternative was to use two groups of fishes in this thesis: a group of pure cachara and other group of pintado hybrids. The minisequencing proved to be useful to differentiate hybrids of the species *P. reticulatum* and *P. corruscans*, but the use of more advanced and accurate tools to differentiate it still necessary.

Keywords: SNP; SNaPshot; Catfish; Sequencing; COI; RAG2; 16S Ribosomal; GLOB; 18S Ribosomal.

3 INTRODUÇÃO

A cachara (*Pseudoplatystoma reticulatum*) e o pintado (*Pseudoplatystoma corruscans*), bagres Neotropicais de água doce, pertencentes à família Pimelodidae (Siluriformes), são espécies relevantes para a aquicultura brasileira. De acordo com Barbosa et al. (2011), no ambiente natural, estas espécies são reofílicas, piscívoras, demersais e de hábitos noturnos. Apesar de apresentarem características zootécnicas favoráveis, a viabilidade econômica de seu cultivo depende do condicionamento alimentar dos alevinos ao arraçoamento, na superfície da água, durante o dia.

Uma prática muito comum entre os criadores de peixes é a hibridação interespecífica (Porto-Foresti et al., 2008). Popularmente conhecidos como “ponto e vírgula”, “pintachara” ou “cachapinta”, os híbridos resultantes do cruzamento entre cachara e pintado adaptam-se com maior facilidade ao manejo no ambiente de cativeiro (Crepaldi et al., 2006), além de apresentarem desejáveis características produtivas superiores aos peixes “puros”, devido à heterose (Bartley et al., 2001). Por outro lado, a hibridação apresenta aspectos negativos para o desenvolvimento sustentável da cachara e do pintado: os híbridos gerados são férteis (Prado et al., 2012); e não podem ser morfologicamente diferenciados dos “puros” na fase juvenil, época em que são comercializados vivos (Hashimoto et al., 2013). Os aspectos citados representam riscos genéticos tanto para os estoques “puros” das pisciculturas, quanto para os estoques naturais de peixes (Porto-Foresti et al., 2008), em decorrência da possibilidade de cruzamentos entre peixes “puros” e híbridos (Prado et al., 2011).

Este cenário impõe sérias dificuldades para o desenvolvimento de programas de conservação e de melhoramento genético “clássico” para essas espécies, uma vez que esses programas necessitam de reprodutores “puros” (que não apresentem introgressão de material genético da outra espécie), seja para a preservação das espécies ou para o desenvolvimento dos processos de avaliação genética das características produtivas. Outro agravante, é que a

produção de híbridos interespecíficos já está amplamente disseminada pelo País (Porto-Foresti et al., 2008). De acordo com Hashimoto et al. (2013), a maioria dos “pintados” juvenis comercializada atualmente no Brasil é híbrida.

Portanto, o monitoramento genético dos estoques de reprodutores “puros” nas pisciculturas é imprescindível. Prado et al. (2011) e Hashimoto et al. (2013) desenvolveram um método que permite a identificação de híbridos inter-específicos F1 e pós-F1 utilizando marcadores moleculares PCR-RFLP e PCR-Multiplex (PCR espécie-específica) nos genes nucleares RAG₂ (Gene Ativador da Recombinase), GLOB (β -Globina), 18S (RNA Ribossomal 18S) e EF1 α (Fator de Elongamento 1 α) e no gene mitocondrial 16S (RNA Ribossomal 16S), baseados em variações de polimorfismos únicos (SNPs). Embora seja uma técnica que não exija uso de equipamentos avançados, é trabalhosa e permite genotipar apenas um loco por ensaio.

Os SNPs (Polimorfismos de Base Única) são marcadores que têm como princípio mutações em bases únicas da cadeia de DNA. Recentes avanços tecnológicos revelaram metodologias de alto desempenho e acurácia, e baixo custo para a prospecção, caracterização e genotipagem de marcadores SNPs, que vêm resultando em novas soluções para aplicações já solidificadas (Caetano, 2009). Um exemplo é o minisequenciamento SNaPshot[®], um método eficiente, rápido e de baixo custo, que permite investigar até 10 SNPs em uma única reação e que tem aceitabilidade na comunidade científica.

Neste estudo, é relatado o desenvolvimento de um ensaio SNaPshot[®] para interrogar SNPs espécie-específicos de *Pseudoplatystoma corruscans* e *Pseudoplatystoma reticulatum* publicamente disponíveis, localizados em 4 genes nucleares (RAG₂, GLOB, 18S e EF1 α) e em dois genes mitocondriais (16S e COI), com o objetivo de identificar animais híbridos no n amostral coletado (n=26).

4 MATERIAL E MÉTODOS

O ensaio SNaPshot foi idealizado para interrogar 9 SNPs espécie-específicos de *P. reticulatum* e *P. corruscans*, localizados em quatro genes nucleares (RAG₂, GLOB, 18S e EF1 α) e em dois genes mitocondriais (16S e COI). Foi desenvolvido com base em informações previamente publicadas sobre marcadores utilizados para diferenciar indivíduos puros de híbridos das espécies *P. reticulatum* e *P. corruscans* (Hashimoto et al., 2013; Prado et al., 2011). O método SNaPshot foi adaptado à essas informações disponíveis de forma a possibilitar a identificação de indivíduos híbridos dentre os 26 indivíduos (morfologicamente identificados como 17 cacharas e 9 pintados) coletados.

4.1 Amostragem e Extração de DNA

A execução desse estudo foi pré-aprovada pela Comissão de Ética no Uso de Animais da Embrapa Recursos Genéticos e Biotecnologia (CEUA/Cenargen, protocolo CEUA-002-2014). As coletas dos peixes foram realizadas em duas viagens diferentes. Dificuldades encontradas nas coletas como necessidade de coletar os peixes em pisciculturas e não na natureza, devido à falta de equipamentos como barcos para a pesca; quantidade insuficiente de peixes adultos para o sacrifício disponibilizadas pelos piscicultores; ausência de recurso financeiro e tempo para a realização de novas viagens para coletas adicionais, acabaram por restringir o “n” amostral para o desenvolvimento deste estudo.

Amostras de nadadeira de 26 exemplares (Tabela 2.1) morfologicamente identificados como 17 cacharas (*P. reticulatum*) e nove pintados (*P. corruscans*), provenientes de duas pisciculturas, porém originalmente capturados em diferentes localidades

da Bacia dos Rios Paraguai (rios Cuiabazinho, Nobres; São Gonçalo, Cuiabá; Paraguai, Parque Nacional do Pantanal e Corumbá) e Paraná (rio Ivinhema), Brasil, foram coletadas. DNA foi extraído das amostras com “*Pure Link Genomic DNA Mini Kit*” (Invitrogen), seguindo o protocolo do fabricante e foi quantificado em gel de agarose 1%, corado com brometo de etídeo.

Tabela 2.1. Informações sobre os peixes coletados

Identificação	Espécie	Nome popular	Sexo	Região da coleta
1	<i>P. reticulatum</i>	Cachara	M	MT
2	<i>P. reticulatum</i>	Cachara	M	MT
3	<i>P. reticulatum</i>	Cachara	F	MT
4	<i>P. reticulatum</i>	Cachara	F	MT
5	<i>P. reticulatum</i>	Cachara	F	MT
6	<i>P. reticulatum</i>	Cachara	F	MT
7	<i>P. reticulatum</i>	Cachara	F	MT
8	<i>P. reticulatum</i>	Cachara	M	MS
9	<i>P. reticulatum</i>	Cachara	M	MS
10	<i>P. reticulatum</i>	Cachara	F	MS
11	<i>P. reticulatum</i>	Cachara	M	MS
12	<i>P. reticulatum</i>	Cachara	M	MS
13	<i>P. reticulatum</i>	Cachara	F	MS
14	<i>P. reticulatum</i>	Cachara	F	MS
15	<i>P. reticulatum</i>	Cachara	F	MS
16	<i>P. reticulatum</i>	Cachara	M	MS
17	<i>P. reticulatum</i>	Cachara	F	MS
18	<i>P. corruscans</i>	Pintado	F	MS
19	<i>P. corruscans</i>	Pintado	M	MS
20	<i>P. corruscans</i>	Pintado	M	MS
21	<i>P. corruscans</i>	Pintado	F	MS
22	<i>P. corruscans</i>	Pintado	F	MS
23	<i>P. corruscans</i>	Pintado	F	MS
24	<i>P. corruscans</i>	Pintado	M	MS
25	<i>P. corruscans</i>	Pintado	F	MS
26	<i>P. corruscans</i>	Pintado	F	MS

4.2 Identificação de SNPs Espécie-Específicos em Genes Nucleares e Mitocondriais de *P. reticulatum* e *P. corruscans* e Desenho de *Primers* *SnaPshot*© Específicos para cada SNP Selecionado

Sequências dos genes nucleares RAG₂, EF1 α , GLOB e 18S e dos genes mitocondriais 16S e COI de *P. reticulatum* e de *P. corruscans* disponíveis no *GenBank* foram alinhadas entre si utilizando a ferramenta *ClustalW Multiple Alignment*, do *software* Bioedit (Hall, 1999), para a localização de SNPs espécie-específicos entre as duas espécies. O

software *Leo's primers selector for SNaPshot*© (<http://sgdp.iop.kcl.ac.uk/leo/cgi-bin/snpshot.cgi>) foi utilizado para desenhar *primers* SNaPshot© específicos para os SNPs selecionados.

4.3 Amplificação dos Fragmentos por PCR Individual

Os pares de *primers* universais descritos na Tabela 2.2 foram utilizados para amplificar os fragmentos dos genes nucleares EF1 α (743 pb), GLOB (567 pb em *P. corruscans* e 569 pb em *P. reticulatum*), 18S (350 pb) e RAG₂ (521 pb) e dos genes mitocondriais 16S (551 pb) e COI (707 pb), por PCR, de forma individual, em cinco reações de PCR com volume final de 10 μ L cada, contendo: 1,5 μ L de DNA genômico (30 ng); 0,5 μ L de dNTP (2,5 mM, Promega); 0,30 μ L de MgCl₂ (25 mM, Promega); 1,0 μ L de Taq DNA Buffer 10x (20 mM Tris-HCl, pH 8.4 e 50 mM KCl, Promega); 0,3 μ L de Taq Polymerase (Phoneutria); 0,2 μ L do *primer* F (10 μ M) e 0,2 μ L do *primer* R (10 μ M). As seguintes condições de PCR foram utilizadas: 94°C (5 minutos), seguido por 35 ciclos de 94°C (1 minuto), 50°C (1 minuto) e 72°C (1 minuto) e uma etapa final de 72°C (7 minutos). Para o gene COI, foi utilizado o protocolo de amplificação do *kit* de PCR da Qiagen, para uma reação de PCR com volume final de 10 μ L: 1,5 μ L de DNA genômico (20 ng); 5,0 μ L de PCRmix; 0,1 μ L do primer COI F (10 μ M); 0,1 μ L do primer COI R (10 μ M); 1,0 μ L de QSolution. As seguintes condições de PCR foram utilizadas: 95°C (15 minutos), seguido por 35 ciclos de 94°C (30 segundos), 50°C (1 minuto) e 72°C (1 minuto) e uma etapa final de 72°C (30 minutos). Os produtos de PCR foram analisados por eletroforese, em gel de agarose 1%, corado com brometo de etídeo, e visualizados em um transiluminador UV, Eagle Eye.

Tabela 2.2. Pares de *primers* universais (F e R) utilizados para amplificação dos fragmentos dos genes nucleares EF1 α , GLOB, 18S, RAG2 e dos genes mitocondriais 16S e COI em *P. reticulatum* e *P. corruscans*

<i>Primer</i> Universal	Sequência (5'-3')	Tm	Comprimento final (pb)	Referências
16S F	ACGCCTGTTTATCAAAAACAT	54,96	21	Kocher et al. (1989)
16S R	CCGGTCTGAACTCAGATCACGT	62,34	22	
*COI F	TCAACCAACCACAAAGACATTGGCAC	68,52	26	Ward et al. (2005)
*COI R	TAGACTTCTGGGTGGCCAAAGAATCA**	66,57	26	
RAG₂ Silu F	CCTGAGTGCTACCTTATTCATGGA	60,31	24	Hashimoto et al. (2013)
RAG₂ Silu R	CTTGGGAGGAAGAGACCATC	57,12	20	
EF1α F	ATTGGAAGTGTACCTGTGG-	51,57	19	Moyer et al. (2004)
EF1α R	CAGCCTTCTGTGCAGACTT	55,01	19	

GLOB Silu F	TCAATATGGTTCACTGGACAGA	56,55	22	Hashimoto et al. (2013)
GLOB Silu R	CCAAGAAGCTGAAAGTAGACAGT	55,62	23	
18S NS1	GTAGTCATATGCTTGTCTC	42,97	19	White et al. (1990)
18S Silu R	CCATCGAAAAGTTGATAGGG	55,68	20	Hashimoto et al. (2013)

*Os primers universais COI F e COI R foram utilizados para amplificar o gene mitocondrial COI (Citocromo Oxidase C Sub-unidade I), nos ensaios SNaPshot, substituindo o gene mitocondrial 16S.

**A sequência do primer COI R apresentada na Tabela 2.2 difere em 1 base da sequência apresentada por Hashimoto et al. (2011): 5-TAGACTTCTGGGTGGCCAAAGGAATCA-3' (27 pb).

4.4 Minisequenciamento SNaPshot® para Interrogação de SNPs Espécie-Específicos por Fragmento Individual

Fragmentos dos 4 genes nucleares e dos dois genes mitocondriais amplificados por PCR individual foram purificados com a adição de 2,0 µL de um mix de ExoSAP-IT (0,4 µL de ExoSAP-IT e 1,6 µL de H₂O milliQ) em 2,0 µL do produto final da PCR, seguida de incubação no termociclador por 45 minutos à 37°C, mais 15 minutos à 80°C. Esta purificação é necessária para eliminar dNTPs e *primers* não consumidos pela reação de PCR.

Seis reações de extensão dos *primers* SNaPshot® foram realizadas, uma para cada fragmento amplificado, em um volume final de 10 µL. Cada reação continha: 2,0 µL do produto de PCR individual purificado com ExoSAP-IT; 6,0 µL do *mix* contendo 0,6 µL *SNaPshot Multiplex Mix* (Applied Biosystems) mais 5,4 µL H₂O; 1 µL do *mix* de *primer* SNaPshot do fragmento analisado, conforme descrito abaixo:

Gene	Primers	Volume Final
Mix 16S	0,78 µL <i>Primer</i> 16S_271F + 0,8 µL <i>Primer</i> 16S_356R + 8,42 µL H ₂ O	10 µL
Mix COI	0,2 µL <i>Primer</i> COI_243R + 0,2 µL <i>Primer</i> COI_409R + 9,6 µL H ₂ O	10 µL
Mix RAG₂	0,78 µL <i>Primer</i> RAG ₂ _243F + 0,8 µL <i>Primer</i> RAG ₂ _254R + 8,42 µL H ₂ O	10 µL
Mix EF1α	0,96 µL <i>Primer</i> EF1α_484R + 1,0 µL <i>Primer</i> EF1α_587F + 8,04 µL H ₂ O	10 µL
Mix GLOB	0,4 µL <i>Primer</i> GLOB_300F + 0,9 µL <i>Primer</i> GLOB_374F + 8,7 µL H ₂ O	10 µL
Mix 18S	0,4 µL <i>Primer</i> 18S_185F + 9,6 µL H ₂ O	10 µL

Cada reação foi submetida à 25 ciclos de 96°C (10 segundos), 50°C (5 segundos) e 60°C (30 segundos). Após a extensão dos primers, ddNTPs não incorporadas foram removidas das reações de SNaPshot através do tratamento com a enzima SAP (Fosfatase Alcalina de Camarão), nas seguintes condições: 0,4 µL de SAP e 1,6 µL de H₂O foram adicionados aos produtos das reações de SNaPshot, seguido por incubação de 1 hora no termociclador, à 37°C, mais 15 minutos à 75°C.

Em seguida, 1,5 µL do produto da reação de SNaPshot® purificado foi combinado com 8,1 µL de formamida HiDi e 0,4 µL de marcador de peso molecular GeneScan120 LIZ (Applied Biosystems), e incubado na temperatura de 95°C durante 5 minutos. Eletroforese foi realizada em um sequenciador automatizado (ABI3100 - Applied Biosystems), e os eletroferogramas foram analisados com o *software* GeneMapper (Applied Biosystems), para a genotipagem dos SNPs.

4.5 Validação do Minisequenciamento SNaPshot® Individual por Sequenciamento

A acurácia da extensão dos *primers* SNaPshot® foi verificada pelo sequenciamento direto dos produtos de PCR amplificados com os *primers* universais em 157 indivíduos (morfologicamente identificados como 100 cacharas e 57 pintados). As reações de sequenciamento foram realizadas com a química do terminador *BigDye* (Applied Biosystems) utilizando 20 ng dos produtos de PCR purificados com ExoSAP-IT (Amersham) seguindo as recomendações do fabricante. Eletroforese de capilaridade foi realizada no sequenciador automático ABI 3100 e as sequências foram analisadas e alinhadas utilizando Phred/Phrap/Consed (Ewing & Green, 1998; Ewing et al., 1998). O sequenciamento inicialmente foi realizado para as 17 cacharas e os nove pintados coletados.

4.6 Amplificação dos Fragmentos por PCR Multiplex

Pares de *primers* universais (Tabela 2.2) foram utilizados para amplificar os fragmentos dos genes nucleares EF1α (743 pb), GLOB (567 pb em *P. corruscans* e 569 pb em *P. reticulatum*), 18S (350 pb) e RAG₂ (521 pb) e do gene mitocondrial COI (707 pb) por PCR multiplex, utilizando o *kit* de PCR Multiplex da Qiagen. A reação de PCR com volume final de 10 µL continha: 1,0 µL de DNA genômico (20 ng); 5,0 µL de PCRMix; 0,18 µL do primer COI F (10 µM); 0,18 µL do primer COI R (10 µM); 0,14 µL do primer RAG₂ Silu F (10 µM); 0,14 µL do primer RAG₂ Silu R (10 µM); 0,12 µL do primer EF1α F (10 µM); 0,12 µL do primer EF1α R (10 µM); 0,09 µL do primer GLOB Silu F (10 µM); 0,09 µL do primer GLOB Silu R (10 µM); 0,05 µL do primer 18S NS1 (10 µM); 0,05 µL do primer 18S Silu R (10

μM); 1,0 μL de QSolution. As seguintes condições de PCR foram utilizadas: 95°C (15 minutos), seguido por 35 ciclos de 94°C (30 segundos), 50°C (1 minuto) e 72°C (1 minuto) e uma etapa final de 72°C (30 minutos). Os produtos de PCR foram analisados por eletroforese, em gel de agarose 1%, corado com brometo de etídeo, e visualizados em um transiluminador UV, Eagle Eye.

4.7 Minisequenciamento SNaPshot® para Interrogação dos Nove SNPs Espécie-Específicos Selecionados

Uma reação de extensão dos *primers* SNaPshot foi realizada para os fragmentos dos 4 genes nucleares (RAG2, EF1 α , GLOB e 18S) e do gene mitocondrial (COI) amplificados por PCR *multiplex* e purificados com ExoSAP-IT (conforme descrito no item 4.4 deste Material e Métodos), em um volume final de 10 μL , contendo: 2,0 μL do produto de PCR multiplex purificado com ExoSAP-IT; 6,0 μL do mix contendo 3,5 μL SNaPshot® Multiplex Mix (Applied Biosystems) mais 2,5 μL de H₂O MilliQ; 19,9 μL do mix de *primers* SNaPshot constituído por:

<i>Primer</i>
1,0 μL <i>Primer</i> COI_243R
1,0 μL <i>Primer</i> COI_409R
1,6 μL <i>Primer</i> RAG ₂ _243F
2,0 μL <i>Primer</i> RAG ₂ _254R
1,5 μL <i>Primer</i> EF1 α _484R
1,8 μL <i>Primer</i> EF1 α _587F
4,0 μL <i>Primer</i> GLOB_300F
3,5 μL <i>Primer</i> GLOB_374F
3,5 μL <i>Primer</i> 18S_185F
Total – 19,9 μL de mix de <i>primers</i> SNaPshot

A reação foi submetida à extensão dos *primers*. Posteriormente, ddNTPs não incorporadas foram removidas da reação e a eletroforese foi realizada em um sequenciador automatizado (ABI3100 - Applied Biosystems). Os eletroferogramas gerados foram analisados com o *software* GeneMapper (Applied Biosystems).

4.8 Minisequenciamento SNaPshot® para Interrogação de Cinco SNPs Espécie-Específicos

Uma reação de extensão dos *primers* SNaPshot foi realizada para os fragmentos dos 4 genes nucleares (RAG2, EF1 α , GLOB e 18S) e do gene mitocondrial (COI) amplificados em *multiplex*, em um volume final de 10 μ L, contendo: 2,0 μ L do produto de PCR multiplex purificado com ExoSAP-IT; 6,0 μ L do mix contendo 2,5 μ L SNaPshot Multiplex Mix (Applied Biosystems) mais 3,5 μ L de H₂O MilliQ; 10,75 μ L do mix de *primers* SNaPshot constituído por:

<i>Primer</i>
0,65 μ L <i>Primer</i> COI_409R
1,8 μ L <i>Primer</i> RAG ₂ _254R
1,2 μ L <i>Primer</i> EF1 α _484R
4,6 μ L <i>Primer</i> GLOB_300F
2,5 μ L <i>Primer</i> 18S_185F
Total – 10,75 μL de mix de <i>primers</i> SNaPshot

A reação foi submetida à extensão dos *primers*. Posteriormente, ddNTPs não incorporadas foram removidas da reação e a eletroforese foi realizada em um sequenciador automatizado (ABI3100 - Applied Biosystems). Os eletroferogramas gerados foram analisados com o *software* GeneMapper (Applied Biosystems).

5 RESULTADOS E DISCUSSÃO

5.1 Identificação de SNPs Espécie-Específicos em Genes Nucleares e Mitocondriais de *P. reticulatum* e *P. corruscans* e Desenho dos Primers SNaPshot

Com o objetivo de localizar SNPs espécie-específicos que pudessem compor um ensaio de minisequenciamento SNaPshot para identificar indivíduos puros e híbridos das espécies *P. reticulatum* e *P. corruscans*, ao todo foram selecionadas no *GenBank* 68 sequências, sendo duas sequências provenientes do gene RAG₂ (uma para cada espécie), dez sequências do gene EF1 α (cinco para cada espécie), dez sequências do gene GLOB (cinco para cada espécie), dez sequências do gene 18S (cinco para cada espécie), duas sequências do gene 16S (uma para cada espécie) e 44 sequências do gene COI (10 sequências para *P. reticulatum* e 34 sequências para *P. corruscans*) (Tabela 2.3).

Tabela 2.3. Informações sobre as sequências dos genes nucleares (RAG₂, EF1 α , GLOB e 18S) e mitocondriais (16S e COI) disponíveis no *GenBank*.

Gene	Nome do Gene	Nº acesso GenBank	Espécie	Tam (pb)	Autor	
16S	RNA ribossomal 16S	HM107837	<i>P. corruscans</i>	551	Prado et al., 2011	
		HM107839	<i>P. reticulatum</i>			
COI	Citocromo C Oxidase Subunidade I	JX462936 à JX462945	<i>P. reticulatum</i>	552	Carvalho et al., 2012	
		GU701553	<i>P. corruscans</i>	652	Pereira et al., 2013	
		GU701555		652	Pereira et al., 2013	
		GU701558		652	Pereira et al., 2013	
		JN989174		702	Pereira et al., 2013	
		HM405206		652	Carvalho et al., 2011b	
		HQ600840		586	Carvalho et al., 2011b	
		HQ600841		591	Carvalho et al., 2011b	
		HQ600842		600	Carvalho et al., 2011b	
		JX462911 à JX462935		552	Carvalho et al., 2012	
		HQ689335		529	Carvalho et al., 2011a	
		RAG ₂		Proteína ativadora	HM107838	<i>P. corruscans</i>

	da recombinação 2	HM107836	<i>P. reticulatum</i>		
EF1α	Fator de alongação 1 α	JF834090	<i>P. corruscans</i>	743	Hashimoto et al., 2013
		JF834091			
		JF834092			
		JF834093			
		JF834094			
		JF834095	<i>P. reticulatum</i>		
		JF834096			
		JF834097			
		JF834098			
		JF834099			
GLOB	Hemoglobina β	JF834080	<i>P. corruscans</i>	567	Hashimoto et al., 2013
		JF834081			
		JF834082			
		JF834083			
		JF834084			
		JF834085	<i>P. reticulatum</i>		
		JF834086			
		JF834087			
		JF834088			
		JF834089			
18S	RNA ribossomal 18S	JF834100	<i>P. corruscans</i>	350	Hashimoto et al., 2013
		JF834101			
		JF834102			
		JF834103			
		JF834104			
		JF834105	<i>P. reticulatum</i>		
		JF834106			
		JF834107			
		JF834108			
		JF834109			

Verificou-se que as sequências cujos números de acesso variam entre JF834100 e JF834109, e apresentam 350 pb, foram depositadas no *GenBank* por Hashimoto et al. (2013) como sequências do gene 16S. Segundo Prado et al. (2011), as sequências do gene 16S de *P. reticulatum* e de *P. corruscans* possuem 551 pb (e não 350 pb) e segundo Hashimoto et al. (2013), as sequências do gene 18S (e não as do gene 16S) de *P. reticulatum* e de *P. corruscans* possuem 350 pb. Foi realizada uma pesquisa BLAST com estas sequências (JF834100 à JF834109) e elas apresentaram elevada homologia com o gene 18S de outras espécies. Essas evidências indicam que as sequências JF834100 e JF834109 foram erroneamente depositadas no GeneBank. Na Tabela 2.3, essas sequências (JF834100 à JF834109) são apresentadas corretamente, como sequências pertencentes ao gene 18S. As sequências HM107837 e HM107839 são as sequências corretas referentes ao gene 16S.

Na Figura 2.1, é mostrado, como exemplo, o resultado do alinhamento das sequências do gene 18S de *P. reticulatum* e de *P. corruscans* disponíveis no *GenBank*. Como pode ser observado, foram identificados SNPs espécie-específicos em duas posições da

sequência de 350 pb do gene 18S: 185 (genótipo CC para *P. corruscans* e TT para *P. reticulatum*) e 187 (genótipo CC para *P. corruscans* e GG para *P. reticulatum*).



Figura 2.1. Alinhamento das 10 sequências do gene 18S de *P. reticulatum* e *P. corruscans* para identificação de SNP espécie-específicos.

O mesmo procedimento foi realizado para os outros cinco genes (RAG₂, EF1 α , GLOB, COI e 16S). Foram identificados, no total, 45 SNPs espécie-específicos: 5 SNPs no gene 16S; 21 SNPs no gene mitocondrial COI; 3 SNPs no gene RAG₂; 4 SNPs no EF1 α ; 10 SNPs no GLOB e 2 SNPs no 18S. Os resultados podem ser visualizados na Tabela 2.4.

Tabela 2.4. Gene, tamanho, posição e genótipo dos SNPs espécie-específicos identificados nas sequências dos genes RAG₂, GLOB, 18S, EF1 α , COI e 16S nas espécies *P. reticulatum* e *P. corruscans*

Gene	Tamanho (pb)	Posição SNP	Genótipo <i>P. reticulatum</i>	Genótipo <i>P. corruscans</i>
16S	551	267	T	C
		271**	C	T
		356**	T	C
		398	T	C
		422	C	T
COI	707	145	C	T
		190	G	A
		208	G	A
		247**	A	C
		283	C	T
		292	A	G
		295	T	G
		327	G	A
		363	T	C
		366	T	C
369	T	C		

		372	T	C
		375	C	T
		390	A	G
		415**	C	T
		438	A	G
		447	G	A
		457	C	T
		499	A	G
		542	A	G
		563	T	C
RAG₂	521	243**	A	G
		254**	C	A
		285	G	C
EF1α	743	206	T	C
		484**	A	T
		587**	T	C
		630	A	G
GLOB	567/569	23	C	A
		33	C	G
		285	A	G
		300**	G	A
		245	A	G
		374**	T	A
		455	C	A
		456	C	A
		457	-	A
458	-	A		
18S	350	185**	C	T
		187	C	G

**SNPs selecionados para compor o minisequenciamento SNaPshot.

Onze SNPs espécie-específicos foram inicialmente selecionados dentre os 45 SNPs identificados nos alinhamentos, com base nos polimorfismos localizados nos marcadores moleculares PCR-RFLP e PCR-multiplex (ou PCR-espécie-específicos) publicamente disponíveis (Prado et al., 2011; Hashimoto et al., 2013), para serem interrogados no minisequenciamento SNaPshot®. Devido à possibilidade de ocorrência de recombinação genética aleatória entre as cromátides irmãs durante a meiose, ao longo do cromossomo, foram selecionados dois SNPs em cada gene para compor o ensaio SNaPshot®. No caso do gene 18S, como o alinhamento das suas sequências revelou a existência de apenas dois SNPs, em posições muito próximas (posição 185 e 187 da sequência), apenas um SNP foi selecionado para esse gene, totalizando assim, onze SNPs espécie-específicos. Esses SNPs encontram-se destacados em negrito na Tabela 2.4.

O princípio do minisequenciamento SNaPshot® é a extensão de um *primer* que se anela na posição adjacente ao sítio do SNP de interesse (neste caso, o SNP espécie-específico selecionado). Um terminador ddNTP (dideoxinucleotídeo) fluorescentemente marcado, correspondente ao alelo do SNP é incorporado, inibindo a continuidade da síntese

da nova fita, devido à ausência de um grupo hidroxil no carbono 3', essencial para que a reação de incorporação de uma nova base nitrogenada ocorra. Os *primers* SNaPshot ou *primers* de extensão de uma única base (SBE), que são diferentes em comprimento (devido a uma cauda (dGACT)_n que é adicionada à sua extremidade 5') e fluorescência, são submetidos à eletroforese por capilaridade e analisados no software *GeneMapper*. Os alelos dos SNPs são diferenciados pelo tamanho e pela cor detectados na eletroforese (Geppert & Roewer, 2012; Alves et al., 2007).

Os *primers* SNaPshot desenhados para cada um dos onze SNPs espécie-específicos selecionados são apresentados na Tabela 2.5.

Tabela 2.5. *Primers* SBE utilizados no minisequenciamento SNaPshot® para identificação de híbridos entre cachara (*P. reticulatum*) e pintado (*P. corruscans*) (T: tamanho; T_m: temperatura de *melting*)

Gene	Primer	Sequência (5'-3')	T (pb)	T _m (°C)
16S	16S_271F	TTTTTTTTTTCACCTATGTCAAGGACCAYAAA	31	53,1
	16S_356R	TTTTTTTTTTTTTTTTTCTTGGTTTTGGGGTTGTT	35	54,1
COI	COI_243R	TTTTTTTTTTTTGTGGGAATGCTATGTCTGG	31	53,77
	COI_409R	TTTTTTTTTTTGCYAGATGTAGGGAGAAAATAGTTA	35	54,39
RAG ₂	RAG ₂ _243F	GCCTGCTTCTGAGCG	15	54,71
	RAG ₂ _254R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCACCATGCAGTTCCAG	47	54,6
EF1 α	EF1 α _484R	TTTTTACAATGGCAGCATCTCC	23	54,11
	EF1 α _587F	TTTTTTTTTTTTTTTTTTTTTTCGTACACATGCCACAACACT	39	53,98
GLOB	GLOB_300F	TTTTTRGGAAACCCCAAGGT	19	53,4
	GLOB_374F	TTTTTTTTTTTTTTTTTTTTTTTTTTTCAACGTTAAGGGCACCT	43	53,8
18S	18S_185F	TTTTTTTTTTTGACCCTCCTCTCCCTC	27	54,4

5.2 Amplificação dos Fragmentos por PCR Individual

O SNaPshot® é uma metodologia que funciona em *multiplex*, permitindo interrogar até 10 SNPs em uma única reação. No entanto, inicialmente, os *primers* SNaPshot desenhados para o ensaio foram testados de forma individual, com o intuito de verificar se haviam sido desenhados corretamente. Para isso, reações de PCR individuais utilizando os *primers* universais descritos na Tabela 2.2 e as condições detalhadas no Material e Métodos deste Capítulo foram otimizadas para cada um dos cinco genes estudados neste ensaio (16S, RAG₂, EF1 α , GLOB e 18S). Os fragmentos amplificados podem ser visualizados nas Figuras 2.2 e 2.3.

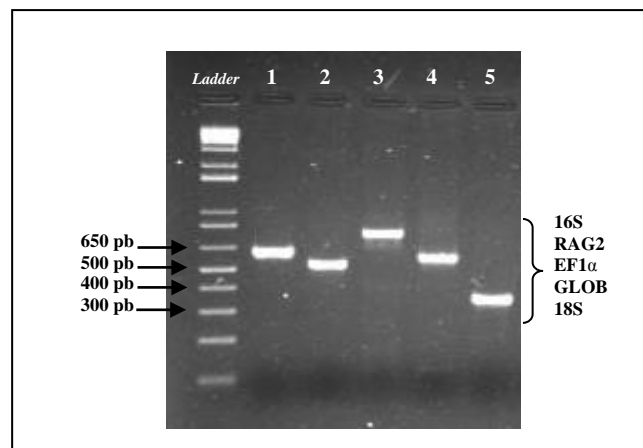


Figura 2.2. Gel de agarose contendo os cinco fragmentos amplificados através de PCR individual: gene 16S (551 pb, poço 1), RAG₂ (521 pb, poço 2), EF1 α (743 pb, poço 3), GLOB (569/567 pb, poço 4) e 18S (350 pb – poço 5). *Ladder* (Marcador de DNA 1 Kb Plus).

5.3 Minisequenciamento SNaPshot para Interrogação de SNPs Espécie-Específicos por Fragmento Individual

Uma análise geral dos picos observados nos eletroferogramas gerados a partir do sequenciamento das reações individuais de extensão dos *primers* SNaPshot dos seis genes estudados neste ensaio revelou que a extensão dos *primers* ocorreu, indicando, portanto, que eles foram desenhados corretamente. Na sequência, o resultado da genotipagem obtida com o ensaio individual foi analisado para cada gene.

5.3.1 Gene mitocondrial 16S

A análise dos eletroferogramas gerados a partir do ensaio SNaPshot[®] para o gene 16S (Figura 2.3) revelou genótipos idênticos para todos os indivíduos morfologicamente identificados como cacharas e pintados amostrados nos SNPs 271 e 356 do gene mitocondrial 16S: C (pico preto no eletroferograma) e A (pico verde no eletroferograma), respectivamente. Os genótipos obtidos eram esperados apenas para a espécie *P. reticulatum* (cachara). Esperava-se que os indivíduos *P. corruscans* (pintado) apresentassem o genótipo T para o SNP 271 e o genótipo G para o SNP 356.

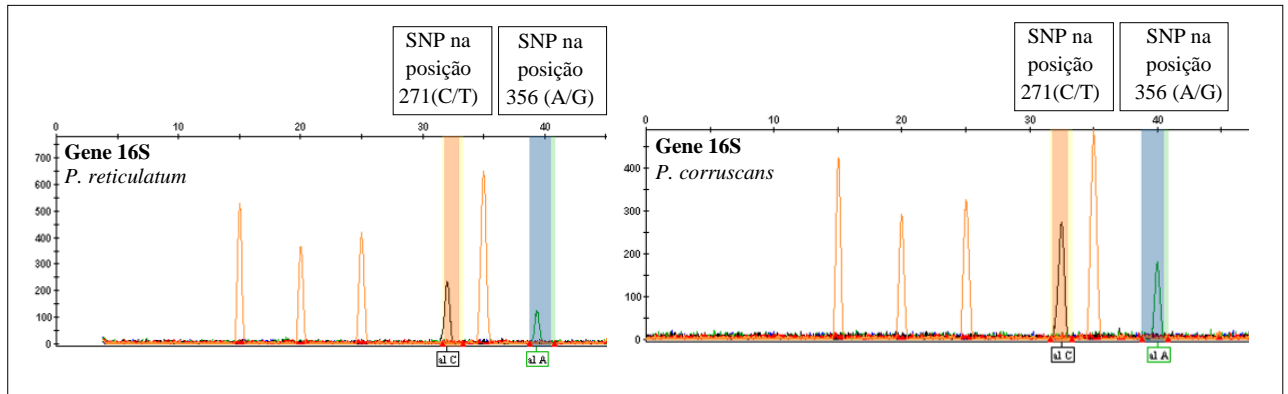


Figura 2.3. Eletroferogramas ilustrando os produtos da reação de SNaPshot dos SNPs 271 (C/T) e 356 (A/G) do gene mitocondrial 16S de *P. reticulatum* e de *P. corruscans*. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50 pb, respectivamente. Pico preto: SNP na posição 271 e pico verde: SNP na posição 356.

Esse resultado foi validado através do sequenciamento do fragmento do gene mitocondrial 16S em 157 animais morfologicamente identificados como 100 cacharas e 48 pintados. As sequências obtidas também revelaram que todos os indivíduos sequenciados possuíam o mesmo genótipo para ambos os SNPs (271 e 356) do gene mitocondrial 16S, confirmando os resultados obtidos com o minisequenciamento SNaPshot®. Sendo assim, os polimorfismos esperados nas posições 271 e 356 do gene 16S não foram observados nos animais coletados. Todos os animais apresentaram o genótipo esperado para a cachara.

Na Figura 2.4 é possível visualizar dois eletroferogramas do gene 16S, o primeiro para *P. reticulatum* (cachara) e o segundo para *P. corruscans* (pintado), evidenciando o resultado para o SNP 271: mesma sequência apresentada pelas duas espécies. O mesmo padrão foi observado para o SNP 356 em ambas as espécies.

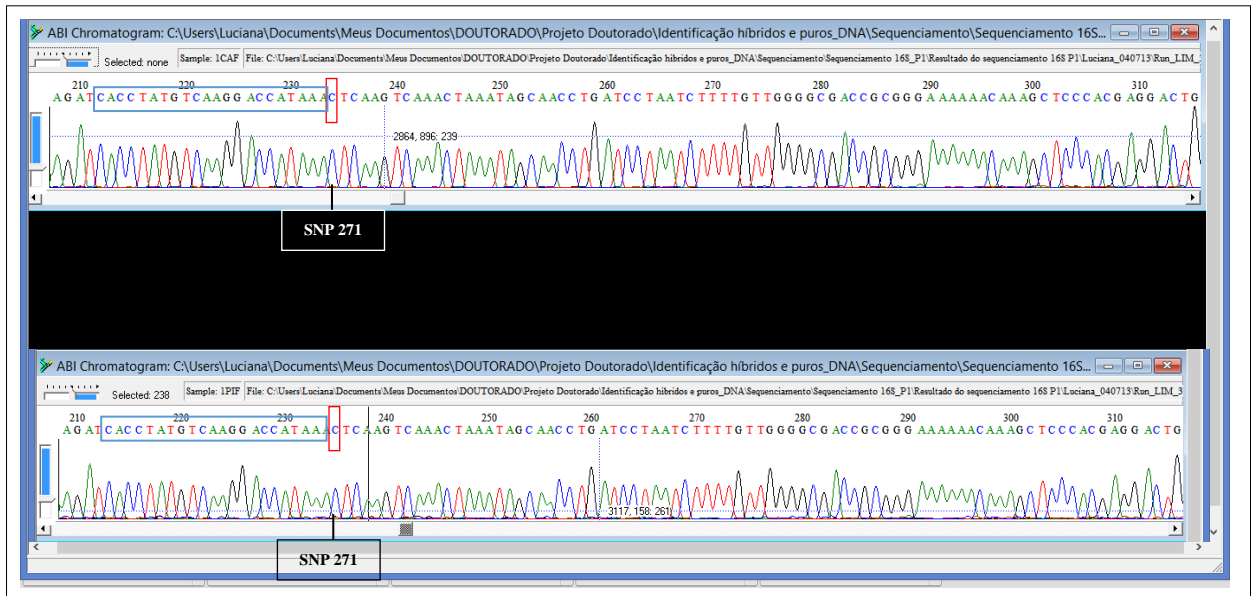


Figura 2.4. Eletroferogramas gerados pelo sequenciamento do gene 16S de *P. reticulatum* e *P. corruscans*, evidenciando a sequência do *primer* SNaPshot 16S_271F (caixa azul na figura) e logo em seguida, na posição 271 (caixa vermelha na figura), o genótipo C do SNP esperado apenas para *P. reticulatum*, mas observado para as duas espécies.

Os resultados obtidos tanto no minisequenciamento quanto no sequenciamento do gene 16S obtidos indicam que todos os pintados (*P. corruscans*) coletados e amostrados neste ensaio possuem mitocôndria de cachara, sendo, conseqüentemente, produtos de cruzamentos interspecíficos, ou seja, são híbridos de pintado. Ou alternativamente, que os SNPs do gene 16S não podem ser utilizados para a identificação das duas espécies, conforme indicado por Hashimoto et al. (2013). Conseqüentemente, o gene 16S foi retirado do ensaio SNaPshot®.

5.3.2. Gene mitocondrial COI

Eletroferogramas gerados a partir do minisequenciamento do gene COI podem ser observados na Figura 2.5. Genótipos idênticos também foram obtidos para todos os 26 indivíduos morfologicamente identificados como cacharas e pintados amostrados nos SNPs 243 e 409 do gene mitocondrial COI: T (pico vermelho no eletroferograma) e G (pico azul), respectivamente. Os genótipos obtidos eram esperados apenas para a cachara. Esperava-se que o pintado apresentasse o genótipo G para o SNP 243 e A para o SNP 409. O resultado confirma que todos os pintados possuem mitocôndria de cachara e são, portanto, híbridos (Tabela 2.6).

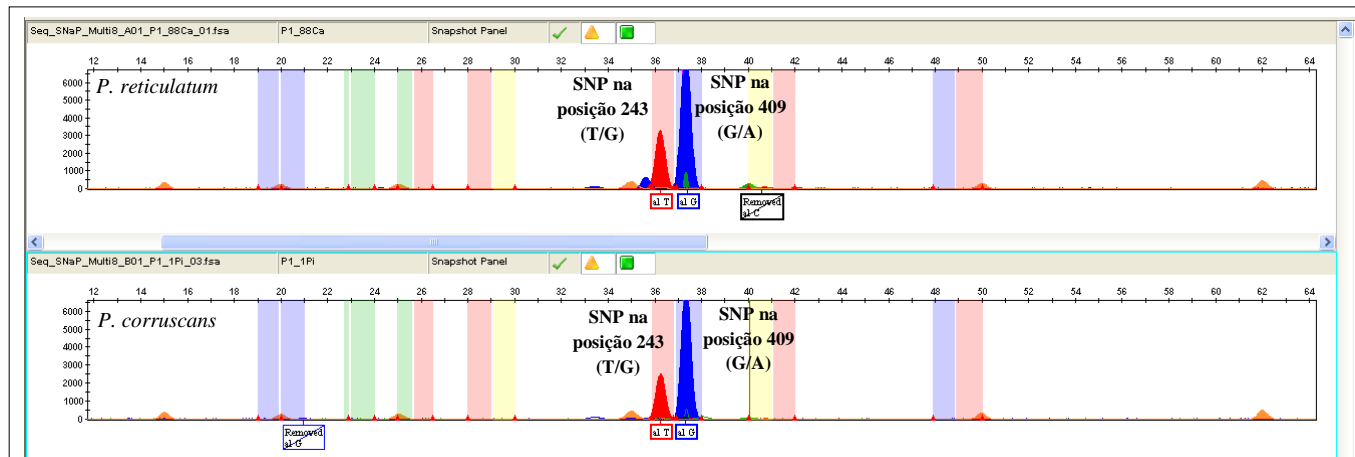


Figura 2.5. Eletroferogramas ilustrando os produtos da reação de SNaPshot dos SNPs 243 (T/G) e 409 (G/A) do gene mitocondrial COI de *P. reticulatum* e de *P. corruscans*. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50 pb, respectivamente. Pico vermelho: SNP na posição 243 e pico azul: SNP na posição 409.

A análise e o alinhamento das sequências geradas para o COI revelou que elas também são semelhantes para as duas espécies, confirmando os resultados obtidos com o minsequenciamento. Concluiu-se, portanto, que todos os 26 animais coletados (morfologicamente identificados como 17 cacharas e 9 pintados) possuem mtDNA de cachara. Isso significa que todos os nove pintados coletados são híbridos e, que pela análise do gene mitocondrial COI, as cacharas são puras.

Os demais 131 animais sequenciados (morfologicamente identificados como 83 cacharas e 48 pintados) apresentaram genótipos de animais puros, ou seja, as cacharas apresentaram os genótipos T e G para os SNPs 243 e 409, respectivamente, enquanto os pintados apresentaram os genótipos G e A. Esse resultado confirma que os SNPs interrogados no ensaio SNaPshop para o gene COI são adequados para diferenciar as duas espécies (Tabela A1, do Anexo A).

5.3.3 Gene nuclear EF1 α

A análise dos SNPs 484 e 587 do gene nuclear EF1 α revelou três possíveis genótipos dentre os 26 animais coletados (Figura 2.6): TT/TT (*P. corruscans* ou pintado), AA/CC (*P. reticulatum* ou cachara) e AT/CT (híbrido). No ensaio SNaPshot com o gene EF1 α foram identificados seis animais híbridos dentre os nove pintados (*P. corruscans*)

coletados, e um híbrido dentre as 16 cacharas coletadas, totalizando sete animais híbridos identificados neste ensaio.

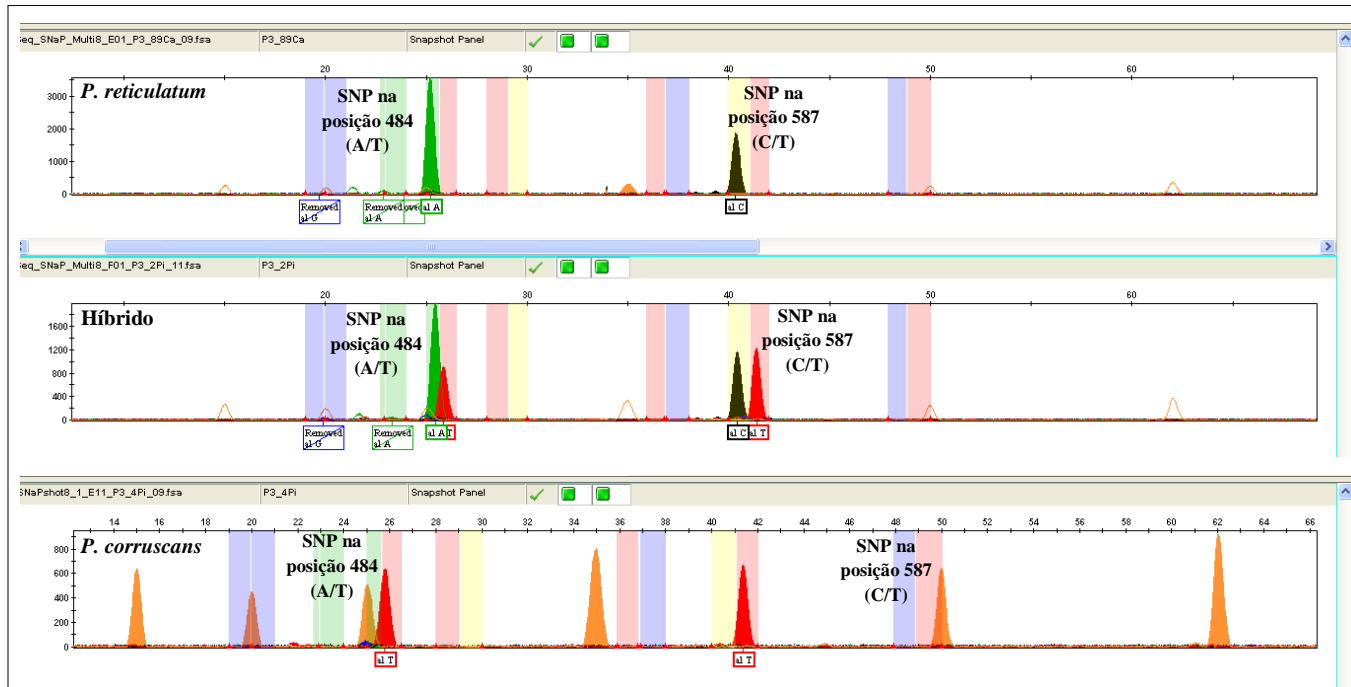


Figura 2.6. Eletroferogramas mostrando os produtos da reação de SNaPshot dos SNPs 484 (A/T) e 587 (C/T) do gene nuclear EF1 α . Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50, 62, 80 pb, respectivamente. Picos verde e vermelho à esquerda: SNPs na posição 484 e picos preto e vermelho à direita: SNPs na posição 587.

Os fragmentos do gene EF1 α sequenciados confirmaram os resultados alcançados com o ensaio SNaPshot, apresentando os mesmos polimorfismos identificados no ensaio SNaPshot para esse fragmento. Foram identificadas a presença de uma cachara e seis pintados híbridos dentre os 26 animais coletados e que o ensaio SNaPshot desenvolvido de forma individual para o gene EF1 α foi otimizado a contento. Os genótipos dos animais para os SNPs interrogados neste ensaio podem ser visualizados na Tabela 2.6.

O sequenciamento do gene EF1 α para as 83 cacharas e 48 pintados restantes, identificou híbridos de cinco pintados e de uma cachara, comprovando que os SNPs 484 e 587 interrogados no minisequenciamento do gene EF1 α são úteis para diferenciar indivíduos híbridos de puros das espécies *P. reticulatum* e *P. corruscans* (resultados podem ser visualizados na Tabela A1, do Anexo A).

5.3.4 Gene nuclear RAG₂

O minisequenciamento do gene nuclear RAG₂ também revelou a presença de híbridos dentre os 26 animais coletados. A análise dos SNPs 254 e 484 do gene nuclear RAG₂ revelou três possíveis genótipos (Figura 2.7): AA/GG (*P. reticulatum* ou cachara), GG/TT (*P. corruscans* ou pintado) e GA/GT (híbrido). Foram identificadas três cacharas híbridas e três pintados híbridos, totalizando seis animais híbridos identificados neste ensaio. Desses, apenas dois pintados e uma cachara híbridos também foram identificados no ensaio com o gene EF1 α . Os outros três são animais diferentes dos identificados no ensaio do gene EF1 α .

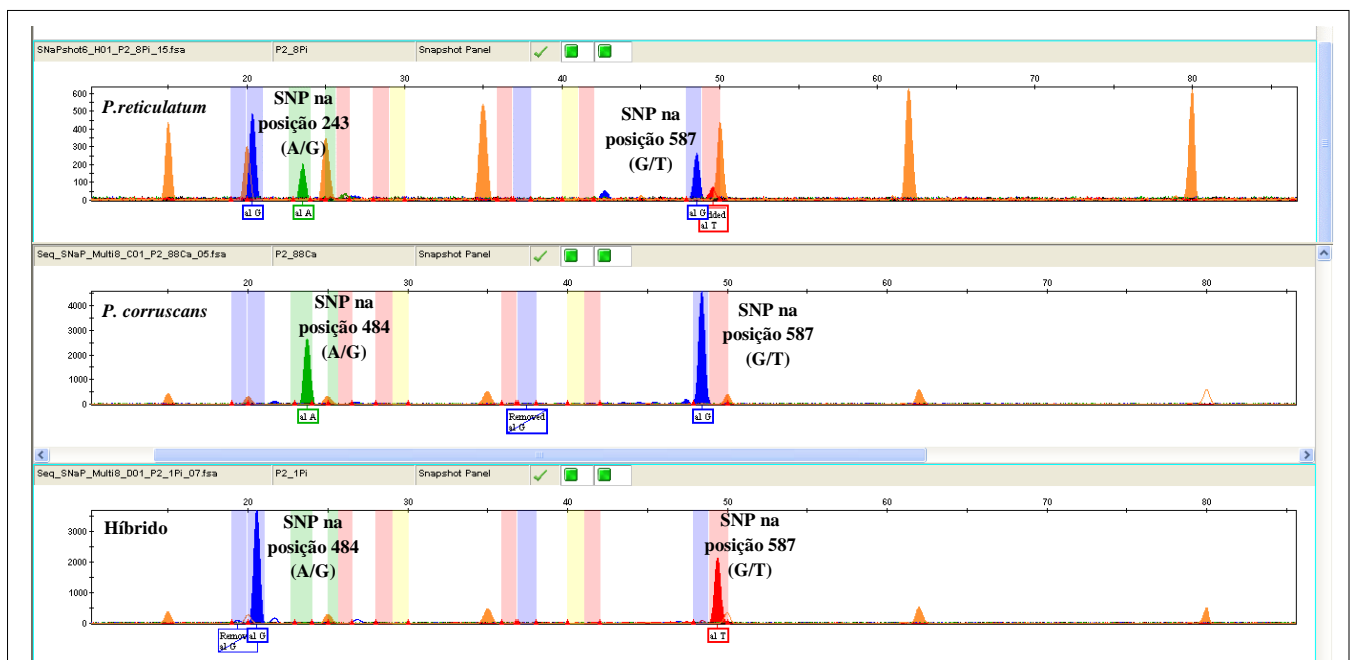


Figura 2.7. Eletroferogramas mostrando os produtos da reação de SNaPshot dos SNPs 243 (A/G) e 254 (G/T) do gene nuclear RAG₂. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50, 62, 80 pb, respectivamente. Picos verde e azul à esquerda: SNPs na posição 243 e picos azul e vermelho à direita: SNPs na posição 254.

O alinhamento das sequências do gene RAG₂ obtidas permitiu a comparação com os resultados da genotipagem realizada no minisequenciamento do gene RAG₂. Resultados semelhantes confirmaram que o ensaio SNaPshot para o gene RAG₂ foi otimizado e os genótipos dos 26 animais foram corretamente identificados (Tabela 2.6).

O sequenciamento do gene RAG₂ das demais 83 cacharas e 48 pintados, identificou cinco pintados e duas cacharas como sendo híbridos, confirmando, assim, que os SNPs 245 e 256 interrogados no ensaio SNaPshot do gene RAG₂ podem ser utilizados para

diferenciar indivíduos híbridos de puros das espécies *P. reticulatum* e *P. corruscans* (resultados podem ser visualizados na Tabela A1, do Anexo A).

5.3.5 Gene nuclear GLOB

O minisequenciamento do gene nuclear GLOB, gerou genótipos para apenas cinco cacharas e dois pintados para o SNP na posição 300 e dez cacharas e seis pintados para o SNP 374. Na Figura 2.8 são mostrados eletroferogramas referentes à dois genótipos de indivíduos híbridos. No entanto, os resultados obtidos são inconsistentes. Os polimorfismos identificados com base no alinhamento das sequências do gene GLOB de *P. reticulatum* e *P. corruscans* disponíveis no GenBank (Tabela 2.1) indicavam que os seguintes genótipos deveriam ser identificados no ensaio SNaPshot: AA para a cachara e GG para o pintado na posição 300 e AA para a cachara e TT para o pintado na posição 374. Verificou-se, porém, que para o SNP 374, das 10 cacharas genotipadas, oito apresentaram o genótipo esperado o pintado: TT (indicando que elas eram híbridas) e apenas duas foram genotipadas como AA (puras). Os seis pintados genotipados foram TT (puros). Esse resultado colocou em dúvida os resultados obtidos com os ensaios para os outros genes (COI, EF1 α e RAG2) que indicavam que apenas duas cacharas eram híbridas e todos os pintados eram híbridas (Tabela 2.6).



Figura 2.8. Eletroferogramas mostrando os produtos da reação de SNaPshot dos SNPs 300 e 347 do gene nuclear GLOB. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50, 62, 80 pb, respectivamente. Picos verde e azul à esquerda: SNPs na posição 300 e pico vermelho à direita: SNP na posição 347.

Os resultados alcançados com o sequenciamento do gene GLOB dos 26 animais coletados são consistentes com os obtidos no minisequenciamento. Na posição 374, 11 das 17 cacharas foram identificadas com o genótipo TT (híbridas), esperado para o pintado e quatro foram identificadas como AT (híbridas também). Todos os pintados foram identificados como TT (puros). Na posição 300, o sequenciamento do fragmento GLOB indicou que quatro das 17 cacharas eram híbridas, uma vez que apresentaram o genótipo esperado para o pintado: GG. As demais eram puras (AA). Todos os pintados eram puros (GG). Esses genótipos podem ser visualizados na Tabela 2.6.

Embora o sequenciamento tenha confirmado o resultado do minisequenciamento SNaPhot, afirmações em relação ao genótipo dos animais através da interrogação desses dois SNPs não são confiáveis. Os resultados indicam que esses SNPs não são espécie-específicos, uma vez que os alelos não encontram-se fixados nas espécies.

O resultado do sequenciamento do gene GLOB para os demais 131 animais (morfologicamente identificados como 83 cacharas e 48 pintados) revelou uma grande quantidade de genótipos heterozigotos para as duas espécies e para os dois SNPs. Das 83 cacharas sequenciadas, 80 foram identificadas como puras (genótipo AA) na posição 300 e mais da metade dos pintados (36 animais) apresentou genótipo AG. Para o SNP 374, o genótipo de mais da metade das cacharas (51 animais) foi AT e 44 dos 48 pintados apresentaram genótipo TT (puros). Os resultados podem ser visualizados na Tabela A1, do Anexo A.

Os genótipos gerados pelo sequenciamento dos 131 animais restantes confirmaram os resultados observados no ensaio SNaPshot e no sequenciamento dos 26 animais. O gene GLOB, portanto, não é um marcador adequado para diferenciar indivíduos puros de híbridos das duas espécies (uma vez que não é espécie-específico), conforme indicado por Hashimoto et al. (2013). Desta forma, o gene GLOB foi excluído do ensaio SNaPshot.

5.3.6 Gene nuclear 18S

Através do ensaio SNaPshot individual para interrogar o SNP 185 do gene 18S foi possível identificar genótipos para 20 dos 26 animais coletados (13 cacharas e 7 pintados). Três possíveis genótipos foram observados: TT (cachara), CC (pintado) e CT (híbridos). Pela

análise deste SNP, duas das 13 cacharas genotipadas foram consideradas híbridas, assim como três dos 7 pintados. Na Figura 2.6, visualiza-se o padrão dos picos no GeneMapper para os genótipos CT (híbrido) e TT (cachara) do ensaio SNaPshot com o gene 18S.

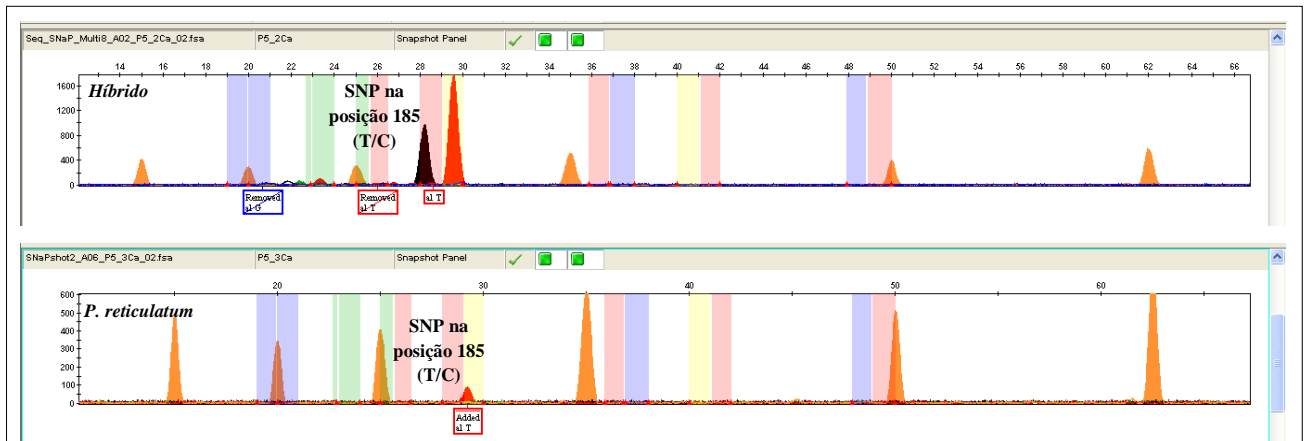


Figura 2.9. Eletroferogramas mostrando os produtos da reação de SNaPshot dos SNPs 185 (T/C) do gene nuclear 18S. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50, 62, 80 pb, respectivamente. Picos preto e vermelho correspondem ao SNP na posição 185.

O alinhamento das sequências geradas para o gene 18S permitiu a identificação de 4 cacharas híbridas (T/C) dentre as 17 coletadas e três pintados híbridos (T/C) dentre os sete sequenciados (para dois pintados não foi possível obter sequências) e foram consistentes com os resultados obtidos no ensaio SNaPshot (com exceção de dois pintados cujos genótipos foram diferentes dos identificados no ensaio SNaPshot) (Tabela 2.6.).

Sequências do gene GLOB geradas paraas 83 cacharas e 48 pintados restantes confirmaram que esses SNPs são adequados para diferenciar indivíduos puros de híbridos das espécies *P. reticulatum* e *P. corruscans* (Tabela A1, do Anexo A).

Através do minisequenciamento individual dos genes mitocondriais e nucleares selecionados para este estudo com base nas informações disponíveis em Hashimoto et al., (2013) e Prado et al. (2011), ficou evidente que os genes 16S e GLOB não são marcadores adequados para identificar indivíduos híbridos nas espécies *P. reticulatum* e *P. corruscans* e foram retirados do ensaio SNaPshot®. Restaram, assim, quatro genes (com 7 posições polimórficas) que apresentaram resultados consistentes para a adequada diferenciação de indivíduos puros e híbridos nessas espécies para o desenvolvimento do ensaio SNaPshot (em *multiplex*): COI, EF1 α , RAG₂ e 18S.

Tabela 2.6. Genótipo dos 26 peixes coletados obtidos para os SNPs interrogados no minisequenciamento SNaPshot e no sequenciamento dos genes mitocondriais COI e nucleares RAG₂, EF1 α , GLOB e 18S

ID ^a	Fenó ^b .	COI				EF1 α				RAG ₂				GLOB**				18S		Sp. ^d
		SNaPshot		Seq. ^c		SNaPshot		Seq. ^c		SNaPshot		Seq. ^c		SNaPshot		Seq. ^c		SNaPshot	Seq. ^c	
		SNP 243*	SNP 409*	SNP 243	SNP 409	SNP 484*	SNP 587	SNP 484	SNP 587	SNP 243	SNP 254*	SNP 243	SNP 254	SNP 300	SNP 374	SNP 300	SNP 374	SNP 185	SNP 185	
Ref. ^e	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	AA	AA	AA	AA	TT	TT	
Ref. ^e	Pintado	GG	AA	GG	AA	TT	TT	AA	TT	GG	TT	GG	AA	GG	TT	GG	TT	CC	CC	
88Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	AA	AA	AA	AT	TT	TT	Cachara
89Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	AA	TT	AA	AT	TT	TT	Cachara
90Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	AA	TT	AA	AT	TT	TT	Cachara
91Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	AA	AA	AA	TT	TT	TT	Cachara
92Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	AA	AT	TT	TT	Cachara
93Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	AA	TT	TT	TT	Cachara
94Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	AA	TT	TT	TT	Cachara
1Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	GG	TT	GG	AA	-	TT	GG	TT	CT	CT	Híbrido
2Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	GA	GT	GA	CA	-	-	GG	TT	CT	CT	Híbrido
3Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	AA	AA	-	TT	Cachara
4Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	AA	TT	-	TT	Cachara
5Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	TT	AA	AA	TT	TT	Cachara
6Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	TT	GG	TT	TT	TT	Cachara
7Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	TT	AA	TT	TT	TT	Cachara
8Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	TT	AA	TT	TT	TT	Cachara
9Ca	Cachara	TT	GG	AA	CC	AA	CC	TT	CC	AA	GG	AA	CC	-	-	GG	TT	-	CT	Híbrido
10Ca	Cachara	TT	GG	AA	CC	AT	CT	AT	CT	GA	GT	GA	CA	AA	TT	AA	TT	-	CT	Híbrido
1Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	GG	TT	GG	AA	GG	TT	GG	TT	CC	CC	Híbrido
2Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	GG	TT	GG	AA	GA	TT	GG	TT	CC	CC	Híbrido
3Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	GG	TT	GG	AA	-	TT	GG	TT	CT	CT	Híbrido
4Pi	Pintado	TT	GG	AA	CC	TT	TT	AA	TT	AG	GG	AG	CA	-	TT	GG	TT	CT	CT	Híbrido
5Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	AG	GG	AG	CA	-	TT	GG	TT	CC	CT	Híbrido
6Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	GG	TT	GG	AA	-	TT	GG	TT	CC	CC	Híbrido
7Pi	Pintado	TT	GG	AA	CC	TT	TT	AA	TT	GG	TT	GG	AA	-	-	GG	TT	CT	CC	Híbrido
8Pi	Pintado	TT	GG	AA	CC	AT	CT	AT	CT	AG	GT	AG	CA	-	-	GG	TT	-	-	Híbrido
9Pi	Pintado	TT	GG	AA	CC	TT	TT	AA	TT	GG	TT	GG	AA	-	-	GG	TT	-	-	Híbrido

^a ID: Identificação do animal; ^b Fenó^t.: Fenótipo; ^c Seq.: Sequenciamento; ^d SP.: Espécie: Cachara (*Pseudoplatystoma reticulatum*); Híbrido (*P. reticulatum* x *P. corruscans*);

^e Referência – Genótipo esperado para a espécie; * *Primers* SNaPshot sentido Reverse.

** Genótipos obtidos para o gene GLOB foram ignorados.

5.4 Amplificação dos Fragmentos por PCR *Multiplex*

Na figura 2.10 podem ser visualizadas cinco bandas por poço do gel de eletroforese, referentes aos cinco genes analisados neste estudo, amplificados em uma única reação de PCR, ou seja, em *multiplex*.

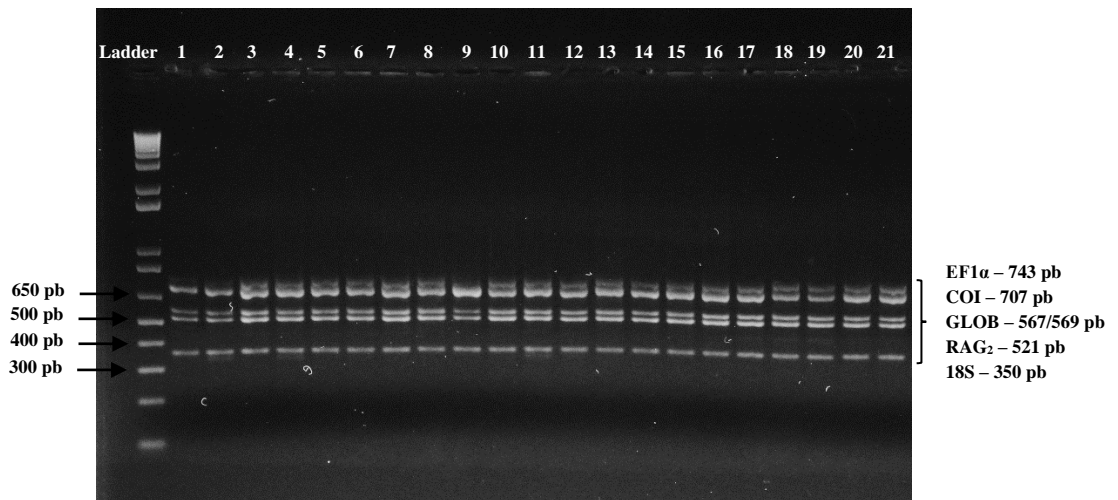


Figura 2.10. Visualização da amplificação dos 5 fragmentos (poços 2 à 20) em *multiplex*: EF1 α (743 pb), COI (707 pb), GLOB (569/567 pb), RAG₂ (521 pb) e 18S (350 pb).

5.5 Minisequenciamento Snapshot para Interrogar Nove SNPs

Como esse ensaio foi desenvolvido concomitantemente com o ensaio SNaPshot individual, não havia conhecimento sobre os resultados inconsistentes obtidos para o *primer* GLOB, e por isso ele não foi excluído desse ensaio, que foi desenvolvido para os nove SNPs inicialmente selecionados para comporem o ensaio. Não foi possível otimizar o ensaio SNaPshot para todos os nove SNPs. Na análise dos eletroferogramas gerados após a reação de extensão dos *primers* SNaPshot no *software* GeneMapper, não era possível observar os picos de todos os nove SNPs que estavam sendo interrogados no ensaio. Vários testes alterando a concentração dos nove *primers* incluídos no ensaio foram realizados, porém, resultados positivos não foram obtidos.

5.6 Minisequenciamento SNaPshot com Cinco SNPs

Ainda sem ter conhecimento dos resultados negativos para o gene GLOB, o ensaio SNaPshot foi testado com apenas um dos SNPs de cada gene. Foram selecionados cinco SNPs cujos picos podiam ser visualizados no ensaio SNaPshot com os nove SNPs: COI_409, RAG₂_254, EF1 α _484, GLOB_300 e 18S_185.

Testes foram realizados para a otimização da reação de extensão destes cinco *primers*. Na Figura 2.11 é apresentado o eletroferograma resultante da reação que apresentou as melhores condições de extensão dos primers.

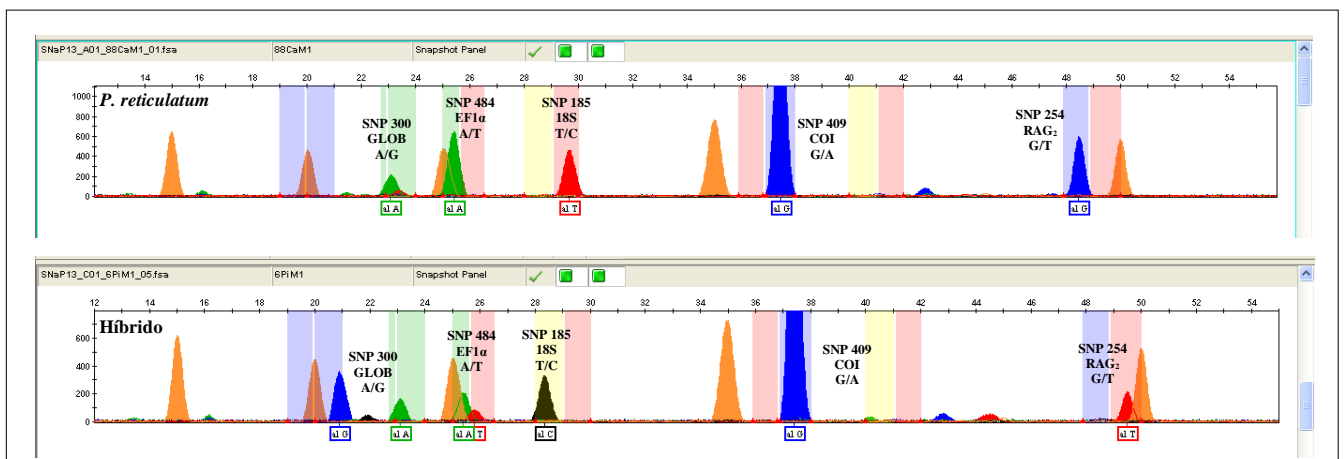


Figura 2.11. Eletroferogramas mostrando os produtos da reação de SNaPshot dos SNPs 300 do gene GLOB (A/G), 484 do EF1 α (A/T), 185 do 18S (T/C), 409 do COI (G/A) e 254 (G/T) do RAG₂. Picos laranja correspondem ao marcador de peso molecular LIZ120, da esquerda para a direita: 15, 25, 35, 50, 62 pb, respectivamente.

Ajustes ainda podem ser feitos para obter condições melhores para a reação final de extensão dos *primers* para este ensaio, como a diminuição da concentração do *primer* COI_409R, cujo pico estava muito alto, o aumento da concentração dos *primers* EF1 α _484R e RAG₂_254R, cujos picos ainda estão baixos. Porém as condições da reação de extensão dos *primers* otimizada já possibilitam a identificação dos genótipos para cada SNP que compõe o ensaio SNaPshot. Como os resultados necessários para o desenvolvimento dos outros experimentos desta Tese já haviam sido alcançados com o ensaio individual e com o sequenciamento (identificação de indivíduos híbridos dentre os 26 animais coletados), optou-se por esperar o conserto do sequenciador ABI3100 para finalizar a otimização do ensaio SNaPshot. Diante dos resultados obtidos com o sequenciamento dos cinco genes, o gene

GLOB precisará ser retirado do ensaio. Assim que o ensaio estiver otimizado, essas informações serão publicadas.

6 CONCLUSÕES

O desenvolvimento do minisequenciamento com base em informações publicamente disponíveis sobre cinco marcadores espécie-específicos (Hashimoto et al., 2013 e Prado et al., 2011) para identificar indivíduos híbridos entre *P. reticulatum* e *P. corruscans* permitiu a identificação de 13 animais híbridos dentre os 26 que foram coletados. Dos 17 animais coletados, morfologicamente identificados como cachara, quatro foram identificados como híbridos no ensaio. Todos os nove animais coletados com fenótipo de pintado foram identificados como híbridos.

Adicionalmente, os resultados obtidos com o ensaio SNaPshot© individual seguido pelo sequenciamento dos cinco genes, permitiram concluir que dois dos cinco marcadores recomendados na literatura (genes 16 S e GLOB), não são adequados para diferenciar indivíduos híbridos de puros. O gene mitocondrial 16S e o gene nuclear GLOB foram, portanto, excluídos do ensaio. Os quatro genes que permaneceram no ensaio SNaPshot© (mitocondrial COI e três nucleares: EF1 α , RAG₂ e 18S) são adequados para identificar híbridos F1. Porém, é importante o desenvolvimento de ferramentas que permitam a identificação de introgressões em escalas menores.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, D.M.T.; PERREIRA, R.W.; LEAL-BERTIOLI, S.C.M. et al. Desenvolvimento e utilização de variações em um único nucleotídeo para o mapeamento de marcadores âncoras e de genes candidatos em *Arachis*. **Boletim de Pesquisa e Desenvolvimento. Embrapa Recursos Genéticos e Biotecnologia**. Brasília, DF. 2007. 15p.
- BARBOSA, O.N.; RAIZER, J.; GONDA, M.F. et al. Desempenho e coeficiente intestinal de alevinos puros e híbridos de pintados em condicionamento alimentar. **Revista Brasileira de Zootecnia**. v. 40, n. 12, p.2621-2627, 2011.
- BARTLEY, D.M.; RANA, K., IMMINK, The use of inter-specific hybrids in aquaculture and fisheries. **Reviews in Fish Biology and Fisheries**. v. 10, p.325-327, 2001.
- CAETANO, A.R. SNP markers: basic concepts, applications in animal breeding and management and perspectives for the future. **Revista Brasileira de Zootecnia**. v. 38, p.64-71, 2009.
- CARVALHO, D.C.; NETO, D.A.P.; BRASIL, B.S.A.F. et al. DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. **Mitochondrial DNA**, v. 22, p.97-105, 2011a.
- CARVALHO, D.C.; OLIVEIRA, D.A.; POMPEU, P.S. et al. Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco River basin. **Mitochondrial DNA**, v. 22, n. 1, p.80-86, 2011b.
- CARVALHO, D.C.; OLIVEIRA, D.A.A.; BEHEREGARAYM L.B. et al. Hidden genetic diversity and distinct evolutionarily significant units in an commercially important Neotropical apex predator, the catfish *Pseudoplatystoma corruscans*. **Conservation Genetics**, v. 13, n. 6, p. 1671-1675, 2012.
- CREPALDI, D.V.; FARIA, P.M.C.; TEIXEIRA, E.A. et al. O surubim na aquacultura do Brasil. **Revista Brasileira de Reprodução Animal** v. 30, n. 3/4, p. 150-158, 2006.

- EWING, B.; GREEN, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**. v. 8, p. 186-194, 1998.
- EWING, B.; HILLIER, L.; WENDL, M. et al. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**. v. 8, p. 175-185, 1998.
- GEPPERT, M.; ROEWER, L. SNaPshot minisequencing analysis of multiple ancestry-informative Y-SNPs using capillary electrophoresis. In: ALONSO, A. (Ed.). **DNA Electrophoresis Protocols for Forensic Genetics, Methods in Molecular Biology**. v. 830. Madrid, Spain, 2012. p.127-140.
- HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl. Acids. Symp. Ser.** v. 41, p. 95-98, 1999.
- HASHIMOTO, D.T.; PRADO, F.D. do; SENHORINI, J.A. et al. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. **Aquaculture Research**. v. 44, p. 876-884, 2013.
- KOCHER, T.D.; THOMAS, W.K.; MEYER, A. et al. Dynamics of Mitochondrial DNA Evolution in Animals: Amplification and Sequencing with Conserved Primers. **Proc. Natl. Acad. Sci.** v. 86, p. 6196-6200, 1989.
- MOYER G.R. ; BURR B.M. ; KRAJEWSKI C. Phylogenetic relationships of thorny catfishes (Siluriformes: Do-radiidae) inferred from molecular and morphological data. **Zoological Journal of the Linnean Society**. v. 140, p. 551–575, 2004.
- PEREIRA, L.H.G.; HANNER, R.; FORESTI, F. et al. Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? **BMC Genetics**, v. 14, n. 20, p. 1-14, 2013.
- PORTO-FORESTI, F.; HASHIMOTO, D.T.; ALVES, A.L. et al. Cytogenetic markers as diagnoses in the identification of the hybrid between Piauçu (*Leporinus macrocephalus*) and Piapara (*Leporinus elongatus*). **Genetics and Molecular Biology**, v. 31, n. 1, p. 195-202, 2008.
- PRADO, F.D. do; HASHIMOTO, D.T.; MENDONÇA, F.F.; SENHORINI, J.A.; FORESTI, F.; PORTO-FORESTI, F. Molecular identification of hybrids between Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum*. **Aquaculture Research**. 2011. p.1-5.
- PRADO, F.D. do; NUNES, T.L.; SENHORINI, J.A. et al. Cytogenetic characterization of F1, F2 and backcross hybrids of the Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Pimelodidae, Siluriformes). **Genetics and Molecular Biology**, v. 35, n. 1, p. 57-64, 2012.
- WARD, R. D., T. S. ZEMLAK, B. H. INNES, et al. DNA barcoding Australia's fish species. **Philosophical Transactions of the Royal Society B**. v. 360, p.1847-1857, 2005.
- WHITE, T.J.; BRUNS, T.; LEE, S.; TAYLOR, J.W. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: **PCR Protocols: A Guide to**

Methods and Applications(ed. by M.A. Innis, D.H. Gelfand, J.J. Sninsky & T.J. White).
New York, USA: Academic Press. 1990. p.315–322.

ANEXO A

Tabela A.1. Genótipos dos 131 animais (83 cacharas e 48 pintados) sequenciados para os genes GLOB, 18S, RAG₂, EF1 α e COI

		GLOB				18S				RAG ₂				EF1 α				COI			
		301		375		189		191		245		256		1151		1254		247		413	
	REF CACHARA	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
	REF PINTADO	G	G	T	T	C	C	C	C	G	G	A	A	A	A	T	T	C	C	T	T
1	Cachara_112902	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
2	Cachara_934000090005414	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
3	Cachara_934000090005417	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
4	Cachara_934000090005434	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
5	Cachara_934000090005455	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
6	Cachara_934000090005458	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
7	Cachara_934000090005470	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
8	Cachara_934000090005474	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
9	Cachara_934000090005476	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
10	Cachara_934000090005478	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
11	Cachara_934000090005479	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
12	Cachara_934000090005483	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
13	Cachara_934000090005484	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
14	Cachara_934000090005485	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
15	Cachara_934000090005490	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
16	Cachara_934000090005496	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
17	Cachara_934000090005498	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
18	Cachara_934000090005500	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
19	Cachara_MT_246	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
20	Cachara_MT_247	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
21	Cachara_MT_248	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
22	Cachara_MT_249	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
23	Cachara_MT_250	A	A	A	A	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C
24	Cachara_MT_252	A	A	A	T	T	T	G	G	A	A	C	C	T	T	C	C	A	A	C	C

25	Cachara_MT_253	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
26	Cachara_MT_254	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
27	Cachara_MT_255	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
28	Cachara_MT_256	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
29	Cachara_MT_257	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
30	Cachara_MT_258	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
31	Cachara_MT_259	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
32	Cachara_MT_260	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
33	Cachara_MT_261	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
34	Cachara_MT_269	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
35	Cachara_MT_270	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
36	Cachara_MT_271	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
37	Cachara_MT_272	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
38	Cachara_MT_273	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
39	Cachara_MT_274	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
40	Cachara_MT_275	A	G		T	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
41	Cachara_MT_276	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
42	Cachara_MT_277	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
43	Cachara_MT_278	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
44	Cachara_MT_279	A	A		A	A		T	T		G	G		A	G		A	C		T	T		C	C		A	A		C	C
45	Cachara_MT_280	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
46	Cachara_MT_281	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
47	Cachara_MT_282	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
48	Cachara_MT_283	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
49	Cachara_MT_284	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
50	Cachara_MT_285	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
51	Cachara_MT_286	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
52	Cachara_MT_287	A	G		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
53	Cachara_MT_320	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
54	Cachara_MT_321	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
55	Cachara_MT_322	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
56	Cachara_MT_323	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
57	Cachara_MT_324	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C

58	Cachara_MT_325	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
59	Cachara_MT_326	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
60	Cachara_MT_327	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
61	Cachara_MT_328	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
62	Cachara_MT_329	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
63	Cachara_MT_330	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
64	Cachara_MT_331	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
65	Cachara_MT_332	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
66	Cachara_MT_805	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
67	Cachara_MT_806	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
68	Cachara_MT_807	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
69	Cachara_MT_808	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
70	Cachara_MT_809	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
71	Cachara_MT_810	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
72	Cachara_MT_811	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
73	Cachara_MT_812	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
74	Cachara_MT_813	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
75	Cachara_MT_814	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
76	Cachara_MT_815	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
77	Cachara_MT_816	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
78	Cachara_MT_817	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
79	Cachara_MT_818	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
80	Cachara_MT_819	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
81	Cachara_MT_820	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
82	Cachara_MT_821	A	A		A	A		T	T		G	G		A	A		C	C		T	T		C	C		A	A		C	C
83	Cachara_MT_822	A	G		A	T		T	T		G	G		A	G		A	C		A	T		C	C		A	A		C	C
84	Pintado_1000	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
85	Pintado_1001	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
86	Pintado_1002	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
87	Pintado_1003	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
88	Pintado_1004	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
89	Pintado_1005	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
90	Pintado_1006	G	G		A	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T

91	Pintado_1007	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
92	Pintado_1008	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
93	Pintado_1009	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
94	Pintado_1010	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
95	Pintado_1011	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
96	Pintado_1012	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
97	Pintado_1013	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
98	Pintado_1014	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
99	Pintado_1015	A	A		A	T		T	T		G	G		A	A		C	C		T	T		C	C		C	C		T	T
100	Pintado_1119	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
101	Pintado_1120	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
102	Pintado_1121	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
103	Pintado_1122	A	G		T	T		C	C		C	C		G	G		A	A		T	T		T	T		C	C		T	T
104	Pintado_1124	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
105	Pintado_1125	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
106	Pintado_263	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
107	Pintado_264	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
108	Pintado_265	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
109	Pintado_382	A	G		A	T		T	T		G	G		A	A		C	C		T	T		C	C		C	C		T	T
110	Pintado_461	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
111	Pintado_463	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
112	Pintado_823	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
113	Pintado_824	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
114	Pintado_825	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
115	Pintado_826	A	G		T	T		T	T		G	G		A	G		A	C		A	A		T	T		C	C		T	T
116	Pintado_827	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
117	Pintado_828	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
118	Pintado_829	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
119	Pintado_830	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
120	Pintado_831	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
121	Pintado_832	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
122	Pintado_833	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
123	Pintado_834	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T

124	Pintado_835	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
125	Pintado_836	A	G		T	T		C	C		C	C		A	G		C	C		A	A		C	T		C	C		T	T
126	Pintado_837	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
127	Pintado_838	A	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
128	Pintado_839	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
129	Pintado_840	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
130	Pintado_998	G	G		T	T		C	C		C	C		G	G		A	A		A	A		T	T		C	C		T	T
131	Pintado_999	A	A		A	T		T	T		G	G		A	G		A	C		A	T		C	C		C	C		T	T

CAPÍTULO 3

PROSPECÇÃO DE POLIMORFISMOS DE BASE ÚNICA (SNP) ESPÉCIE- ESPECÍFICOS NOS TRANSCRIPTOMAS DA CACHARA (*Pseudoplatystoma reticulatum*) E DE HÍBRIDOS DE PINTADO (*Pseudoplatystoma corruscans*)

1 RESUMO

Por apresentar aspectos negativos para o desenvolvimento sustentável do sistema produtivo de *Pseudoplatystoma reticulatum* (cachara) e *P. corruscans* (pintado), a hibridação realizada pelos piscicultores entre essas espécies precisa ser monitorada. No presente estudo, foi realizado o sequenciamento Illumina (HiSeq 2000 e MiSeq v3) e a montagem *de novo* (*software* Trinity) dos transcriptomas da cachara (*Pseudoplatystoma reticulatum*) e de híbridos de pintado (*Pseudoplatystoma reticulatum* x *Pseudoplatystoma corruscans*) Os transcriptomas foram obtidos através do sequenciamento de RNA de uma variedade de tecidos (músculo branco, músculo vermelho, brânquias, gônadas, rim, fígado e hipófise), coletados de 12 cacharas e nove híbridos de pintado. Os dois transcriptomas foram mapeados separadamente da montagem do RNA-seq, utilizando o *software* BWA. A identificação dos SNPs foi realizada através do *software* SAMtools mpileup e a anotação funcional dos transcritos contendo os candidatos à SNPs espécie-específicos foi realizada utilizando o pacote Trinotate. 321 polimorfismos candidatos à SNPs espécie-específicos foram identificados em 292 transcritos. A razão transição/transversão dos SNPs espécie-específicos identificados foi 2,11. Dos 292 transcritos, 266 transcritos (91,1%) apresentaram *hit* significativo para proteína no banco de dados não redundante (nr) e 222 transcritos

puderam ser anotados para os termos GO, sendo que quase metade desses transcritos estava relacionada à processos biológicos. O alinhamento dos 292 transcritos contendo os SNPs espécie-específicos candidatos com um genoma referência (*Danio rerio*) revelou que 64 transcritos não possuíam regiões de junção de *splice* em suas regiões flangeadoras e, que após validação, esses SNPs poderão compor um painel de SNPs espécie-específicos para diferenciação de indivíduos puros e híbridos das espécies *P. reticulatum* e *P. corruscans*. As informações geradas neste estudo para ambas espécies poderão ser utilizadas em diferentes tipos de estudos genéticos para a cachara e para o pintado, além de servir como referência para o desenvolvimento de estudos semelhantes para outras espécies de bagres e de peixes em geral.

Palavras-Chave: Sequenciamento de nova geração; RNA-Seq; Transição; Transversão; Anotação funcional; Ontologia Gênica

SINGLE NUCLEOTIDE POLYMORPHISMS PROSPECTION (SNP) SPECIE-SPECIFIC IN “CACHARA” (*Pseudoplatystoma reticulatum*) AND “PINTADO” HYBRIDS (*Pseudoplatystoma corruscans*) TRANSCRIPTOMES

2 ABSTRACT

This study prospected specie-specific SNPs on the transcriptome of cachara (*Pseudoplatystoma reticulatum*) and pintado hybrids (*Pseudoplatystoma reticulatum* x *Pseudoplatystoma corruscans*) generated by Illumina sequencing on platforms HiSeq 2000 and MiSeq v3 and assembled *de novo* using the software Trinity. The transcriptomes were obtained by RNA sequencing of a variety of tissues collected in both species. The two transcriptomes were mapped separately of the RNA-seq assembly using the software BWA. The SNPs were identified using SAMtools mpileup. We identified 321 polymorphisms candidates for specie-specific SNPs to differentiate pure and hybrid individuals of the species *P. reticulatum* and *P. corruscans* amongst 292 transcript. Transition/transversion ratio were 2.11. Amongst the 292 transcript, 266 (91.1%) presented significant hit for protein on non-redundant databank (nr) and 222 transcript could be annotated to GO terms, and almost half part of these transcript due to biological processes. The alignment of the 292 transcript containing the specie-specific SNPs candidates for a reference genome (*Danio rerio*) revealed that 64 transcript do not have splice regions in their flanking regions and after validation can compose a specie-specific SNPs panel to differentiate pure and hybrid individuals of *P. reticulatum* and *P. corruscans*. These are unprecedented information about the studied species and can be used for cachara and pintado different types of genetic studies, besides serving as reference to the development of similar studies to other catfish and fish species in general.

Keywords: New generation sequencing; RNA-Seq; transition; transversion; functional annotation; gene ontology.

3 INTRODUÇÃO

Embora seja uma prática bastante comum entre os criadores de peixes das espécies *Pseudoplatystoma reticulatum* (cachara) e *Pseudoplatystoma corruscans* (pintado), a geração de híbridos apresenta diversos aspectos negativos para o desenvolvimento sustentável do sistema produtivo dessas espécies. Os híbridos gerados são férteis, o que coloca em risco tanto os estoques naturais de peixes, quanto os estoques “puros” nas pisciculturas (Godinho, 2007; Prado et al., 2011), dificultando o desenvolvimento de programas de melhoramento genético “clássico” para essas espécies, uma vez que eles necessitam de estoques de reprodutores “puros” para que os processos de avaliação das características produtivas possam ser desenvolvidos.

Ferramentas que identifiquem animais “puros” dessas espécies são imprescindíveis. Hashimoto et al. (2013) e Prado et al. (2011) desenvolveram marcadores baseados em variações de polimorfismos únicos (SNPs), em quatro genes nucleares: Gene Ativador da Recombinase (RAG₂), β -Globina (GLOB), RNA Ribossomal 18S (18S) e Fator de Elongamento 1 α (EF1 α) e um gene mitocondrial (RNA ribossomal 16S), utilizando as técnicas de PCR-RFLP e PCR-Multiplex (PCR espécie-específica). Esses marcadores permitem a identificação de híbridos inter-específicos F1 e pós-F1. De acordo com Boecklen & Howard (1997), uma quantidade de cinco marcadores nucleares é suficiente para diferenciar indivíduos puros de indivíduos híbridos F1 indivíduos com um nível de retrocruzamento mais simples. No entanto, devido à fertilidade dos híbridos resultantes do cruzamento entre cachara e pintado, os níveis de retrocruzamento desses híbridos podem ser muito mais avançados. Para a identificação desses níveis de introgressão, esses marcadores não são suficientes. Conforme afirmam Boecklen & Howard (1997), em torno de 70 marcadores seriam necessários.

Os marcadores moleculares SNPs (Polimorfismos de Base Única) têm como princípio mutações em bases únicas da cadeia de DNA, são extremamente abundantes e encontram-se homoganeamente distribuídos pelo genoma. O desenvolvimento das tecnologias de seqüenciamento de nova geração (NGS) na última década, como o seqüenciamento Illumina e 454 Life Sciences, revelou metodologias de alto desempenho e acurácia, e baixo custo para a prospecção, caracterização e genotipagem de milhares desses marcadores. Com isso, as tendências metodológicas atuais para a geração de dados moleculares de genotipagem têm convergido para o uso de Polimorfismos de Base Única (Caetano, 2009).

Utilizando as NGS, os SNPs podem ser identificados não apenas em indivíduos com genoma completo seqüenciado. Uma alternativa econômica e eficiente para a prospecção de polimorfismos em espécies que ainda não possuem genoma seqüenciado, e que está se tornando o principal método para a descoberta de SNPs é o seqüenciamento do transcriptoma completo ou RNA-Seq (Yu et al., 2014; Helyar et al., 2012; Xu et al., 2012). O RNA-Seq é um método muito acurado para a detecção de SNPs (Quinn et al., 2013), além de permitir a caracterização do transcriptoma, a descoberta de novos genes, transcritos, junções de *splice* alternativo, o estudo da expressão gênica diferencial, dentre outros (Shen et al., 2012, Jeukens et al., 2010).

Apesar de produzir fragmentos curtos, aliada à evolução das ferramentas de bioinformática que permitem a montagem, o alinhamento e a análise dos fragmentos seqüenciados, a tecnologia de seqüenciamento Illumina vem sendo empregada com sucesso para a mineração de milhares de SNPs em transcriptomas de diferentes espécies, inclusive em espécies aquícolas como os bagres *Ictalurus punctatus* e *I. funcaus* (Liu et al., 2011); carpas *Cyprinus carpio* (Xu et al., 2012); arenque do Atlântico, *Clupea harengus* (Helyar et al., 2012); camarão branco do Pacífico, *Litopenaeus vannamei* (Yu et al., 2014; etc.). Os SNPs identificados no transcriptoma podem ainda ser associados com genes ou regiões funcionais do genoma (Xu et al., 2012).

No presente estudo, foram utilizadas cacharas e híbridos de cachara e pintado além de uma diversidade de tecidos (músculo branco, músculo vermelho, fígado, gônada, hipófise, rim e brânquias) para a construção de bibliotecas de cDNA. A partir do seqüenciamento e montagem desses fragmentos, foram prospectados SNPs espécie-específicos utilizando tecnologias mistas de seqüenciamento Illumina (MiSeq e HiSeq), para a diferenciação entre indivíduos puros e híbridos das espécies *P. reticulatum* e *P. corruscans*, com níveis mais avançados de introgressão genética.

4 MATERIAL E MÉTODOS

4.1 Amostragem, Extração, Confecção dos *Pools* de RNA Total e RNA-Seq

Amostras de músculo branco (B), músculo vermelho (V), brânquia (Q), fígado (F), rim (R), hipófise (H) e gônada (G) foram coletadas de 12 indivíduos identificados através de um minisequenciamento SnaPshot (Capítulo 2), como cachara (*Pseudoplatystoma reticulatum*) e 9 indivíduos identificados como híbridos de pintado (*P. reticulatum* x *P. corruscans*). As amostras de tecidos foram mantidas em RNAlater, à 80°C. RNA total foi extraído com Trizol[®] (Ambion), seguindo o protocolo do fabricante.

Após controle de qualidade e quantificação, as amostras de RNA total foram agrupadas em quantidades equimolares em 16 diferentes *pools*: um *pool* de RNA total de cada um dos sete tecidos amostrados para cada espécie (ou seja, sete *pools* de tecidos individuais para *P. reticulatum* e sete *pools* para os híbridos, totalizando 14 *pools*), e um *pool* de RNA total contendo todos os tecidos para cada espécie (totalizando dois *pools* de tecido total, um para *P. reticulatum* e um para os híbridos).

4.1.1 Construção das Bibliotecas de cDNA e Sequenciamento Illumina HiSeq 2000

Bibliotecas de cDNA foram construídas a partir de cada um dos 16 *pools* de tecidos individuais, utilizando o kit “TruSeq Stranded RNA Sample Preparation” (Illumina, San Diego, CA), seguindo os protocolos do fabricante. Resumidamente, RNA mensageiro foi isolado de 1 µg de RNA total de alta qualidade de cada *pool*. A primeira fita foi sintetizada

com iniciadores aleatórios com o kit *SuperScript II* (Life Technologies, Grand Island, NY). O cDNA dupla fita foi fragmentado, aparado nas pontas e uma Adenina adicionada à extremidade 3'. Adaptadores indexados foram ligados ao cDNA.

O cDNA dupla fita ligado aos adaptadores foi amplificado por PCR com o kit “*Kapa HiFi Polymerase*” (Kapa Biosystems, Woburn, MA) por apenas 10 ciclos, para reduzir problemas de representatividade das sequências devido à amplificação preferencial. As bibliotecas foram quantificadas e o tamanho médio dos fragmentos determinado. O tamanho médio dos fragmentos de cDNA nas bibliotecas foi de 250 nt (variando entre 100 e 580 nt).

As 14 bibliotecas de tecidos individuais foram sequenciadas em duas canaletas da plataforma Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA), utilizando o protocolo para gerar leituras de 100 pb pareadas (*paired-end*), com *kit* de sequenciamento SBS versão 3. Os arquivos .bcl gerados foram convertidos em arquivos .fastq com o *software* Casava 1.8.2 (Illumina). Foram gerados 14 arquivos FASTQ (7 R₁ e 7 R₂).

4.1.2 Construção das Bibliotecas de cDNA e Sequenciamento Illumina MiSeq V3

Duas bibliotecas de cDNA foram construídas a partir dos dois *pools* de tecido total, construídos para as duas espécies, seguindo o mesmo protocolo descrito no item 4.1.1, com uma modificação: o cDNA não foi fragmentado, produzindo assim fragmentos de cDNA mais longos, com tamanho variando entre 230 e 880 nt (tamanho médio: 450 nt). Essas bibliotecas de cDNA foram sequenciadas com 301 ciclos a partir de cada extremidade dos fragmentos, em uma corrida no equipamento Illumina MiSeq V3 (Illumina Inc., San Diego, CA, USA), utilizando 600 ciclos do *kit* de sequenciamento MiSeq versão 3. Os fragmentos obtidos apresentaram comprimento de 300 pb e os arquivos .bcl gerados foram processados seguindo os mesmos procedimentos descritos anteriormente.

A construção das 16 bibliotecas de cDNA e o sequenciamento de RNA (RNA-Seq) Illumina foram realizados pela equipe do W. M. Keck Center for Comparative and Functional Genomics, na Universidade de Illinois, em Urbana-Champaign (licença exportação das amostras de RNA total de *Pseudoplatystoma reticulatum* enviadas para os Estados Unidos nº 15BR016615/DF).

4.2 Montagem *de Novo* dos Transcriptomas da Cachara e dos Híbridos

A montagem *de novo* dos transcriptomas do *Pseudoplatystoma reticulatum* e dos híbridos (*Pseudoplatystoma reticulatum* x *P. corruscans*) foi realizada pela equipe do Laboratório Multiusuário de Bioinformática da Embrapa Informática Agropecuária (CNPTIA), Campinas/SP, Brasil. As sequências obtidas nos sequenciamentos Illumina MiSeq (300 pb) e HiSeq (100 pb) foram processadas com o programa Trimmomatic (Bolger et al., 2014) para a remoção dos adaptadores e para controle de qualidade, no qual foi exigido que as extremidades tivessem qualidade ≥ 25 . Para as demais posições foi aplicada uma janela deslizante com a mesma exigência de qualidade mínima.

As sequências MiSeq bidirecionais de 300 pb foram combinadas para a obtenção das *super-reads*. O programa FLASH (Magoc & Salzberg, 2011) foi empregado para essa finalidade com parâmetros padrão. Como não há genoma referência para a espécie em questão, a montagem *de novo* do transcriptoma foi realizada com o montador Trinity (Haas et al., 2013, Grabherr et al., 2011), com parâmetros padrões.

4.3 Alinhamento das Sequências de cDNA NGS e Identificação dos SNPs

Após a montagem *de novo*, os fragmentos de cada uma das espécies foram mapeados separadamente da montagem do RNA-seq, usando o software BWA (Li, 2009). A identificação dos SNPs foi realizada usando o Samtools Mpileup (Li, 2011). Dentre os SNPs identificados, a seleção dos melhores SNPs candidatos à SNPs espécie-específicos, foi realizada com base nos seguintes filtros de qualidade:

- a) SNPs que contivessem pelo menos 200 fragmentos de cobertura;
- b) SNP com qualidade associada ao genótipo de pelo menos 90 (essa é uma medida estabelecida pelo Bcftools, ferramenta do programa Samtools);
- c) O SNP selecionado deveria estar a uma distância de pelo menos 100 pb do próximo SNP, para garantir a presença de uma região flanqueadora de 100 pb à sua direita e 100 pb à sua esquerda;
- d) O SNP deveria ser homozigoto na cachara;

e) o SNP deveria ser heterozigoto nos híbridos, com $0,48 \leq \text{MAF} \leq 0,50$ (MAF - Frequência do Alelo Menor).

4.4 Anotação Funcional dos Transcritos

A anotação funcional automática dos transcritos contendo os polimorfismos candidatos à SNPs espécie-específico foi realizada usando o pacote Trinotate (*Transcriptome Functional Annotation and Analysis* - <https://trinotate.github.io/>), que faz essa anotação principalmente para transcriptomas montados *de novo*. Uma pesquisa por homologia de sequências foi realizada contra bancos de dados de sequências conhecidas (BLAST+/SwissProt/Uniref90). Os resultados do BLAST foram utilizados para anotar funcionalmente os transcritos para as três categorias de termos GO: processos biológicos, funções moleculares e componentes celulares.

4.5 Alinhamento dos SNPs identificados com genoma referência do *Danio rerio*

Como não existe um genoma referência para as espécies do gênero *Pseudoplatystoma*, os transcritos contendo polimorfismos candidatos à SNPs identificados foram alinhados ao genoma referência do *Danio rerio* (espécie modelo utilizada em estudos genéticos de vertebrados, cujo genoma completo disponível no GenBank – GRCz10, é o mais próximo das espécies *Pseudoplatystoma*) para identificação de eventuais regiões de junção de *splice* nas regiões flangeadoras dos SNPs. Foi definido que, se pelo menos 20 pb em cada região flangeadora (à direita e à esquerda do SNP) alinhassem com a sequência do *Danio rerio*, o SNP candidato à espécie-específico estaria contido dentro de um exon simples. Por outro lado, o não alinhamento indicaria a presença de uma região de junção de *splice* (limite intron-exon) na região flangeadora do SNP (100 pb à direita e 100 pb à esquerda do SNP).

5 RESULTADOS E DISCUSSÃO

5.1 RNA-Seq e Processamento das Sequências

Considerando as 16 bibliotecas sequenciadas nas duas plataformas Illumina (MiSeq e HiSeq), foram gerados 997.043.038 de fragmentos e aproximadamente 120 trilhões de pares de bases, com cobertura das sequências de 39,95, o que garantiu a descoberta de SNPs com alta acurácia. Segundo Gautier et al. (2013), o sequenciamento de *pools* de DNA de indivíduos diplóides envolve menor esforço de sequenciamento e de construção de bibliotecas e menor custo, garantindo resultados similares aos obtidos com o sequenciamento de genomas individuais.

O sequenciamento MiSeq da biblioteca de cDNA de tecidos total de *P. reticulatum* produziu 47.214.078 sequências pareadas de 300 pb de comprimento com qualidade para passar nos testes de qualidade. Um total de 14.211.437.478 pb foram sequenciados. Após controle de qualidade, restaram 28.721.880 sequências de alta qualidade, o que corresponde à 60,83% dos fragmentos sequenciados. A combinação das sequências MiSeq produziu aproximadamente 12 milhões de *super-reads* (52% do total), com comprimento médio de 392 pb (Figura 3.1). Dessas, 1.331.155 (10,83%) apresentaram comprimento ≥ 500 pb.

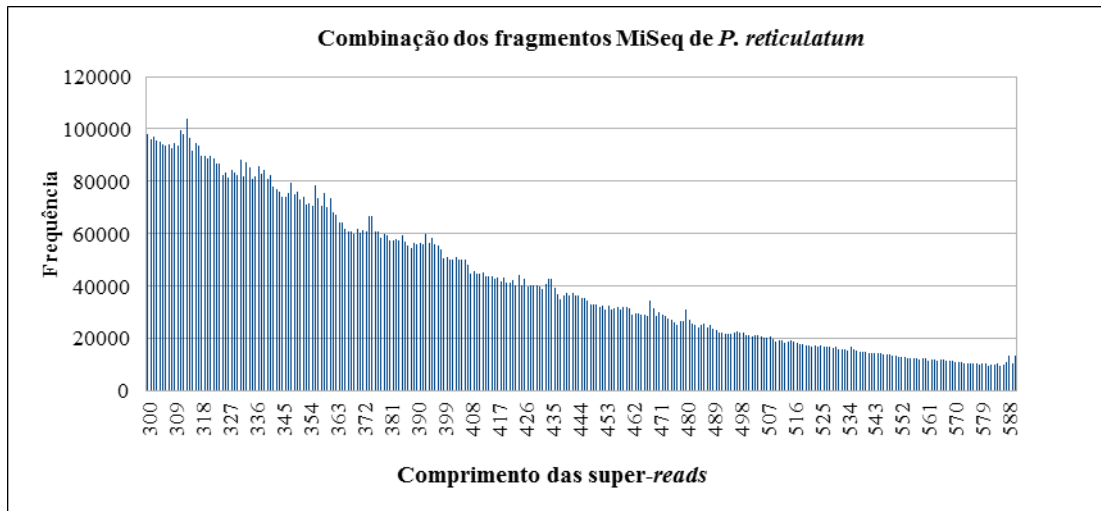


Figura 3.1. Distribuição do comprimento médio das super-reads de *P. reticulatum* geradas após a combinação das sequências bidirecionais MiSeq.

Para os híbridos (*P. reticulatum* x *P. corruscans*), o sequenciamento MiSeq da biblioteca de cDNA de tecidos total produziu 48.600.236 sequências pareadas de 300 pb de comprimento, o que corresponde à 14.628.671.036 pb sequenciados. Após controle de qualidade, restaram 32.802.878 sequências de alta qualidade (67,50% dos fragmentos sequenciados). A combinação das sequências MiSeq produziu aproximadamente 13 milhões de super-reads (56% do total), com comprimento médio de 440 pb (Figura 3.2). 2.014.376 de super-reads apresentaram comprimento ≥ 500 pb.

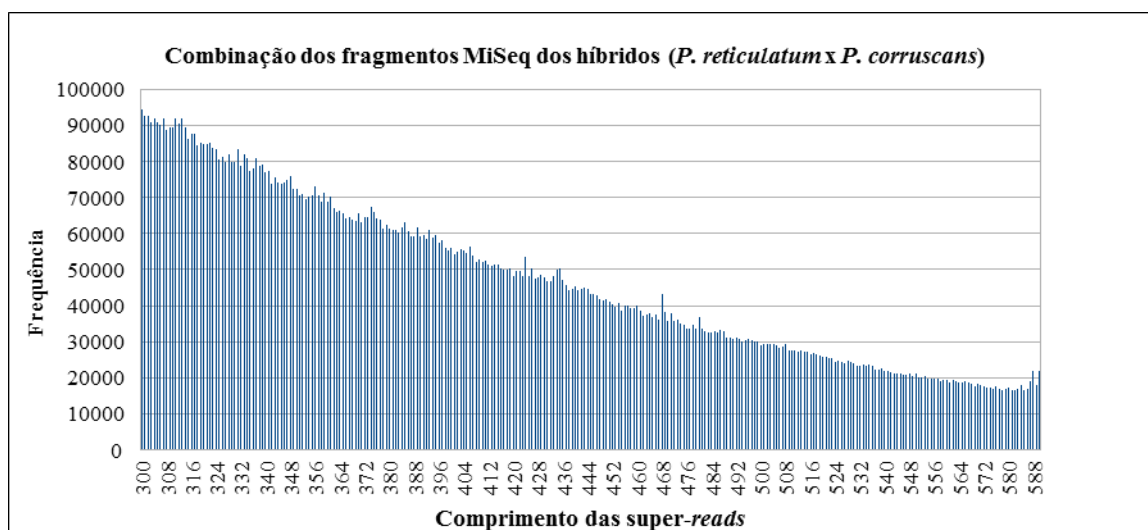


Figura 3.2. Distribuição do comprimento médio das super-reads dos híbridos *P. reticulatum* x *P. corruscans* geradas após a combinação das sequências bidirecionais MiSeq.

O sequenciamento *paired-end* Illumina HiSeq 2000 das 14 bibliotecas de cDNA de tecidos individuais produziu aproximadamente 900 milhões de sequências, sendo

444.717.156 de sequências com 100 pb de comprimento provenientes das sete bibliotecas de cDNA de *P. reticulatum* e 456.511.568 de sequências provenientes das sete bibliotecas de cDNA dos híbridos. As sequências apresentaram escore de qualidade acima de 30. Após controle de qualidade, restaram em média 870.774.578 de sequências de alta qualidade, o que corresponde a 96,6% dos fragmentos sequenciados. Dessas, 429.706.374 de sequências eram provenientes de *P. reticulatum* (Tabela 3.1) e 441.068.204 dos híbridos (Tabela 3.1).

Tabela 3.1. Número de fragmentos com 100 pb produzidos no sequenciamento Illumina HiSeq 2000 das 14 bibliotecas de cDNA de tecidos individuais de *Pseudoplatystoma reticulatum* e dos híbridos de (*P. reticulatum* x *P. corruscans*)

Cachara (<i>Pseudoplatystoma reticulatum</i>)				
Tecido	Nº fragmentos (pb) sequenciados	Nº pb	Nº fragmentos (pb) após controle de qualidade	% pb úteis
Músc. Branco	59.226.886	5.981.915.486	57.333.620	96,80
Músc. Vermelho	61.205.154	6.181.720.554	59.254.808	97,09
Fígado	64.303.546	6.494.658.146	62.431.032	97,09
Gônada	68.642.886	6.932.931.486	66.292.590	96,58
Hipófise	59.577.552	6.017.332.752	57.390.966	96,33
Brânquia	63.773.342	6.441.107.542	61.011.338	95,67
Rim	67.987.790	6.866.766.790	65.992.020	97,06
TOTAL	444.717.156	44.916.432.756	429.706.374	Média: 96,6
Híbridos de pintado (<i>Pseudoplatystoma reticulatum</i> x <i>P. corruscans</i>)				
Tecido	Nº fragmentos (pb) sequenciados	Nº pb	Nº fragmentos (pb) após controle de qualidade	% pb úteis
Músc. Branco	69.457.646	7.015.222.246	66.680.026	96,00
Músc. Vermelho	56.623.010	5.718.924.010	55.038.974	97,20
Fígado	66.313.544	6.697.667.944	64.303.214	96,97
Gônada	69.258.888	6.995.147.688	66.652.716	96,24
Hipófise	62.749.918	6.337.741.718	60.566.572	96,52
Brânquia	69.448.730	7.014.321.730	67.095.472	96,61
Rim	62.659.832	6.328.643.032	60.731.230	96,92
TOTAL	456.511.568	46.107.668.368	441.068.204	Média: 96,6

5.2 Montagem *de Novo* dos Transcriptomas da Cachara e do Híbrido

A montagem *de novo* dos transcriptomas da cachara e dos híbridos gerou 389.562 *contigs* com comprimento médio de 1.158 pb e N50 de 2.269 pb (Tabela 3.2).

Tabela 3.2. Estatística da montagem *de novo* dos transcriptomas da cachara e dos híbridos

	Todos os contigs
Nº de contigs	389.562
Tamanho total dos contigs	451.169.903
Contig mais longo	30.166
Contig mais curto	201
Nº contigs > 1 K nt	134.659 (34,6%)
Nº contigs > 10 K nt	566 (0.1%)
Tamanho médio dos contigs	1.158
N50	2.269
%A no contig	28,56
%C no contig	21,46
%G no contig	21,75
%T no contig	28,23

5.3 Identificação dos SNPs

Foram identificados 647.954 polimorfismos nos 451.169.903 pb montados, ou seja, 1 SNP a cada 700 pb, aproximadamente, valor menor que as frequências encontradas em *Litopenaeus vannamei* (camarão branco do Pacífico: 1SNP a cada 476 pb) (Yu et al., 2014) e *Gadus morhua* (bacalhau do Atlântico: 1SNP a cada 516 pb) (Hubert et al., 2010). Segundo Quinn et al. (2013), o melhor método para alcançar elevadas sensibilidade (probabilidade de o RNA-Seq detectar e expressar um SNP se ele estiver presente em um gene transcrito) e especificidade (probabilidade de um SNP detectado por RNA-Seq ser uma variação verdadeira na sequência de DNA), além de grande número de SNPs é através da remoção de fragmentos duplicados após o alinhamento e da utilização do SAMtools (*software* utilizado neste estudo para a chamada de SNPs).

Dentre os 647.954 SNPs identificados, foram identificados 321 SNPs de alta qualidade, candidatos à SNPs espécie-específicos, que obedeciam aos critérios estabelecidos pelos filtros de qualidade aplicados. Polimorfismos com alelos fixados nas sequências das cacharas e alelos com frequências próximas de 50% no *pool* de híbridos foram considerados candidatos à SNPs espécie-específicos entre cacharas e pintados. Todas as sequências apresentaram 201 pb, ou seja, o SNP mais 100 pb de base à sua direita e 100 pb à sua esquerda, de forma a permitir o desenho de *primers* nessa região para o desenvolvimento de futuros ensaios de genotipagem.

Dentre os 321 SNPs identificados, as substituições do tipo transição foram mais comuns (67,9%) que as do tipo transversão (32,1%). Transição também foi o tipo de substituição encontrada com maior frequência nos SNPs identificados em *Litopenaeus vannamei* (66,8%) (Yu et al., 2014) e em bovinos da raça Limousin (73%) (Djari et al., 2013). Analisando especificamente as substituições de transição, a quantidade de transições G/A foi maior que C/T. Em termos de substituições de transversão, a frequência de A/T foi maior, seguida por /T, A/C e G/C (Tabela 3.3). A razão estimada de transição/transversão foi 2,11, maior que as encontradas para *L. vannamei* (2,0) (Yu et al., 2014), *Scophthalmus maximus* (1,354) e menor que a encontrada em bovinos (2,65) (Djari et al., 2013).

Tabela 3.3. Tipos de substituições presentes nos 321 SNPs espécie-específicos identificados nos transcriptomas de *P. reticulatum* e dos Híbridos (*P. reticulatum* x *P. corruscans*)

	Transição		Transversão			
	GA	CT	GC	GT	AC	AT
Número	168	50	11	29	22	41
Porcentagem	52,3%	15,6%	3,4%	9,0%	6,9%	12,8%
Total	218		103			

Os 321 SNPs candidatos à SNPs espécie-específicos estavam distribuídos em 292 transcritos. Desses 292 transcritos, a maioria (92,1%) apresentou um SNP em sua sequência (92,1%). Também foram encontrados transcritos com mais de um SNP na sua sequência, porém em menor frequência: 6,2% dos transcritos continham 2 SNPs na sua sequência; 1,4% continha 3 SNPs e 0,3% com 4 SNPs na sua sequência. Para *L. vannamei* (Yu et al., 2014), unigenes contendo 1 SNP na sua sequência também foram mais comuns e 93% dos unigenes continham até 10 SNPs na sua sequência.

5.4 Anotação

Entre os 292 transcritos contendo polimorfismos candidatos à SNPs espécie-específicos, 266 (91,1%) apresentaram *hit* significativo para proteína no banco de dados não redundante (nr).

A Ontologia Gênica (GO) categoriza os produtos gênicos e padroniza sua representação entre as espécies (Consortium, 2008), fornecendo um vocabulário controlado para descrevê-los em três categorias ontológicas independentes: função molecular, processo

biológico e componente celular (Ashburner et al., 2000). Após a anotação de Ontologia Gênica (GO), 222 (76,0%) transcritos foram assinados com um ou mais termos GO (Figura 3.3).

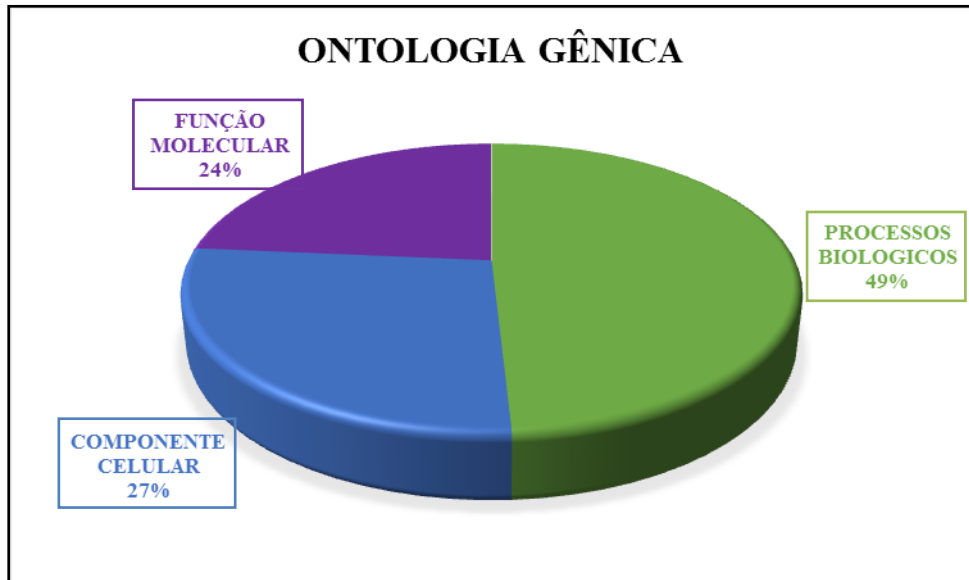


Figura 3.3. Designações de Ontologia Gênica (GO) para 222 dos transcritos que contêm polimorfismos candidatos à SNPs espécie-específico para diferenciar *P. reticulatum* e *P. corruscans* de híbridos identificados nos transcriptomas da cachara e dos híbridos.

A categoria Processos Biológicos apresentou 796 designações GO (49%); Componentes Celulares 442 designações GO (27%) e Função Molecular, 382 designações GO (24%). No 2º nível das três áreas da Ontologia Gênica, na categoria Processos Biológicos 4,7% dos transcritos contendo SNPs candidatos à espécie-específicos estavam relacionados com transcrição DNA-dependente; 2,6% com regulação da transcrição DNA-dependente e 2,3% com processo apoptótico. Na categoria Componentes Celulares, 41% dos transcritos estavam relacionados ao núcleo, 37,2% estavam relacionados com citoplasma e 20,9% com integral à membrana. Finalmente, na categoria Função Molecular, 22,7% dos genes estavam associados com ligação íon metal; 17,4% ligação íon zinco e 14,5% com ligação ATP. A Figura 3.4 apresenta o 2º nível de designações de cada uma das três categorias GO: Processo Biológico, Componente Celular e Função Molecular. São apresentados os 10 termos mais frequentes em cada categoria GO.

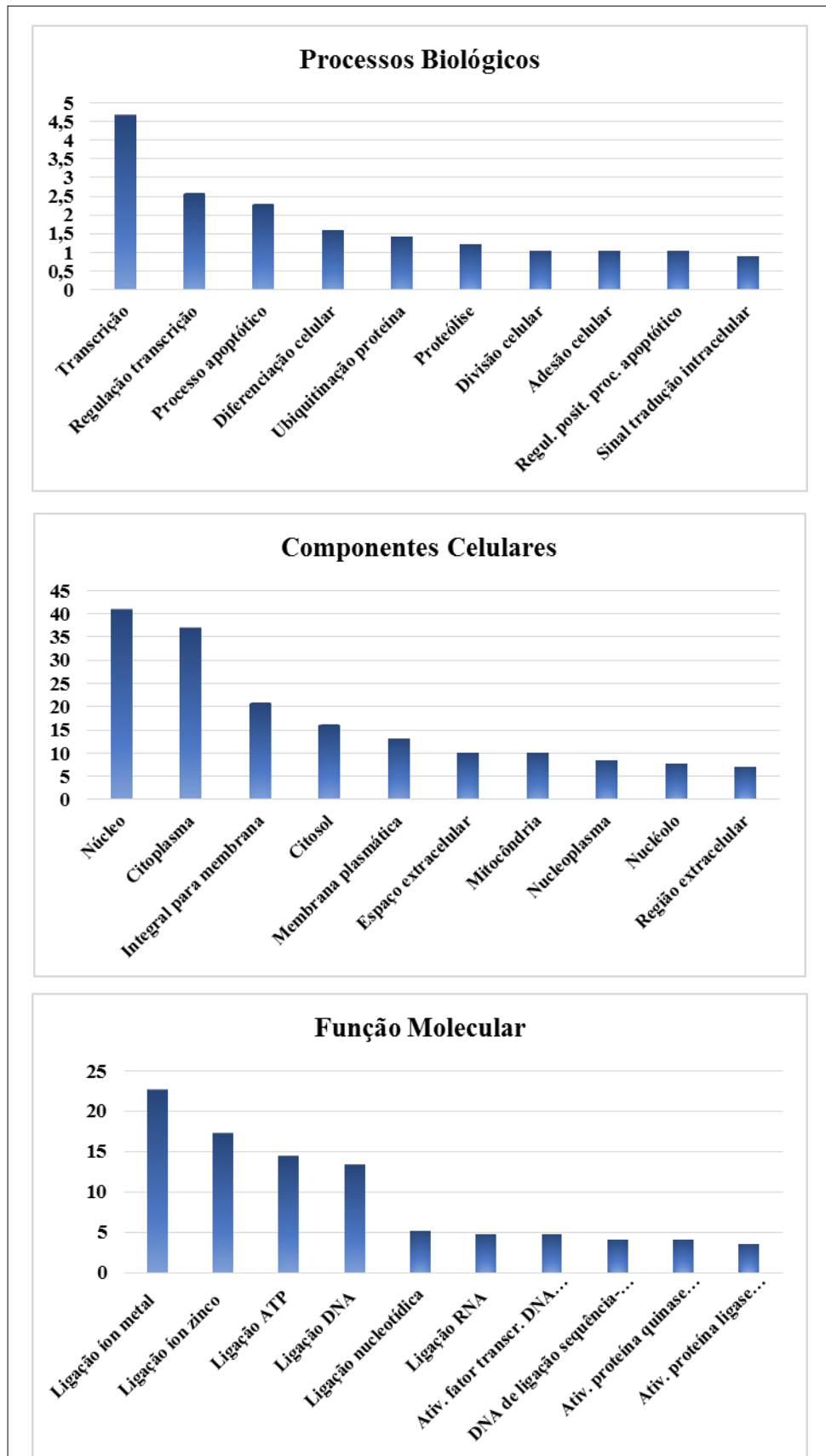


Figura 3.4. Designações do 2º nível de Ontologia Gênica (GO) para 222 transcritos com candidatos à SNPs espécie-específico para diferenciar *P. reticulatum* e *P. corruscans* de híbridos, identificados nos transcriptomas da cachara e dos híbridos.

5.5 Alinhamento dos SNPs identificados com genoma referência do *Danio rerio*

É importante ressaltar que as sequências que deram origem aos 321 SNPs espécie-específicos são de RNA. Portanto, para o desenvolvimento de ensaios diagnósticos esse fato deve ser levado em consideração e só ser aplicado em RNA. Caso seja aplicado em DNA, há uma chance dos *primers* desenhados para o teste se localizarem em regiões de junção de *splice* (limite entre intron-exon), o que significa que eles não anelariam na sequência de DNA.

Neste sentido, as 321 sequências dos transcritos contendo candidatos à SNPs espécie-específicos foram alinhadas contra o genoma referência do *Danio rerio*. Exigiu-se que este alinhamento contivesse o ponto de fusão 101 (o SNP) e pelo menos 20 pb à direita e à esquerda do SNP. Das 321 sequências dos transcritos inicialmente identificadas no transcriptoma da cachara e dos híbridos (pintado x cachara), apenas 64 alinharam com a sequência do *Danio rerio*. Isso significa que nesses 64 SNPs não há região de junção de *splice* nas sequências flangeadoras dos SNPs (pelo menos em 20 pb à direita e à esquerda do SNP) e eles podem ser selecionados para compor um painel de SNPs espécie-específicos para diferenciar indivíduos puros das espécies *P. reticulatum* e *P. corruscans* de indivíduos híbridos em ensaios de genotipagem (Tabela 3.4).

Tabela 3.4. Transcritos contendo os SNPs candidatos à SNPs espécie-específicos entre *P. reticulatum* e *P. corruscans* que alinharam com o genoma referência do *Danio rerio*.

	Identificação do transcrito	Posição SNP	Alelos do SNP	Tamanho sequência (pb)	Posição inicial do alinhamento	Posição final do alinhamento
1	comp118086_c0_seq1	709	[A/T]	148	37	184
2	comp122524_c0_seq1	467	[A/G]	141	13	153
3	comp123004_c0_seq1	666	[G/A]	187	1	187
4	comp123469_c0_seq1	3684	[T/C]	162	22	183
5	comp124502_c0_seq1	697	[G/A]	95	36	130
6	comp126607_c0_seq1	390	[A/C]	71	63	133
7	comp128430_c0_seq1	1949	[T/C]	124	60	183
8	comp129749_c0_seq1	1443	[A/G]	64	76	139
9	comp130330_c0_seq1	1364	[G/A]	118	65	181
10	comp130330_c0_seq1	2191	[A/G]	154	15	168
11	comp131296_c0_seq1	1764	[G/A]	124	78	201
12	comp131468_c3_seq1	1431	[A/G]	151	1	151
13	comp131753_c0_seq2	1147	[T/C]	84	59	142

14	comp131897_c0_seq1	2046	[G/A]	132	6	136
15	comp133209_c1_seq1	352	[G/A]	147	56	201
16	comp134413_c2_seq1	1215	[G/A]	79	54	132
17	comp134495_c0_seq1	2414	[A/G]	94	55	148
18	comp134829_c0_seq1	2485	[G/A]	72	60	130
19	comp134835_c0_seq1	810	[T/C]	104	53	156
20	comp137291_c0_seq1	976	[G/A]	116	77	192
21	comp138130_c0_seq1	1861	[T/C]	183	4	186
22	comp138311_c0_seq1	1132	[T/C]	121	6	126
23	comp138311_c0_seq1	1132	[T/C]	161	3	163
24	comp139083_c2_seq1	466	[T/C]	164	27	190
25	comp139173_c0_seq1	765	[G/A]	122	3	124
26	comp139764_c1_seq1	2444	[G/T]	142	60	201
27	comp141395_c0_seq1	770	[A/C]	145	57	201
28	comp141395_c0_seq1	770	[A/C]	139	63	201
29	comp141898_c0_seq1	1169	[T/C]	164	37	199
30	comp141943_c0_seq1	3816	[T/C]	157	12	168
31	comp141943_c0_seq1	3816	[T/C]	125	6	130
32	comp142287_c3_seq1	766	[T/A]	127	1	127
33	comp142287_c3_seq1	766	[T/A]	127	1	127
34	comp142287_c3_seq1	766	[T/A]	143	1	143
35	comp142287_c3_seq1	766	[T/A]	119	9	127
36	comp142312_c6_seq1	738	[A/G]	134	66	199
37	comp142312_c6_seq1	738	[A/G]	130	72	201
38	comp142312_c6_seq1	738	[A/G]	122	78	199
39	comp142724_c1_seq4	933	[T/C]	64	75	138
40	comp143224_c1_seq1	1567	[G/A]	69	65	133
41	comp143380_c5_seq1	479	[T/C]	109	78	186
42	comp143592_c1_seq5	418	[A/G]	88	64	151
43	comp143592_c1_seq5	418	[A/G]	65	63	127
44	comp144984_c2_seq1	316	[A/G]	74	65	138
45	comp144984_c2_seq1	316	[A/G]	77	65	141
46	comp144984_c2_seq1	316	[A/G]	77	65	141
47	comp144984_c2_seq1	316	[A/G]	77	65	141
48	comp144984_c2_seq1	316	[A/G]	78	65	142
49	comp144984_c2_seq1	316	[A/G]	77	65	141
50	comp144984_c2_seq1	316	[A/G]	77	65	141
51	comp145254_c0_seq1	10728	[A/G]	148	3	150
52	comp145471_c0_seq2	829	[G/C]	166	31	194
53	comp146582_c2_seq1	1272	[A/G]	120	15	134
54	comp146981_c0_seq1	477	[G/T]	119	59	176

55	comp146981_c0_seq1	477	[G/T]	119	59	176
56	comp147024_c1_seq1	2360	[T/G]	180	1	180
57	comp147841_c3_seq2	421	[T/A]	128	17	144
58	comp148553_c2_seq1	1763	[A/G]	156	2	157
59	comp148553_c2_seq1	1763	[A/G]	145	13	157
60	comp83908_c2_seq1	3370	[A/G]	165	1	163
61	comp83908_c2_seq1	3370	[A/G]	165	1	163
62	comp83908_c2_seq1	3370	[A/G]	165	1	163
63	comp83908_c2_seq1	3370	[A/G]	165	1	163
64	comp83908_c2_seq1	3370	[A/G]	163	1	162

A cachara e o pintado são espécies de bagres cujo genoma ainda não foi sequenciado. Portanto, seu tamanho ainda é desconhecido. Com base no valor C (quantidade de DNA contida dentro de um núcleo haplóide de um organismo eucarionte) é possível estimar o tamanho do genoma de uma espécie. No entanto, o valor C dessas duas espécies também é desconhecido. De acordo com Smith & Gregory (2009), a média do valor C de algumas espécies da ordem Siluriformes, família Pimelodidae, ordem e família a que pertencem pintado e cachara, é 1,12 pg. Considerando que o valor C da cachara e do pintado seja igual à 1,12 pg, seu genoma teria um tamanho aproximado de 1 Gpb. Para um genoma desse tamanho (1 Gpb), 64 SNPs (ou seja, 1 SNP a cada 17Mpb) seriam suficientes para detectar introgressões da ordem de 1,56% (equivalente à 5 gerações de retrocruzamento de um híbrido F1). Os 64 SNPs espécie-específicos prospectados nos transcriptomas da cachara e dos híbridos seriam suficientes portanto, para a detecção desse nível de introgressão. Para que possam ser utilizados em futuros ensaios de genotipagem, esses SNPs precisam ser validados, através do desenho de *primers* nas suas regiões flangeadoras.

Em relação aos 257 transcritos que não apresentaram alinhamento significativo com o genoma do *Danio rerio*, não há como afirmar com certeza se isso ocorreu devido à uma divergência entre as espécies alinhadas (*P. reticulatum* e *P. corruscans* x *Danio rerio*) que impediu ao alinhamento ou se os transcritos estavam realmente interrompidos por introns (localizados em regiões de junção de *splice*), o que impossibilitaria seu uso nos ensaios de genotipagem. No entanto, eles não devem ser excluídos do painel sem antes passar por um ensaio de validação. Não havendo região de junção de *splice* na região flangeadora dos SNPs, eles poderão ser incluídos no painel, aumentando a quantidade de SNPs espécie-específicos identificados neste experimento.

6 CONCLUSÕES

O RNA-Seq permitiu a prospecção de 64 polimorfismos candidatos à SNPs espécie-específico no transcriptoma de duas espécies de bagres importantes para a aquicultura nacional (cachara e pintado), e que ainda não possuem o genoma sequenciado. A validação dos SNPs é necessária antes de poder utilizá-los em ensaios de genotipagem. Após a validação, um número maior de SNPs espécie-específicos pode vir a compor esse painel, uma vez que 257 SNPs que não alinharam com a sequência do genoma referência do *Danio rerio*, podem não conter regiões de intron em suas sequências e ser incluídos no painel.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ASHBURNER, M.; BALL, C.A.; BLAKE, J.A. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p.25–29. 2000.
- BOECKLEN, W.J.; HOWARD, D.J. Genetic analysis of hybrid zones: numbers of markers and power of resolution. **Ecology**, v. 78, n. 8, p.2611-2616, 1997.
- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. **Bioinformatics**, btu170. 2014.
- CAETANO, A.R. SNP markers: basic concepts, applications in animal breeding and management and perspectives for the future. **Revista Brasileira de Zootecnia**, v. 38, p.64-71, 2009.
- CONSORTIUM, 2008. The gene ontology project in 2008. **Nucleic Acids Research**, v. 36, p.D440–D444, 2008.
- DJARI, A.; ESQUERRÉ, D.; WEISS, B. et al. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. **BMC Genomics**, v. 14, 307. 2013.
- GAUTIER, M.; FOUCAUD, J.; AL, K.G. et al. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. **Molecular Ecology**, v. 22, p.3766–3779, 2013.
- GODINHO, H.P. Estratégias reprodutivas de peixes aplicadas à aquicultura: bases para o desenvolvimento de tecnologias de produção. **Revista Brasileira Reprodução Animal**, v. 31, p.351-360, 2007.
- GRABHERR, M.G.; HAAS, B.J.; YASSOUR M. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. **Nature Biotechnology**, v. 29, n. 7, p.644-652, 2011.
- HAAS, B.J.; PAPANICOLAOU, A.; YASSOUR, M. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols**, v. 8, n. 8, p.1494-1512, 2013.

- HASHIMOTO, D.T.; PRADO, F.D. DO; SENHORINI, J.A. et al. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. **Aquaculture Research**, v. 44, p.876-884, 2013.
- HELYAR, S.J.; LIMBORG, M.T.; BEKKEVOLD, B. et al. SNP discovery using next generation transcriptomic sequencing in Atlantic Herring (*Clupea harengus*). **Plos One**, v. 7, n. 8, p. 1-11, 2012.
- HUBERT, S.; HIGGINS, B.; BORZA, T.; BOWMAN, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). **BMC Genomics**, v. 11, 191. 2010.
- JEUKENS, J.; RENAUT, S.; ST-CYR, J. et al. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. **Molecular Ecology**, v. 19, n. 24, p.5389-5403, 2010.
- LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, n. 21, p. 2987-2993, 2011.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p.1754-1760, 2009.
- LIU, S.; ZHOU, Z., LU, J. et al. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. **BMC Genomics**, v. 12, n. 53, p.1-13, 2011.
- MAGOC, T.; SALZBERG, S. FLASH: Fast length adjustment of short reads to improve genome assemblies. **Bioinformatics** v. 27, n. 21, p.2957-2963, 2011.
- PRADO, F.D.; HASHIMOTO, D.T.; MENDONÇA, F.F. et al. Molecular identification of hybrids between Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum*. **Aquaculture Research**, v. 42, p.1890-1894, 2011.
- QUINN, E.M.; CORMICAN, P.; KENNY, E.M. et al. Development of strategies for SNP detection in RNA-Seq data: Application to lymphoblastoid cell lines and evaluation using 1000 genomes data. **Plos One**, v. 8, n. 3, p.1-9, 2013.
- SHEN, Y.; CATCHEN, J.; GARCIA, T. et al. Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F₁ interspecies hybrids. **Comparative Biochemistry Physiology, Part C**. v. 155, p.102-108, 2012.
- SMITH, E.M.; GREGORY, T.R. Patterns of genome size diversity in the ray-finned fishes. **Hydrobiologia**. v. 625, p.1-25, 2009.
- XU, J.; JI, P.; ZHAO, Z. et al. Genome-wide SNP discovery from transcriptome of four common carp strains. **Plos One**. v. 7, n. 10, p.1-9. 2012.

YU, Y.; WEI, J.; ZHANG, X. et al. SNP discovery in the transcriptome of white pacific shrimp *Litopenaeus vannamei* by next generation sequencing. **Plos One**. v. 9, n.1, p.1-9, 2014.

CAPÍTULO 4

CARACTERIZAÇÃO DO TRANSCRIPTOMA DA CACHARA (*Pseudoplatystoma reticulatum*): SEQUENCIAMENTO, MONTAGEM E ANÁLISE DE EXPRESSÃO TECIDO-ESPECÍFICA

1 RESUMO

A cachara (*Pseudoplatystoma reticulatum*), economicamente importante nas regiões onde ocorre no Brasil, é um bagre de água doce, pertencente à família Pimelodidae (ordem Siluriforme). Informações moleculares poderiam ajudar a alavancar o desenvolvimento da espécie, porém, ainda são relativamente escassas nas bases de dados públicas. Adicionalmente, a ausência de um genoma referência da cachara dificulta o desenvolvimento de tecnologias moleculares que poderiam ser aplicadas nas pisciculturas, de forma a contribuir com o melhoramento de sua produção. Diante deste cenário, o sequenciamento do transcriptoma constitui uma alternativa rápida e eficiente para a geração de informações moleculares (descoberta de genes, identificação de polimorfismos de base única (SNPs), desenvolvimento de marcadores moleculares, estudos de expressão gênica diferencial, estudos de expressão tecido específica, etc.) do que o sequenciamento do genoma. Um transcriptoma completo e bem anotado pode prover os bancos de dados públicos com informações até então inexistentes para a espécie em questão. Sendo assim, o objetivo neste estudo foi sequenciar o transcriptoma da cachara utilizando tecnologias Illumina (Mi Seq e HiSeq) de sequenciamento de nova geração (NGS). O sequenciamento MiSeq V3 Illumina gerou 47.214.078 fragmentos com 300 pb de comprimento e o sequenciamento HiSeq 2000

Illumina produziu 444.717.156 fragmentos com 100 pb. A montagem de novo do transcriptoma da cachara foi realizada com o *software* Trinity e gerou 312.766 contigs e 187.216 genes. Foram identificados 93.674 genes expressos (tamanho mínimo de 200 pb e FPKM \geq 1). De 11.456 à 58.494 genes foram expressos nos sete tecidos analisados. 37.917 genes tecido-específicos e 7.456 *housekeeping* foram identificados em uma análise conjunta envolvendo seis dos tecidos coletados (músculo branco, músculo vermelho, fígado, rim, hipófise e brânquias). Músculo branco foi o tecido que apresentou a menor quantidade de genes expressos (11.456 genes expressos) e genes tecido-específico (754 genes). A maioria dos genes expressos no músculo branco é expressa também nos outros tecidos. Músculo branco e músculo vermelho constituem um modelo interessante para a caracterização funcional de genes e de vias metabólicas envolvidas em seus processos biológicos.

Palavras chave: RNA-Seq; Illumina; Genes *housekeeping*; músculo branco; músculo vermelho

**CHARACTERIZATION OF CACHARA (*Pseudoplatystoma reticulatum*)
TRANSCRIPTOME: SEQUENCING, ASSEMBLY AND TISSUE-SPECIFIC
EXPRESSION ANALYSIS**

2 ABSTRACT

Cachara (*Pseudoplatystoma reticulatum*) it is a freshwater catfish, belonging to Family Pimelodidae (Order Siluriforms), and has economical importance in the Brazilian regions of its occurrence. Molecular information could help to boost the development of the specie, although it is relatively scarce on public databases. Additionally, the lack of genomic reference of cachara hardens the development of molecular technologies that could contribute to production improvement in the fish farms. In this scenario, the transcriptome sequencing is a faster and efficient alternative to provide molecular information (gene discovery, single nucleotide polymorphisms (SNPs), development of molecular tracers, diferential gene expression. studies, tissue-specific expression. studies, etc.) when compared with genome sequencing. A complete and well annotated transcriptome can provide the public databases with unprecedented information about the studied specie. Therefore, the aim of this study was the adoption of new generation sequencing Illumina technologies (MiSeq and HiSeq) for transcriptome sequencing of cachara. The sequencing MiSeq V3 Illumina generated 47,214,078 fragments with 300 pb in length and the sequencing HiSeq 2000 Illumina produced 444,717,156 fragments with 100 pb in lenght. For the *de novo* assembly of the cachara's transcriptome we used the software Trinity, generating 312,766 contigs and 187,216 genes. We identified 93,674 expressed genes (minimum size of 200 pb and FPKM ≥ 1). Were expressed between 11,456 and 58,494 genes in the seven analysed tissues. We identified 37,917 tissue-specific and 7,456 housekeeping in six tissues (white muscle, red muscle, liver,

kidney, hypophysis and gill). The white muscle tissue showed less quantity of expressed genes (11,456), tissue-specific genes (754). The majority of the expressed genes in the white muscle was also expressed in other tissues. White muscle and red muscle tissues compose an interesting model for the functional characterization of the gene and metabolic pathways related to its biological processes.

Keywords: RNA-Seq; Illumina; housekeeping genes; white muscle; red muscle.

3 INTRODUÇÃO

A cachara (*Pseudoplatystoma reticulatum*) é uma espécie de bagre de água doce encontrada nas bacias dos rios Paraguai-Paraná, embora existam também relatos de sua presença na bacia do rio Amazonas (Buitrago-Suárez & Burr, 2007; Torrico et al., 2009). A espécie pertence à família Pimelodidae, ordem Siluriforme, e apresenta grande importância econômica para a pesca nas regiões onde ocorre naturalmente (Buitrago-Suárez & Burr, 2007). A criação em cativeiro da cachara apresentou um aumento significativo nos últimos anos, reflexo do aumento da produção de diferentes bagres nativos e seus híbridos (*P. reticulatum*, *P. corruscans*, *P. punctifer*, etc) observado entre os anos de 2007 e 2011, que atingiu marcas superiores a 300% (Boletim da Pesca e Aquicultura de 2010 e 2011). A biologia da espécie ainda é pouco conhecida, e conhecimentos básicos sobre processos relacionados com sua reprodução, crescimento, engorda e adaptação ao ambiente de produção em cativeiro são necessários para o desenvolvimento de novas tecnologias que possam contribuir para melhorar a produtividade da cachara em sistemas aquícolas (Bignotto et al., 2009).

O *P. reticulatum* ainda não teve seu genoma sequenciado. Na ausência de um genoma referência, o sequenciamento de partes menores do genoma, como o transcriptoma, constitui uma alternativa rápida e eficiente para gerar informações moleculares que poderão ser aplicadas nos estudos e desenvolvimento de novas tecnologias para melhorar a produtividade da cachara criada em cativeiro e, conseqüentemente ajudar a alavancar o desenvolvimento da aquicultura no Brasil.

O sequenciamento de RNA (RNA-Seq) com tecnologias de sequenciamento de nova geração (NGS) possibilita a caracterização e o estudo do transcriptoma de uma espécie com cobertura, profundidade e eficiência elevadas (Nagalakshmi et al., 2010). Apesar de a maioria das plataformas de NGS atuais gerar seqüências curtas, o que implica em dificuldades

específicas para a montagem *de novo* de um transcriptoma (Lal & Seshasayee, 2014; Alkan et al., 2011), estratégias laboratoriais e algoritmos computacionais específicos foram gerados para permitir a montagem de transcriptomas de espécies que não possuem genomas referência (Yang et al., 2012). A caracterização do transcriptoma de diferentes tecidos também proporciona a identificação de genes *housekeeping* e de genes expressos em tecidos específicos (Butte et al., 2001; She et al., 2009), permitindo assim a identificação de processos e a formulação de estudos para aprofundar conhecimentos específicos sobre a biologia e o desenvolvimento de um organismo (Salem et al., 2015;).

Este estudo relata os resultados da montagem *de novo* e a caracterização inicial do transcriptoma da cachara (*P. reticulatum*) com NGS, através da identificação de genes *housekeeping* e genes com expressão tecido-específica.

4 MATERIAL E MÉTODOS

4.1 Amostragem, Extração de RNA Total, Confeção dos *Pools* de RNA Total e das Bibliotecas de cDNA, Sequenciamento e Montagem *de novo* do Transcriptoma da Cachara

A amostragem das 12 cacharas, extração de RNA total e confecção dos sete *pools* de RNA total de cada tecido da cachara e do *pool* com amostras de RNA de todos os tecidos, a construção das oito bibliotecas de cDNA das cacharas, os sequenciamentos Illumina HiSeq 2000 e Illumina MiSeq V3 e a montagem *de novo* do transcriptoma da cachara foram descritas em Material e Métodos, no Capítulo 3.

4.2 Identificação de Genes Constitutivos e Genes com Expressão Tecido-Específica

Após a montagem *de novo* do transcriptoma da cachara, foram considerados genes expressos nos sete tecidos analisados aqueles transcritos com valor FPKM (fragmentos por quilobase de transcritos por milhões de fragmentos mapeados) ≥ 1 e com tamanho mínimo de 200 pb foram selecionados como genes expressos.

Genes tecido-específico e genes constitutivos ou *housekeeping* foram identificados utilizando o *software* Jvenn (<http://bioinfo.genotoul.fr/jvenn/example.html>). Gráficos de Venn foram construídos para realizar as comparações dos genes tecido-específicos e *housekeeping* entre os tecidos. Foram considerados genes tecido-específicos os

genes que só apareciam em um tecido específico. Genes *housekeeping* foram aqueles que apareceram em todos os tecidos analisados.

5 RESULTADOS E DISCUSSÃO

5.1 RNA-Seq e Montagem do Transcriptoma de *Pseudoplatystoma reticulatum*

Considerando as oito bibliotecas de cDNA da cachara sequenciadas nas duas plataformas Illumina (MiSeq e HiSeq), um total de 491.931.234 sequências e 591.278.702.34 pb foram gerados. O sequenciamento MiSeq (Illumina) da biblioteca de cDNA gerada a partir de amostras de todos os sete tecidos coletados produziu 28.721.880 de sequências (300 pb de comprimento) de alta qualidade, o que corresponde à 60,83% do total de fragmentos sequenciados. A combinação dessas sequências produziu aproximadamente 12 milhões de *super-reads*, com comprimento médio de 392 pb (Figura 3.1). Já o sequenciamento *paired-end* Illumina HiSeq 2000 das sete bibliotecas de cDNA de tecidos individuais produziu 429.706.374 de sequências (100 pb) de alta qualidade, o que corresponde a 96,6% dos fragmentos sequenciados. O resultado do sequenciamento de cada uma das sete bibliotecas de cDNA da cachara foi fornecido na Tabela 3.1. Informações mais detalhadas sobre os resultados referentes ao sequenciamento das bibliotecas de cDNA da cachara foram apresentadas em Resultados e Discussão, no Capítulo 3.

A montagem *de novo* do transcriptoma da cachara gerou 312.766 *contigs* e 187.216 genes (Tabela 4.1).

Tabela 4.1. Estatística da montagem *de novo* do transcriptoma da cachara

	Todos os contigs	Contigs ≥ 200 pb e FPKM ≥ 1
Nº de contigs	312.766	128.581
Tamanho total dos contigs	391.280.046	140.153.805
Contig mais longo	26.274	26.274
Contig mais curto	201	201
Nº contigs > 1 K nt	106.043 (33,9%)	42.462 (33,0%)
Nº contigs > 10 K nt	1.663 (0,5%)	157 (0,1%)
Tamanho médio dos contigs	1.251	1.090
N50	2.765	2.230
%A no contig	28,26	28,30
%C no contig	21,74	21,66
%G no contig	22,00	21,84
%T no contig	27,99	28,19

As partes do genoma transcritas de uma célula ou de um tecido refletem os processos biológicos e as funções que ela realiza. Neste sentido, o RNA-Seq tem ajudado a mapear e a anotar transcriptomas de peixes, e a entender melhor alguns de seus processos biológicos (Qian et al., 2014). Vem sendo empregado com sucesso na caracterização do transcriptoma de diferentes espécies de peixes, como a truta arco-íris (*Oncorhynchus mykiss*, Salem et al., 2015), o guppy selvagem (*Poecilia reticulata*, Fraser et al., 2011), o peixe-serra do Atlântico (*Sarda sarda*, Sarropoulou et al., 2014), a carpa comum (*Cyprinus carpio*, Ji et al., 2012), dentre outras espécies.

5.5. Identificação de Genes Constitutivos e com Expressão Tecido-Específica

5.5.1 Genes expressos nos tecidos analisados

Foram identificados 93.674 genes expressos (transcritos com FPKM ≥ 1 e comprimento ≥ 200 pb) no transcriptoma de *Pseudoplatystoma reticulatum*, nos sete diferentes tecidos avaliados (músculo branco, músculo vermelho, fígado, hipófise, gônada, rim e brânquia). Para um genoma dessa complexidade (em torno de 1Gpb), esperava-se encontrar entre 20 à 30 mil genes expressos. Diversos fatores (ou a somatória deles) podem ser responsáveis por esse número mais elevado de genes expressos identificados: a) a utilização de um *pool* de indivíduos (e não apenas um indivíduo). A variação entre indivíduos

pode resultar em variantes alélicas de um mesmo gene, o que pode resultar em um número maior de genes expressos; b) para o Trinity, diferentes isoformas de um gene são consideradas diferentes transcritos. Essa pode ser uma das razões do elevado número de genes expressos encontrado; c) deve-se levar em consideração também a duplicação de genes expressos nos peixes ligados ao metabolismo.

A Figura 4.1 apresenta o número de genes expressos em cada tecido analisado. De 11.456 à 58.494 genes foram expressos nos sete tecidos analisados.

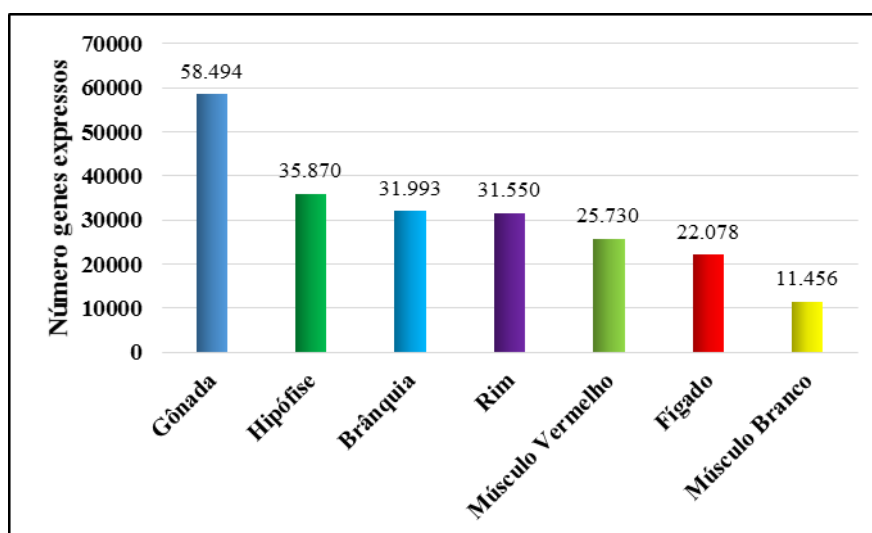


Figura 4.1. Número de genes expressos em cada tecido analisado de *Pseudoplatystoma reticulatum* com FPKM ≥ 1 .

De acordo com a Figura 4.1, gônada foi o tecido que apresentou a maior quantidade de genes expressos (58.494 genes expressos), seguido por hipófise (35.870 genes expressos), brânquia (31.993 genes expressos) e rim (31.550 genes expressos). Em contrapartida, músculo branco foi o tecido que apresentou a menor quantidade de genes expressos (11.456 genes expressos), seguido por fígado (22.078 genes expressos) e músculo vermelho (25.730 genes expressos). A expressão de transcritos por tecido apresentou tendência parecida com a identificada em truta arco-íris (*Oncorhynchus mykiss*, Salem et al., 2015), para a qual cérebro e testículo apresentaram maior quantidade de genes expressos e músculo branco, fígado e músculo vermelho apresentaram menor quantidade de genes expressos.

5.6 Identificação de Genes Constitutivos e com Expressão Tecido-Específica

Genes cujas funções e expressões são favorecidas em um determinado tecido ou tipo celular são conhecidos como genes tecido-específico. Sua expressão é fundamental para manter a especificidade de um determinado tecido, ainda que possam ser expressos em níveis muito baixos ou mesmo não ser expressos em outros tecidos. Sua identificação ajuda a entender a relação existente gene e tecido (Xiao et al., 2010; She et al., 2009). Por outro lado, genes expressos constitutivamente (em todos os tecidos ou na maioria deles), normalmente têm o objetivo de manter as funções básicas de uma célula e são conhecidos como genes *housekeeping* (Butte et al., 2001; She et al., 2009).

Neste estudo, o sequenciamento do transcriptoma de diferentes tecidos da cachara possibilitou a identificação de genes *housekeeping* e tecido-específicos expressos seis dos tecidos analisados: músculo branco, músculo vermelho, rim, fígado, brânquia e hipófise. As informações sobre os genes expressos na gônada não foram incluídas, uma vez que o *pool* continha amostras de ovário e testículo juntas, o que poderia ser responsável pela maior quantidade de genes expressos neste tecido quando comparada aos demais.

Foram identificados nos seis tecidos um total de 37.917 genes tecido-específicos e 7.456 genes *housekeeping* com FPKM ≥ 1 e tamanho mínimo de 200 pb. Pelo gráfico de Venn (Figura 4.2) é possível visualizar os 7.456 genes *housekeeping* identificados entre estes seis tecidos, além dos genes tecido-específicos para cada um destes seis tecidos.

A quantidade de genes tecido-específicos expressos por tecido, dentre os seis tecidos considerados nesta análise, pode ser observada na Tabela 4.2.

Hipófise expressou o maior número de genes tecido-específicos (32,3%), seguido por brânquia (20,0%) e rim (19,4%). Por outro lado, músculo branco expressou a menor quantidade de genes tecido-específicos (2,0%), fígado (13,0%) e músculo vermelho (13,3%). Em relação aos genes *housekeeping*, foram encontrados em maior quantidade nas brânquias (20,2%), seguido por rim (20,0%) e hipófise (19,5%). Músculo branco expressou a menor quantidade de genes *housekeeping* (8,9%), seguido por fígado (14,2%) e músculo vermelho (17,1%).

Dos 93.674 genes expressos, 37.917 genes foram considerados tecido-específico e 7.456 foram considerados *housekeeping* entre os seis tecidos analisados. Os demais genes, foram expressos em dois ou mais tecidos, mas não em todos os tecidos.

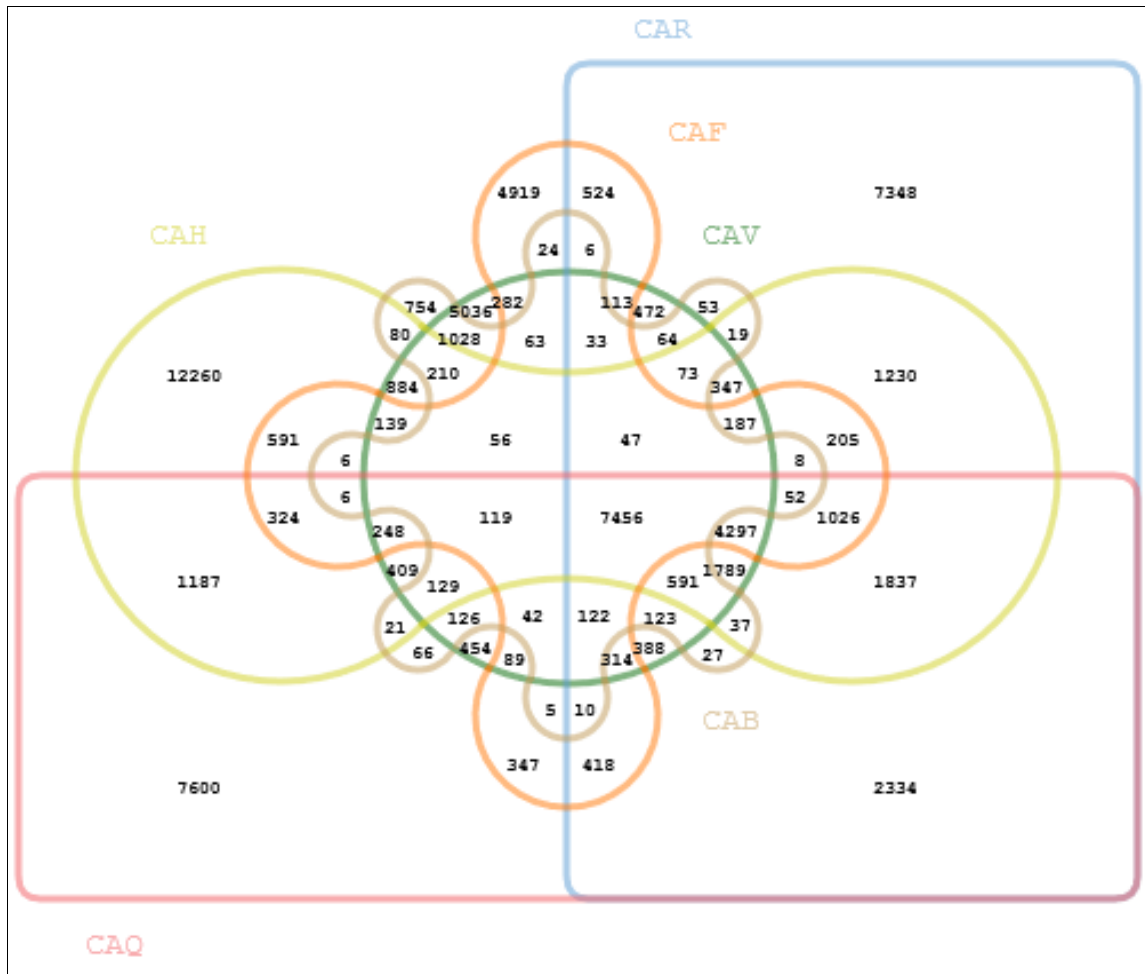


Figura 4.2. Diagrama de Venn evidenciando a expressão de genes tecido-específico e *housekeeping* nos seis tecidos analisados do transcriptoma da cachara. As siglas CAR, CAF, CAV, CAB, CAQ e CAH na figura correspondem aos tecidos rim, fígado, músculo vermelho, músculo branco, brânquias e hipófise de cachara, respectivamente.

Tabela 4.2. Genes tecido-específico identificados nos seis tecidos analisados

	Tecido-Específico
Hipófise	12.260
Brânquia	7.600
Rim	7.348
Músculo Vermelho	5.036
Fígado	4.919
Músculo Branco	754
TOTAL	37.917

5.6.1 Genes expressos no músculo branco e no músculo vermelho

Comparado aos outros cinco tecidos analisados, músculo branco foi o tecido que apresentou as menores quantidades de genes expressos (11.456) e genes tecido-específico (754). Os diagramas de Venn comparando os genes expressos no músculo branco com os genes expressos nos outros quatro tecidos analisados podem ser visualizados na Figura 4.3.

A comparação entre os genes expressos no músculo branco e os genes expressos no músculo vermelho é interessante. Dos 11.456 genes expressos no músculo branco, 10.282 genes (praticamente 90%) também foram expressos no músculo vermelho. Apenas 1.174 genes expressos (aproximadamente 10%) foram expressos apenas no músculo brancogenes tecido-específico. Em contrapartida, dos 25.730 genes expressos no músculo vermelho, 15.448 genes (60%) eram expressos apenas no músculo vermelho.

A maioria dos genes expressa no músculo branco também é expressa nos outros tecidos analisados neste estudo (Figura 4.3). Esse resultado indica que a maioria dos genes expressos pelo músculo branco possui um papel biológico relacionado às funções celulares básicas, necessárias para sustentação da vida, incluindo metabolismo, processos celulares e regulação biológica. O músculo branco é um tecido que apresenta poucos genes com funções específicas para este tecido. Provavelmente, os poucos genes tecido-específicos expressos pelo músculo branco (células especializadas), são altamente expressos. Em estudo desenvolvido por Salem et al. (2015) com a truta arco-íris, miosina e troponina estavam entre os transcritos tecido-específicos mais altamente expressos no músculo branco. Ambas têm importante papel na contração muscular. Transcritos característicos do músculo lento foram identificados no músculo vermelho (Cadeia de miosina leve, Troponina-I, Troponina-T lenta, Cadeia pesada da miosina lenta) (Salem et al., 2015).

Neste sentido, o tecido muscular esquelético dos peixes constitui grande parte da sua massa corporal (Bone, 1978). É formado por três tipos de fibras musculares ou músculos: branco, vermelho e rosa (ou intermediário), que, com base na sua velocidade de contração muscular são classificadas em fibras de contração rápida (músculo branco) e fibras de contração lenta (músculos vermelho e rosa). As fibras de contração rápida compreendem a maior parte da massa muscular (aproximadamente 60%), e são recrutadas para movimentos mais bruscos de natação, associados com movimentos de captura de presa e alimento e escape. Constituídas por fibras musculares maiores, possuem metabolismo glicolítico; apresentam reduzido conteúdo de mioglobina, mitocôndrias e lipídeos. Em relação às fibras

de contração lenta, especificamente a musculatura vermelha compreende aproximadamente 30% do volume da massa muscular, está restrita à região da linha lateral e é formada por fibras menores. Apresenta metabolismo oxidativo; alta concentração de mioglobina, mitocôndrias e lipídeos e excelente suprimento sanguíneo. É utilizada na realização de movimentos lentos e de sustentação da natação. A musculatura rosa é encontrada entre as musculaturas branca e vermelha, e apresenta velocidade de contração e características metabólicas intermediárias (Carani et al., 2008; Mareco et al., 2015; Coughlin 1996).

Os peixes constituem, portanto, um modelo interessante para o estudo dos processos biológicos, regulatórios e metabólicos envolvidos na locomoção e no crescimento, uma vez que as diferentes fibras que compõem seu tecido muscular esquelético encontram-se anatômica e funcionalmente separadas (e não entremeadas, como nos mamíferos) e o mecanismo de hiperplasia nos peixes que atingem tamanhos maiores não cessa logo após o nascimento, como ocorre com os mamíferos (Altringham e Johnston, 1990; Coughlin 1996). O crescimento muscular envolve processos de síntese e degradação de proteínas. Nestes processos, centenas de genes estruturais e reguladores estão envolvidos.

A montagem e a caracterização do transcriptoma de uma espécie permitem identificar sequências específicas de genes envolvidos com processos biológicos importantes para a produção animal (Salem et al., 2012; Wang et al., 2009). Os resultados da expressão tecido-específico serão importantes para revelar como os padrões de expressão são regulados nos diferentes tecidos, além de contribuir com o estudo das vias metabólicas nas quais eles estão envolvidos.



Figura 4.3. Gráficos/Diagramas de Venn apresentando a comparação da expressão tecido-específica e *housekeeping* do músculo branco e outros cinco tecidos de *Pseudoplatystoma reticulatum*.

6 CONCLUSÕES

O sequenciamento com NGS e a montagem do transcriptoma da cachara permitiu a identificação de genes expressos nos sete diferentes tecidos. 37.917 genes tecido-específicos (transcritos com FPKM ≥ 1 , tamanho mínimo de 200 pb) e 7.456 genes *housekeeping* foram identificados em seis dos tecidos analisados. O padrão de expressão dos genes tecido-específicos e *housekeeping* no músculo branco e no músculo vermelho da cachara é interessante para o desenvolvimento de estudos de caracterização funcional dos genes e vias metabólicas envolvidos nos processos biológicos, regulatórios e metabólicos da locomoção e do crescimento.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALKAN, C.; SAJJADIAN, S.; EICHLER, E.E. Limitations of next-generation genome sequence assembly. **Nature Methods**, v. 8, n. 1, 2011.
- ALTRINGHAM, J.D.; JOHNSTON, I.A. Modelling muscle power output in a swimming fish. **Journal of Experimental Biology**, v. 148, p.395-402, 1990.
- BIGNOTTO, T.S.; PRIOLI A.J.; PRIOLI, S.M.A.P. et al. Genetic divergence between *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Siluriformes: Pimelodidae) in the Paraná River Basin. **Brazilian Journal of Biology**, v. 69, Suppl. 2, p.681-689, 2009.
- BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA. BRASIL 2010. Ministério da Pesca e Aquicultura. 2011.
- BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA. BRASIL 2011. Ministério da Pesca e Aquicultura. 2011.
- BONE, Q.; KICENIUK, J.; JONES, D.R.. On the role of the different fibre types in fish myotomes at intermediate swimming speeds. **Fishery Bulletin Fish**, v. 76, p.691-699, 1978.
- BUITRAGO-SUÁREZ, U.A.; BURR, B.M. Taxonomy of the catfish genus *Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. **Zootaxa**, v. 1512, p.1-38, 2007.
- BUTTE, A.J.; DZAU, V.J.; GLUECK, S.B. Further defining housekeeping, or “maintenance,” genes Focus on “A compendium of gene expression in normal human tissues”. **Physiological Genomics**, v. 7, p.95-96, 2001.
- CARANI F.R.; AGUIAR, D.H.; ALMEIDA, F.L.A. de et al. Morfologia e crescimento do músculo estriado esquelético no pirarucu *Arapaima gigas* Cuvier, 1817 (Teleostei, Arapaimidae). **Acta Scientiarum Biological Sciences**, v. 30, n. 2, p.205-211, 2008.
- COUGHLIN, D.J. The roles of pink and red muscle in powering steady swimming in scup, *Stenotomus chrysops*. **American Zoologist**, v. 36, p.666-677, 1996.

- FRASER, B.A.; WEADICK, C.J.; JANOWITZ, I. et al. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. **BMC Genomics**, v. 12, p.202, 2011.
- JI, P.; LIU, G.; XU, J. et al. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. **Plos One**, v.7, n. 4, p.1-9, 2012.
- LAL, A.; SESHASAYEE, A.S.N. The impact of next-generation sequencing technology on bacterial genomics. KULKARNI, V.V. et al. (eds.). A systems theoretic approach to systems and synthetic biology II: Analysis and design of cellular systems. DOI: 10.1007/978-94-017-9047-5_2, 2014.
- MARECO, E.A.; SERRANA, D.G. de la; JOHNSTON, I.A. et al. Characterization of the transcriptome of fast and slow muscle myotomal fibres in pacu (*Piaractus mesopotamicus*). **BMC Genomics**, v. 16, p.182, 2015.
- NAGALAKSHMI, U.; WAERN, K.; SNYDER, M. RNA-Seq: A Method for Comprehensive Transcriptome Analysis. **Current Protocols in Molecular Biology**, 2010.
- QIAN, X.; BA, Y.; ZHUANG, Q. et al. RNA-Seq technology and its application in fish transcriptomics. **OMICS A Journal of Integrative Biology**, v. 18, n. 2, 2014.
- SALEM, M.; VALLEJO, R.L.; LEEDS, T.D. et al. RNA-Seq identifies snp markers for growth traits in rainbow trout. **Plos One**, v. 7, n. 5, p.1-13, 2012.
- SALEM, M.; PANERU, B.; AL-TOBASEI, R. et al. Transcriptome assembly, geneannotation and tissue gene expression atlas of the rainbow trout. **Plos One**, DOI: 10.1371/journal.pone.0121778, p. 1-27, 2015.
- SARROPOULOU, E.; MOGHADAM, H.K.; PAPANDROULAKIS, N. et al. The Atlantic Bonito (*Sarda sarda*, Bloch 1793) transcriptome and detection of differential expression during larvae development. **Plos One**, v. 9, n. 2, p.1-13, 2014.
- SHE, X.; ROHL, C.A.; CASTLE, J.C. et al. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. **BMC Genomics**, v. 10, p.269, 2009.
- TORRICO, J. P.; HUBERT, N.; DESMARAIS, E. et al. Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): biogeographic and evolutionary implications. **Molecular Phylogenetics and Evolution**, v. 51, n. 3, p.588-594, 2009.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. The Review provides a comprehensive overview of recent advances and challenges in techniques that are used in transcriptome profiling methods that use NGS technologies (RNA-seq). **Nature Reviews Genetic**, v. 10, p.57-63, 2009.
- XIAO, S.J.; ZHANG, C.; ZOU, Q. et al. TiSGeD: a database for tissue-specific genes. **Bioinformatics applications note**, v. 26, n. 9, p. 1273-1275, 2010.

CAPÍTULO 5

SEQUENCIAMENTO, MONTAGEM *DE NOVO* E ANÁLISE DESCRITIVA DO GENOMA MITOCONDRIAL COMPLETO DA CACHARA (*Pseudoplatystoma reticulatum*)

1 RESUMO

A cachara (*Pseudoplatystoma reticulatum*) é uma espécie de peixe Neotropical de água doce, pertencente à família Pimelodidae (Siluriformes) e de relevante importância econômica para o Brasil, porém, pouco ainda se conhece sobre sua biologia. A geração de informações genômicas, como a disponibilização da sequência completa do genoma mitocondrial (mitogenoma) de *P. reticulatum*, poderão ajudar a aprofundar o conhecimento biológico. O mitogenoma completo da cachara foi obtido através do sequenciamento Illumina (MiSeq e HiSeq) de oito bibliotecas de cDNA, montadas a partir de RNA total extraído de sete diferentes tecidos coletados de 12 exemplares. Foi realizada a montagem *de novo* do transcriptoma da cachara utilizando o software Trinity. Três transcritos que alinhavam com o mtDNA de outras espécies de peixes foram montados em uma sequência única utilizando o programa SSAKE. A região do D-Loop que não foi transcrita, foi obtida por sequenciamento Sanger. O mtDNA completo da cachara apresentou 16.576 pb de comprimento, 13 genes codificadores de proteína, dois genes de rRNAs, 22 de tRNAs, além da região controle (D-Loop). O mitogenoma é composto por 24,63% de T; 28,47% de C; 31,45% de A e 15,44% de G e apresentou a mesma estrutura básica, ordem e organização gênica observada em outras espécies de vertebrados. Com exceção do gene codificador de proteína *NAD6* e oito genes

tRNAs, todos os demais genes mitocondriais localizaram-se na fita H. Este é o primeiro relato da sequência completa do mtDNA de *P. reticulatum*. Espera-se que a disponibilização desse mitogenoma contribua para a realização de estudos adicionais em diferentes áreas, como filogeografia, sistemática filogenética, genética de populações, conservação e melhoramento genético.

Palavras-chave: Mitogenoma; Siluriformes; RNA-Seq; Bagres; Recursos genéticos animais.

**SEQUENCING, *DE NOVO* ASSEMBLY AND COMPLETE DESCRIPTIVE
MITOCHONDRIAL GENOME ANALYSIS OF CACHARA (*Pseudoplatystoma
reticulatum*)**

2 ABSTRACT

Cachara (*Pseudoplatystoma reticulatum*) is a freshwater Neotropical fish specie belonging to Family Pimelodidae (Siluriforms). It has relevant economical importance to Brazil, but little is known about its biology. This is one of the factors that delay its effective use in Brazil. Genomical information can boost the biological knowledge of the specie and consequently enhance its production performance. Therefore, this study presents the mitochondrial genome (mitogenome) complete sequence of *P. reticulatum*. The Cachara complete mitogenome measured 16.576 pb in length and contains 13 encoding protein genes, 2 rRNA genes, 22 tRNA and control region (D-Loop). The mitogenome is composed by 24.63% of T; 28.47% of C; 31.45% of A and 15.44% of G and showed the same basic structure, order and genic organization observed in other vertebrate species. Excepting protein NAD6 encoding gene and eight tRNA, all other mitochondrial genes were located in H tape. The *P. reticulatum* mitogenome showed the same basic structure, order and genic organization observed in other vertebrate species. This is the first relate of the complete mtDNA sequence of *P. reticulatum*. We hope that the availability of this mitogenome contribute to further studies in different fields like phylogeography, phylogenetic systematics, population genetics, conservation and breeding.

Key-words: Mitogenome, Siluriforms, RNA-Seq; Catfish; Animal genetic resources.

3 INTRODUÇÃO

Algumas espécies de bagres neotropicais de água doce, apresentam alto valor comercial e crescente importância para a aquicultura. Entre elas, destaca-se a cachara ou surubim (*Pseudoplatystoma reticulatum*), que, apesar de apresentar relevância econômica para o setor produtivo aquícola, ainda precisa ter sua sua biologia melhor estudada (Bignotto et al., 2009), fato que contribuiria para a expansão de seu cultivo.

As espécies do gênero *Pseudoplatystoma* (família Pimelodidae, ordem Siluriforme) são migratórias e distribuem-se ao longo das principais bacias hidrográficas da América do Sul. O *P. reticulatum* Eigenmann e Eigenmann 1889 pode ser encontrado exclusivamente nas bacias dos rios Paraguai-Paraná, embora existam relatos de sua ocorrência na bacia Amazônica (Buitrago-Suárez & Burr, 2007; Torrico et al., 2009).

Informações genéticas publicamente disponíveis para a cachara ainda são escassas e baseiam-se na caracterização da diversidade genética, através do uso dos marcadores moleculares clássicos: microssatélites (Prado et al., 2014) e sequências parciais de genes nucleares e mitocondriais (Hashimoto et al., 2013; Pereira et al., 2013; Carvalho et al., 2012; Carvalho et al., 2011; Carvalho-Costa et al., 2011; Prado et al., 2011; Bignotto et al., 2009; Torrico et al., 2009). Espera-se alterar esse cenário com a aplicação de tecnologias de sequenciamento de alto desempenho ou de nova geração (NGS). Dados genômicos amplos gerados para a espécie possibilitarão o desenvolvimento de ferramentas não apenas para atender estudos de filogenia, filogeográficos, de estrutura e diversidade genética de populações selvagens, mas também para alavancar a área de produção aquícola, permitindo a avaliação e o gerenciamento de plantéis de reprodutores, a estruturação de programas de melhoramento genético que foquem avanços na produtividade, e a manutenção de bancos de germoplasma caracterizado (Yáñez et al., 2015).

Genes e regiões específicas do DNA mitocondrial (mtDNA) vêm sendo amplamente utilizados nos estudos de filogenia, evolução das espécies, filogeografia, genética de populações, e genética da conservação (Borba et al., 2013a, b; Hahn et al., 2013; Sosa et al., 2012), devido às suas características estruturais e funcionais. A grande maioria dos estudos foi realizada com base no sequenciamento e análise de uma combinação de genes mitocondriais, como forma de melhorar a acurácia dos dados. Contudo, uma nova tendência se iniciou nos últimos oito anos, com análises utilizando o sequenciamento completo do mtDNA (e.g., Alam et al., 2014; Powell et al., 2013; Yu et al., 2007).

Milhares de espécies já têm seu genoma mitocondrial (mitogenoma) completamente sequenciado e disponibilizado em bases de dados públicas (Hahn et al., 2013). Dos 91 mitogenomas de peixes da ordem Siluriformes publicados no *GenBank* (Agosto, 2015), apenas quatro pertencem à espécie da família Pimelodidae. No presente trabalho é relatada a sequência completa do genoma mitocondrial de *P. reticulatum* (Pimelodidae), obtida a partir de uma montagem *de novo* gerada a partir do sequenciamento de RNA (RNA-seq) de diferentes tecidos da espécie.

4 MATERIAL E MÉTODOS

4.1 Amostragem, Extração de RNA Total, Confeção dos *Pools* de RNA Total e das Bibliotecas de cDNA, Sequenciamento e Montagem *de novo* do Transcriptoma da Cachara

A amostragem das 12 cacharas, extração de RNA total e confecção dos sete *pools* de RNA total de cada tecido da cachara e do *pool* com amostras de RNA de todos os tecidos, a construção das oito bibliotecas de cDNA das cacharas, os sequenciamentos Illumina HiSeq 2000 e Illumina MiSeq V3 e a montagem *de novo* do transcriptoma da cachara foram descritas em Material e Métodos, no Capítulo 3.

Após a montagem *de novo* do transcriptoma da cachara, utilizando o programa BLAST, foram identificados três transcritos que alinhavam com o mtDNA de outras espécies de peixes. Para montar esses transcritos em uma única sequência, foram utilizadas as super-*reads* geradas a partir das *reads* de MiSeq e o programa SSAKE (Warren et al., 2007). Como o sequenciamento foi feito a partir de cDNA, não foi possível recuperar toda a região controle (*D-Loop*). Sua sequência e tamanho foram determinados por sequenciamento Sanger.

4.2 Sequenciamento da Região Controle (*D-Loop*)

Parte da sequência da região controle (*D-Loop*) do mtDNA de *P. reticulatum* foi obtida com o sequenciamento de um fragmento amplificado a partir de DNA genômico isolado dos 12 indivíduos utilizados para amostragem de RNA com os iniciadores 5'-

CTCTTCCTGCACCCTACCAA-3' e 5'-ATTGAGGGCATTCTCACAGG-3'. Os produtos de PCR obtidos foram purificados enzimaticamente com ExoSAP-IT e sequenciados nos dois sentidos com um sequenciador automatizado (ABI Prism 3100) com química *BigDye Terminator* v3, segundo as instruções do fabricante.

As sequências obtidas foram processadas, montadas e alinhadas automaticamente utilizando os programas Phred (Ewing & Green, 1998; Ewing et al., 1998) e Phrap. Os *contigs* gerados foram observados com o programa Consed (Gordon et al., 2004). As sequências obtidas de cada indivíduo foram alinhadas com ClustalW para identificação de polimorfismos e obtenção da sequência consenso do D-Loop.

4.3 Anotação dos Genes Mitocondriais

Os genes codificadores de proteína do mtDNA de *P. reticulatum* foram anotados inicialmente de forma manual, utilizando a ferramenta NCBI/BLASTx em uma pesquisa contra o genoma mitocondrial de *P. magdaleniatum* (NC_026526.1) e *P. corruscans* (NC_026846.1). Genes tRNAs e suas estruturas secundárias foram identificadas com o *software* tRNAScan-SE 1.21 (Lowe & Eddy, 1997), utilizando parâmetros padrão. Os genes rRNA e tRNA^{Ser(GCT)} foram identificados por similaridade contra os mitogenomas de *P. magdaleniatum* e *P. corruscans*. Para confirmar a anotação manual do mitogenoma de *P. reticulatum*, foram utilizados os anotadores automáticos MitoAnnotator (Iwasaki et al., 2013), DOGMA (Wyman et al., 2004) e MITOS (Bernt et al., 2013).

Composição de bases nucleotídicas dos mitogenomas e uso de códon foram determinados utilizando os *softwares* Bioedit v 7.2.5 e MEGA 6.0 (Tamura et al., 2013), respectivamente. A origem da replicação da fita L (O_L) foi definida por alinhamento com sequências anotadas de outras espécies, assim como os elementos regulatórios da região controle não codificadora.

4.4. Comparações entre os mitogenomas de Cinco Bagres da Família Pimelodidae

Comparações entre os mitogenomas de *P. reticulatum* e de outros quatro bagres da família Pimelodidae disponíveis no *GenBank*: *P. magdaleniatum* (acesso *GenBank*:

NC_026526.1), *P. corruscans* (acesso *GenBank*: NC_026846.1), *Pimelodus pictus* (acesso *GenBank*: NC_015797.1) e *Sorubim cuspicaudus* (acesso *GenBank*: NC_026211.1) foram realizadas.

5 RESULTADOS E DISCUSSÃO

5.1 RNA-Seq e Montagem do Genoma Mitocondrial de *P. reticulatum*

O sequenciamento MiSeq (Illumina) da biblioteca de cDNA gerada a partir de amostras de todos os sete tecidos coletados produziu 28.721.880 de sequências (300 pb de comprimento) de alta qualidade (60,83% do total de fragmentos sequenciados). A combinação dessas sequências produziu aproximadamente 12 milhões de *super-reads*, com comprimento médio de 392 pb (Figura 3.1). Já o sequenciamento *paired-end* Illumina Hiseq 2000 das sete bibliotecas de cDNA de tecidos individuais produziu 429.706.374 de sequências (100 pb) de alta qualidade (96,6% dos fragmentos sequenciados). O resultado do sequenciamento de cada uma das sete bibliotecas de cDNA da cachara é apresentado na Tabela 3.1. Informações detalhadas sobre os resultados referentes ao sequenciamento das bibliotecas de cDNA da cachara foram apresentadas em Resultados e Discussão, no Capítulo 3.

A montagem *de novo* do transcriptoma da cachara gerou 312.766 *contigs* e 187.216 genes (Tabela 4.1). Três desses transcritos, contendo 9.069 pb, 3.798 pb e 1.692 pb, apresentaram elevada homologia com o mtDNA de outras espécies de peixes. A montagem desses três transcritos resultou em uma única sequência de 16.281 pb.

A região controle (*D-Loop*) apresentou 481 pb. Seu alinhamento com a sequência de cDNA montada (16.281 pb), resultou em uma sequência final completa do mtDNA de *P. reticulatum* com 16.576 pb, revelando que apenas 295 pb do genoma mitocondrial não foram efetivamente transcritos. Essa sequência será depositada no *GenBank*.

Genomas mitocondriais completos geralmente são obtidos através do sequenciamento de DNA. Nos últimos cinco anos, tem se intensificado o uso das tecnologias de sequenciamento de nova geração para essa finalidade, na maioria das espécies animais,

inclusive para as espécies piscícolas, como por exemplo *Rhincodon typus* (Alam et al., 2014); *Petrochromis trewavasae* e *Tropheus moorii* (Fischer et al., 2013); *Pseudoplatystoma magdaleniatum* (Rangel-Medrano et al., 2015); *Sorubim cuspicaudus* (Restrepo-Escobar et al., 2014), dentre outras. Por outro lado, a montagem de mitogenomas a partir de DNA mitocondrial transcrito, ou seja, a partir do sequenciamento de mRNA (RNA-Seq), como a realizada no presente estudo, ainda não é uma prática comum. Apenas dois estudos seguindo essa estratégia puderam ser identificados na literatura (Mu et al., 2015, Heidtmann, 2014). Sendo assim, a montagem e caracterização do DNA mitocondrial da cachara a partir de RNA-Seq revelou-se um trabalho pioneiro, em especial para as espécies piscícolas no Brasil, além de ser o primeiro mitogenoma completo relatado para a espécie *P. reticulatum*.

5.2 Anotação do Genoma Mitocondrial de *P. reticulatum*

As anotações do mtDNA de *P. reticulatum* obtidas de forma manual e com os anotadores automáticos DOGMA (Tamura et al., 2013), MITOS (Bernt et al., 2013) e MitoAnnotator (Iwasaki et al., 2013) podem ser visualizadas no Arquivo Suplementar 1.

A comparação da anotação manual com as anotações geradas pelos anotadores automáticos revelou diferenças de até 23 pb no tamanho dos genes rRNAs (*12S* e *16S*) e de até 96 pb no tamanho de oito genes codificadores de proteínas (*NAD1*, *NAD2*, *COX1*, *COX2*, *NAD3*, *NAD4L*, *NAD4* e *NAD5*). Consequentemente, as posições iniciais e/ou finais desses mesmos genes também diferiram entre os resultados obtidos com os diferentes métodos utilizados. Particularmente, o gene *NAD6* quando anotado pelo MitoAnnotator apresentou o mesmo tamanho da anotação manual, porém sua posição inicial e final foi diferente. As diferenças comentadas podem ser visualizadas no Anexo A.

Genes tRNAs foram anotados de forma semelhante com o *software* tRNAScan (utilizado na anotação manual) e pelos anotadores automáticos MitoAnnotator e MITOS, com exceção do gene tRNA^{Ser(GCT)} entre as posições 11.823 e 11.889 do mtDNA, que não foi identificado na anotação manual. Os genes tRNA^{Phe(GAA)} (5.087-5.157) e tRNA^{Met(TGG)} (15.690-15.789) foram identificados apenas na anotação realizada com o DOGMA. O DOGMA também identificou os genes tRNA^{Phe(AAA)} (7.039-7.109) e tRNA^{Asp(AGG)} (7.892-7.965) ao invés dos genes tRNA^{Ser(TGA)} e tRNA^{Lys(TTT)}, respectivamente, identificados pelos outros métodos nestas mesmas posições.

Genes codificadores de proteína foram anotados de forma idêntica ao se utilizar os anotadores automáticos MitoAnnotator e DOGMA. Os genes rRNAs foram anotados de forma semelhante com o MitoAnnotator e com a anotação manual. A região codificadora (*D-Loop*) foi anotada apenas pelo anotador automático MitoAnnotator, além da anotação manual.

Como as diferenças encontradas entre a anotação manual e a anotação realizada pelo anotador automático MitoAnnotator foram pequenas, e como o MitoAnnotator fornece o mapa circular do DNA mitocondrial analisado e um arquivo de saída com os resultados no formato *GenBank*, sua anotação foi a selecionada para ser depositada no GenBank e para ser descrita no presente trabalho (Tabela 5.1).

5.3 Organização do Genoma Mitocondrial de *P. reticulatum*

O mitogenoma do *P. reticulatum* possui todos os 37 genes estruturais mitocondriais observados na maioria dos vertebrados (Tabela 5.1, Figura 5.1). Os genes mitocondriais localizam-se nas duas fitas do mitogenoma, designadas como L (*light*) e H (*heavy*). Seguindo o padrão previamente estabelecido para os vertebrados, apenas o gene NADH desidrogenase subunidade 6 (*NAD6*) e 8 genes tRNAs (tRNA^{Gln}(*TGG*), tRNA^{Ala}(*TGC*), tRNA^{Asn}(*GTT*), tRNA^{Cys}(*GCA*), tRNA^{Tyr}(*GTA*), tRNA^{Ser}(*TGA*), tRNA^{Glu}(*TTC*) e tRNA^{Pro}(*TGG*)) estão localizados na fita L do mitogenoma de *P. reticulatum*. Os demais genes são codificados pela fita H.

Tabela 5.1. Anotação do genoma mitocondrial completo de *Pseudoplatystoma reticulatum* (16.576 pb) gerada com o anotador automático MitoAnnotator (Iwasaki et al., 2013)

Genes	Fita ^a	Posição (pb)		Tamanho (pb) ^{c,d}	Nucleotídeos Intergênicos ^e	Códon		Anti códon
		Início ^b	Fim ^b			Start	Stop ^f	
tRNA ^{Phe}	H	1	69	69	0			GAA
rRNA 12S	H	70	1030 (5)	961(5)	0			
tRNA ^{Val}	H	1031	1102	72	0			TAC
rRNA 16S	H	1103	2782	1680	0			
tRNA ^{Leu (TAA)}	H	2783	2857	75	1			TAA
NAD1	H	2859	3830	972	1	ATG	TAA	
tRNA ^{Ile}	H	3832	3903	72	-1			GAT
tRNA ^{Gln}	L	3903	3973	71	-1			TTG
tRNA ^{Met}	H	3973	4042	70	0			CAT
NAD2	H	4043	5087 (96)	1045 (948)	0	ATG	T	
tRNA ^{Trp}	H	5088	5158	71	18			TCA
tRNA ^{Ala}	L	5177	5245	69	1			TGC
tRNA ^{Asn}	L	5247	5319	73	0			GTT
O ^L	-	5320	5348	29	0			
tRNA ^{Cys}	L	5349	5415	67	1			GCA
tRNA ^{Tyr}	L	5417	5486	70	1			GTA
COX1	H	5488	7038	1551	0	GTG	TAA	
tRNA ^{Ser (TGA)}	L	7039	7109	71	4			TGA
tRNA ^{Asp}	H	7114	7186	73	14			GTC
COX2	H	7201	7891	691	0	ATG	T	
tRNA ^{Lys}	H	7892	7965	74	1			TTT
ATP8	H	7967	8134	168	-10	ATG	TAA	
ATP6	H	8125	8807	683	0	ATG	TA	
COX3	H	8808	9591	784	0	ATG	T	
tRNA ^{Gly}	H	9592	9664	73	0			TCC
NAD3	H	9665 (-51)	10013	349 (51)	0	ATG	T	
tRNA ^{Arg}	H	10014	10084	71	0			TCG
NAD4L	H	10085	10381 (18)	297 (18)	-7	ATG	TAA	
NAD4	H	10375	11752	1378	0	ATG	T	
tRNA ^{His}	H	11753	11822	70	0			GTG
tRNA ^{Ser}	H	11823	11889	67	5			GCT
tRNA ^{Leu}	H	11895	11967	73	0			TAG
NAD5	H	11968 (-54)	13794 (42)	1827 (96)	-4	ATG	TAA	
NAD6	L	13791 (-3)	14306 (-3)	516	0	ATG	TAA	
tRNA ^{Glu}	L	14307	14375	69	1			TTC
CYTb	H	14377	15514	1138	0	ATG	T	
tRNA ^{Thr}	H	15515	15586	72	-2			TGT
tRNA ^{Pro}	L	15585	15654	70	0			TGG
D-Loop	-	15655	16576	922				

^a Fita: H = heavy (+); L = light (-).

^b Números entre parênteses indicam diferenças nas posições iniciais ou finais em relação às posições observadas com a anotação manual.

^c Genes codificadores de proteína com stop códon.

^d Números entre parênteses indicam diferenças no tamanho dos genes em relação ao tamanho observado com a anotação manual.

^e Números positivos indicam nucleotídeos separando dois genes adjacentes. Números negativos indicam sobreposição de nucleotídeos.

^f T ou TA indicam stop códon incompleto.

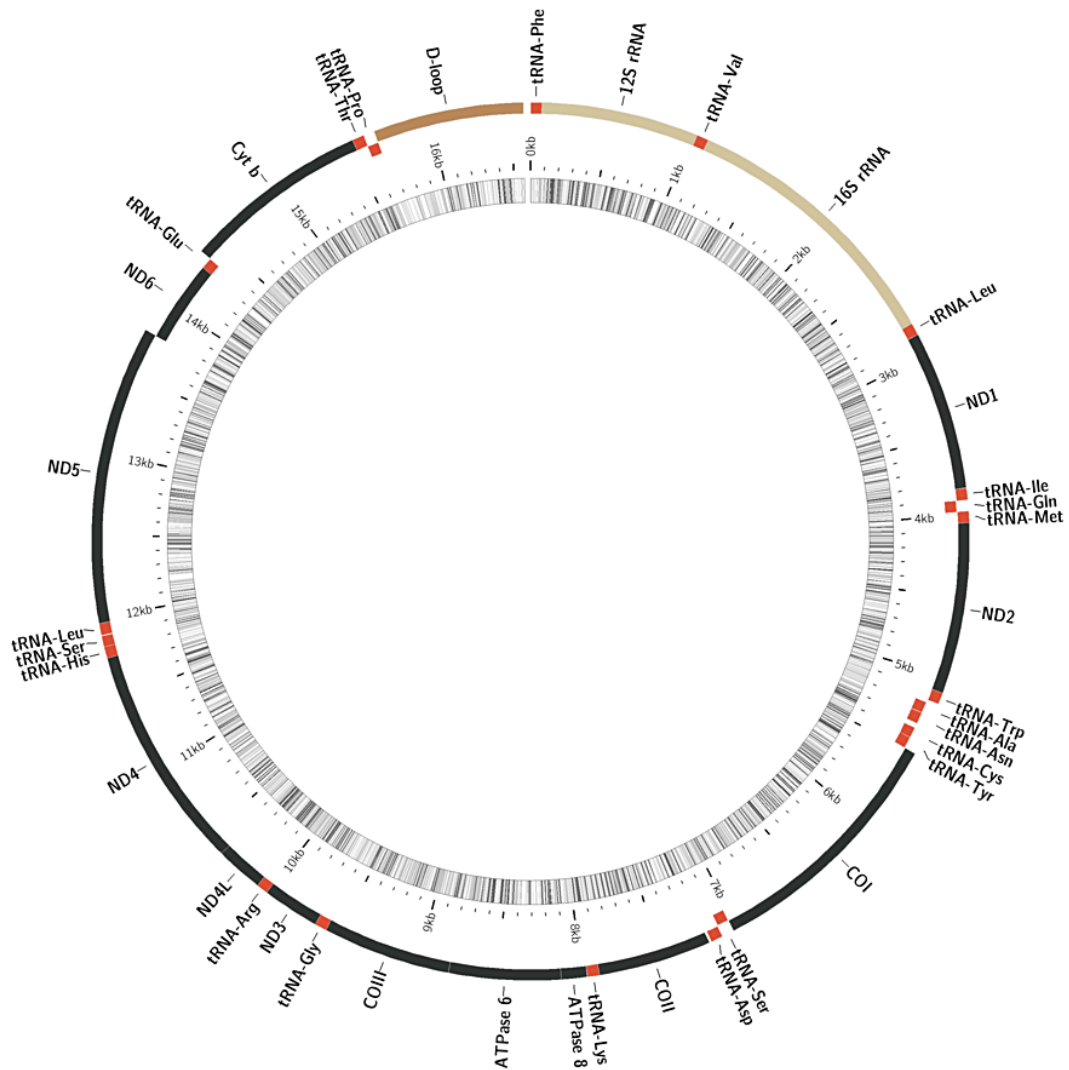


Figura 5.1. Mapa de organização dos genes mitocondriais de *Pseudoplatystoma reticulatum* gerado com o MitoAnnotator. Genes codificadores de proteína, tRNAs, rRNAs e a região não codificadora são mostrados em diferentes cores. Genes localizados dentro do círculo no mapa indicam genes codificados na fita *light* (L) do mitogenoma.

O comprimento das regiões espaçadoras intergênicas observado foi de 48 pb em 11 regiões (Tabela 5.1). As sobreposições entre os genes codificadores *ATP8* e *ATP6* (10 nucleotídeos), *NAD4L* e *NAD4* (7 nucleotídeos) e *NAD5* e *NAD6* (4 nucleotídeos) observadas no mitogenoma de *P. reticulatum* são comuns entre os vertebrados, no entanto, sua extensão pode variar de uma espécie para outra, inclusive nos peixes (Broughton et al., 2001; Guo et al., 2008; Fischer et al., 2013; Hrbek & Farias, 2008; Jiang et al., 2014; Kim & Lee et al., 2004; Mu et al., 2015; Prosdocimi et al., 2011).

Como na maioria dos vertebrados, apenas o gene codificador de proteína *COXI* de *P. reticulatum* apresentou GTG como *start* códon, e não o convencional ATG apresentado pelos demais genes codificadores de proteínas. Sete genes apresentaram *stop* códon incompleto: T (*NAD2*, *COX2*, *COX3*, *NAD3*, *NAD4* e *CYTB*) e TA (*ATP6*). Os demais genes apresentaram o *stop* códon completo TAA (Tabela 5.1). A presença de *stop* códons incompletos nos genes codificadores de proteínas é uma característica comum dos genes mitocondriais dos vertebrados e de inúmeras espécies de peixes (por exemplo: Alam et al., 2014; Jiang et al., 2014; Kartavtsev et al., 2007; Mu et al., 2015).

A composição nucleotídica dos genes codificadores de proteína do genoma mitocondrial completo de *P. reticulatum* observada foi A: 30,26%; C: 30,12%; T: 25,36%; e G: 14,26% (Tabela 5.2). O conteúdo AT foi maior que o conteúdo GC. Os resultados observados refletem o tradicional viés contra o nucleotídeo guanina (G) na fita L (fita senso para todos os genes codificadores de proteína, com exceção de *NAD6*), que foi mais pronunciado na 3ª posição dos códons (5,12%). O viés negativo em relação a presença do nucleotídeo G na 3ª posição do códon dos genes codificadores de proteínas do mitogenoma também foi reportado em *Danio rerio* (Broughton et al., 2001), *Oreochromis niloticus* e *O. aureus* (He et al., 2011), *Arapaima gigas* (Hrbek & Farias, 2008), *Macropodus opercularis* (Mu et al., 2015), dentre outras espécies.

De acordo com Fonseca et al. (2008), Meyer (1993) e Clayton (1991), os genes codificadores de proteína são caracterizados por um viés na composição nucleotídica: a fita H é rica em G e T (viés contra C e A), enquanto a fita L é rica em A e C (viés contra G e T). Esse viés na composição nucleotídica pode ser explicado pela exposição da fita H como fita simples de DNA por um período de tempo mais longo durante a replicação assimétrica do genoma mitocondrial circular. O DNA fita simples é mais susceptível à mutações, especialmente a desaminação do nucleotídeo Citosina na fita H, que resulta em um forte viés contra C na fita H e, conseqüentemente, um forte viés contra G e a favor do nucleotídeo A na fita L.

Tabela 5.2. Composição de bases nucleotídicas dos genes codificadores de proteínas, rRNAs, tRNAs e da região controle encontrados no mitogenoma de *Pseudoplatystoma reticulatum*

Gene	Tamanho (pb)	Composição de Bases Nucleotídicas				Conteúdo GC (%)	Conteúdo AT (%)
		T(U) (%)	C (%)	A (%)	G (%)		
Codificador de proteína*	11.399	25,36 (40,70)	30,12 (9,30)	30,26 (13,76)	14,26 (36,24)	44,38 (45,54)	55,62 (54,46)
1ª posição do códon	3.804	20,59 (33,14)	27,69 (8,72)	26,97 (12,79)	24,75 (45,35)	52,44 (54,07)	47,56 (45,93)
2ª posição do códon	3.798	40,52 (45,93)	27,44 (16,28)	19,15 (12,21)	12,88 (25,58)	40,32 (41,86)	59,68 (58,14)
3ª posição do códon	3.797	14,96 (43,02)	35,24 (2,91)	44,68 (16,28)	5,12 (37,79)	40,36 (40,70)	59,64 (59,30)
tRNA	1.562	27,21	21,45	28,94	22,41	43,85	56,15
rRNA	2.641	20,94	25,03	19,99	20,94	45,02	54,98
Região Controle	922	33,51	21,69	31,45	13,23	34,92	64,97

* Composição de bases nucleotídicas dos 12 genes codificadores de proteína localizados na fita H. A composição de bases para o gene NAD6, localizado na fita L é mostrada entre parênteses.

Os códons complementares aos tRNAs foram os mais frequentemente utilizados no mitogenoma de *P. reticulatum*, com exceção da metionina, alanina e treonina. O códon CUA (leucina) foi o mais utilizado, enquanto GCG (códon da alanina) foi o menos utilizado. O conteúdo AT foi sempre maior que o GC em qualquer das 3 posições do códon (Tabela 5.3). Broughton et al. (2001), mencionam que dos possíveis códons de um aminoácido (especificamente códons terminando em A ou C, com exceção da metionina, cujo códon termina em G), apenas um (ocasionalmente dois) é complementar ao anticódon do tRNA que o serve.

Entre os aminoácidos *fourfold* de *P. reticulatum*, os códons complementares aos tRNAs que terminavam em A foram os mais frequentes. Entre os aminoácidos *twofold*, C foi o nucleotídeo mais usado na 3^a posição do códon. Consistente com o viés contra G na fita H, o nucleotídeo G foi o menos frequente na 3^a posição do códon. Esse padrão é semelhante ao identificado em *Danio rerio* (Broughton et al., 2001) e na maioria dos vertebrados.

Tabela 5.3. Padrão de utilização de códon do mitogenoma de *P. reticulatum*

Aminoácido	Códon	Número ^a	Frequência (%)	RSCU ^b	Aminoácido	Códon	Número ^a	Frequência (%)	RSCU ^b
Fenilalanina (Phe/F)	UUU	81	1,47	0,76	Tirosina (Tyr/Y)	UAU	102	1,85	0,94
	UUC*	132	2,39	1,24		UAC*	114	2,06	1,06
Leucina (Leu/L)	UUA*	113	2,05	1	Stop	UAA	86	1,56	1,54
	UUG	39	0,71	0,35		UAG	52	0,94	0,93
	CUU	107	1,94	0,95	Histidina (His/H)	CAU	67	1,21	0,64
	CUC	93	1,68	0,83		CAC*	143	2,59	1,36
	CUA*	269	4,87	2,39	Glutamina (Gln/Q)	CAA*	131	2,37	1,43
	CUG	54	0,98	0,48		CAG	52	0,94	0,57
Isoleucina (Ile/I)	AUU	156	2,82	0,99	Asparagina (Asn/N)	AAU	113	2,05	0,83
	AUC*	158	2,86	1,01		AAC*	160	2,90	1,17
Metionina (Met/M)	AUA	141	2,55	1,39	Lisina (Lys/K)	AAA*	179	3,24	1,42
	AUG*	62	1,12	0,61		AAG	74	1,34	0,58
Valina (Val/V)	GUU	46	0,83	0,83	Ácido Aspártico (Asp/D)	GAU	44	0,80	0,68
	GUC	45	0,81	0,81		GAC*	86	1,56	1,32
	GUA*	99	1,79	1,78	Ácido Glutâmico (Glu/E)	GAA*	94	1,70	1,31
	GUG	32	0,58	0,58		GAG	50	0,91	0,69
Serina (Ser/S)	UCU	67	1,21	0,9	Cisteína (Cys/C)	UGU	34	0,62	0,89
	UCC	96	1,74	1,29		UGC*	42	0,76	1,11
	UCA*	106	1,92	1,42	Triptofano (Trp/W)	UGA*	96	1,74	1,44
	UCG	21	0,38	0,28		UGG	37	0,67	0,56
Prolina (Pro/P)	CCU	122	2,21	1,11	Arginina (Arg/R)	CGU	23	0,42	0,66
	CCC	141	2,55	1,28		CGC	41	0,74	1,18
	CCA*	142	2,57	1,29		CGA*	60	1,09	1,73
	CCG	34	0,62	0,31		CGG	15	0,27	0,43
Treonina (Thr/T)	ACU	132	2,39	1,11	Serina (Ser/S)	AGU	45	0,81	0,6
	ACC	153	2,77	1,29		AGC*	113	2,05	1,51
	ACA*	150	2,72	1,26	Stop	AGA	38	0,69	0,68
	ACG	41	0,74	0,34		AGG	48	0,87	0,86
Alanina (Ala/A)	GCU	58	1,05	0,76	Glicina (Gly/G)	GGU	40	0,72	0,64
	GCC	123	2,23	1,62		GGC	76	1,38	1,22
	GCA*	111	2,01	1,46		GGA*	86	1,56	1,38
	GCG	12	0,22	0,16		GGG	47	0,85	0,76

Tabela 5.3. continuação.

Aminoácido	Códon	Número ^a	Frequência (%)	RSCU ^b	Aminoácido	Códon	Número ^a	Frequência (%)	RSCU ^b
Fenilalanina (Phe/F)	UUU	81	1,47	0,76	Tirosina (Tyr/Y)	UAU	102	1,85	0,94
	UUC*	132	2,39	1,24		UAC*	114	2,06	1,06
Leucina (Leu/L)	UUA*	113	2,05	1	Stop	UAA	86	1,56	1,54
	UUG	39	0,71	0,35		UAG	52	0,94	0,93
	CUU	107	1,94	0,95	Histidina (His/H)	CAU	67	1,21	0,64
	CUC	93	1,68	0,83		CAC*	143	2,59	1,36
	CUA*	269	4,87	2,39	Glutamina (Gln/Q)	CAA*	131	2,37	1,43
	CUG	54	0,98	0,48		CAG	52	0,94	0,57
	Isoleucina (Ile/I)	AUU	156	2,82	0,99	Asparagina (Asn/N)	AAU	113	2,05
AUC*		158	2,86	1,01	AAC*		160	2,90	1,17
Metionina (Met/M)	AUA	141	2,55	1,39	Lisina (Lys/K)	AAA*	179	3,24	1,42
	AUG*	62	1,12	0,61		AAG	74	1,34	0,58
Valina (Val/V)	GUU	46	0,83	0,83	Ácido Aspártico (Asp/D)	GAU	44	0,80	0,68
	GUC	45	0,81	0,81		GAC*	86	1,56	1,32
	GUA*	99	1,79	1,78	Ácido Glutâmico (Glu/E)	GAA*	94	1,70	1,31
	GUG	32	0,58	0,58		GAG	50	0,91	0,69
Serina (Ser/S)	UCU	67	1,21	0,9	Cisteína (Cys/C)	UGU	34	0,62	0,89

	UCC	96	1,74	1,29			UGC*	42	0,76	1,11
	UCA*	106	1,92	1,42		Triptofano (Trp/W)	UGA*	96	1,74	1,44
	UCG	21	0,38	0,28			UGG	37	0,67	0,56
Prolina (Pro/P)	CCU	122	2,21	1,11		Arginina (Arg/R)	CGU	23	0,42	0,66
	CCC	141	2,55	1,28			CGC	41	0,74	1,18
	CCA*	142	2,57	1,29			CGA*	60	1,09	1,73
	CCG	34	0,62	0,31			CGG	15	0,27	0,43
Treonina (Thr/T)	ACU	132	2,39	1,11		Serina (Ser/S)	AGU	45	0,81	0,6
	ACC	153	2,77	1,29			AGC*	113	2,05	1,51
	ACA*	150	2,72	1,26		Stop	AGA	38	0,69	0,68
	ACG	41	0,74	0,34			AGG	48	0,87	0,86
Alanina (Ala/A)	GCU	58	1,05	0,76		Glicina (Gly/G)	GGU	40	0,72	0,64
	GCC	123	2,23	1,62			GGC	76	1,38	1,22
	GCA*	111	2,01	1,46			GGA*	86	1,56	1,38
	GCG	12	0,22	0,16			GGG	47	0,85	0,76

^a Número de códons; ^b RSCU – Utilização Relativa de Códons Sinônimos (*Relative Synonymous Codon Usage*).

* Códons complementares aos genes tRNAs.

A composição de bases de cada gene codificador de proteína do mtDNA de *P. reticulatum* é apresentada na Tabela 4.4. Os 12 genes codificados na fita L (que serve de molde para o gene NAD6 e codifica todos os outros genes codificadores de proteína) apresentaram composição nucleotídica mais rica em Adenina (A) e citosina (C). O viés esperado contra os nucleotídeos guanina (G) e timina (T) também foi observado. Os resultados alcançados estão de acordo com os relatos de Fonseca et al. (2008) e Clayton (1991). Na fita H (que codifica apenas o gene NAD6 e serve de molde para os outros genes), observou-se maior presença de T (40,70%) e G (36,24%) e um acentuado viés contra o nucleotídeo C (9,30%). Esse resultado está de acordo com as afirmações de diversos autores que também identificaram um acentuado viés contra o nucleotídeo C (citosina) no gene NAD6 (Fonseca et al., 2008; Clayton, 1991; Broughton et al., 2001; Fischer et al., 2013; Hrbek & Farias, 2008; Kartavtsev et al., 2007; Kim & Lee et al., 2004, dentre outros).

Tabela 5.4. Composição de bases nucleotídicas (% do número total) dos genes codificadores de proteína do mitogenoma de *P. reticulatum*

Gene	Tamanho (pb)	Composição de Bases Nucleotídicas				Conteúdo GC (%)	Conteúdo AT (%)
		T(U) (%)	C (%)	A (%)	G (%)		
NAD1	972	25,00	30,45	30,76	13,79	44,24	55,76
NAD2	1045	20,77	33,88	33,30	12,06	45,93	54,07
COX1	1551	27,21	26,76	28,18	17,86	44,62	55,38
COX2	691	25,33	27,21	31,84	15,63	42,84	57,16
ATP8	168	25,00	27,38	35,71	11,90	39,29	60,71
ATP6	683	25,77	30,45	31,33	12,45	42,90	57,10
COX3	784	26,15	29,46	27,81	16,58	46,05	53,95
NAD3	349	29,51	30,37	25,79	14,33	44,70	55,30
NAD4L	297	23,91	35,02	24,58	16,50	51,52	48,48
NAD4	1378	25,76	28,74	31,20	14,30	43,03	56,97
NAD5	1827	24,90	30,27	32,24	12,59	42,86	57,14
NAD6	516	40,70	9,30	13,76	36,24	45,54	54,46
CYTb	1138	25,66	31,81	29,00	13,53	45,34	54,66

5.4 Genes rRNA e tRNA

Obedecendo ao padrão verificado no mtDNA da maioria dos vertebrados, dentre os genes tRNA que compõem o mitogenoma de *P. reticulatum*, foram identificadas as duas serinas (TGA e GCT) e as duas leucinas (TAA e TAG). Sobreposições foram observadas entre os genes tRNA^{Ile} e tRNA^{Gln} (1 pb), tRNA^{Gln} e tRNA^{Met} (1 pb) e tRNA^{Thr} e tRNA^{Pro} (2 pb). Os anticódons dos genes tRNAs do

mitogenoma de *P. reticulatum* correspondem aos anticódons do código genético mitocondrial dos vertebrados (Tabela 5.1). Com exceção do gene tRNA^{Ser(GCT)}, todos os outros dobraram-se nas típicas estruturas secundárias de folha de trevo, que foram previstas pelo *software* tRNAScan. Normalmente, elas apresentam 7 pb no braço do aminoácido e no braço do anticódon do *loop*, 5 pb nos braços TΨC e anticódon e 4 pb no braço DHU (BROUGHTON et al., 2001; HRBEK e FARIAS, 2008; KIM e LEE et al., 2004). No entanto, variações de tamanho e posição ou composição de bases nucleotídicas e/ou resíduos dos braços e diversos pareamentos não complementares foram identificados na maioria das estruturas secundárias dos genes tRNAs de *P. reticulatum* (Anexo B).

5.5 Região Controle (D-Loop)

A região controle de *P. reticulatum* está localizada entre os genes tRNA^{Pro} e tRNA^{Phe}, e apresentou 922 pb de comprimento, sendo que um total de 295 pb dessa sequência foi identificado por sequenciamento Sanger. A região controle do mitogenoma de *P. reticulatum* foi mais rica em AT (64,97%, Tabela 5.2) do que o mitogenoma completo (56,08%).

Dois elementos estruturais foram identificados na região central (a menos polimórfica) da região controle de *P. reticulatum* com base no alinhamento com as sequências da região controle de *Macropodus opercularis* (Mu et al., 2015) e *Danio rerio* (Broughton et al., 2001): a região central CSB2 (TAAACCCCCCTACCCCC), que apresentou 94% de similaridade com CSB2 de *Macropodus opercularis* e 88% com CBS de *Danio rerio* e CSB3 (TGTCAAACCCCTAAACCA), que apresentou 100% de similaridade com CSB3 de *Danio rerio* e 70% com o de *Macropodus opercularis*. A região controle que contém a origem da replicação da fita L (O_L) foi identificada entre os genes tRNA^{Asn} e tRNA^{Cys}, possuindo 29 pb de comprimento (5'-CTT TCC CCG CCG CCT TAA AAA GGC GGG GA-3').

5.6 Comparação dos Genomas Mitocondriais Completos de Cinco Espécies de Bagres da Família Pimelodidae

A análise comparativa do mitogenoma de *P. reticulatum* montado neste estudo com quatro mitogenomas completos de bagres da família Pimelodidae disponíveis no GeneBank (*Pseudoplatystoma corruscans*, *Pseudoplatystoma magdaleniatum*, *Sorubim cuspicaudus* e *Pimelodus pictus*) revelou que a estrutura, o conteúdo e a ordem dos genes estruturais mitocondriais desses cinco mitogenomas é semelhante (Tabela 5.5 e Anexo C).

A composição nucleotídica do genoma mitocondrial completo de *P. reticulatum* foi A: 31,45%; C: 28,47%; T: 24,63%; e G: 15,44%. A composição total de bases dos mitogenomas de *P. corruscans*, *P. magdaleniatum*, *Pimelodus pictus* e *S. cuspicaudus* foi similar (Tabela 5.6). De forma semelhante, o conteúdo GC (43,92%) do mitogenoma de *P. reticulatum* foi menor que o conteúdo AT (56,08%). O mesmo padrão foi observado no mitogenoma das outras quatro espécies analisadas, o que confirma o viés contra G existente na composição nucleotídica dos mitogenomas da maioria dos vertebrados (Broughton et al., 2001; Fischer et al., 2013; Guo et al., 2008).

Com exceção de *P. corruscans* (16.123 pb), as outras quatro espécies apresentaram mitogenomas com tamanhos semelhantes ao de *P. reticulatum*, variando entre 16.544 pb e 16.576 pb de comprimento (Tabela 5.5). O comprimento total de nucleotídeos dos genes codificadores de proteína, rRNAs, tRNAs e D-Loop de *P. reticulatum* foi 11.399 pb, 2.641 pb, 1.562 pb e 922 pb, correspondendo à 68,8%, 15,9%, 9,4% e 5,6% do genoma, respectivamente. Esses tamanhos são semelhantes aos das outras quatro espécies analisadas (Tabela 5.5), com exceção do tamanho dos genes ribossomais de *P. corruscans*, cujo comprimento total de nucleotídeos (2.193 pb) foi menor quando comparado ao das demais espécies (2.626 pb, em média).

Essa variação está relacionada ao menor tamanho do gene rRNA *12S* apresentado por *P. corruscans* (NC_026846.1; PERINI et al., não publicado): 515 pb, enquanto as outras espécies apresentaram rRNA *12S* com um tamanho médio de 959 pb (Anexo C). De acordo com sequenciamento realizado por Lundberg et al. (2011), o gene ribossomal *12S* de *P. corruscans* apresentou 944 pb (JF_898657.1), valor, ou seja, 429 pb a mais que o rRNA *12S* sequenciado por Perini et al. (não publicado) e um valor mais próximo do determinado para as outras espécies. Esse resultado indica que o

tamanho de 515 pb encontrado por Perini et al. (não publicado, NC_026846.1) para o gene ribossomal 12S pode estar relacionado a um erro de montagem desse genoma mitocondrial.

A análise da similaridade total do mitogenoma de *P. reticulatum* com o das outras quatro espécies de bagres foi elevada, confirmando os resultados obtidos: 95% de nucleotídeos idênticos aos de *P. corruscans*; similaridade de 93% com *P. magdaleniatum*; 88% com *Sorubim cuspicaudus* e 87% com *Pimelodus pictus*. Essa similaridade demonstra a proximidade entre as espécies *P. reticulatum* e *P. corruscans*, já indicada por Buitrago-Suárez & Burr (2007), Torrico et al. (2009) e Carvalho-Costa et al. (2011).

A espécie *P. reticulatum* fazia parte do complexo de espécies “*P. fasciatum*”, que foi desmembrado em 8 espécies de acordo com caracteres morfológicos e distribuição geográfica (Buitrago-Suárez & Burr, 2007). No entanto, os dados moleculares são controversos para suportar essa divisão de *P. fasciatum* em 8 espécies de acordo com a baixa divergência genética entre as espécies com base em genes mitocondriais e nucleares (Torrico et al., 2009; Carvalho-Costa et al., 2011). Nesse sentido, o genoma mitocondrial completo poderia trazer maiores informações na busca da validação da hipótese de divisão de *P. fasciatum* e no suporte de *P. reticulatum* como espécie endêmica das bacias Paraguai–Paraná.

A disponibilidade do genoma mitocondrial completo de *P. reticulatum* pode, portanto, ajudar a desvendar a filogenia do clado “*fasciatum*”, um complexo de espécies da família Pimelodidae, ainda não claramente definida (Buitrago-Suarez & Burr, 2007; Torrico et al., 2009; Carvalho-Costa et al., 2011). No entanto, uma análise filogenética mais robusta é necessária para poder afirmar com certeza se o mitogenoma completo pode esclarecer essa filogenia.

Tabela 5.5. Tamanho e estrutura dos genomas mitocondriais completos de *Pseudoplatystoma reticulatum* e das cinco espécies de bagres da família Pimelodidae disponíveis no *GenBank*.

Espécie	ID NCBI <i>GenBank</i>	Autor	Tamanho (pb)	Genes codificadores de proteína* (pb)	Genes rRNA (pb)	Genes tRNA (pb)	D-Loop (pb)
<i>P. reticulatum</i>		Este estudo	16.576	11.399	2.641	1.562	922
<i>P. corruscans</i>	NC_026846.1	PERINI et al. (não publicado)	16.123	11.400	2.193	1.561	912
<i>P. magdaleniatum</i>	NC_026526.1	RANGEL-MEDRANO et al., 2015	16.568	11.401	2.647	1.559	912
<i>Pimelodus pictus</i>	NC_015797.1	NAKATANI et al., 2011	16.575	11.402	2.588	1.614	923
<i>Sorubim cuspicaudus</i>	NC_026211.1	RESTREPO-ESCOBAR et al., 2014	16.544	11.403	2.630	1.558	909

*Stop códon incluído nas sequências.

Tabela 5.6. Composição total de bases nucleotídicas do genoma mitocondrial completo de *Pseudoplatystoma reticulatum* de cinco bagres da família Pimelodidae com mtDNA completo disponível no *Genbank*.

Espécie	T(U) (%)	C (%)	A (%)	G (%)	Conteúdo GC (%)	Conteúdo AT (%)
<i>Pseudoplatystoma reticulatum</i>	24,63	28,47	31,45	15,44	43,92	56,08
<i>Pseudoplatystoma magdaleniatum</i>	24,41	28,78	31,40	15,41	44,19	55,81
<i>Pseudoplatystoma corruscans</i>	24,71	28,48	31,59	15,21	43,69	56,30
<i>Pimelodus pictus</i>	26,04	27,17	31,39	15,40	42,57	57,43
<i>Sorubim cuspicaudus</i>	25,40	27,32	32,39	14,89	42,21	57,79

6 CONCLUSÕES

A partir da montagem de RNA-Seq do transcriptoma da cachara, foi possível obter a sequência completa do genoma mitocondrial deste bagre (*Pseudoplatystoma reticulatum*), e realizá-la sua anotação. Em termos de organização, estrutura, conteúdo e ordem dos genes estruturais mitocondriais, o mitogenoma de *P. reticulatum* não difere da maioria dos mitogenomas dos vertebrados e das quatro espécies de bagres da família Pimelodidae, que já possuem seu genoma mitocondrial completo publicado. Além de enriquecer as bases de dados públicas, as informações moleculares sobre *P. reticulatum* disponibilizadas pelo presente trabalho constituem importante fonte de informação para estudos em diversas áreas, como genética, melhoramento genético, conservação, filogenética, dentre outras, como forma de aumentar o conhecimento biológico da espécie, contribuindo para o desenvolvimento da sua criação em cativeiro.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALAM, M.T.; PETIT III, R.A.; READ, T.D. et al. The complete mitochondrial genome sequence of the world's largest fish, the whale shark (*Rhincodon typus*), and its comparison with those of related shark species. **Gene**, v. 539, p. 44-49, 2014.
- AVISE, J.C. Animal Mitochondrial DNA. In: AVISE, J.C. (Ed.) **Molecular markers, natural history and evolution**. New York: Chapman & Hall, 1994. p.60-68.
- BERNT, M.; DONATH, A.; JÜHLING, F. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. **Molecular Phylogenetics and Evolution**, v. 69, p.313–319, 2013.
- BIGNOTTO, T.S.; PRIOLI A.J.; PRIOLI, S.M.A.P. et al. Genetic divergence between *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Siluriformes: Pimelodidae) in the Paraná River Basin. **Brazilian Journal of Biology**, v. 69, Suppl. 2, 681-689, 2009.
- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. **Bioinformatics**, p.1-7, 2014.
- BORBA, R.S. de; ZAWADZKI, C.H.; OLIVEIRA, C.; et al. Phylogeography of *Hypostomus strigaticeps* (Siluriformes: Loricariidae) inferred by mitochondrial DNA reveals its distribution in the upper Paraná River basin. **Neotropical Ichthyology**, v. 11, n. 1, p.111-116, 2013a.
- BORBA, R.F. de; SILVA, E.L. da; PONZETTO, J.M. et al. Genetic structure of the ornamental tetra fish species *Piabucus melanostomus* Holmberg, 1891 (CHARACIDAE, IGUANODECTINAE) in the Brazilian Pantanal wetlands inferred by mitochondrial DNA sequences. **Biota Neotropica**, v. 13, n. 1, p.42-4, 2013b.
- BROUGHTON, R.E.; MILAM, J.E.; ROE, B.A. The Complete Sequence of the Zebrafish (*Danio rerio*) Mitochondrial Genome and Evolutionary Patterns in Vertebrate Mitochondrial DNA. **Genome Research**, v. 11, n. 11, p.1958-1967, 2001.

- BUITRAGO-SUÁREZ, U.A.; BURR, B.M. Taxonomy of the catfish genus *Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. **Zootaxa**, v. 1512, p.1-38, 2007.
- CARVALHO, D.C.; NETO, D.A.P.; BRASIL, B.S.A.F. et al. DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. **Mitochondrial DNA**, v. 22, p.97-105, 2011.
- CARVALHO, D.C.; OLIVEIRA, D.A.A.; BEHEREGARAYM L.B. et al. Hidden genetic diversity and distinct evolutionarily significant units in an commercially important Neotropical apex predator, the catfish *Pseudoplatystoma corruscans*. **Conservation Genetics**, v. 13, n. 6, p.1671-1675, 2012.
- CARVALHO-COSTA, L.F.; PIORSKI, N.M.; WILLIS, S.C. et al. Molecular systematics of the neotropical shovelnose catfish genus *Pseudoplatystoma* Bleeker 1862 based on nuclear and mtDNA markers. **Molecular Phylogenetics and Evolution**, v. 59, n. 1, p.177-194, 2011.
- CLAYTON, D.A. Replication and transcription of vertebrate mitochondrial DNA. **Annual Review of Cell Biology**, v. 7, p.453-478, 1991.
- EWING, B.; GREEN, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, p.186-194, 1998.
- EWING, B.; HILLIER, L.; WENDL, M.C. et al. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v. 8, p.175-185, 1998.
- FISCHER, C.; KOBLMULLER, S.; GÜLLY, C. et al. Complete mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the blunthead cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in cichlid fishes. **Plos One**, v. 8, p.1-14, 2013.
- FONSECA, M.M.; POSADA, D.; HARRIS, D.J. Inverted replication of vertebrate mitochondria. **Molecular Biology and Evolution**, v. 25, n. 5, p.805-808, 2008.
- GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**, v. 8, p.195-202, 1998.
- GUO, Y.; WANG, Z.; LIU, C. et al. Sequencing and analysis of the complete mitochondrial DNA of Russell's snapper (*L. russellii*). **Progress in Natural Science**, v. 18, n. 10, p.1233-1238, 2008.
- HAAS, B.J.; PAPANICOLAOU, A.; YASSOUR, M. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols**, v. 8, n. 8, p.1494-1512, 2013.

- HAHN, C.; BACHMANN, L.; CHEVREUX, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. **Nucleic Acids Research**, p.1-9, 2013.
- HASHIMOTO, D.T.; PRADO, F.D. do; SENHORINI, J.A. et al. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. **Aquaculture Research**, v. 44, p.876-884, 2013.
- HE, A.; LUO, Y.; YANG, H. et al. Complete mitochondrial DNA sequences of the Nile tilapia (*Oreochromis niloticus*) and Blue tilapia (*Oreochromis aureus*): genome characterization and phylogeny applications. **Molecular Biology Reports**, v. 38, n. 3, p.2015-2021, 2010.
- HEIDTMANN, Laura Moretti. **Caracterização do genoma mitocondrial de onça-pintada (*Panthera onca*) e elucidação da filogenia mitogenômica do gênero pantera**. Porto Alegre: Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul. 2014. 58 p. Dissertação (Mestrado em Zoologia) - Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul. 2014.
- HRBEK, T.; FARIAS, I.P. The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). **Genetics and Molecular Biology**, v. 1, n. 31, p.293-302, 2008.
- IWASAKI, W.; FUKUNAGA, T.; ISAGOZAWA, R. et al. MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. **Molecular Biology and Evolution**, v. 30, p.2531-2540, 2013.
- JIANG, M.; YANG, C.; WEN, H. The complete mitochondrial genome of *Aspiorhynchus laticeps* and its phylogenetic analysis. **Meta Gene**, v. 2, p.218-225, 2014.
- KARTAVTSEV, Y.P.; JUNG, S.O.; LEE, Y.M. et al. Complete mitochondrial genome of the bullhead torrent catfish, *Liobagrus obesus* (Siluriformes, Amblycipididae): Genome description and phylogenetic considerations inferred from the Cyt b and 16S rRNA genes. **Gene**, v. 396, n. 1, p.13-27, 2007.
- KIM, II C.; LEE, J.S. The complete mitochondrial genome of the rockfish *Sebastes schlegeli* (Scorpaeniformes, Scorpaenidae). **Molecules and Cells**, v. 17, n. 2, p.322-328, 2004.
- LOWE, T.M.; EDDY, S.R. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Research**, v. 25, n. 5, p.955-964, 1997.

- LUNDBERG, J.G.; SULLIVAN, J.P.; HARDMAN, M. Phylogenetics of the South America catfish family Pimelodidae (Teleostei: Siluriformes) using nuclear and mitochondrial gene sequences. **Proceedings of the Academy of Natural Sciences of Philadelphia**, v. 161, p.153-189, 2011.
- MAGOC, T.; SALZBERG, S. FLASH: Fast length adjustment of short reads to improve genome assemblies. **Bioinformatics**, v. 27, n. 21, p. 2957-2963, 2011.
- MEYER, A. Evolution of mitochondrial DNA in fishes. In: MOMMSEN, T.P.; HOCHACHKA, P.W. (Ed.) **Biochemistry and Molecular Biology of Fishes**. 2ed. Hardcover: Elsevier Science Publishers, 1993. Cap. 1. p.1-38.
- MU, X.; LIU, Y.; LAI, M. et al. Characterization of the *Macropodus opercularis* complete mitochondrial genome and family Channidae taxonomy using Illumina-based *de novo* transcriptome sequencing. **Gene**, v. 559, p.189-195, 2015.
- NAKATANI, M.; MIYA, M.; MABUCHI, K. et al. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeon origin and Mesozoic radiation. **BMC Evolutionary Biology**, v. 11, n. 1, p.1-25, 2011.
- PEREIRA, L.H.G.; HANNER, R.; FORESTI, F. et al. Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? **BMC Genetics**, v. 14, n. 20, p.1-14, 2013.
- POWELL, A.F.L.A.; BARKER, F.K.; LANYON, S.M. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: An avian case study. **Molecular Phylogenetics and Evolution**, v. 66, n. 1, p.69-79, 2013.
- PRADO, F.D.; HASHIMOTO, D.T.; MENDONÇA, F.F. et al. Molecular identification of hybrids between Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum*. **Aquaculture Research**, v. 42, n. 12, p.1890-1894, 2011.
- PRADO, F.D.; PARDO, B.G.; GUERRA-VARELA, J. et al. Development and characterization of 16 microsatellites for the Neotropical catfish *Pseudoplatystoma reticulatum* and cross species analysis. **Conservation Genetic Resources**, v. 6, n. 3, p.679-681, 2014.
- PROSDOCIMI, F.; de CARVALHO, D.C; de ALMEIDA, R.N. et al. The complete mitochondrial genome of two recently derived species of the fish genus *Nannoperca* (Perciformes, Percichthyidae). **Molecular Biology Reports**, v. 39, n. 3, p.2767-2772, 2011.

- RANGEL-MEDRANO, J.D.; ALZATE, J.F.; MÁRQUEZ, E.J. Complete mitochondrial genome of the Neotropical catfish *Pseudoplatystoma magdaleniatum* (Siluriformes, Pimelodidae). **Mitochondrial DNA**, p.1-2, 2015.
- RESTREPO-ESCOBAR, N.; ALZATE, J.F.; MÁRQUEZ, E.J. Mitochondrial genome of the Trans -Andean shovelnose catfish *Sorubim cuspicaudus* (Siluriformes, Pimelodidae). **Mitochondrial DNA**, p.1-2, 2014.
- SOSA, M.X.; SIVAKUMAR, I.K.A.; MARAGH, S. et al. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. **Plos Computational Biology**, v. 8, n. 10, p.1-11, 2012.
- TAMURA, K.; STECHER, G.; PETERSON, D. et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. **Molecular Biology and Evolution**, v. 30, p.2725-2729, 2013.
- TORRICO, J. P.; HUBERT, N.; DESMARAIS, E. et al. Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): biogeographic and evolutionary implications. **Molecular Phylogenetics and Evolution**, v. 51, n. 3, p.588-594, 2009.
- WARREN, R.L.; SUTTON, G.G.; JONES, S.J.M. et al. Assembling millions of short DNA sequences using SSAKE. **Bioinformatics**, v. 23, n. 4, p.500-501, 2007.
- WYMAN, S.K., JANSEN, R.K.; BOORE, J.L. Automatic annotation of organellar genomes with DOGMA. **Bioinformatics**, v. 20, p.3252–3255, 2004.
- YÁÑEZ, J.M.; NEWMAN, S.; HOUSTON, R.D. Genomics in Aquaculture to better understand species biology and accelerate genetic progress. **Frontiers in Genetic**, v. 6, p.1-3, 2015.
- YU, L.; LI, Y.-W.; RYDER, O.A. et al. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. **BMC Evolutionary Biology**, v. 7, p.198-209, 2007.
- YUE, G.H.; LIEW, W.C.; ORBAN, L. The complete mitochondrial genome of a basal teleost, the Asian arowana (*Scleropages formosus*, Osteoglossidae). **BMC Genomics**, v. 7, p.242-255, 2006.

ANEXO A

Tabela A.1. Anotações do genoma mitocondrial completo de *P. reticulatum* (16.576 pb) obtidas pelo método manual e pelos anotadores automáticos DOGMA (Wyman et al., 2004), MITOS (Bernt et al., 2013) e MitoAnnotator (Iwasaki et al., 2013)

Genes ^{a,b}	MANUAL			DOGMA			MITOS			MITOANNOTATOR		
	Posição (pb)		Tamanho (pb) ^c	Posição (pb)		Tamanho (pb) ^{c,e}	Posição (pb)		Tamanho (pb) ^{c,e}	Posição (pb)		Tamanho (pb) ^{e,f}
	Início	Fim		Início ^d	Fim ^d		Início ^d	Fim ^d		Início ^d	Fim ^d	
tRNA ^{Phe}	1	69	69	1	69	69(0)	1	69	69(0)	1	69	69(0)
rRNA 12S	70	1025	956	85(15)	1030(5)	946(-10)	70	1030(5)	961(5)	70	1030(5)	961(5)
tRNA ^{Val}	1031	1102	72	1031	1102	72(0)	1031	1102	72(0)	1031	1102	72(0)
rRNA 16S	1103	2782	1680	1103	2759(-23)	1657(-23)	1103	2784(2)	1682(2)	1103	2782	1680(0)
tRNA ^{Leu}	2783	2857	75	2783	2857	75(0)	2783	2857	75(0)	2783	2857	75(0)
NAD1	2859	3827	969	2859	3827	969(0)	2862(3)	3824(-3)	963(-6)	2859	3830	972(0)
tRNA ^{Ile}	3832	3903	72	3832	3903	72(0)	3832	3903	72(0)	3832	3903	72(0)
tRNA ^{Gln}	3903	3973	71	3903	3973	71(0)	3903	3973	71(0)	3903	3973	71(0)
tRNA ^{Met}	3973	4042	70	3973	4042	70(0)	3973	4042	70(0)	3973	4042	70(0)
NAD2	4043	4990	948	4043	5086(96)	1044(96)	4043	5080(90)	1038(90)	4043	5087(96)	1045(948)
tRNA ^{Phe}				5087	5157	71						
tRNA ^{Trp}	5088	5158	71	5088	5158	71(0)	5088	5158	71(0)	5088	5158	71(0)
tRNA ^{Ala}	5177	5245	69	5177	5245	69(0)	5177	5245	69(0)	5177	5245	69(0)
tRNA ^{Asn}	5247	5319	73	5247	5319	73(0)	5247	5319	73(0)	5247	5319	73(0)
tRNA ^{Cys}	5349	5415	67	5349	5415	67(0)	5349	5415	67(0)	5349	5415	67(0)
tRNA ^{Tyr}	5417	5486	70	5417	5486	70(0)	5417	5486	70(0)	5417	5486	70(0)
COX1	5488	7035	1548	5488	7035	1548(0)	5494(6)	7026(-9)	1533(-15)	5488	7038	1551(0)
tRNA ^{Phe}				7039	7109	71						
tRNA ^{Ser}	7039	7109	71				7039	7109	71(0)	7039	7109	71(0)
tRNA ^{Asp}	7114	7186	73	7114	7186	73(0)	7114	7186	73(0)	7114	7186	73(0)
COX2	7201	7890	690	7201	7890	690(0)	7201	7884(-6)	684(-6)	7201	7891	691(0)
tRNA ^{Asp}				7892	7965	74						
tRNA ^{Lys}	7892	7965	74				7892	7965	74(0)	7892	7965	74(0)
ATP8	7967	8131	165	7967	8131	165(0)	7967	8131	165(0)	7967	8134	168(0)
ATP6	8125	8805	681	8125	8805	681(0)	8125	8805	681(0)	8125	8807	683(0)
COX3	8808	9590	783	8808	9590	783(0)	8808	9590	783(0)	8808	9591	784(0)
tRNA ^{Gly}	9592	9664	73	9592	9664	73(0)	9592	9664	73(0)	9592	9664	73(0)
NAD3	9716	10012	297	9665(-51)	10012	348(51)	9665(-51)	10012	348(51)	9665(-51)	10013	349(51)
tRNA ^{Arg}	10014	10084	71	10014	10084	71(0)	10014	10084	71(0)	10014	10084	71(0)
NAD4L	10085	10360	276	10085	10378(18)	294(18)	10085	10378(18)	294(18)	10085	10381(18)	297(18)
NAD4	10375	11751	1377	10375	11751	1377(0)	10375	11745(-6)	1371(-6)	10375	11752	1378(0)

tRNA ^{His}	11753	11822	70	11753	11822	70(0)	11753	11822	70(0)	11753	11822	70(0)
tRNA ^{Ser}				11823	11889	67(0)	11823	11889	67(0)	11823	11889	67(0)
tRNA ^{Leu}	11895	11967	73	11895	11967	73(0)	11895	11967	73(0)	11895	11967	73(0)
NAD5	12022	13749	1728	11968(-54)	13791(42)	1824(96)	11980(-42)	13782(33)	1803(75)	11968(-54)	13794(42)	1827(96)
NAD6	13794	14306	513	13794	14306	513 (0)	13794	14306	513(0)	13791(-3)	14306(-3)	516(0)
tRNA ^{Glu}	14307	14375	69	14307	14375	69 (0)	14307	14375	69(0)	14307	14375	69(0)
CYTb	14377	15513	1137	14377	15513	1137 (0)	14377	15513	1137(0)	14377	15514	1138(0)
tRNA ^{Thr}	15515	15586	72	15515	15586	72 (0)	15515	15586	72(0)	15515	15586	72(0)
tRNA ^{Pro}	15585	15654	70	15585	15654	70 (0)	15585	15654	70(0)	15585	15654	70(0)
tRNA ^{Met}				15690	15789	100 (0)						
D-Loop	15655	16576	922							15655	16576	922(0)

^a Tanto na anotação manual quanto na realizada com os anotadores automáticos DOGMA, MITOS e MitoAnnotator, o gene *NAD6* e os genes tRNA^{Gln(TTG)}, tRNA^{Ala(TGC)}, tRNA^{Asn(GTT)}, tRNA^{Cys(GCA)}, tRNA^{Tyr(GTA)}, tRNA^{Ser(TGA)}, tRNA^{Glu(TTC)} e tRNA^{Pro(TGG)} estão localizados na fita L. Os demais genes localizam-se na fita H.

^b Os genes tRNAs apresentaram os tradicionais anticódons dos vertebrados nas quatro anotações realizadas.

^c Genes codificadores de proteína sem *stop* códon.

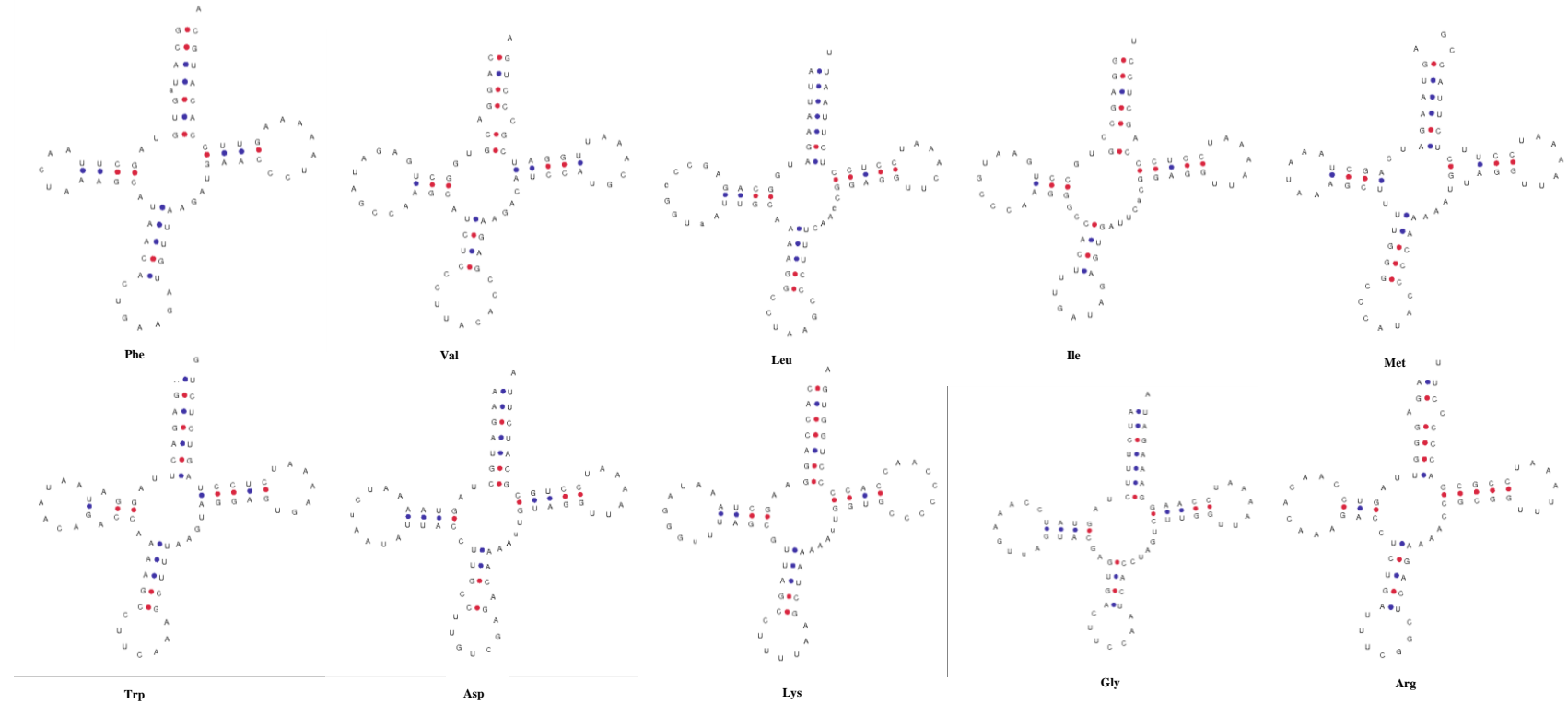
^d Números entre parênteses indicam diferenças nas posições iniciais ou finais em relação às posições observadas com a anotação manual.

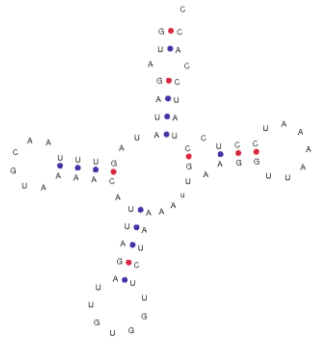
^e Números entre parênteses indicam diferenças no tamanho dos genes em relação ao tamanho observado com a anotação manual.

^f Genes codificadores de proteína com *stop* códon.

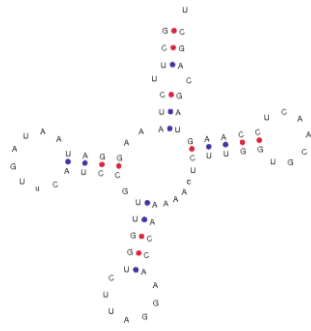
ANEXO B

Figura B.1. Estruturas secundárias dos genes tRNAs do mitogenoma de *Pseudoplatystoma reticulatum* obtidas com o software tRNAScan-SE 1.21 (Lowe & Eddy, 1997)

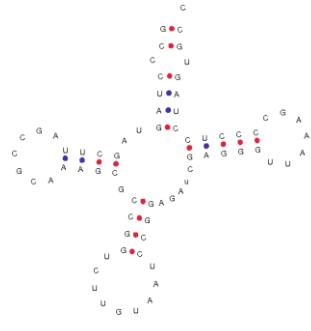




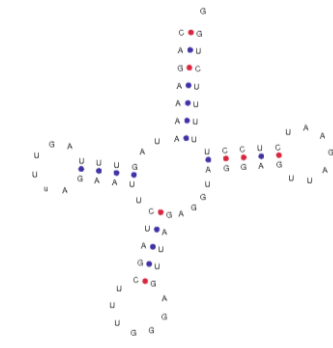
His



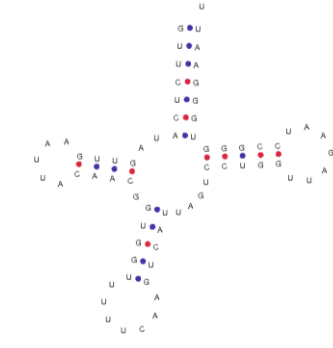
Leu



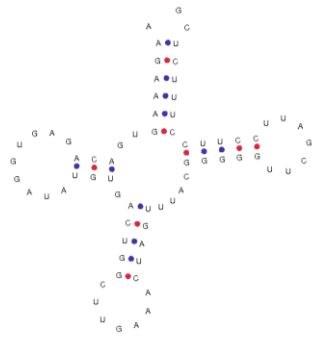
Thr



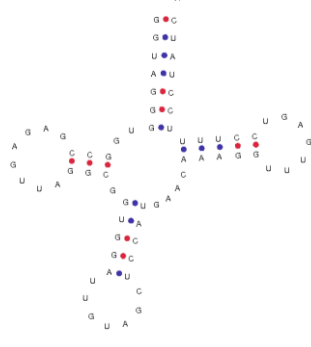
Pro



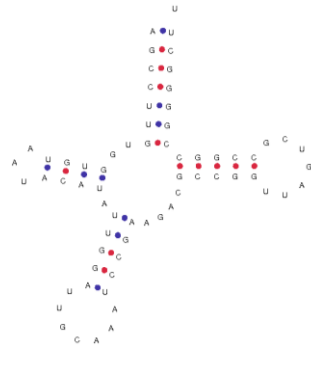
Glu



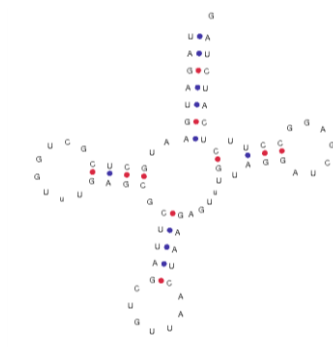
Thr



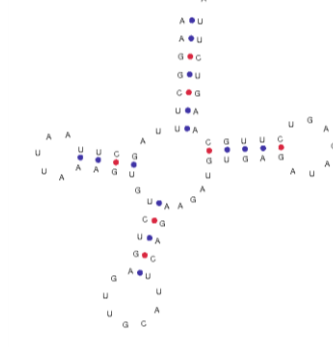
Tyr



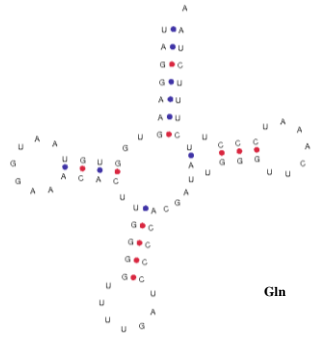
Cys



Asn



Ala



ANEXO C

Tabela C.1. Organização dos genomas mitocondriais de quatro espécies de bagres da família Pimelodidae disponíveis no Genbank: *Pseudoplatystoma magdaleniatum* (NC_026526.1), *Pseudoplatystoma corruscans* (NC_026846.1), *Pimelodus pictus* (NC_015797.1) e *Sorubim cuspicaudus* (NC_026211.1)

Genes	Fita	Anticódon	<i>P. magdaleniatum</i> ^a			<i>P. corruscans</i>			<i>P. Pictus</i>			<i>S. cuspicaudus</i> ^a		
			Posição (pb)		Tamanho (pb) ^b	Posição (pb)		Tamanho (pb) ^b	Posição (pb)		Tamanho (pb) ^b	Posição (pb)		Tamanho (pb) ^b
			Início	Fim		Início	Fim		Início	Fim		Início	Fim	
tRNA ^{Phe}	H	GAA	1	68	68	1	68	68	1	68	68	1	69	69
rRNA <i>12S</i>	H		69	1027	959	72	586	515	69	1024	956	70	1024	955
tRNA ^{Val}	H	TAC	1028	1099	72	594	665	72	1025	1096	72	1025	1096	72
rRNA <i>16S</i>	H		1100	2787	1688	666	2343	1678	1097	2728	1632	1097	2771	1675
tRNA ^{Leu}	H	TAA	2788	2862	75	2342	2416	75	2777	2851	75	2772	2846	75
NAD1	H		2864	3835	972	2418	3389	972	2853	3824	972	2848	3819	972
tRNA ^{Ile}	H	GAT	3837	3908	72	3391	3462	72	3826	3897	72	3821	3892	72
tRNA ^{Gln}	L	TTG	3908	3978	71	3462	3532	71	3897	3967	71	3892	3962	71
tRNA ^{Met}	H	CAT	3978	4047	70	3532	3601	70	3967	4036	70	3962	4031	70
NAD2	H		4048	5092	1045	3602	4646	1045	4037	5081	1045	4032	5076	1045
tRNA ^{Trp}	H	TCA	5093	5163	71	4647	4717	71	5028	5152	125	5077	5148	72
tRNA ^{Ala}	L	TGC	5179	5247	69	4734	4802	69	5167	5235	69	5161	5229	69
tRNA ^{Asn}	L	GTT	5249	5321	73	4804	4876	73	5237	5309	73	5231	5303	73
tRNA ^{Cys}	L	GCA	5351	5417	67	4906	4972	67	5340	5406	67	5333	5398	66
tRNA ^{Tyr}	L	GTA	5419	5488	70	4974	5043	70	5408	5477	70	5399	5468	70
COX1	H		5490	7040	1551	5045	6595	1551	5479	7029	1551	5470	7020	1551
tRNA ^{Ser}	L	TGA	7042	7110	69	6596	6666	71	7030	7100	71	7021	7091	71
tRNA ^{Asp}	H	GTC	7116	7188	73	6671	6743	73	7105	7177	73	7096	7165	70
COX2	H		7203	7893	691	6758	7448	691	7192	7882	691	7180	7870	691
tRNA ^{Lys}	H	TTT	7894	7967	74	7449	7522	74	7883	7956	74	7871	7944	74
ATP8	H		7969	8136	168	7524	7691	168	7958	8125	168	7946	8113	168

<i>ATP6</i>	H		8127	8810	684		7682	8365	684		8116	8798	683		8104	8787	684
<i>COX3</i>	H		8810	9594	785		8365	9148	784		8799	9582	784		8787	9570	784
tRNA ^{Gly}	H	TCC	9594	9666	73		9149	9221	73		9583	9655	73		9571	9643	73
<i>NAD3</i>	H		9667	10015	349		9222	9570	349		9656	10004	349		9644	9992	349
tRNA ^{Arg}	H	TCG	10016	10086	71		9571	9641	71		10005	10074	70		9993	10062	70
<i>NAD4L</i>	H		10087	10383	297		9642	9938	297		10075	10371	297		10063	10359	297
<i>NAD4</i>	H		10377	11754	1378		9932	11309	1378		10365	11745	1381		10353	11733	1381
tRNA ^{His}	H	GTG	11755	11824	70		11310	11379	70		11746	11815	70		11734	11803	70
tRNA ^{Ser}	H	GCT	11825	11891	67		11380	11446	67		11816	11882	67		11804	11870	67
tRNA ^{Leu}	H	TAG	11897	11969	73		11452	11524	73		11893	11965	73		11876	11948	73
<i>NAD5</i>	H		11970	13796	1827		11525	13351	1827		11966	13792	1827		11949	13775	1827
<i>NAD6</i>	L		13793	14308	516		13348	13863	516		13789	14304	516		13772	14287	516
tRNA ^{Glu}	L	TTC	14309	14377	69		13864	13932	69		14305	14373	69		14288	14356	69
<i>CYTb</i>	H		14379	15516	1138		13934	15071	1138		14375	15512	1138		14358	15495	1138
tRNA ^{Thr}	H	TGT	15517	15588	72		15072	15143	72		15513	15584	72		15496	15567	72
tRNA ^{Pro}	L	TGG	15587	15656	70		15142	15211	70		15583	15652	70		15566	15635	70
<i>D-Loop</i>			15657	16568	912		15212	16123	912		15653	16575	923		15635	16543	909

^amtDNA foi recentralizado para que a 1ª base coincidissem com a 1ª base do gene tRNA^{Phe}.

^bGenes codificadores de proteína com *stop* códon.

CAPÍTULO 6

CONSIDERAÇÕES FINAIS

Este trabalho envolveu o desenvolvimento de tecnologias moleculares inovadoras para espécies de bagres com pouco conhecimento biológico e a geração de informações moleculares, que irão compor bancos de dados públicos e que poderão ser utilizadas para estudos genéticos envolvendo essas espécies. Foram utilizadas tecnologias de sequenciamento de nova geração Illumina para o sequenciamento de RNA de diversos tecidos coletados de cachara e de híbridos de pintado. Os dados gerados a partir do RNA-Seq permitiram a montagem do transcriptoma da cachara e dos híbridos para a prospecção de SNPs espécie-específicos e a caracterização inicial do transcriptoma da cachara.

Um resultado não esperado gerado a partir do transcriptoma da cachara e que irá contribuir com estudos na área de filogenia, filogeografia, genética de populações, etc., foi a montagem do mitogenoma da cachara e sua posterior caracterização. As informações geradas serão depositadas em bases de dados públicas.

A hibridação interespecífica sem controle praticada pela maioria dos produtores de cachara (*P. reticulatum*) e pintado (*P. corruscans*) coloca em risco tanto os estoques de reprodutores nas próprias pisciculturas quanto os estoques nos ambientes naturais. Programas de melhoramento genético e de conservação para essas espécies, que poderiam melhorar sua produtividade só podem ser implementados com reprodutores puros. Por isso o desenvolvimento de ferramentas que possam identificar, de forma precisa, níveis de introgressão menores é fundamental. Embora a estratégia que precisou ser adotada neste estudo para o desenvolvimento de um painel de marcadores SNPs espécie-específicos não tenha sido a mais adequada (foram utilizados híbridos de pintado e não pintados puros, uma

vez que todos os animais coletados e identificados morfologicamente como pintado, apresentaram no minisequenciamento SNaPshot desenvolvido neste estudo, genótipo de híbrido: pintado x cachara), ela permitiu a identificação de 64 SNPs espécie específicos no transcriptoma da cachara e do híbrido, para a identificação de níveis de até 1,65% de introgressão genética cachara e pintado. Seria importante repetir esse experimento utilizando amostras coletadas de animais adultos das duas espécies, na natureza e puros. Com os animais puros, esperava-se identificar uma quantidade maior de SNPs espécie-específicos nos transcriptomas da cachara e do pintado.

O sequenciamento do transcriptoma possibilitou a anotação funcional dos transcritos que continham SNPs, para a termos de ontologia GO. E também a identificação de genes com expressão tecido-específica e genes *housekeeping* no transcriptoma da cachara. Análises adicionais permitirão a caracterização completa do transcriptoma da cachara, como a descrição funcional dos transcritos, além da identificação de genes e vias metabólicas envolvidas com processos biológicos importantes para a produção da espécie em cativeiro.

O ensaio SNaPshot desenvolvido, juntamente com o sequenciamento de quatro genes nucleares e um mitocondrial foi essencial para a identificação de indivíduos híbridos dentre as amostras coletadas. Se não houvesse sido realizado, estaríamos gerando informações incorretas a respeito dos pintados. Essa metodologia já foi utilizada para identificar híbridos em outras espécies de bagres e também foi adaptada para os peixes redondos, como tambaqui e pacu.

Além de fornecer informações moleculares novas para as duas espécies de bagres, proporcionar o treinamento da doutoranda, o estudo desenvolvido poderá servir como base para futuros estudos sobre genômica comparativa e funcional da cachara (*P. reticulatum*) e de outras espécies de peixes.