

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Paulo Henrique Jorge

**Sequenciamento do transcriptoma e caracterização de microssatélites na
pirapitinga *Piaractus brachypomus* para análises de variabilidade genética**

MESTRADO

Botucatu, 2016

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Paulo Henrique Jorge

**Sequenciamento do transcriptoma e caracterização de microssatélites na
pirapitinga *Piaractus brachypomus* para análises de variabilidade genética**

Dissertação apresentada ao Instituto de
Biociências de Botucatu, Universidade Estadual
Paulista – UNESP, como parte dos requisitos
para a obtenção do título de Mestrado em
Ciências Biológicas. Área de Concentração:
Zoologia

Orientador: Dr. Diogo Teruo Hashimoto

Botucatu, 2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Jorge, Paulo Henrique.

Sequenciamento do transcriptoma e caracterização de microssatélites na pirapitinga *Piaractus brachypomus* para análises de variabilidade genética / Paulo Henrique Jorge.
- Botucatu, 2016

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Diogo Teruo Hashimoto

Capes: 20400004

1. Microssatélites (Genética). 2. Marcadores genéticos. 3. Peixe - Genética. 4. Pirapitinga (Peixe). 5. Peixe - Genética. 6. Sequenciamento de nucleotídeo.

Palavras-chave: Marcadores moleculares; Peixes; Serrasalminidae.

*Eu dedico esta Dissertação aos meus pais, minha
família e ao Laboratório de Genética em Aquicultura
e Conservação – LaGeAC*

Epígrafe

“... E, para que não me exaltasse pela excelência das revelações, foi-me dado um espinho na carne, a saber, um mensageiro de Satanás para me esbofetear, a fim de não me exaltar.

Acerca do qual três vezes orei ao Senhor para que se desviasse de mim.

E disse-me: A minha graça te basta, porque o meu poder se aperfeiçoa na fraqueza. De boa vontade, pois, me gloriarei nas minhas fraquezas, para que em mim habite o poder de Cristo.

Por isso sinto prazer nas fraquezas, nas injúrias, nas necessidades, nas perseguições, nas angústias por amor de Cristo. Porque quando estou fraco então sou forte...”.

2 Coríntios 12:7-10

Agradecimentos

Gostaria de expressar meus sinceros agradecimentos a todos que me ajudaram com a elaboração desta Dissertação de Mestrado, contei com inúmeros incentivos, que estarei eternamente grato pela força e o respeito que tiveram comigo e com o meu projeto.

A Nossa Senhora Aparecida, por me aparar na dor e na alegria, sempre que lhe fui pedir e veio em meu socorro, muito obrigado MÃE por não me deixar desistir e por ajudar a superar as dificuldades.

Ao Doutor, Professor e orientador Diogo Teruo Hashimoto, pela orientação, dedicação, disponibilidade, colaboração e pela confiança, sem dúvida esta Dissertação não seria possível sem seu grande esforço.

À minha Psicóloga Karina Cestari, sua amizade e seus ensinamentos foram para mim de valor inestimável.

Ao Professor de Inglês Alcides Antônio Doretto Cintra (Tidão), que além de me ensinar inglês, entendeu minhas dificuldades e me ajudou muito.

Aos meus amados pais, Paulo Sergio Jorge e Paula de Barros Vieira Jorge que sempre fizeram e me deram de tudo, se dedicaram à minha educação como ser humano, me deram muito amor. Vocês fizeram de mim a pessoa que hoje sou, e eu só tenho motivos para agradecer.

As minhas avós Dirce de Barros Vieira e Erothildes Caraski de Jorge, agradeço pelas orações e pelos almoços e jantas deliciosos que só as senhoras sabem fazer.

A oportunidade de ser aluno do Instituto de Biociências de Botucatu, ao programa de pós-graduação em Ciências Biológicas com área de concentração em Zoologia e também ao CAUNESP – Centro de Aquicultura da UNESP.

Ao meu Laboratório do CAUNESP, LaGeAC – Laboratório de Genética em Aquicultura e Conservação, Lucas, Vito, Milene, Bruna, Marcos, Raquel, Milena, Natália e Paola, pelo companheirismo, pela contribuição, pelas risadas e pelos grandes momentos que guardo em meu coração, à todos vocês que me ajudaram nesta etapa de superação, desejo muito sucesso, porque vocês são inesquecíveis, agradeço por entrarem em minha vida e sempre terão o meu carinho e respeito.

Não posso deixar de ressaltar quatro pessoas que foram de extrema importância para a condução da minha dissertação, que deixaram suas vidas e sempre estavam dispostos para me ajudar em todos os momentos, não importando a hora, o meu reconhecimento para Prof. Diogo Teruo Hashimoto, Vito Antônio Mastrochirico Filho, Milene Elissa Hata e Natália Jade Mendes, a todos vocês, muito obrigado pelo apoio.

A todos meus amigos dos outros laboratórios da CAUNESP, técnicos e funcionários, que estiveram comigo nesta caminhada, o meu muito obrigado.

Aos pesquisadores Giovanni Vitti Moro e Leandro Kanamaru, da Embrapa Pesca e Aquicultura, pelo auxílio nas coletas de material biológico realizadas no Tocantins.

A FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo, pelo auxílio da bolsa concebida (PROCESSO FAPESP 2014/05732-2) e auxílio à pesquisa (PROCESSO FAPESP 2014/03772-7).

A CAPES – Coordenação de Aperfeiçoamento de Pessoal Nível Superior, pelo auxílio concebido, como bolsa inicial, durante a vigência de maio a junho de 2014.

RESUMO

A pirapitinga, *Piaractus brachypomus* (Characiformes, Serrasalminidae), é um peixe de ocorrência na bacia Amazônica e uma das principais espécies nativas utilizadas para a produção em aquicultura. O principal objetivo deste estudo foi realizar o sequenciamento do transcriptoma de *P. brachypomus* por meio de Sequenciamento de Nova Geração (NGS) e, em seguida, caracterizar um conjunto de marcadores microssatélites para esta espécie. Além disso, marcadores microssatélites polimórficos foram usados para realizar a análise da variabilidade genética em estoques cultivados de pirapitinga. O sequenciamento do transcriptoma foi realizado por meio da tecnologia 454/Roche, que resultou em 3.696 contigs não reduntantes (nr). Destes 2.568 genes com correspondência no banco de proteínas (nr) foram caracterizados nas categorias de Gene Ontology (GO), sendo que 2.075 sequências (80,8%) foram anotadas em termos GO. Dos 30 *loci* de microssatélites, após o processo validação, 8 *loci* de microssatélites demonstraram polimorfismo. A análise desses marcadores polimórficos em estoques cultivados revelou que as pisciculturas do Norte apresentaram uma maior variabilidade genética (riqueza de alelos e heterozigosidade) do que as Pisciculturas do Sudeste. Além disso, os resultados da AMOVA demonstraram que a maior variação estava presente dentro das populações (62,5%). Entretanto, quando comparado os grupos de acordo com a origem (Selvagem, Pisciculturas do Norte e Pisciculturas do Sudeste), foi verificada uma variação considerável (31,76%) entre os grupos. Estes dados foram corroborados com os valores de F_{st} , pois houve alta estruturação genética entre os estoques cultivados e a população selvagem, bem como também entre as Pisciculturas do Norte e do Sudeste. A estratégia de sequenciamento do transcriptoma por NGS se demonstrou eficaz na prospecção de marcadores moleculares, além de gerar um alto volume de recursos genéticos para serem explorados em diversas áreas da biologia. Os microssatélites gerados nesse estudo são importantes ferramentas para serem empregadas no manejo genético de estoques de reprodutores cultivados, o que poderá aumentar a produtividade em sistemas de produção e gerar um alto valor agregado do pescado com qualidade genética.

ABSTRACT

The pirapitinga, *Piaractus brachypomus* (Characiformes, Serrasalminidae) is an fish occurrence in the Amazon basin and one of the main native species used for production in aquaculture. The aim of this study was performed transcriptome sequencing of *P. brachypomus* through Next Generation Sequencing (NGS) and then characterize a set of microsatellite markers for this species. The sequencing of the transcriptome was conducted through technology 454/Roche, which resulted in 3,696 contigs non redundant (nr). Of these 2,568 genes with matching proteins in a database (nr) were characterized in the categories of Gene Ontology (GO), and sequences 2,075 (80.8%) were noted in GO terms. 30 microsatellite loci, after the validation process, 8 microsatellite loci showed polymorphism. The analysis of these polymorphic markers in cultured stocks revealed that the northern fish farms had a higher genetic diversity (richness of alleles and heterozygosity) than the fish farms in the Southeast. In addition, the AMOVA results demonstrated that the highest variation was present in the population (62.5%). However, when comparing the groups according to the source (Wild, North Fish farms and fish farms in the Southeast) considerable variation (31.76%) between the groups was observed. These data were corroborated with *F*_{st} values, as there was high genetic structuring among cultivated stocks and the wild population, and also between fish farms in North and Southeast. The sequencing strategy of the transcriptome by NGS has been shown effective in the exploration of molecular markers, and generate a high volume of genetic resources to be exploited in many areas of biology. Microsatellites generated in this study are important tools to be used in the genetic management of cultured broodstocks, which could increase productivity in production systems and generate a high added value of the fish with genetic quality.

SUMÁRIO

1. INTRODUÇÃO	2
1.1 Caracterização da espécie <i>Piaractus brachypomus</i>	3
1.2 Problemas decorrentes da endogamia.....	4
1.3 Marcador molecular do tipo microssatélite	7
1.4 Sequenciamento de nova geração para descoberta de microssatélites	9
2. MATERIAIS E MÉTODOS.....	12
2.1 Materiais	12
2.2 Métodos	15
2.2.1 Análise de pureza genética dos indivíduos de <i>Piaractus brachypomus</i>	15
2.2.2 Caracterização do transcriptoma e descoberta dos microssatélites.....	15
2.2.3 Extração de DNA para validação dos microssatélites	18
2.2.4 Validação dos microssatélites por PCR	19
2.2.5 Genotipagem dos microssatélites em sequenciador automático	20
2.2.6 Análise estatística para estudos de variabilidade genética.....	21
3. RESULTADOS.....	22
3.1 Sequenciamento do Transcriptoma	22
3.2 Descoberta de microssatélites e anotação funcional.....	27
3.3 Validação dos marcadores microssatélites	34
3.4 Análise da diversidade genética em estoques cultivados de pirapitinga	40
4. DISCUSSÃO	48
5. CONCLUSÕES	56
6. REFERÊNCIAS BIBLIOGRÁFICAS	57

1. INTRODUÇÃO

No Brasil, apesar da aquicultura comercial representar uma atividade zootécnica recente (estabelecido efetivamente na década de 1990), este setor tem crescido de forma expressiva nos últimos anos. Entre o período de 2007 a 2010, a produção de pescado pela aquicultura teve um avanço de 87%, cujos valores passaram de 210.644 t para 394.340 t, segundo estatísticas do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis e Ministério da Pesca e Aquicultura (IBAMA, 2007; MPA, 2012). Assim, somado ao crescente consumo de peixes (MPA, 2012) e as vantagens nutricionais do pescado para alimentação humana, a adoção de tecnologias genéticas por pisciculturas são fundamentais para suprir a demanda do mercado consumidor, bem como para garantir a produção de peixes com qualidade e de forma sustentável.

A piscicultura de água doce no Brasil é dividida em dois setores: a produção de alevinos e o cultivo de peixes (engorda) (Suplicy, 2007). É possível encontrar empresas com uma produção de alevinos de pequena a grande escala em quase toda parte do país, embora o número exato de produtores seja difícil de determinar devido a um deficiente serviço de estatística neste setor. A escala de produção varia de 5 mil a 50 milhões de alevinos/ano. Entretanto, a maioria das pisciculturas produz cerca de 500 mil a 10 milhões de alevinos/ano. O Estado de SP é um dos que apresenta maior número de instalações de pisciculturas produtoras de alevinos (Suplicy, 2007).

Na década de 1980, a indústria do comércio de alevinos foi fortemente focada para fornecer peixes a empreendimentos de lazer ou repovoamento de represas. A produção de alevinos para a piscicultura comercial visando o abastecimento de restaurantes, supermercados e fábricas de processamento e filetagem só foram estabelecidos na década de 1990, o que trouxe consideráveis mudanças na demanda do consumidor (Suplicy, 2007).

Logo, no Brasil há uma crescente exigência por peixes de melhor qualidade, o que mostra a necessidade do estabelecimento de estratégias de manejo genético para as pisciculturas nacionais, principalmente para os produtores de alevinos, que é considerado a base deste setor.

A partir dos anos 80, a utilização de metodologias genéticas passou a contribuir de maneira efetiva nos programas de criação de peixes de pisciculturas mundiais (Foresti, 2000; Hulata, 2001). Entretanto, no Brasil, é insignificante o percentual de pisciculturas que se utilizam de tecnologias genéticas para a produção de alevinos com qualidade. Para as espécies de peixes nativos do país, poucas tecnologias genéticas estão disponíveis para ser utilizado em nível de produção, principalmente o uso de marcadores moleculares, que poderiam ser aplicados no monitoramento da endogamia ou direcionamento de cruzamentos dos estoques de reprodutores.

1.1 Caracterização da espécie *Piaractus brachipomus*

De acordo com Jégu (2003), a subfamília Serrasalminae inclui 15 gêneros e 80 espécies, com uma distribuição estritamente na região Neotropical. Atualmente, este grupo é classificado em nível de família, denominado Serrasalmidae (Calcagnotto et al., 2005; Mirande, 2010), e englobam os peixes conhecidos popularmente conhecidos como pacus, tambaqui e as famosas piranhas. Com base em dados moleculares, os serrasalmídeos são compreendidos por três grupos principais: “Piranha”, “Myleus” e “Pacu” (Ortí et al., 1996; Ortí et al., 2008).

O clado “Pacu” inclui espécies herbívoras dos gêneros *Mylossoma*, *Colossoma* e *Piaractus*. As espécies deste clado são as mais importantes economicamente da família, pois são consideradas alimentos de alta qualidade na América do Sul. No Brasil, o cultivo de

representantes deste grupo corresponde à maior parcela da produção de espécies nativas na aquicultura continental (IBAMA, 2007). O aproveitamento destas espécies pela piscicultura tem sido uma boa alternativa para suprir a demanda do mercado e reduzir o impacto nos estoques das populações naturais provocados pelo alto nível de exploração.

A pirapitinga *Piaractus brachypomus* é um peixe nativo das bacias dos rios Amazonas e Orinoco e pode alcançar até 20 kg de peso corporal (Alcântara et al., 1990). É considerada uma espécie rústica, valorizada pela sua carne, com crescimento rápido, e que apresenta excelente condição para a criação de peixes (Fresneda et al., 2004). Em condições naturais, a pirapitinga migra entre Junho a Outubro e desova entre Novembro a Fevereiro, quando a temperatura da água atinge 27°C após as primeiras chuvas (Hernández, 1994).

No Brasil, o cultivo de *P. brachypomus* alcança o terceiro maior valor de produção de pescado (aproximadamente 10 mil toneladas) entre as espécies nativas puras (MPA, 2013a). Além disso, esta espécie tem sido amplamente utilizada para a produção de híbridos interespecíficos, em especial a tambatinga (fêmea de tambaqui x macho de pirapitinga), cujo valor de produção é superior a 14 mil toneladas por ano. As unidades produtivas de pirapitinga concentraram-se particularmente nas regiões Centro-Oeste e Norte (87%), seguido do Sudeste (9%), Nordeste (3%) e Sul (1%) (MPA, 2013b). Esta espécie também tem considerável importância econômica para a aquicultura de outros países da América do Sul (Colômbia, Peru e Venezuela), e também em países asiáticos (China, Mianmar, Tailândia e Vietnã) (Flores Nava, 2007; Honglang 2007; FAO 2010).

Entretanto, apesar desta representatividade na aquicultura mundial, poucos dados biológicos e genéticos estão disponíveis para *P. brachypomus*, além de que nenhum programa de manejo ou melhoramento genético tem sido desenvolvido.

1.2 Problemas decorrentes da endogamia

Infelizmente, existem poucos estudos de endogamia com peixes, o que faz ser o aspecto menos estudado na genética de peixes. A maioria dos estudos foi conduzida com salmonídeos, particularmente com truta arco-íris. Enquanto a endogamia é uma importante forma de manipulação genética que pode ser usada para melhorar um estoque quando bem planejado e dirigido, processos de endogamia sem planejamento e descontrolado podem arruinar um estoque com a depressão por endocruzamento, o que causa uma diminuição no crescimento e na viabilidade, acoplada com um aumento de anormalidades (Tave, 1999).

Com aumento da consanguinidade em programas de piscicultura, aumenta-se a chance de ocorrer à depressão por endocruzamento, que resulta do fato de indivíduos biologicamente aparentados terem maior chance de possuírem genes recessivos deletérios. Assim, a probabilidade de anomalias biológicas e zootécnicas aumenta quando os pais são aparentados e, quanto maior for o nível de consanguinidade, maiores serão os danos biológicos, tais como o crescimento e a sobrevivência, causando também aumento na frequência de deformidades de coluna vertebral e opérculos (Kincaid, 1983; Galhardo, 1989; Su et al., 1996; Imsland et al., 2001).

Desta forma, nos estoques utilizados em piscicultura, os níveis de consanguinidade devem ser mantidos abaixo dos limites críticos, sob aspecto da biologia e genética (Toledo-Filho et al., 1998). Entretanto, especialmente na piscicultura brasileira, há uma tendência dos plantéis de matrizes serem formados por peixes que possuem alto grau de parentesco. Isto se deve principalmente aos seguintes fatores: formação de lotes de reprodutores a partir de estoques já endogâmicos de outras pisciculturas; falta de controle da reprodução e desconhecimento da origem dos animais; falta de recursos para obtenção de matrizes de boa qualidade genética; utilização de baixo número de reprodutores; carência de profissionais especializados na piscicultura; e ausência de programas de monitoramento de endogamia (Hashimoto et al., 2012).

Devido aos prejuízos que podem acarretar para os produtores de peixes, há uma imediata necessidade para o monitoramento da variabilidade genética dos plantéis de matrizes. Com a informação do perfil genético desses animais, em conjunto com práticas adequadas de manejo genético, pode-se reduzir e até evitar os principais problemas resultantes do uso inadequado de animais em programas de produção. Mesmo quando os estoques cultivados são formados por peixes selvagens, o processo de endogamia deve ser monitorado e evitado, pois geralmente uma baixa amostragem de peixes coletados não representa a variabilidade genética natural e os acasalamentos são realizados sem qualquer critério de seleção (Hashimoto et al., 2012).

O outro motivo pela qual se torna imprescindível o monitoramento da variabilidade genética em pisciculturas é o uso das espécies cultivadas como potenciais bancos genéticos *ex-situ* para repovoamentos de populações selvagens. O objetivo destes programas de repovoamento tem sido contribuir para a conservação do potencial biológico de populações selvagens de peixes, cujo hábitat tenha sido alterado e que corram, portanto, risco de redução ou até mesmo de extinção (Toledo-Filho et al., 1992; Koljonen et al., 2002; Machado-Schiaffino et al., 2007).

Existem três métodos mais usados para estimar os níveis de consanguinidade em piscicultura: método da genealogia, método do número de parentais e método dos marcadores genéticos. Dentre estes, os marcadores genéticos mostram-se mais vantajosos por possibilitarem cálculos relativos de endocruzamentos acumulados, sem que haja necessidade de conhecimento prévio a respeito da genealogia ou do número de reprodutores efetivamente usados por geração (Toledo-Filho et al., 1998).

A aplicação de tecnologias moleculares para monitoramento da endogamia requer metodologias polimórficas e com alto poder de resolução para revelar as variações genéticas entre os indivíduos (Liu e Cordes, 2004). Os microssatélites tem se mostrado eficiente para

esta finalidade e representam a principal classe de marcadores genéticos aplicados em peixes comerciais da piscicultura (Koljonen et al., 2002; Lehoczky et al., 2005; Machado-Schiaffino et al., 2007).

1.3 Marcador molecular do tipo microssatélite

Inferências sobre fluxo gênico, variabilidade e estrutura genética em populações podem ser feitas por meio de várias técnicas moleculares disponíveis para a detecção de polimorfismos genéticos em nível de DNA. O desenvolvimento da tecnologia da reação de polimerase em cadeia (*PCR – Polymerase Chain Reaction*) causou uma revolução na biologia, por possibilitar análises moleculares de grande número de indivíduos de qualquer organismo (Ferreira e Grattapaglia, 1998).

Microssatélites, também conhecidos como sequências simples repetidas (SSRs - *Single Sequence Repeats*) ou repetições curtas em *tandem* (STRs - *Short Tandem Repeats*), representam um tipo único de sequências genômicas de DNA não codificante repetidas em *tandem*, as quais são distribuídas abundantemente pelo genoma eucariótico e demonstram altos níveis de polimorfismo alélico (Oliveira et al., 2006). São marcadores codominantes onde as sequências repetidas são de tamanho relativamente pequeno (sequências curtas de 1 – 6 pb) que podem ser amplificadas facilmente pela PCR. Os microssatélites geralmente estendem-se de vinte a algumas centenas de bases (Chistiakov et al., 2006).

Os microssatélites ocorrem em alta frequência e são aleatoriamente distribuídos pelo genoma. As estimativas da taxa de mutação de microssatélites em *E. coli* em sistemas *in vivo* são elevadas (10^{-2}), quando comparadas à taxa de mutação de ponto que ocorre na ordem de 10^{-9} e 10^{-10} (Hancock 2000). Vários mecanismos têm sido sugeridos para explicar a alta taxa de mutação dos microssatélites gerando a variação no número de repetições em *tandem*: erros

durante a recombinação, eventos de *crossing-over* desiguais e deslizamento da DNA polimerase durante a replicação ou reparo (Oliveira et al., 2006).

A distribuição dos microssatélites no indivíduo é função da história evolutiva no genoma e dos efeitos da pressão de seleção. Os locos SSR parecem ser geneticamente estáveis, são multialélicos e apresentam segregação mendeliana. Em uma população, todos os alelos daquele loco podem ser detectados e discriminados, fazendo com que os microssatélites estejam entre os marcadores moleculares mais apropriados para mapeamento genético e físico de genomas, para a identificação e discriminação de genótipos, análise de parentesco e estudos de genética de populações (Ferreira e Grattapaglia, 1998; Fungaro e Vieira, 2001; Oliveira et al., 2006).

Quando sequências únicas que flanqueiam um dado microssatélite são conhecidas, *primers* complementares podem ser sintetizados e utilizados como iniciador loco-específico em reações de PCR. Os produtos de amplificação de diferentes indivíduos, visualizados após a eletroforese, geralmente diferem em tamanho devido às diferenças no número de unidades de repetições no loco de microssatélite. Cada unidade de microssatélite assim detectada é um loco polialélico e, em um indivíduo diploide, a amplificação de um loco resultará em uma ou duas bandas, respectivamente, dependendo se é um homozigoto ou heterozigoto (Fungaro e Vieira, 2001).

Os microssatélites apresentam vantagens sobre os métodos clássicos da genética na análise da estrutura genética de populações, tais como: a herança das características detectadas pode ser facilmente demonstrada, a maioria dos locos é codominante e as frequências alélicas podem ser calculadas diretamente sem a necessidade de delineamentos genéticos; as estimativas da variabilidade genética podem ser comparadas diretamente entre populações e um conjunto de locos pode ser acessado usando-se uma pequena quantidade de material. A utilização de microssatélite, pela sua neutralidade, apresenta como principal

vantagem à possibilidade de caracterização de forças evolutivas sem a influência da seleção, permitindo maior compreensão do comportamento das populações naturais, em termos de deriva genética e migração (Chistiakov et al., 2006).

Entretanto, esta classe de marcadores ainda não está disponível para a maioria das espécies, em especial para *P. brachypomus*, considerada um organismo não modelo da aquicultura. A geração de dados genéticos para a pirapitinga é um dos aspectos prioritários para ser incorporada dentro do cenário de pesquisa agropecuária no país, frente à sua importância econômica para piscicultura nacional e mundial, a crescente demanda do consumidor por produtos melhorados e a expansão da piscicultura brasileira nos últimos anos. Somente nestas condições será possível realizar um programa de manejo e melhoramento genético acelerado para a pirapitinga, como tem sido feito para outras espécies da piscicultura mundial, como salmão, catfish, carpa e tilápia (Liu et al., 2011; Lien et al., 2011; Jin et al., 2012; Guyon et al., 2012).

1.4 Sequenciamento de nova geração para descoberta de microssatélites

Para suprir esta demanda, as tecnologias de sequenciamento de nova geração (NGS) estão causando uma revolução na biologia nos últimos cinco anos, pois têm possibilitado à descoberta de milhões de marcadores moleculares em organismos não modelos (Seeb et al., 2011; Davey et al., 2011). Para identificar um maior número de marcadores moleculares, é necessário um maior rendimento das leituras de sequências. Neste caso, sequenciadores de nova geração como as plataformas Roche/454, Solexa/Illumina e ABI/SOLID são particularmente adaptadas para produzir uma alta cobertura de sequências com precisão.

Com a utilização de tecnologias NGS, a ausência de um genoma referência é uma das maiores dificuldades para montagem de sequências em organismos não modelos. Nestes

casos, as leituras individuais podem ser montadas em sequências consenso (*contigs*), que poderão servir como um genoma de pseudoreferência (Waldbieser, 2011). Além disso, para descoberta de marcadores em organismos sem genoma de referência, uma etapa de redução genoma para adquirir conjuntos de *contigs* redundantes é outra estratégia que deve ser realizada. Para isto, o sequenciamento do transcriptoma (RNA-seq), especificamente o RNA mensageiro de determinados tecidos, é o mais utilizado e considerado fundamental para garantir o sucesso na busca de marcadores moleculares (Seeb et al., 2011).

Uma das principais vantagens é que os marcadores moleculares derivados do transcriptoma estão associados a genes atuais, com informação funcional, evitando íntrons e regiões extragênicas que podem dificultar a análise de dados (Parkinson e Blaxter, 2009). Além disso, RNA-seq permite o uso de marcadores para mapeamento genético e subsequente uso em estudos comparativos (Sarropoulou et al., 2008; Xu et al., 2012), que é importante para espécies da aquicultura sem o genoma completo. A análise do transcriptoma tem demonstrado ser muito eficaz para identificação de SNPs e microssatélites em peixes, tanto para espécies modelos como catfish e carpa (Liu et al., 2011; Ji et al., 2012), como no caso de espécies não-modelos (Long et al., 2013; Gross et al., 2013).

Além da possibilidade de caracterizar a expressão gênica e identificar novos genes (Morozova et al., 2009; Zheng et al., 2013), o RNA-seq permite estudar funcionalmente os SNPs gene-associados, uma vez que estes podem apresentar diferentes implicações dependendo da exata localização genômica (regiões não traduzidas ou regiões codificantes - cds). As mutações em cds causam efeitos particularmente sobre a composição de aminoácidos e, por conseguinte, a funcionalidade pode ser avaliada por meio da diferenciação de SNPs sinônimos ou não-sinônimos (Xu et al., 2012). Já os microssatélites, quando estão presentes em regiões gênicas, podem apresentar uma maior pressão de seleção e divergência populacional e, conseqüentemente, também podem gerar informações diferenciadas em

relação aos marcadores neutros (Vera et al., 2013). Além disso, também é possível realizar a anotação funcional dos *contigs* para verificar quais as classes funcionais (função molecular, componente celular e processo biológico) e vias metabólicas em que estão inseridos (Xu et al., 2012), resultando em melhores inferências sobre os efeitos que SNPs e microssatélites podem causar.

2. MATERIAIS E MÉTODOS

2.1 Materiais

Para realizar o sequenciamento do transcriptoma e prospectar os marcadores moleculares, foram coletadas amostras de fígado de 10 exemplares de pirapitinga *Piaractus brachypomus*, provenientes de diferentes pisciculturas e ambiente natural, o que possibilitou a obtenção de maior variabilidade genética possível (Tabela 1, Figuras 1 e 2). Para a coleta do material, os peixes foram sacrificados por overdose do anestésico benzocaína. O sequenciamento do transcriptoma foi realizado a partir do fígado porque este órgão desempenha um papel crítico na coordenação de vários processos fisiológicos, incluindo a digestão, metabolismo, desintoxicação, biossíntese de proteínas séricas, sistema endócrino e resposta imune (Zheng et al., 2013). Todas as coletas foram feitas sob autorização da Licença Permanente para Coleta de Material Zoológico (nº 33435-1), emitida pelo ICMBio (MMA).

Tabela 1 - Amostras de peixes coletados para o sequenciamento de nova geração.

Localidade	<i>n</i>
Fazenda São Paulo, Brejinho de Nazaré, TO	5
CAUNESP, Jaboticabal, SP	3
Projeto Peixe, Sales de Oliveira, SP	1
Rio Tocantins, Lajeado, TO	1

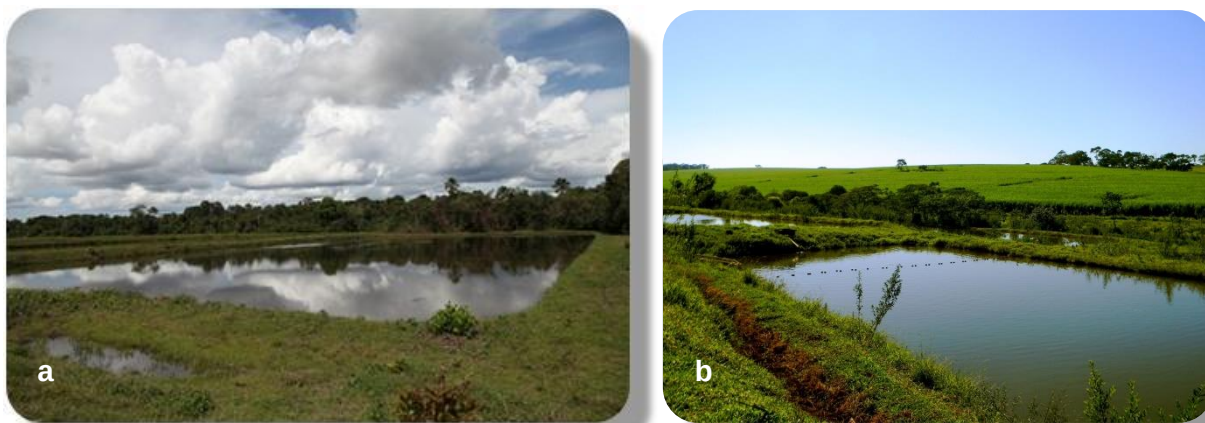


Figura 1 - Em (a), piscicultura da Fazenda São Paulo, Brejinho de Nazaré, TO. Em (b), piscicultura do Projeto Peixe, Sales de Oliveira, SP.



Figura 2 - Em (a), piscicultura do CAUNESP, Jaboticabal, SP. Em (b), rio Tocantins, Lajeado, TO.

Para a caracterização e validação dos microssatélites, as análises foram realizadas em 22 indivíduos coletados no Rio Tocantins (Lajeado, TO), que estão mantidos no banco de material biológico e DNA do Laboratório LaGeAC (Laboratório de Genética em Aquicultura e Conservação), Centro de Aquicultura (CAUNESP, Jaboticabal - SP) (Tabela 2). Além disso, para o estudo de variabilidade genética e análise de parentesco, foram coletadas amostras (fragmentos de nadadeira) em quatro pisciculturas comerciais (Tabela 2 e Figura 3), sendo duas da região Norte (Estado do TO) e duas do Sudeste (Estado de SP). Para manter a integridade destas pisciculturas, uma vez que o objetivo do estudo é a validação e aplicação de marcadores para estudos de variabilidade, e não uma análise comercial ou prestação de

serviços, os nomes e informações das pisciculturas foram preservados. Desta forma, as pisciculturas no presente estudo foram descritas como: SP1 e SP2 (provenientes do Estado de SP); TO1 e TO2 (do Estado do TO). Os animais foram individualmente marcados com pit-tags (*passive integrated transponder tag*, modelo *full-duplex* FDX-B, 134.2-kHz) e mantidos vivos nas instalações das pisciculturas para que seja realizado o posterior manejo genético dos reprodutores, por meio da análise de parentesco. As coletas foram feitas com animais anestesiados (benzocaína 60 mg/l, Sigma), sob autorização da Licença Permanente para Coleta de Material Zoológico (nº 33435-1), emitida pelo ICMBio (MMA).



Figura 3 - Foto de um exemplar de pirapitinga (*Piaractus brachypomus*) coletado em piscicultura.

Tabela 2 - Amostras de peixes coletados para posterior análise de variabilidade genética.

Pisciculturas	Número amostral
SP1	36
SP2	20
TO1	25
TO2	26
Rio Tocantins / Palmas, TO	22

2.2 Métodos

2.2.1 Análise de pureza genética dos indivíduos de *Piaractus brachypomus*

De acordo com Hashimoto et al. (2014), híbridos interespecíficos tem sido detectados em lotes de reprodutores de pisciculturas brasileiras. A pirapitinga pode ser cruzada com pacu *Piaractus mesopotamicus* ou tambaqui *Colossoma macropomum*, resultando em híbridos viáveis e férteis (Hashimoto et al., 2012). Portanto, no presente projeto, atenção especial foi direcionada para que sejam analisadas pirapitingas puras, e não híbridos interespecíficos. A análise da pureza genética em todos os animais aqui estudados foi realizada com o marcador do gene nuclear *tpm1* (Hashimoto et al., 2011), por meio de Multiplex-PCR. O ensaio Multiplex-PCR foi conduzido de acordo com os protocolos de Hashimoto et al. (2011). O tamanho dos fragmentos de DNA foi determinado em eletroforese de agarose 1,5%, corado com Nancy (Sigma, 5ul a cada 50ml de TBE). A visualização do gel foi em transiluminador com luz ultravioleta e capturada pela câmera CCD (3.1 megapixels), Locus Biotecnologia L-Pix STi. Os animais identificados como híbridos interespecíficos foram excluídos das análises posteriores deste estudo.

2.2.2 Caracterização do transcriptoma e descoberta dos microsatélites

O sequenciamento do transcriptoma foi realizado com a tecnologia da Roche/454, que tem sido o método mais utilizado para transcriptoma de espécies de peixes não modelos. A escolha desta plataforma deve-se as características das sequências geradas por esta tecnologia, que podem alcançar fragmentos de até 800 pb (pares de bases) em um rendimento de 400 Mb

(Milhões de Bases) por corrida. Estas características são ideais para organismos não modelos, cujos genomas de referência ainda não estão disponíveis (Renaut et al., 2010).

Para a construção das bibliotecas não normalizadas de cDNA foram seguidos os protocolos de Renaut et al. (2010). Fragmentos de cerca de 100 mg de fígado fixado em RNAlater foram extraídos com Rneasy Mini Kit (Qiagen), segundo recomendações do fabricante. Cada amostra foi quantificada por espectrofotometria no equipamento NanoDrop ND-1000 e verificada a qualidade (integridade) no equipamento 2100 Bioanalyzer. Sucedeu o preparado um pools equimolar de amostras de RNA total (proveniente de 10 indivíduos) para sequenciamento no equipamento 454 GS FLX+.

O RNA total do pool foi utilizado para enriquecimento de mRNA com o kit μ MACS mRNA Isolation Kit (Miltenyi Biotec). As bibliotecas de cDNA não-normalizadas foram preparadas com o kit cDNA Synthesis System Kit, com primer randômico, o GS Rapid Library Prep Kit e o GS Rapid Library MID Adaptors Kit (Roche). Resumidamente, 200 ng de mRNA da amostra foram utilizados para a fragmentação por tampão contendo ZnCl_2 por 20s a 70°C. O mRNA fragmentado foi purificado e selecionado por tamanho por SPRI (solid phase reversible immobilization) pelo kit RNAClean XP Reagent (Beckman Coulter). Após a purificação, o mRNA fragmentado foi submetido à síntese de cDNA por um protocolo two-step. O cDNA fragmentado foi purificado com AMPure XP beads (Beckman Coulter), polido e suas overhangs foram preenchidas para a obtenção de fragmentos blunt-ended. Rapid Library MID Adaptors (MID) foram ligados aos fragmentos de cDNA, com subsequente passo de purificação com AMPure XP beads.

Foi utilizado o High Sensitivity DNA LabChip Kit (Agilent Technologies) com 2100 Bioanalyzer para a análise de qualidade da biblioteca de cDNA. A concentração de moléculas/uL da amostra foi obtida pelo fluorímetro QuantiFluor™-ST (Promega) com as Minicell Borosilicate Glass Cuvettes. Para tal, foi gerada uma curva padrão e a fluorescência

do fluorocromo FAM conjugados aos adaptadores RL MID da amostra foi detectada. As concentrações das bibliotecas foram calculadas por meio do Rapid Library Quantitation Calculator (www.454.com/my454).

A titulação com emPCR (PCR de emulsão) foi com o GS FLX Titanium SV emPCR Kit (Lib-L) (Roche), de acordo com o emPCR Amplification Method Manual - LibL SV, GS FLX+ Series, para identificação do número ótimo de moléculas de DNA por bead (cpb = copies per bead). O número ideal de cpb é definido como aquele que resulta em 5 - 20% de beads enriquecidas carregando moléculas de DNA após a emPCR, sendo 8% a porcentagem ideal. Após a titulação com emPCR, foi realizada uma emPCR com o GS FLX Titanium LV emPCR Kit (Lib-L) (Roche), de acordo com o emPCR Amplification Method Manual - LibL LV, GS FLX+ Series. O sequenciamento foi realizado com GS FLX Titanium Sequencing Kit XL+ e o GS FLX Titanium PicoTiterPlate Kit (Roche), de acordo com Sequencing Method Manual for the GS FLX+ Instrument, GS FLX+ Series, na empresa Helixxa (Campinas, SP). O processamento das imagens e filtragem foi realizado pelo *pipeline* padrão para bibliotecas shotgun. Para tal, foram utilizados os pacotes GS Run Processor v2.6 e GS Run Reporter v2.6 (Roche).

Para a construção dos *contigs*, foi utilizado os programas e parâmetros descritos em Renaut et al. (2010). A filtragem da qualidade inicial das sequências 454 foi realizada por meio do programa da Roche Newbler. As sequências dos adaptadores foram retiradas do conjunto de *reads* e, em seguida, para subsequente montagem, foram excluídos os *reads* de baixa qualidade (*quality score* $Q < 20$ e nucleotídeos ambíguos 'N') e com tamanho menor que 15 pb. A anotação funcional foi realizada pelo programa Blast2GO (Conesa et al., 2005). Em síntese, os transcritos não redundantes foram utilizados para uma busca no banco de dados do NCBI (zebrafish e fugu RefSeq protein database), usando BLASTX com $E\text{-value} \leq 1e^{-3}$. Em seguida, foram classificados nas categorias de GO (processo biológico, função molecular e

componente celular), com um limiar de *E-value* 1e-3. Os *contigs* foram analisados para identificar microssatélites usando o software msatcommander (Faircloth, 2008). Os primers flanqueadores dos microssatélites foram desenhados com o software primer3 (Rozen e Skaletsky, 2000).

2.2.3 Extração de DNA para validação dos microssatélites

O protocolo do kit comercial Wizard Genomic DNA Purification Kit (Promega) foi usado para a extração de DNA. O protocolo descrito abaixo foi empregado para isolamento e purificação de DNA de tecidos sólidos (nadadeira). A integridade e a quantidade das amostras de DNA foram analisadas através de eletroforese em gel de agarose 1,0%. As amostras de DNA obtidas estão armazenadas no LaGeAC (Laboratório de Genética em Aquicultura e Conservação) do CAUNESP – Jaboticabal/SP, formando uma biblioteca para as análises genético-moleculares.

O protocolo seguiu os procedimentos abaixo:

1. numerar duas séries de tubos de microcentrífuga (1,5 ml);
2. retirar um pequeno pedaço de tecido, adicionar nos tubos de microcentrífuga e colocar na estufa por aproximadamente 10 min;
3. adicionar 600 µl de “Nuclei Lysis Solution” ao tecido;
4. adicionar 5 µl de “Proteinase K” (10mg/ml);
5. colocar no banho-seco com agitação a 60° C por aproximadamente 2 h;
6. retirar do banho-seco e deixar voltar a temperatura ambiente;
7. adicionar 3 µl de “RNAse” e colocar no banho-seco a 37° C por 30 min;
8. retirar e deixar voltar a temperatura ambiente;
9. adicionar 200 µl de “Protein Precipitation Solution” e misturar;

10. levar ao gelo por 5 min.
11. centrifugar por 4 min a 13.000 rpm;
12. transferir o sobrenadante para a segunda série de tubo de microcentrífuga contendo 600 µl de Isopropanol e misturar por inversão;
13. centrifugar por 4 min a 13.000 RPM e descartar o sobrenadante;
14. adicionar 600 µl de Etanol 70% a temperatura ambiente e centrifugar por 4 min a 13.000 rpm;
15. descartar o sobrenadante e deixar secar em temperatura ambiente por 1 h;
16. adicionar 50 µl de “DNA Rehydration Solution”.

2.2.4 Validação dos microssatélites por PCR

A validação dos microssatélites foi feita a partir de 30 *loci*, escolhidos conforme o *motif* (di, tri, tetra e pentanucleotídeo) e anotação funcional dos *contigs*. As reações foram realizadas em termociclador, com volume final de 25 µl: 1X de tampão da Taq polimerase; 1.5 µM de MgCl₂; 100 µM de cada dNTP; 0.1 µM de cada primer (F e R); 10-50 ng de DNA genômico e 0.5 U de Taq polimerase (Invitrogen). Os ensaios de PCR foram realizados no equipamento ProFlex™ PCR System (Life Technologies) com 35 ciclos: 95°C/30s, 55 a 60°C/30s (ajustado para cada primer), e 72°C/20s. Os fragmentos de DNA foram aplicados em um gel de agarose 2%, corado com Nancy (Sigma, 5ul a cada 50ml de TBE), visualizados em luz ultravioleta e capturados pela câmera CCD (3.1 megapixels), Locus Biotecnologia L-Pix STi. Os *loci* de microssatélites considerados monomórficos pelo gel de agarose 2%, foram confirmados em gel de poliacrilamida 6%, que foram corados com nitrato de prata e revelados com solução de hidróxido de sódio e formaldeído. Os géis de poliacrilamida foram visualizados em transiluminador de luz branca e capturados pela câmera CCD (3.1 megapixels), Locus

2.2.5 Genotipagem dos microssatélites em sequenciador automático

Após o reconhecimento inicial do polimorfismo de oito *loci* de microssatélites em géis de agarose 2% e poliacrilamida 6%, procedeu-se a genotipagem desses marcadores em sequenciador automático para obter uma melhor acurácia na determinação dos alelos. A estratégia de sequenciamento adotada neste estudo foi a partir dos protocolos de Schuelke (2000), em que um dos *primers* de cada *loci* recebeu na extremidade 5' uma cauda (*tag*) de nucleotídeos conhecidos (5'-CAGTCGGGCGTCATCA-3'), tendo a finalidade de gerar complementaridade com um *primer* adicional (*primer* CAGtag: 5'-CAGTCGGGCGTCATCA-3'). O *primer* CAGtag foi marcado com fluorocromos HEX ou FAM na extremidade 5', e em seguida, utilizado no ensaio de PCR com o par de *primers* padrão para cada *loci*. As reações foram feitas de forma a combinar um *loci* incorporado com a marcação HEX e outro com a marcação FAM, para que ambos fossem genotipados em um mesmo capilar do sequenciador. A PCR para genotipagem no sequenciador foi realizada com os seguintes reagentes: 1X de tampão da Taq polimerase; 1.5 µM de MgCl₂; 100 µM de cada dNTP; 0.1 µM de cada *primer* (F e R); 0.01 µM do *primer* CAGtag; 10-50 ng de DNA genômico e 0.5 U de Taq polimerase (Invitrogen). Os ensaios de PCR foram realizados no equipamento ProFlex™ PCR System (Life Technologies) com 9 ciclos de 95°C/30s, 55 a 60°C/30s (ajustado para cada *loci* de microssatélite), e 72°C/20s; e 30 ciclos de 95°C/30s, 50°C/30s e 72°C/20s. Os produtos de PCR foram analisados em eletroforese capilar no sequenciador automático 3130xl (Applied Biosystem Life Technologies), utilizando a matriz DS-30, com o marcador GeneScan 500 ROX dye Size Standard (Thermo). Foi utilizado o

programa *GeneMapper* 3.7 (Applied Biosystems), para analisar e padronizar os tamanhos de fragmentos de alelos.

2.2.6 Análise estatística para estudos de variabilidade genética

Para determinar os parâmetros estatísticos de diversidade genética em estoques cultivados e natural de pirapitinga (*Piaractus brachipomus*), inicialmente foi utilizado o software *GenAlex Analysis* 6.1 (Peakall e Smouse 2006), para a construção das tabelas matrizes e exportação no formato adequado para os programas estatísticos posteriores. A heterozigosidade observada (H_{obs}) e esperada (H_{esp}), e a análise da variância molecular (AMOVA) (Excoffier et al., 1992), foram calculados a partir do programa *Arlequin* 3.5 (Excoffier et al., 2010). Por meio do programa *GENEPOP* 4.0.11 (Rousset, 2008), foram determinados o coeficiente de endogamia (F_{is}) e conformidade para o Equilíbrio de Hardy-Weinberg (EHW). Os parâmetros determinantes para F_{is} foram estimados de acordo com Weir e Cockerham (1984). Os níveis de significância para os testes de EHW foram ajustados com a correção de Bonferroni ($P < 0,05$) (Rice, 1989). O software *Fstat* 9.3.2 foi usado para calcular o desequilíbrio de ligação. Para avaliação do grau de parentesco, foi calculado o coeficiente r (Wang, 2002) pelo programa SPAGeDI 1.3 (Hardy e Vakemans, 2002), que permitiu estimar os prováveis irmãos completos ($r > 0,375$), meio-irmãos ($0,125 \leq r \leq 0,375$), e não aparentados ($r < 0,125$).

3. RESULTADOS

3.1 Sequenciamento do Transcriptoma

A tecnologia de sequenciamento de nova geração (NGS) permitiu a leitura das bases do transcriptoma de *Piaractus brachipomus* que, após a construção da biblioteca de cDNA obtida pela extração de RNA das amostras de fígado, resultou um total de 192.373 *reads* com tamanho médio de 395,5 pb mediante ao emprego da plataforma Roche/454 (Figura 4).

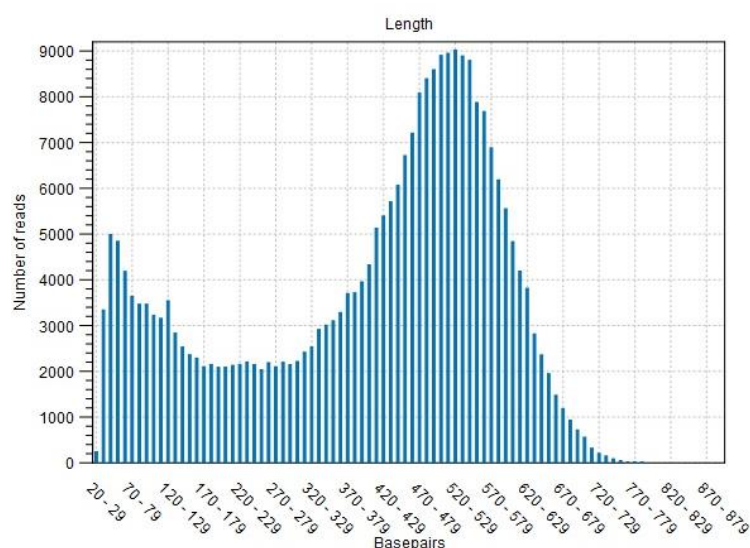


Figura 4 - Frequência de comprimento das sequências (*reads*).

Depois do sequenciamento, foram realizadas as análises de bioinformática utilizando o *software* CLC Genomics Workbench. Inicialmente, as sequências passaram por um processo de *trimming* (exclusão de bases ou sequências com baixa qualidade), cujos parâmetros determinaram eliminar bases ambíguas, descartar sequências menores que 15 pb e maiores que 1.000 pb, (*quality score* $Q < 20$ e nucleotídeos ambíguos ‘N’), correspondendo assim, a permanência de 69.544.759 nucleotídeos distribuídos em 192.077 *reads* (Figura 5). Ainda, o processo de *trimming* acarretou o encurtamento de *reads* (Figura 5), dos quais o tamanho médio final foi de 362,1 pb.

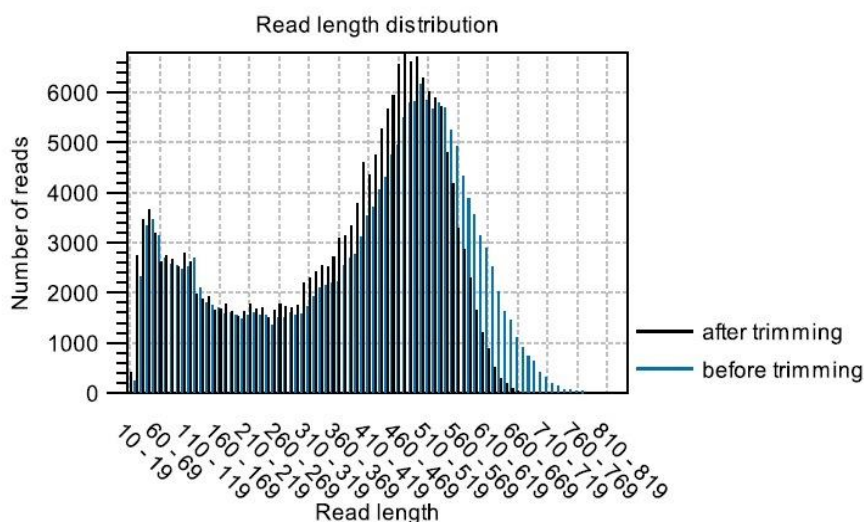


Figura 5 - Comparação entre os tamanhos de *reads* antes e depois da execução do processo de *trimming*.

Com relação à qualidade dos nucleotídeos sequenciados, cerca de 75% tiveram $Q \geq 30$ (Figura 6) e uma grande porcentagem com Q próximo a 40, significando alta acurácia e qualidade do sequenciamento (estimativa de 1 erro a cada 10.000 bases sequenciadas - 99,99% de acurácia), ou seja, quanto maior o valor de Q , maior é a probabilidade de uma base ter sido sequenciada corretamente (Ewing e Green, 1998; Ewing et al., 1998). Apesar disso, é aceitável adotar um limiar de Q menor nas avaliações de transcriptomas de espécies não-modelos, principalmente naquelas em que a quantidade de *reads* produzidos tenha sido pequena (MacManes, 2014).

Tabela 3. Processo de *trimming* dos *reads* do transcriptoma de fígado de pirapitinga (*Piaractus brachypomus*).
pb = pares de base

Dados de sequenciamento	
Número de <i>reads</i>	192.373
Média do comprimento dos <i>reads</i> (pb)	395.5
Número de <i>reads</i> após <i>trim</i>	192.077
Média do comprimento dos <i>reads</i> após <i>trim</i> (pb)	362.1
Total de nucleotídeos	69.544.759

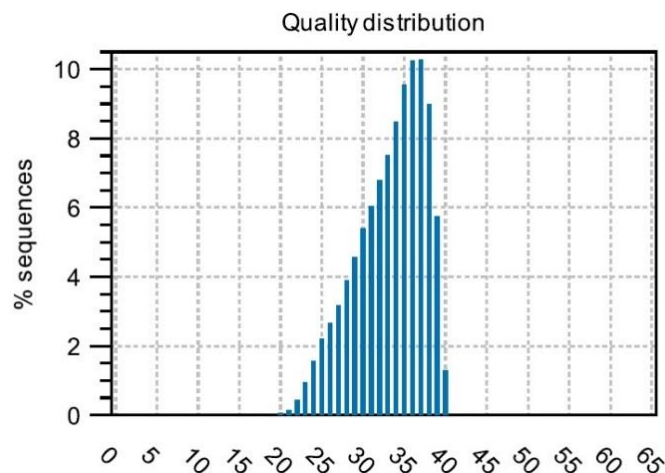


Figura 6 - Distribuição do valor da qualidade (*phred*) das sequências de *reads* após o processo de *trimming*.

A montagem *de novo* dos transcritos resultou na formação de 3.797 *contigs* como consequência da sobreposição de 174.272 *reads* (63.460.229 pb). O comprimento médio, mínimo e máximo dos *contigs* foi: 790 pb, 202 pb e 7.812 pb, respectivamente, totalizando 2.999.680 nucleotídeos de cobertura. Na Figura 7 observa-se que a maioria dos *contigs*, aproximadamente 3.000, englobou entre 1 e 1,9 mil pares de bases (kb) em suas sequências. Também, ao aferir o tamanho dos *contigs* em proporção, foram obtidos os valores de N25 = 1.383 pb, N50 = 861 pb, e N75 = 593 pb. Ao total, 17.805 *reads* (6.084.530 pb) remanescentes foram considerados como *singletons* e, por isso, não foram utilizados para análises de prospecção de marcadores moleculares.

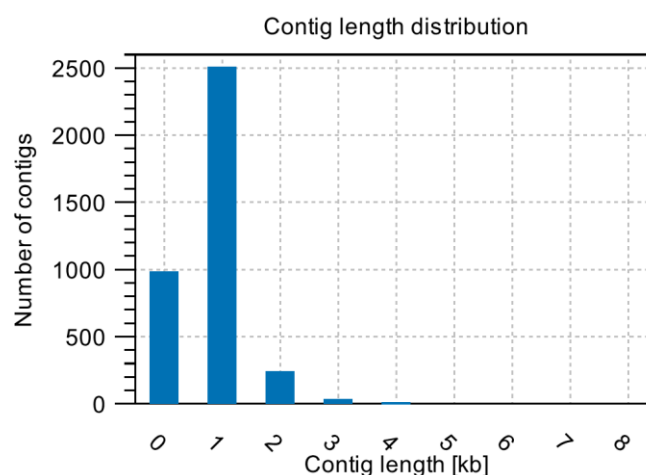


Figura 7 - Organização de *contigs* conforme a quantidade de nucleotídeos nas sequências.

Tabela 4. Dados da montagem *de novo* do transcriptoma de pirapitinga (*Piaractus brachypomus*).

Reads utilizados na montagem	174.272
Total de nucleotídeos na montagem	63.460.229
Número de contigs	3.797
Total de nucleotídeos dos contigs	2.999.680
Comprimento mínimo dos contigs (pb)	202
Comprimento máximo dos contigs (pb)	7.812
Comprimento médio dos contigs (pb)	790
N75 (pb)	593
N50 (pb)	861
N25 (pb)	1.383

Após a montagem dos transcritos, os contigs foram sujeitos ao programa cd-hit-est com um limiar de 90% de identidade para remover redundância (Duan et al., 2012; Li et al., 2012). Além disso, a fim de eliminar contaminação de DNA mitocondrial e RNA ribossômico, as sequências foram comparadas contra o genoma mitocondrial do pacu e o RNA ribossomal de zebrafish (ambos do banco de dados do NCBI), usando o CLC Genomic Workbench (versão 8.0.3). Após esses processos, obteve-se 3.696 contigs para serem utilizados nas análises de anotação funcional e prospecção de marcadores.

Para avaliar a diversidade funcional do transcriptoma de fígado estudado, os 3.696 *contigs* (sequências consenso) da pirapitinga (*Piaractus brachypomus*) foram comparados com o banco de dados de proteínas não-redundantes (*nr*) do NCBI, e especificamente com o banco de proteínas das espécies modelo zebrafish (*Danio rerio*) e fuga (*Takifugu rubripes*), por meio da ferramenta BlastX (*e-value* de $1e^{-3}$) implementado no Blast2GO. Um total de 2.568 transcritos (69,4%) apresentou similaridade significativa no banco de dados *nr* de proteínas. Além disso, ao total, 67,6% (2.498 contigs) das sequências de pirapitinga tiveram similaridade em relação ao Refseq de *zebrafish*, enquanto que 65,4% das sequências (2.419 contigs) exibiram similaridade em relação ao *fugu*, ou seja, praticamente a mesma proporção de genes anotados. Contudo, observou-se um maior número de *hits* quando as sequências

foram confrontadas com o Refseq de *zebrafish*, que pode ser explicado pela maior proximidade filogenética com esta espécie.

Os 2.568 genes com correspondência no banco de proteínas *nr* foram caracterizados nas categorias de Gene Ontology (GO), sendo que 2.075 sequências (80,8%) foram anotados em termos GO. Considerando as categorias de GO mais frequentes, foram encontradas 1.831 atribuições a Processos Biológicos (88,2%), seguido por 1.757 de Função Molecular (84,6%) e 1.378 de Componente Celular (66,4%). Nas subcategorias de GO, os termos mais abundantes foram referentes à *metabolic process*, *cellular process* e *single-organism process* para a categoria de Processo Biológico; *binding*, *catalytic activity* e *transporter activity* para Função Molecular; *cell*, *organelle* e *membrane* para Componente Celular. No presente estudo, destaca-se que foram encontrados genes atribuídos a processos do sistema imune, crescimento e reprodução, o que será de subsídio para futuros estudos que possibilitarão um melhor conhecimento sobre as características importantes na aquicultura da pirapitinga.

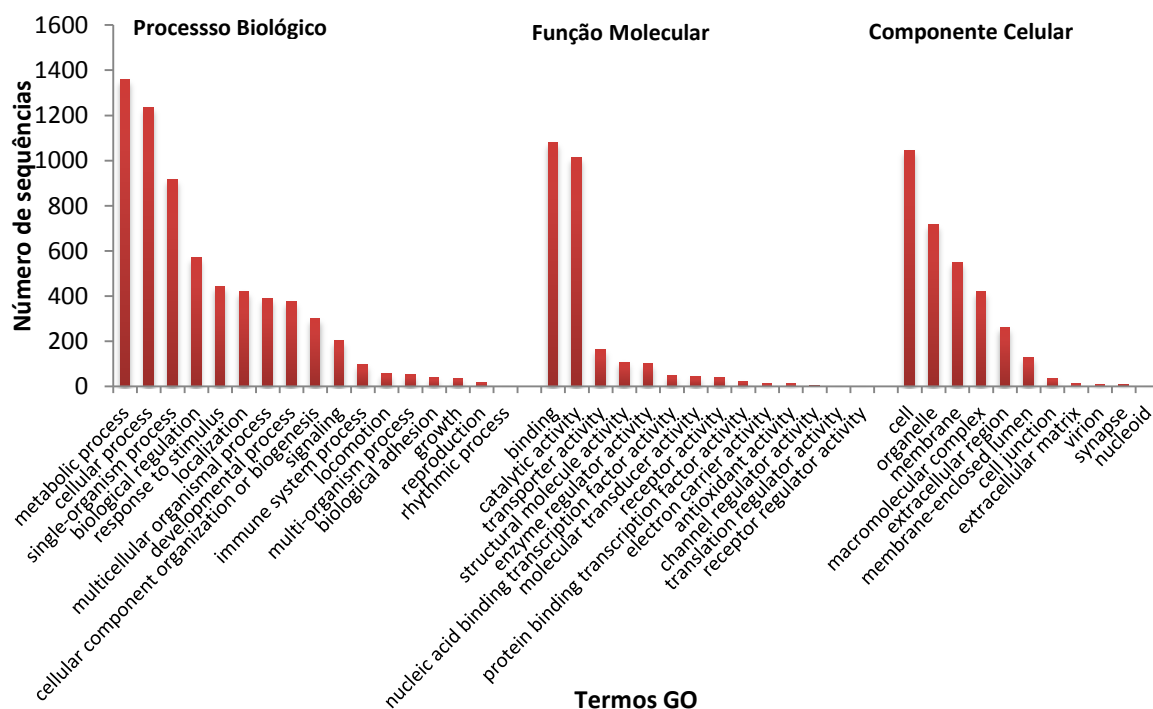


Figura 8 - Resultados da anotação funcional e a atribuição dos genes nas categorias e subcategorias de GO.

Após os resultados de mapeamento de *Gene Ontology*, os transcritos foram caracterizados no banco de dados de vias metabólicas e enzimas KEGG (*Kyoto Encyclopedia of Genes and Genomes*). Ao total, 1.122 transcritos foram identificados em 106 vias metabólicas. Destas, destacam-se os contigs envolvidos na biossíntese de antibióticos, metabolismo de purina e glicólise/gluconeogênese (Figura 9).

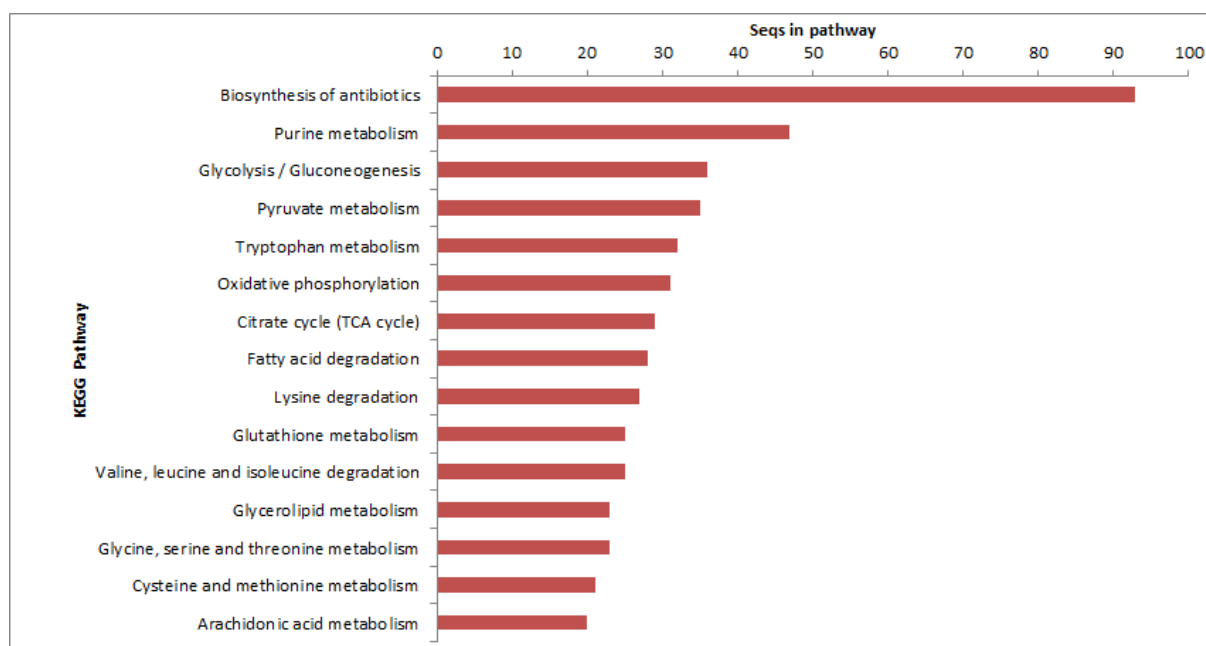


Figura 9 - Resultados da anotação funcional e a atribuição dos genes nas categorias e subcategorias de GO.

3.2 Descoberta de microssatélites e anotação funcional

Para o processo de descoberta de microssatélites, as análises por sequências repetitivas curtas (SSR) foram realizadas em 3.696 *contigs*. Por meio do software Msatcommander, foram detectados 130 marcadores microssatélites distribuídos em 95 *contigs*. No total, 75 pares de *primers* foram desenhados adjacentes aos *loci* de microssatélites, que incluiu sequências dos tipos 56 di, 13 tri, 4 tetra e 2 pentanucleotídeos (Tabela 5). Entre os dinucleotídeos, 42,85% tiveram o *motif* GT/AC; 32,15% foram do tipo AG/CT; e 25% do tipo AT/TA. Em relação aos trinucleotídeos, não tivemos um padrão de maior ocorrência de um

determinado *motif*, sendo que foram observados oito tipos (GAT, AGC, ATT, AAT, AAG, GCT, ACG e CGG). A mesma situação também foi encontrada para sequências de tetranucleotídeos (ATCT, AAAG, AATG e AAAT) e pentanucleotídeos (ACTAT e ATAGT), com a presença de quatro e dois tipos de *motifs*, respectivamente.

Tabela 5 - Microsatélites de genes associados (di, tri, tetra e penta nucleotídeos) obtidos a partir dos resultados de transcriptoma de fígado.

Contig	Descrição das sequências	Primers	Posição	Anelamento
C4	tetraspanin-8-like	F: TCTGAATAGGGTTTCGGTTCAGT R: TGACCGCTGCACTTGAGTAG	1108,(AT) ⁵ ,1118	60°C
C6	sem identificação gênica	F: ACACTCTGCTCCAGATGCAC R: CGGTATGCCATTCGACCAGA	172,(CT) ⁶ ,184	60°C
C13	dihydroxyvitamin d24-mitochondrial-like	F: TCTCTTCAAGCCTCCTCTGC R:ATGCTGCAGCTCCTCCTGT	1880,(AGC) ⁶ ,1898	60°C
C18	scavenger receptor cysteine-rich type 1 protein 130-like	F: GTGACCCCTGAACACCTCTG R: AGGTCCTCTGCTGGAGTGAT	1392,(AG) ⁵ ,1402	60°C
C23	s-adenosylmethionine synthase isoform type - 1	F: GGAAACCGAGACCAGATTGA R: TCCAAAGTGTCCAATGCAAA	1679,(AT) ¹³ ,1705	60°C
C25	cytosolic 5-nucleotidase 3a-like isoform x 1	F: CTTTGTCTGCTTTGGGTCGT R: CTTAGAAGAATGTGCAAAATTGAAA	1345,(AT) ¹¹ ,1367	60°C
C34	fibronectin isoform x1	F: GCAGGTCCACTCTCCTTTCC R:ATCAGCTGCACCATAGCGAA	7533,(AC) ⁵ ,7543	60°C
C38	tryptophan-dioxygenase a-like	F: GCTCCATTAGTGAACCTCCCTCT R: CCTTCCTCTACACGGCTGAG	318,(CT) ⁷ ,332	60°C
C43	se-cephalotoxin-like isoform x1	F: TGGTGCTGGTGTAACACAAA R: CGGAGTGTAAGGTAACAGCCA	1124,(AT) ⁵ ,1134	60°C
C56	solute carrier family 13 member 5-like	F: GGCTCATCTGTCTGTGAGG R: ATGTGTAAGGGCTTGACGCT	1447,(AC) ⁵ ,1457	60°C
C64	sodium-coupled neutral amino acid transporter	F: CAAAGCAAACCTCAAAAAGGAAAA R: TGGGAACGTTTAGCATCTCA	880,(AAAG) ⁷ ,908	55°C
C72	insulin-like growth factor-binding protein 2-B-like isoform X1	F: AGAGGCCTTGCTCCTGTCT R: GCAGAGGTCTATCAGCTGCC	1710,(GT) ⁶ ,1722	60°C
C73	claudin-19-like	F:GTCTGCTCTGGGCCTCAAAT R: GTACCAGGAAACAGCGGTCA	442,(GT) ⁵ ,452	60°C
C82	complement c1r subcomponent-like	F: CTGAACCTGCCCATACAGCA R: GGTCACCACCACAGTGCTC	2397,(AC) ⁵ ,2407	55°C
C107	alcohol dehydrogenase class -3-like	F: CATCAGGTCAAAGGCCTGGT R: GGAAGGGTGTGGACTCAGTG	1250,(GT) ⁵ ,1260	60°C
C119	adp-ribosylation factor 4	F: CGTCCCTCAGCTCATCTTCC R: GCCAGGACAAGATTCTGTCCA	1532,(CT) ⁵ ,1542	60°C

C140	glycogen liver from-like isoform 3	F: GGAGAAAGCCCGCCTGAATA R: ATAGGGCATGGCTAGGACCA	1642,(AG) ⁵ ,1652	60°C
C195	protein disulfide-isomerase a-like	F: CTCGCTTAGTCGGTGTGAGG R: TCTCCTCTTTGACCTCAGTGC	331,(ACG) ⁵ ,346	60°C
C211	syndecan-2-like	F: TCAATGAGTTCACATTACCTTGC R: CCAGCTCCCCAGATATTACAG	890,(AG) ⁹ ,908	60°C
C289	sem identificação gênica	F: TGGTTTTTCAGGTTTTCTTTTG R: TTTAAGGCAATGCATAATTTCG	26,(AG) ⁹ ,44	60°C
C369	si:ch73-protein	F: ACCACCGTGATCTGGACATG R: GCAGAGTTGAAGTGGGACGA	1131,(AT) ⁵ ,1141	60°C
C380	toxin - 1 precursor	F: CCCATAGCACTGGATTTTG R: CTGCTGAAACGTCAAGGACA	732,(AC) ¹⁰ ,752	60°C
C386	transferrin	F: TCACGTCCGCATTTAACCCA R: GAGCAAGACACCTAACCCCC	2870,(AC) ⁵ ,2880	60°C
C410	apolipoprotein e	F: CGCACAGGTCTAAAGGCACT R: CTCCCACACAGTGAAAAGCA	928,(AG) ⁸ ,944	60°C
C491	protein Tob1-like	F: TTGAACATCGCAGTGGAGCT R: TCTCTCTATGGCCTTGGCCT	211,(GCT) ⁶ ,229	60°C
C494	delta-like protein b-like	F: TCCGGTACAAAAGAATATGCAG R: CACAGCTCTTCATCACGGTCT	1318,(ATT) ⁶ ,1336	60°C
C524	eukaryotic translation initiation factor 2 subunit 2	F: AAGAAGGAAAGAGCCGTCGG R: CCCTTCATCCAGCTCGTTGT	324,(AAG) ⁵ ,339	60°C
C532	titin isoform x7	F: TGCAGCAACATCTCTGTCTT R: ATGAGCTGCACAGAGAAGCT	387,(CT) ⁵ ,397	60°C
C718	transposable element tcb 1 transposase	F: CCCTGTGACCCAGAAGGATA R: GGGATATTGCGCAAAGTGTA	716,(GT) ¹⁴ ,744	60°C
C739	claudin-19-like	F: CGCACACACTCTCACACTGA R: CCCAGTACCCAGCTCAAACCT	69,(GT) ⁵ ,79	60°C
C781	se-cephalotoxin-like isoform x1	F: TGGTGCTGGTGACGACAAA R: CGGAGTGTAAGTAACAGCCA	1026,(AT) ⁵ ,1036	60°C
C821	plasma protease c1 inhibitor-like	F: AACTGGAGCTGCATGTCTG R: CGTCAACCCAGAGAAAAGAGG	217,(CT) ⁷ ,231	60°C
C827	granulins	F: GCGCACAGCTGACTAAATGA R: GCTGCTCTTGATCTTCCTG	137,(AT) ⁷ ,151	60°C
C838	ribosomal protein I4	F: TGTTTGAACCATTTGCCAGA R: GGGTTAGGAAGGCCAAGAAG	283,(GCT) ⁵ ,298	60°C
C862	60s ribosomal protein I13a	F: ACACCGATGTCTGAAGCAGT R: GAAGCTCCATACTGGCTGCA	762,(AAT) ⁶ ,780	55°C
C877	sem identificação gênica	F: CGAACATCAGCTTGACTTCTTT R: GTGAATTCAATGCAGGCAAA	1051,(ATAGT) ⁵ ,1076	60°C
C902	protein disulfide-isomerase-like	F: ATGAGTTATAAGGGCGGGCG R: CAGCTGTGGGTTGACTAGG	755,(AG) ⁵ ,765	60°C
C937	complement factor i	F: AAGGTACGGGCACAAACGTA R: CGGTATCTTCGCTTTCAGGT	1171,(GT) ⁵ ,1181	60°C
C951	60s ribosomal protein I35	F: CGCAGATCTCTGGCCTTGAT	415,(GT) ⁵ ,425	60°C

		R: ACACACAGCAAGCCCAAGT		
C964	chymotrypsin-like elastase family member 2a-like	F: GCCATGAGCAAATTAGGGAAAAGA R: TGGATCAACAATGTGATCGCC	184,(AT) ⁵ ,194	60°C
C1003	sem identificação gênica	F: TGGCCTGAGTGTCTTCACTG R: AACGGATGGGAAACAAGAGA	401,(GT) ⁹ ,419	60°C
C1005	apolipoprotein e	F: AGTTGTTGCACCAAATGCAG R: CTTGTTCCCTCCCACACAGT	471,(AG) ⁷ ,485	60°C
C1076	sem identificação gênica	F: ACTGACCCACATCCCGTAAA R: CCTTCCTGCGTGTATGTGTG	91,(AC) ¹⁰ ,111	60°C
C1147	granulin 1	F: TGAAGTCGCAGGAACATAATGC R: GCTCTGACTGTTCTGCTGCT	140,(AT) ⁶ ,152	60°C
C1155	c-type lectin domain family 4 member e	F: GTGGTGTAGTCGGAGCAGAG R: GTGGTGTAGTCGGAGCAGAG	690,(AG) ⁶ ,702	60°C
C1208	complement factor i-like	F: AAGGTACGGGCACAAACGTA R: GAATCTGTGGCGGTATCTTCG	826,(GT) ⁵ ,836	60°C
C1341	ribosomal protein I4	F: TGTTTGAAACCATTGCCAGA R: GGGTTAGGAAGGCCAAGAAG	81,(GCT) ⁵ ,96	60°C
C1376	fructose-biphosphatase 1-like	F: GTGTTACATGGCAGGCGTTT R: CAAGTGAGACCAAATCCAAGG	69,(AC) ¹⁰ ,89	60°C
C1407	vitelline membrane outer layer protein 1 homolog	F: GCACATCCTGCTGCATTCA R: ACCGCTCTCAACGACGTAA	61,(AG) ⁵ ,71	60°C
C1486	complement component c8 alpha chain-like	F: GTGAATTCAATGCAGGCAAA R: CGAACATCAGCTTGACTTCTTTT	1270,(ACTAT) ⁵ ,1295	60°C
C1709	cyclin-i isoform x 2	F: ATGGCGCGTTTCACACTAAT R: CATGGCTCGACCATTGTTTA	167,(AC) ⁸ ,183	60°C
C1716	vacuolar atp synthase 16 kda proteolipid subunit	F: AACCGAAGAGAGGGGAGTGT R: GCATTTACAAGGGGACGCAC	779,(GT) ⁷ ,793	60°C
C1718	sem identificação gênica	F: GCGCATAGGAAATGAGCATA R: TAAAAGCGTGCACCAGAAGA	901,(AATG) ⁶ ,925	60°C
C1728	sem identificação gênica	F: AGTGCTAGTCTTGAAACGGATGT R: ACTGCAGATTGAGAGCGAACA	143,(AT) ⁵ ,153	60°C
C1731	solute carrier organic anion transporter family	F: AGAATCAAGCCAAGGACACG R: CACATGTGAATCCAGGCAGT	203,(CGG) ⁵ ,218	60°C
C1802	diacylglycerol o-acyltransferase 2-like	F:GTACGCAACAGAAGAGCCCT R: AGCTCTCCGCATGGTGAAAT	111,(AT) ⁵ ,121	60°C
C1827	sem identificação gênica	F: TTGCTGAGAAATGTGCACTG R:CCCGAACCTGCAATCATAAC	349,(AAAT) ⁵ ,369	60°C
C1830	complement c1r subcomponent-like	F: AGGGGATAATCAGGGGAGTG R: TTGAAACGTTTGGTTTGGTG	1206,(AGC) ⁵ ,1221	60°C
C1832	tetraspanin-3 isoform x1	F: GGTGCTATGTCGTAGAGGCC R: AGGAAGGCATGACCAGTGTG	842,(AC) ⁶ ,854	60°C
C1872	sem identificação gênica	F: AATGCTTGGGCACAGAAAAG R: GAGGCCGTTGTGTGTATGTG	147,(GAT) ⁸ ,171	50°C
C1923	T-cell leukemia homeobox protein 1-like isoform X2	F: TTGAGGGCGTAGGGTTTGAA R: TGTCCAGTGTGCTGTTTCGT	974,(GT) ⁶ ,986	60°C

C1940	calnexin precursor	F: AGCTGAAGAGCCAGAGGAGA R: TCGGATTTATCTTCCGCTTC	634,(AAG) ⁶ ,652	55°C
C1945	junctional adhesion molecule a-like	F: TCTCCTGCTGTGCTGATGTC R: CAGCATTAGGGTCAGTGCAA	1208,(AT) ⁹ ,1226	60°C
C2043	dolichyl-phosphate beta-glucosyltransferase-like	F: GCAGTAAGAGGAGCAGAGCC R: GAGCTGTGAGTGACTGTCCG	460,(CT) ⁵ ,470	60°C
C2063	sem identificação gênica	F: ACAACATCAGGCTAGTGCTT R: GCATGGCTTTGTTGCACACA	612,(CT) ⁶ ,624	60°C
C2157	excitatory amino acid transporter 1 isoform x 1	F: TGCTGGTACTCTGTGTGTCTG R: CAGACGGACGGACAGACG	559,(GT) ⁵ ,569	60°C
C2193	s-adenosylmethionine synthase isoform type - 1	F: TCCAAAGTGCCAATGCAAA R: GGAAACCGAGACCAGATTGA	148,(AT) ¹³ ,174	60°C
C2447	signal transducer and activator of transcription 3	F: TTCTGGCTCTGTGGAACGAG R: CCCAGGTGAATGTGATGCCT	200,(AG) ⁵ ,210	60°C
C2503	adp-ribosylation factor 1	F: CGTGCGTGTGAGATGAGAGT R: GGACGACCAGAGACAGGAGA	393,(CT) ⁷ ,407	60°C
C2687	sem identificação gênica	F: GACTGCTGTTACAAGTGCACA R: CTGAGCACCTGACTCACTCG	320,(GT) ⁵ ,330	60°C
C2710	mitoferrin-2	F: ATAGAAGTAGCCGCTCGCAC R: GCAGTGAGCCGGAGAATGTA	318,(CT) ⁶ ,330	60°C
C5348	shc-transforming protein 2-like isoform x 1	F: GCTGCTACAATTGTCCATTATCTC R: TCTTACATCAAAACAGACAAAATACA	1085,(AC) ⁵ ,1095	60°C
C5358	shc-transforming protein 2-like isoform x 1	F: CTCGGCCATTGGTCCTCAAG R: AGACCTGAGAGCCCCAAGAA	982,(ATCT) ⁸ ,1014	60°C
C5441	solute carrier family 25 member 33-like	F: AGAAATGGGATGCAGGTTTG R: GGATGCAGAATAAACTTAAGCAA	1190,(AC) ¹⁰ ,1210	60°C

Em relação à anotação funcional destas sequências, 77,3% dos contigs foram identificados pela busca de similaridade com a ferramenta BlastX, a partir do banco de dados de proteínas não-redundante (*nr*) do NCBI, dos quais 94,8% foram funcionalmente anotados nas categorias de ontologia gênica: Processos Biológicos (52%), Componente Celular (19,6%) e Função Molecular (28,4%). Na Figura 10, destacamos as três principais subcategorias de ontologia gênica para cada classe funcional.

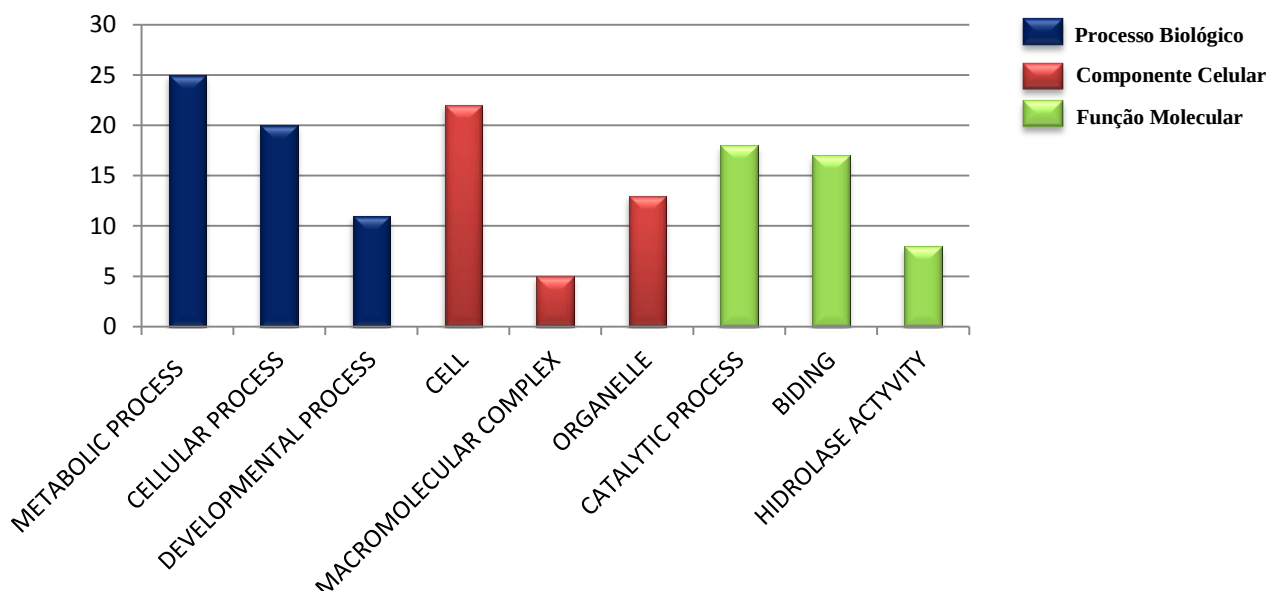


Figura 10 - Principais subcategorias de GO e suas porcentagens que representam entre as classes.

Para a fase inicial de validação dos marcadores microssatélites, dentre os 75 *loci* que podem ser isolados por meio de PCR, foram selecionados 30 marcadores de acordo com a anotação funcional, número de repetições e tipo de *motif*. Ou seja, foram priorizados microssatélites localizados em regiões não traduzidas (UTR), pois se utilizou o princípio de que microssatélites em sequências codificantes (cds) tenham maior probabilidade de serem monomórficos (Tabela 6). Do mesmo modo, a escolha dos marcadores foi fundamentada em sequências repetitivas com maior número de cópias, pois foi adotado o princípio de que uma maior quantidade de repetições apresente maior chance de ser polimórfico. Além disso, também se priorizou *motifs* dos tipos tri, tetra e pentanucleotídeos para evitar erros de genotipagem, como por exemplo, a ocorrência de artefatos de PCR (*stutter bands*) que ocorrem com maior frequência para sequências dinucleotídeos (Schlötterer, 2004; Sekar et al., 2009).

Tabela 6 - Anotação funcional das sequências utilizadas para validação.

Contig	Identidade gênica (Blastx)	Subcategorias de GO (F: função molecular, P: processo biológico, C: componente celular)	Posição gênica
C25	cytosolic 5-nucleotidase 3a-like isoform x 1	F: hydrolase activity; P: metabolic process; F: binding; F: transferase activity; P: nucleobase-containing compound metabolic process; C: endoplasmic reticulum; C: mitochondrion; F: nucleotide binding	3'UTR
C1832	sem identificação gênica	-	-
C2193	s-adenosylmethionine synthase isoform type-1	F: transferase activity; F: binding; P: biosynthetic process; P: nucleobase-containing compound metabolic process; P: metabolic process; F: nucleotide binding	5'UTR
C5441	solute carrier family 25 member 33-like	C: mitochondrion; P: transport; C :cell	3'UTR
C1376	fructose-biphosphatase 1-like	P: carbohydrate metabolic process; F: hydrolase activity; P: metabolic process; P: response to endogenous stimulus	5'UTR
C410	apolipoprotein e	P: transport; C: extracellular region; P: metabolic process; F: lipid binding	3'UTR
C1709	cyclin-i isoform x 2	F: enzyme regulator activity; P: regulation of biological process; P: cell cycle	5'UTR
C1716	vacuolar atp synthase 16 kda proteolipid subunit	F: hydrolase activity; F: transporter activity; P: ion transport; C: cell; C: cytoplasmic membrane-bounded vesicle; C: protein complex; F: ion channel activity; C: cellular component	3'UTR
C2503	adp-ribosylation factor 1	C: Golgi apparatus; F: nucleotide binding; P: protein transport; P: nucleobase-containing compound metabolic process; P: catabolic process; C: cytosol; P: transport; P: signal transduction; P: multicellular organismal development; P: biosynthetic process; P: lipid metabolic process; C: cytoplasm; F: hydrolase activity	3'UTR
C38	tryptophan-dioxygenase a-like	P: metabolic process; P: catabolic process; F: binding; F: oxygen binding; F: catalytic activity	5'UTR
C1005	apolipoprotein e	P: transport; C: extracellular region; P: metabolic process; F: lipid binding	3'UTR
C821	plasma protease c1 inhibitor-like	P: negative regulation of endopeptidase activity; C: extracellular space; F: serine-type endopeptidase inhibitor activity	5'UTR
C827	granulins	P: regulation of biological process	5'UTR
C23	s-adenosylmethionine synthase isoform type-1-like	F: transferase activity; F: binding; P: biosynthetic process; P: nucleobase-containing compound metabolic process; P: metabolic process; F: nucleotide binding	3'UTR
C6	sem identificação gênica	-	-
C82	complement c1r subcomponent-like	F: peptidase activity; P: protein metabolic process; P: biological process	3'UTR
C369	si:ch73-protein	P: negative regulation of endopeptidase activity; C: extracellular space; F: cysteine-type endopeptidase inhibitor activity	3'UTR
C902	protein disulfide-isomerase-like	P: metabolic process; F: catalytic activity; C: cytoplasm; P: cellular homeostasis; P: regulation of biological process; F: electron carrier activity; C: cell; P: cellular protein modification process; C: endoplasmic reticulum	3'UTR
C13	dihydroxyvitamin d24-like	F: binding; F: catalytic activity; F: electron carrier activity; P: metabolic process	3'UTR

C494	delta-like protein b-like	P: anatomical structure morphogenesis; P: multicellular organismal development; P: embryo development; F: calcium ion binding; C: cell; P: cell differentiation; P: regulation of biological process; F: protein binding; P: signal transduction; C: cytoplasm; C: plasma membrane	3'UTR
C862	60s ribosomal protein I13a	F: structural molecule activity; C: ribosome; P: translation	3'UTR
C1940	calnexin precursor	P: protein metabolic process; C: endoplasmic reticulum; F: protein binding; F: calcium ion binding	3'UTR
C1830	complement c1r subcomponent-like	P: proteolyses; F: serine-type endopeptidase activity; F: calcium ion binding; F: hydrolase activity; F: catalytic activity; F: peptidase activity; P: complement activation; C: extracellular region; C: membrane; P: antigen processing and presentation of endogenous peptide antigen via MHC class I	3'UTR
C195	protein disulfide-isomerase a4-like	P: protein metabolic process; C: endoplasmic reticulum; F: electron carrier activity; P: cellular homeostasis; P: regulation of biological process; P: metabolic process; F: catalytic activity	5'UTR
C524	eukaryotic translation initiation factor 2 subunit 2	F: translation factor activity, nucleic acid binding; P: translation; C: cytosol; C: protein complex; C: cytoplasm; F: binding; P: reproduction; P: multicellular organismal development; F: protein binding; P: embryo development	3'UTR
C1731	solute carrier organic anion transporter family	C: membrane; P: transport; F: transporter activity	3'UTR
C1341	ribosomal protein I4	C: nucleolus; F: structural molecule activity; C: ribosome; C: cytosol; P: translation	5'UTR
C1872	sem identificação gênica	-	-
C64	sodium-coupled neutral amino acid transporter	C: cell	5'UTR
C1486	complement component c8 alpha chain-like	C: membrana attack complex; P: immune response; F: molecular function	3'UTR

3.3 Validação dos marcadores microsatélites

Para a etapa de validação dos marcadores microsatélites, as análises foram inicialmente realizadas nas 22 amostras de pirapitinga *Piaractus brachipomus* coletadas do ambiente natural (Rio Tocantins, Lajeado, TO). Na Figura 11, é possível encontrar um resumo dos dados de bioinformática, demonstrando a quantidade de sequências repetitivas que foram utilizadas para esta etapa do projeto.

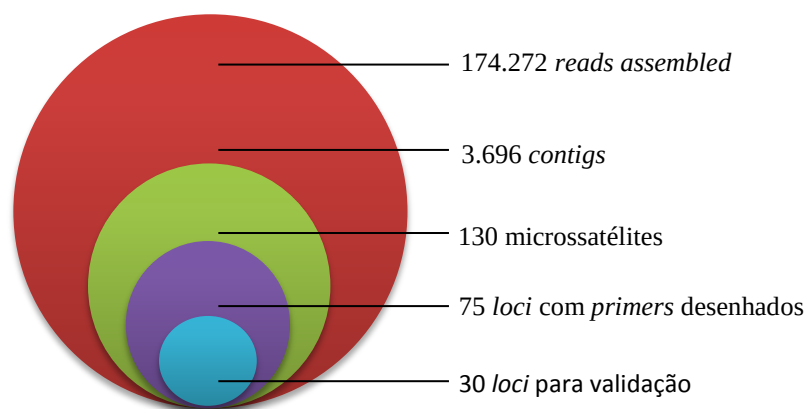


Figura 11 – Volume de dados para descoberta e validação dos microssatélites em pirapitinga.

Após o processo de extração e purificação de DNA genômico das amostras (Figura 12), foram realizados ensaios de padronização da PCR com o intuito de verificar se há ocorrência de amplificação e polimorfismo (detectado por fragmentos de diferentes tamanhos).

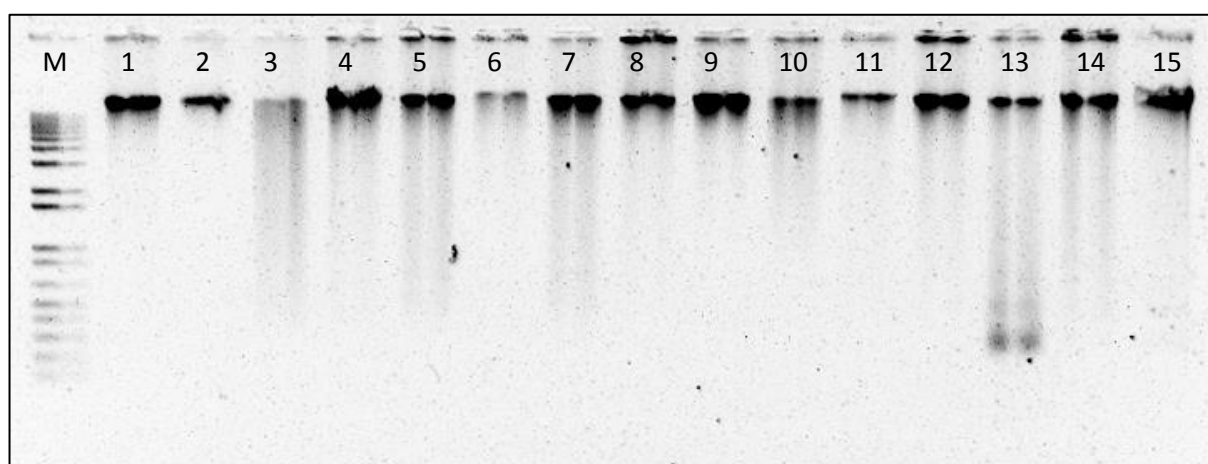


Figura 12 - Gel de agarose: Extração e purificação do DNA genômico de exemplares de pirapitinga coletados em ambiente natural (1-15). M, peso molecular Invitrogen 1Kb Plus DNA Ladder.

Do total de 30 microssatélites avaliados, em 2 loci (C23 e C2193) não foram possíveis de obter fragmentos de amplificação (Tabela 6 e Figura 13), mesmo com diferentes ensaios de otimização, como alteração de temperatura de anelamento e concentração de reagentes (Taq polimerase e $MgCl_2$). Diante destas informações, 93,3% (28) dos loci de microssatélites avaliados deste conjunto de dados resultaram positivamente em produtos amplificados.

Entretanto, as análises de amplificação de 20 *loci* não resultaram em informações úteis que permitam aplicá-los para estudos de variabilidade genética (Tabela 7 e Figuras 13 e 14). Ou seja, 12 marcadores microsatélites foram considerados monomórficos por não apresentar bandas de diferentes tamanhos entre os indivíduos analisados (genótipo único para a população inteira), mesmo quando analisados em géis de poliacrilamida. Além disso, 4 *loci* apresentaram fragmentos amplificados com tamanhos diferentes do esperado, ou seja, resultaram em bandas maiores que 700 pb quando analisados em géis de agarose (tamanho esperado era de no máximo 300 pb, com base no desenho dos *primers*), que inviabiliza a genotipagem para detectar polimorfismos. Adicionalmente, quando amplificados por PCR, outros 4 marcadores demonstraram bandas múltiplas e de mesmo padrão para todos os indivíduos, mesmo após ensaios de otimização, o que também inviabiliza utilizá-los como microsatélites para estudos de diversidade genética.

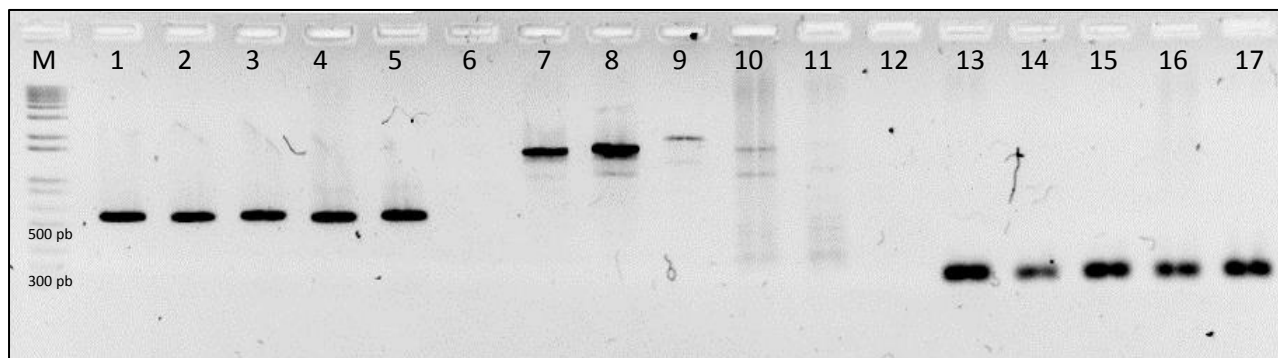


Figura 13 - Gel de agarose: PCR de *loci* de microsatélites de pirapitinga coletado no ambiente natural. 1-5: *locus* C1830 (fragmento amplificado com tamanho acima do esperado); 7-11: *locus* C827 (bandas múltiplas); 13 - 17: *locus* C524 (monomórfico); 6, 12: vazio; M: peso molecular Invitrogen 1Kb Plus DNA Ladder.

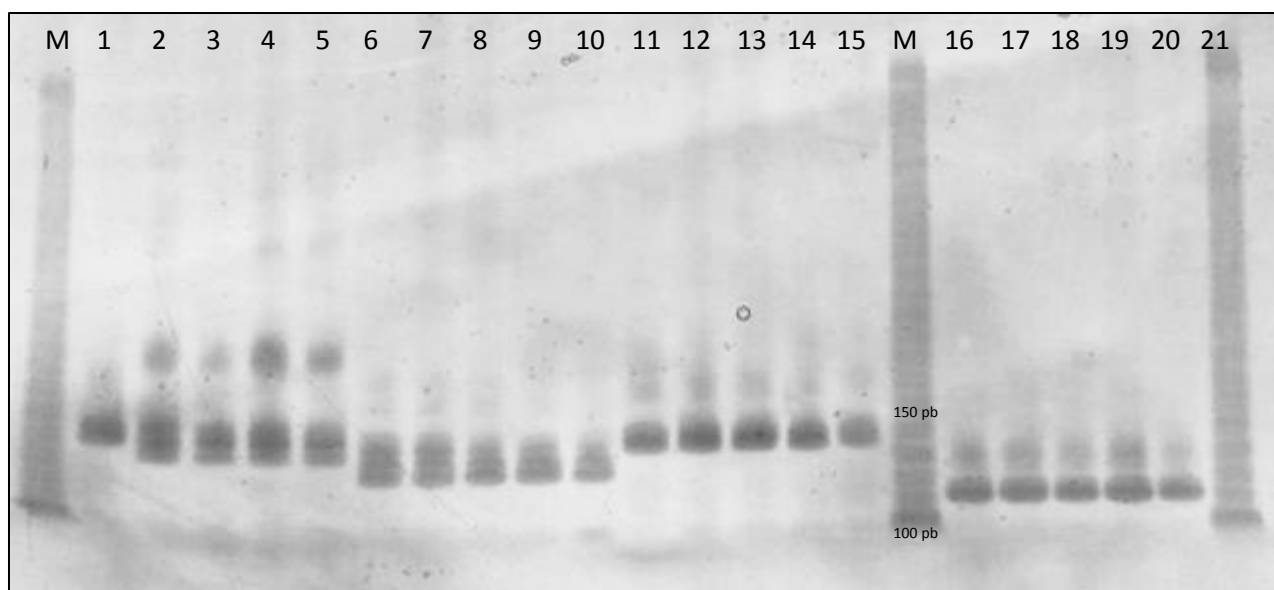


Figura 14 - Gel de poliacrilamida: PCR de *loci* de microssatélites de pirapitinga coletado no ambiente natural. 1-5: *locus* C13 (microssatélite polimórfico); 6-10: *locus* C902 (monomórfico); 11-15: *locus* C1709 (monomórfico); 16-20: *locus* C494 (monomórfico); M: peso molecular Invitrogen 10 bp DNA Ladder.

Por fim, 8 *loci* de microssatélites demonstraram polimorfismo por apresentarem bandas de diferentes tamanhos (com pequenas diferenças) em géis de agarose (Figura 15), o que significa a presença de distintos alelos para estas regiões repetitivas. Estes padrões de polimorfismo foram confirmados também em géis de poliacrilamida. O tamanho dos alelos nos diferentes *loci* variou de 100 a 250 pb.

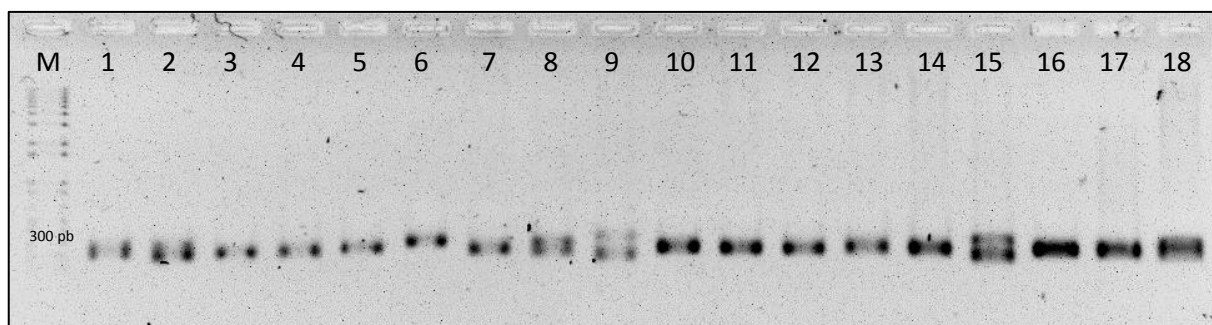


Figura 15 - Gel de agarose: PCR de *locus* de microssatélites de pirapitinga coletado no ambiente natural. 1-18: *locus* C1716 (polimórfico); M: peso molecular Invitrogen 1Kb Plus DNA Ladder.

Tabela 7 - Resultados da validação dos microssatélites.

Contig	Sequência (Blastx)	Motif	Tamanho do produto de PCR (pb)	Caracterização
C25	cytosolic 5-nucleotidase 3a-like isoform x 1	(AT) ¹¹	200	polimórfico
C1832	tetraspanin-3 isoform x1	(AC) ⁶	150	polimórfico
C2193	s-adenosylmethionine synthase isoform type - 1	(AT) ¹³	-	-
C5441	solute carrier family 25 member 33-like	(AC) ¹⁰	100	monomórfico
C1376	fructose-biphosphatase 1-like	(AC) ¹⁰	250	polimórfico
C410	apolipoprotein e	(AG) ⁸	200	polimórfico
C1709	cyclin-i isoform x 2	(AC) ⁸	150	monomórfico
C1716	vacuolar atp synthase 16 kda proteolipid subunit	(GT) ⁷	200	polimórfico
C2503	adp-ribosylation factor 1	(CT) ⁷	200	monomórfico
C38	tryptophan-dioxygenase a-like	(CT) ⁷	200	monomórfico
C1005	apolipoprotein e	(AG) ⁷	200	polimórfico
C821	plasma protease c1 inhibitor-like	(CT) ⁷	800	-
C827	granulins	(AT) ⁷	bandas múltiplas	-
C23	s-adenosylmethionine synthase isoform type-1-like	(AT) ¹³	-	-
C6	sem identificação gênica	(CT) ⁶	100	monomórfico
C82	complement c1r subcomponent-like	(AC) ⁵	bandas múltiplas	-
C369	si:ch73-protein	(AT) ⁵	200	monomórfico
C902	protein disulfide-isomerase-like	(AG) ⁵	100	monomórfico
C13	dihydroxyvitamin d24-like	(AGC) ⁶	100	polimórfico
C494	delta-like protein b-like	(ATT) ⁶	100	monomórfico
C862	60s ribosomal protein I13a	(AAT) ⁶	bandas múltiplas	-
C1940	calnexin precursor	(AAG) ⁶	bandas múltiplas	-
C1830	complement c1r subcomponent-like	(AGC) ⁵	600	-
C195	protein disulfide-isomerase a4-like	(ACG) ⁵	1500	-
C524	eukaryotic translation initiation factor 2 subunit 2	(AAG) ⁵	200	monomórfico
C1731	solute carrier organic anion transporter family	(CGG) ⁵	200	monomórfico
C1341	ribosomal protein I4	(GCT) ⁵	100	monomórfico
C1872	sem identificação gênica	(GAT) ⁸	150	monomórfico
C64	sodium-coupled neutral amino acid transporter	(AAAG) ⁷	200	polimórfico
C1486	complement component c8 alpha chain-like	(ACTAT) ⁵	1000	-

Em seguida, para inferir os parâmetros de genética populacional e diversidade genética nos oito *loci* de microssatélites polimórficos, as análises de genotipagem foram realizadas no sequenciador de capilar automático. O número de alelos encontrado foi baixo, pois variou de

2 (*loci* C25, C64, C410 e C1832) a 5 (C1376). A heterozigosidade esperada (H_{esp}) e observada (H_{obs}) na população natural teve uma média de 0,171 e 0,354, respectivamente. Além dos baixos níveis de heterozigosidade, a H_{obs} foi menor que H_{esp} para a maioria dos *loci*, demonstrado também pelo valor de F_{is} positivo. Três *loci* de microsatélites (C13, C25 e C1716) foram considerados em desacordo com o Equilíbrio de Hardy-Weinberg (EHW). Entretanto, após a correção de Bonferroni para múltiplos *loci* ($p = 0,00625$), a população selvagem de pirapitinga estava em concordância com EHW ($p\text{-value} > 0,05$), mesmo a partir da análise de microsatélites gene-associados que está sujeitos a maior pressão de seleção.

Tabela 8. Parâmetros de diversidade de oito microsatélites polimórficos de pirapitinga (*Piaractus brachypomus*). A população selvagem analisada corresponde a 22 indivíduos coletados no Rio Tocantins. T_A °C = Temperatura de anelamento, A = número de alelos por locos, H_{obs} = heterozigosidade observada, H_{esp} = heterozigosidade esperada, P(HW) = p-value do equilíbrio de Hardy-Weinberg e o F_{is} = coeficiente de endogamia.

Locus	Motifs	Primers 5' - 3'	T_A °C	Tamanho Padrão	A	H_{obs}	H_{esp}	P(HW)	F_{is}
C13	(AGC) ₆	F: TCTCTTCAAGCCTCCTCTGC R: ATGCTGCAGCTCCTCCTGT	60°C	143 - 149	3	0,545	0,669	0,000	0,188
C25	(AT) ₁₁	F: CTTTGTCTGCTTTGGGTCGT R: CTTAGAAGAATGTGCAAATTGAAA	60°C	117 - 120	2	0,000	0,169	0,001	1,000
C64	(AAAG) ₇	F: CAAAGCAAACCTCAAAAAGGAAAA R: TGGGAACGTTTAGCATCTCA	55°C	143 - 151	2	0,545	0,474	0,650	-0,156
C410	(AG) ₈	F: CGCACAGGTCTAAAGGCACT R: CTCCCACACAGTGAAAAGCA	60°C	125 -127	2	0,273	0,359	0,271	0,245
C1005	(AG) ₇	F: AGTTGTTGCACCAAATGCAG R: CTTGTTCCCTCCCACACAGT	60°C	137 -141	3	0,318	0,369	0,225	0,140
C1376	(AC) ₁₀	F: GTGTTACATGGCAGGCGTTT R: CAAGTGAGACCAAATCCAAGG	60°C	157 - 175	5	0,500	0,608	0,185	0,180
C1716	(GT) ₇	F: AACCGAAGAGAGGGGAGTGT R: GCATTTACAAGGGGACGCAC	60°C	155 - 171	3	0,545	0,659	0,000	0,175
C1832	(AC) ₆	F: GGTGCTATGTCTAGAGGCC R: AGGAAGGCATGACCAGTGTG	60°C	159 - 169	2	0,111	0,529	0,036	0,800

3.4 Análise da diversidade genética em estoques cultivados de pirapitinga

Após o processo de validação dos oito *loci* de microssatélites, procedeu-se a etapa de aplicação desses marcadores. Com o intuito de analisar o perfil genético dos reprodutores de pirapitinga utilizados na piscicultura brasileira, foram realizadas análises de variabilidade genética em quatro estoques cultivados de reprodutores: Piscicultura SP1, Piscicultura SP2, Piscicultura TO1 e Piscicultura TO2.

Inicialmente, por meio da análise de pureza genética, nas pisciculturas do Estado de SP foram detectados indivíduos híbridos, sendo 22 amostras na Piscicultura SP1 e 1 exemplar na SP2. Esses animais foram descartados da análise de diversidade genética.

Para a maioria dos *loci* de microssatélites, a riqueza de alelos foi maior para os indivíduos pertencentes aos estoques das Pisciculturas do Norte (TO1 e TO2). Além disso, nos estoques das Pisciculturas TO1 e TO2, foram maiores os valores médios de H_{esp} ($0,601 \pm 0,096$ e $0,503 \pm 0,110$, respectivamente) e H_{obs} ($0,520 \pm 0,171$ e $0,456 \pm 0,104$, respectivamente). Já a população da Piscicultura SP1 apresentou menor polimorfismo genético, com um valor médio de H_{esp} significativamente mais baixo em relação as demais populações ($0,258 \pm 0,179$) ($p < 0,05$) e com no máximo três alelos na maioria dos *loci* de microssatélites analisados.

Com exceção da Piscicultura SP1, os *loci* de microssatélites analisados nos estoques de reprodutores foram caracterizados com deficiência de heterozigotos ($H_{\text{obs}} < H_{\text{esp}}$) e valores de F_{is} positivos, como é o caso do *locus* C25 nos estoques SP2 e TO1. Esses resultados interferiram na conformidade de Equilíbrio de Hardy-Weinberg (EHW) de ambas as pisciculturas no *locus* C25 ($P < 0,05$). Por outro lado, no *locus* C1376, o lote da piscicultura SP1 apresentou desvio do EHW devido a um evento de excesso de H_{obs} (0,857). Todavia, todos os demais marcadores microssatélites estiveram em EHW, após a correção de

Bonferroni ($p > 0,00625$). Adicionalmente, quando os marcadores microssatélites foram testados simultaneamente para todos os *loci*, todas as populações estiveram em EHW com significantes *p-values*. Nenhum desequilíbrio de ligação foi detectado entre os pares de *loci*, mesmo após a correção de Bonferroni ($p = 0,00625$).

Em relação à variabilidade genética avaliada em cada população por meio dos valores de F_{is} (coeficiente de endogamia), as pisciculturas SP1 e SP2 foram caracterizadas com predomínio de F_{is} negativos na maioria dos *loci*, enquanto que as demais populações apresentaram predomínio de F_{is} positivos ($TO1 = 6$, $TO2 = 6$). Na maioria dos casos, os valores médios mais próximos de zero, sugerindo uma homozigose existente, porém pouco representativa dentro de cada população.

Os resultados da análise hierárquica de variância molecular (AMOVA) mostrou que a maior proporção de variância esteve presente dentro das populações (62,5%). Na análise de AMOVA, também foi testada a hipótese de agrupamento das populações de acordo com sua procedência (grupo 1: Rio Tocantins; grupo 2: Pisciculturas do TO; e grupo 3: Pisciculturas de SP), o que mostrou uma variação de somente 5,74% entre as populações dentro dos grupos e de 31,76% entre os grupos, o que já demonstra uma certa diferença genética entre estoques de diferentes regiões do Brasil.

Uma alta diferenciação genética entre todas as populações foi detectada por análises de F_{st} sobre todos os oito *loci* de microssatélites ($F_{st} = 0,308$). Análises de F_{st} par a par ainda detectaram maior diferenciação genética entre a população selvagem e todas as populações cultivadas, com maior diferenciação quando relacionada com populações das pisciculturas SP1 ($F_{st} = 0,521$) e SP2 ($F_{st} = 0,498$). Adicionalmente, alta diferenciação genética entre populações de pirapitinga provenientes da região Norte e Sudeste do Brasil foi verificada, como por exemplo, nas populações das pisciculturas TO2 com SP1 ($F_{st} = 0,431$) e TO1 com SP2 (0,342) (tabela 10).

Tabela 9. Valores de diversidade genética de 8 *loci* de microsatélites de pirapitinga (*Piaractus brachypomus*). TO = população selvagem do Rio Tocantins, TO1 e TO2 = Pisciculturas do Estado do Tocantins, SP1 e SP2 = Pisciculturas do Estado de São Paulo. A = número de alelos, H_{obs} = heterozigosidade observada, H_{esp} = heterozigosidade esperada, P(HW) = P value do Equilíbrio de Hardy-Weinberg, F_{is} = coeficiente de endogamia.

Locus	A					Hobs					Hesp					P(EHW)					Fis				
	TO	TO1	TO2	SP1	SP2	TO	TO1	TO2	SP1	SP2	TO	TO1	TO2	SP1	SP2	TO	TO1	TO2	SP1	SP2	TO	TO1	TO2	SP1	SP2
PbC13	3	4	3	2	2	0,545	0,600	0,500	0,143	0,368	0,669	0,536	0,528	0,137	0,308	0,000	0,650	0,840	1,000	1,000	0,188	-0,121	0,055	-0,040	-0,200
PbC25	2	4	2	0	2	0,000	0,160	0,380	0,000	0,000	0,169	0,704	0,315	0,000	0,193	0,001	0,001	1,000	0,000	0,002	1,000	0,776	-0,212	0,000	1,000
PbC64	2	3	2	0	0	0,545	0,560	0,346	0,000	0,000	0,474	0,495	0,382	0,000	0,000	0,650	0,365	0,626	0,000	0,000	-0,156	-0,133	0,096	0,000	0,000
PbC410	2	3	2	3	3	0,272	0,520	0,384	0,287	0,157	0,359	0,537	0,506	0,264	0,152	0,271	0,267	0,256	1,000	1,000	0,245	0,034	0,244	-0,083	-0,038
PbC1005	3	4	2	2	3	0,318	0,560	0,384	0,500	0,157	0,369	0,562	0,506	0,494	0,152	0,225	0,421	0,256	1,000	1,000	0,140	0,004	0,244	-0,011	-0,385
PbC1376	5	3	6	3	5	0,500	0,440	0,538	0,857	0,625	0,608	0,609	0,632	0,634	0,790	0,185	0,010	0,020	0,000	0,112	0,180	0,282	0,151	-0,368	0,214
PbC1716	3	6	5	3	3	0,545	0,760	0,461	0,143	0,222	0,659	0,782	0,515	0,140	0,207	0,000	0,545	0,030	1,000	1,000	0,175	0,028	0,107	-0,019	-0,070
PbC1832	2	5	5	3	6	0,111	0,560	0,653	0,143	0,473	0,529	0,580	0,638	0,203	0,486	0,036	0,388	0,670	0,109	0,387	0,800	0,035	-0,253	0,306	0,027

Tabela 10. Estruturação genética (F_{st}) de 8 *loci* de microssatélite entre populações de pirapitinga (*Piaractus brachypomus*). TO = população selvagem do Rio Tocantins, TO1 e TO2 = Pisciculturas do Estado do Tocantins, SP1 e SP2 = Pisciculturas do Estado de São Paulo. Valores limiares de F_{st} = pouca diferenciação genética (0 – 0,05); diferenciação genética moderada (0,05 – 0,25); altos níveis de diferenciação genética (> 0,25).

TO1	0,367			
TO	0,521	0,164		
TO2	0,431	0,062	0,183	
SP2	0,149	0,342	0,498	0,380
	SP1	TO1	TO	TO2
Fst sobre todos os <i>loci</i> = 0,308				

O coeficiente de Wang (r) foi usado para avaliar o grau de parentesco em 14 indivíduos da piscicultura SP1 (figura 16), 25 indivíduos da TO1 (figura 17), 26 indivíduos da TO2 (figura 18) e 19 indivíduos da SP2 (figura 19), a fim de avaliar a aplicabilidade dos oito *loci* de microssatélites analisados em futuros programas de manejo e melhoramento genético. Foram detectados que as populações das pisciculturas TO1 e TO2 apresentaram uma maior proporção de indivíduos não aparentados, com frequências de 65,3% e 63,1%, respectivamente, e podem ser acasalados sem nenhum risco de endogamia (coeficiente de Wang na cor verde). Porém, na população da piscicultura SP1, a alta quantidade de irmãos completos (68,1%) (coeficiente de Wang na cor vermelha) pode levar a um maior risco de acasalamentos endogâmicos, dos quais não são interessantes para aquicultura.

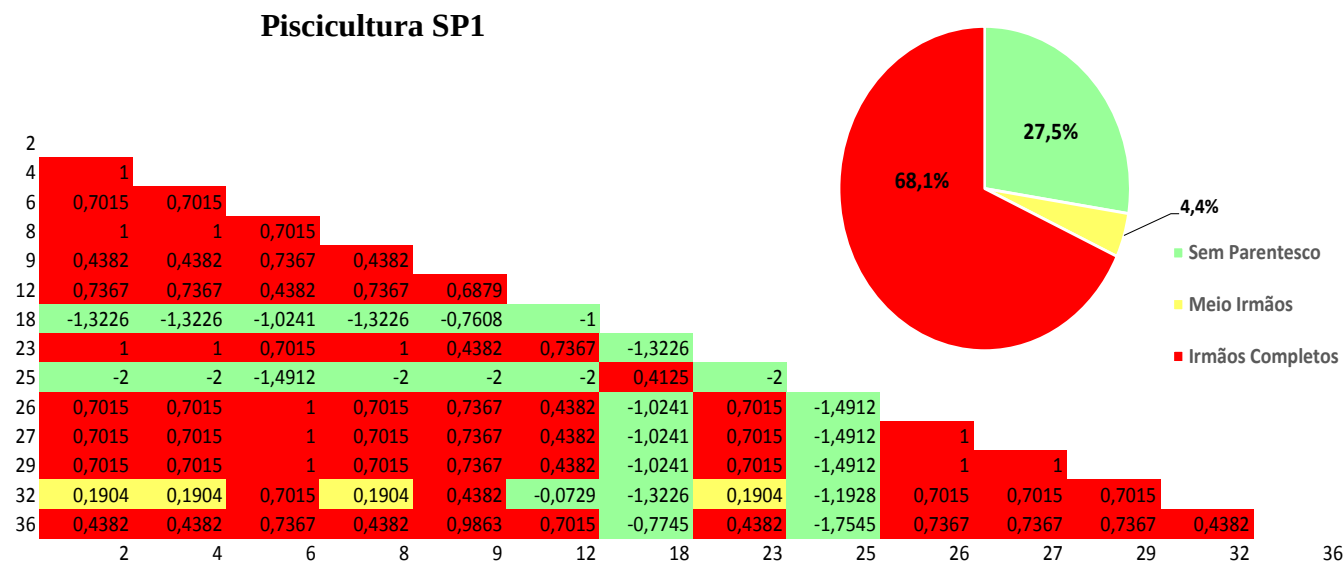


Figura 16. Análise de parentesco da população piscicultura SP1. Valores em verde representam indivíduos não-aparentados, valores em amarelo corresponde indivíduos meio-irmãos e valores em vermelho correspondem a irmãos completos. Irmãos completos ($r > 0,375$), meio-irmãos ($0,125 \leq r \leq 0,375$), e não aparentados ($r < 0,125$).

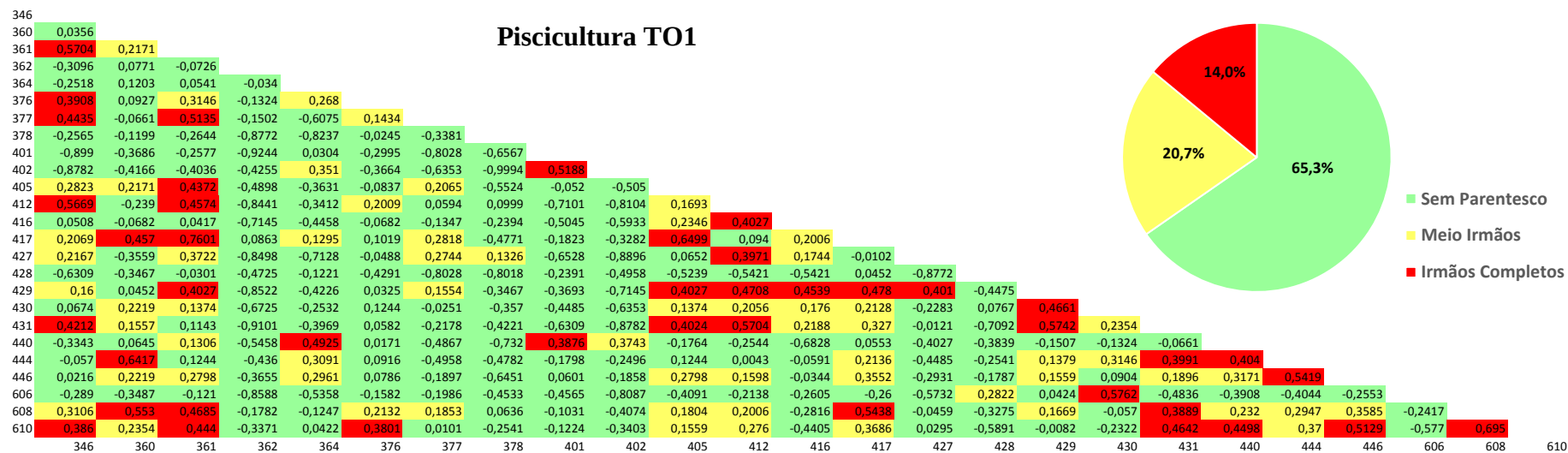


Figura 17. Análise de parentesco da população piscicultura TO1. Valores em verde representam indivíduos não-aparentados, valores em amarelo corresponde indivíduos meio-irmãos e valores em vermelho correspondem a irmãos completos. Irmãos completos ($r > 0,375$), meio-irmãos ($0,125 \leq r \leq 0,375$), e não aparentados ($r < 0,125$).

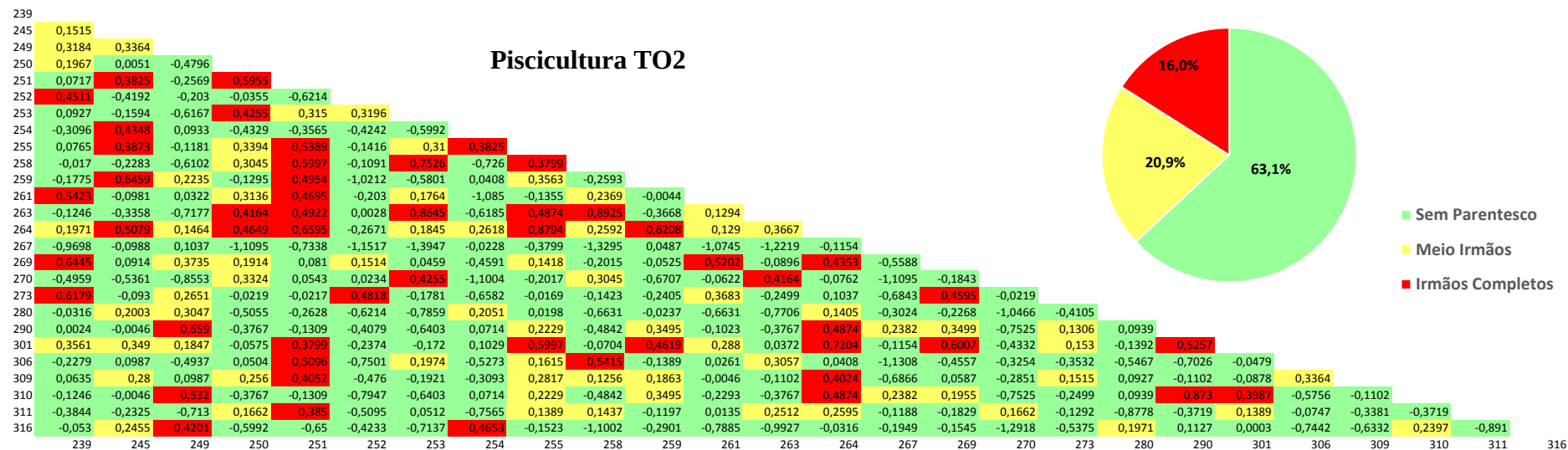


Figura 18. Análise de parentesco da população piscicultura TO2. Valores em verde representam indivíduos não-aparentados, valores em amarelo corresponde indivíduos meio-irmãos e valores em vermelho correspondem a irmãos completos. Irmãos completos ($r > 0,375$), meio-irmãos ($0,125 \leq r \leq 0,375$), e não aparentados ($r < 0,125$).

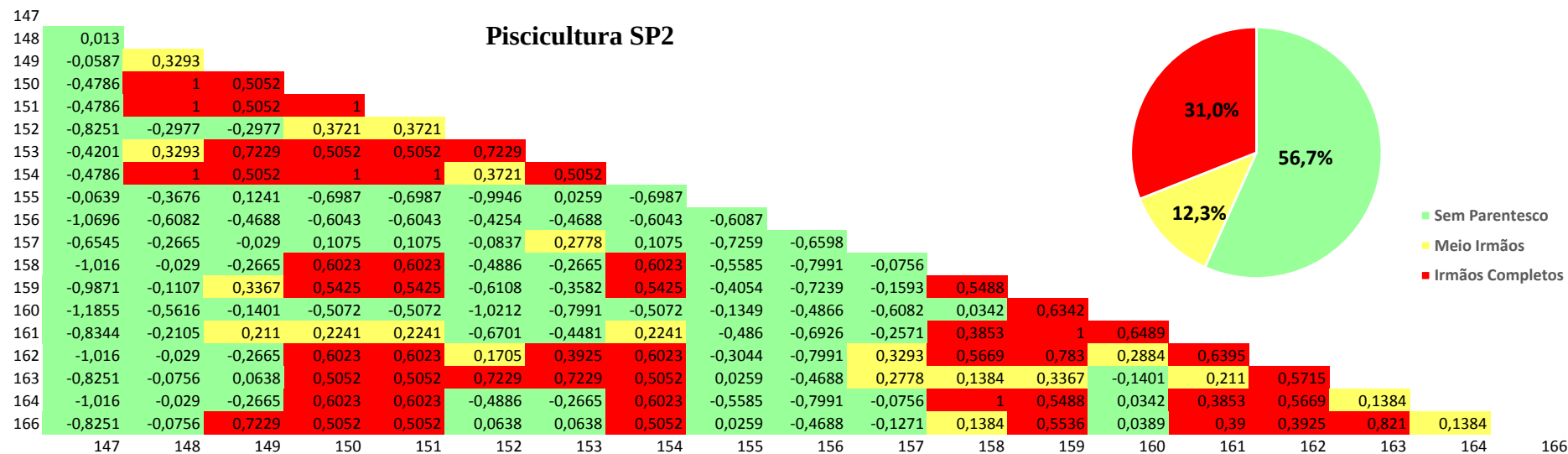


Figura 19. Análise de parentesco da população piscicultura SP2. Valores em verde representam indivíduos não-aparentados, valores em amarelo corresponde indivíduos meio-irmãos e valores em vermelho correspondem a irmãos completos. Irmãos completos ($r > 0,375$), meio-irmãos ($0,125 \leq r \leq 0,375$), e não aparentados ($r < 0,125$).

4. DISCUSSÃO

A pirapitinga é uma das mais importantes espécies Neotropicais de interesse para a produção, sendo cultivada em várias partes do mundo e a quinta espécie nativa mais produzida no Brasil (IBGE, 2014). Porém, informações genéticas relacionadas a espécies Neotropicais são praticamente inexistentes, o que leva a limitação do desenvolvimento de programas de melhoramento genético e consequente limitação do seu potencial produtivo. Pesquisas genéticas em pirapitingas estão limitadas apenas ao sequenciamento do genoma mitocondrial (Chen et al., 2016), sendo que estudos relacionados a características interessantes para a aquicultura, como crescimento, reprodução e sistema imunológico ainda são inexistentes.

Uns dos principais resultados deste estudo foram os dados gerados por meio do sequenciamento do transcriptoma. Estas sequências gênicas são essenciais porque pouco ou nenhum conhecimento sobre os genes desta espécie está disponível, frente aos milhares de genes que apresentam em seu genoma. Atualmente, as tecnologias de sequenciamento de nova geração (NGS) têm revolucionado a forma de explorar o genoma de qualquer tipo de organismo, independente de ser uma espécie-modelo ou não. Há cerca de quinze anos atrás, praticamente nenhum genoma de peixes tinha sido sequenciado com um bom nível de cobertura. Os primeiros genomas de peixes a serem disponibilizados foram utilizando a tecnologia de Sanger, somente para espécies modelo e com genoma pequeno, como *medaka* e *baiacu* (Aparicio et al., 2002; Jaillon et al., 2004; Kasahara et al., 2007). Entretanto, com o advento das tecnologias NGS, inúmeros genomas de peixes estão sendo obtidos nos últimos três anos, iniciando-se com o sequenciamento do genoma de *sticklebacks* (Jones et al., 2012). No período de 2012 a 2014, mais de 10 genomas de peixes já foram publicados, inclusive para espécies não-modelos, resultando em uma enormidade de dados genéticos que ainda

precisam ser explorados (Xu et al., 2013; Smith et al., 2013; Scharl et al., 2013; Howe et al., 2013; Amemiya et al., 2013; Brawand et al., 2014; Chen et al., 2014; Venkatesh et al., 2014).

A maioria dos sequenciamentos de genomas de peixes tem enfoque direcionado para genômica comparativa, especialmente com a finalidade de entender a evolução dos genomas de peixes. Poucas espécies com interesse econômico, sobretudo para aquicultura, tiveram os genomas estudados de forma mais detalhada e aplicada. Somente no sequenciamento do genoma da carpa comum *Cyprinus carpio* foi gerado uma grande quantidade de marcadores genéticos (Xu et al., 2012), já disponíveis para aplicação direta no melhoramento e cultivo desta espécie, visando aumentar a produtividade em aquicultura.

Um dos entraves de sequenciamentos de genomas completos é o trabalho intenso e prolongado de bioinformática, particularmente de predição de genes, montagem de sequências repetitivas e anotação funcional. Por isso, de acordo com Seeb et al. (2011), estratégias de redução de genoma para o sequenciamento NGS (ex., sequenciamento de transcriptoma) são mais viáveis quando se objetiva prospectar marcadores e informações genéticas para uso em aquicultura, de forma direta e rápida, além do baixo custo econômico.

No presente estudo, a plataforma 454/Roche foi utilizada com eficiência no sequenciamento do transcriptoma de fígado de pirapitinga (*Piaractus brachypomus*), mostrando ser uma ferramenta poderosa, pois resultou na descoberta de dezenas de microssatélites, mesmo sem algum conhecimento prévio sobre a sequência de referência (genoma de referência), já que a organização do genoma completo desta espécie ainda não foi estudada. Desta forma, os dados deste estudo confirmam a aplicabilidade da tecnologia 454/Roche em pesquisas que envolvem caracterizar o transcriptoma de espécies não-modelos (Renaut et al., 2010; Shin et al., 2012), propiciando melhor compreensão da genômica funcional por meio da análise do DNA complementar (cDNA), sintetizado a partir do RNA mensageiro (mRNA). Logo, entende-se que o sequenciamento do transcriptoma por meio de

tecnologias de NGS viabiliza a descoberta de maior quantidade de microssatélites em menor tempo e custo, uma vez que realiza o sequenciamento evitando íntrons e regiões extragênicas que podem dificultar a análise de dados, como o que ocorre quando se realiza bibliotecas de WGS (*whole genome shotgun*) a partir do DNA genômico (Parkinson e Blaxter, 2009).

Para obtenção deste transcriptoma de pirapitinga por NGS, foi realizada uma corrida de poucas horas, o que resultou em cerca de 70 milhões de nucleotídeos sequenciados. Desta forma, fica evidente a principal característica dos sequenciadores de nova geração: velocidade, rendimento e preço do sequenciamento. Esta capacidade de realizar maior número de leitura de base por corrida, com um custo e período menor, se comparados ao sequenciamento por Sanger (uma corrida pode gerar poucas milhares de bases), possibilita uma exploração melhor dos genomas de espécies não modelos da aquicultura mundial. Os dados de transcriptoma obtidos para a pirapitinga neste estudo estão similares com outros trabalhos prévios em outras espécies de peixes utilizando a mesma tecnologia, o que corrobora a tendência em utilizar esta metodologia para descoberta de marcadores moleculares em larga escala (Renaut et al., 2010; Helyar et al., 2012; Shin et al., 2012).

A eficiência do sequenciador 454/Roche na análise de genômica funcional de *P. brachypomus* também pode ser constatada em virtude da identificação dos potenciais genes, nas quais 3.797 sequências transcritas foram geradas neste estudo. Essa quantidade de *contigs* foi semelhante ao produzido por Shin et al. (2012) em um sequenciamento de mRNA do fígado de uma espécie de peixe Antártico, o *Chaenocephalus aceratus*, mostrando a similaridade no número de sequências neste órgão, apesar de pertencerem à classificação taxonômica distinta. Em um levantamento realizado em Janeiro de 2016 no banco de dados *Genbank*, por exemplo, foram constatadas apenas 187 sequências nucleotídicas, aproximadamente 14% são sequências nucleotídeos nucleares e o restante, 86% são sequências mitocondriais disponíveis para a pirapitinga, o que demonstra o avanço que o

presente estudo pode proporcionar para as pesquisas genéticas nesta espécie.

Assim, está evidente que as tecnologias NGS vêm proporcionando progresso nas metodologias de biologia molecular, especialmente porque estão adaptadas a produzirem uma elevada cobertura de sequências com precisão (Mardis, 2008), contudo, a escolha de uma plataforma adequada é essencial para cumprir o objetivo de cada pesquisa devido às particularidades individuais. Por exemplo, as tecnologias Illumina/HiSeq e ABI/SOLID, por produzirem fragmentos de *reads* mais curto em relação à 454/Roche, apresentam menor precisão no processo de montagem *de novo* (Morozova e Marra, 2008). Além disso, ter conhecimento sobre a bioinformática é imprescindível para analisar os dados e ponderar os riscos em longo prazo, ao estabelecer critérios menos rigorosos para os parâmetros.

Neste trabalho, ao admitir limiar de $Q > 20$ para o parâmetro de *trimming*, observa-se que não houve redução brusca na quantidade nem no tamanho médio de *reads*, equiparando-se ainda mais com os resultados apresentados por Shin et al. (2012), nos quais, além do número de *contigs* de *Chaenocephalus aceratus* ter sido semelhante, o número de *reads* ($n = 220.146$) e o comprimento médio ($\bar{x} = 339$ pb) também foram similares. Este resultado conduz a deduzir que o valor de *phred* adotado nesse estudo foi adequado e que o sequenciamento gerou fragmentos de qualidade.

Os resultados da anotação funcional mostraram que as sequências de pirapitinga apresentaram praticamente a mesma proporção de genes anotados quando comparadas com o banco de proteínas de *zebrafish* e *fugu*. Porém, observou-se uma maior similaridade com *zebrafish*, do que em relação ao *fugu*, pode ser explicado pela maior proximidade filogenética com esta espécie. A anotação gênica permitiu identificar regiões responsáveis por processos de desenvolvimento ontogênico, regulação biológica, resposta imunológica, e regiões envolvidas em processos de crescimento e reprodução. Consequentemente, estes dados poderão servir como subsídio para estudos da biologia da espécie ou para futuros programas

de melhoramento genético.

De forma aplicada, o sequenciamento do transcriptoma pela plataforma 454 é um excelente método para o desenvolvimento de marcadores moleculares do tipo microssatélites (Novaes et al., 2008; Bai et al., 2011). Até o presente momento, nossos resultados também corroboram com esta perspectiva, uma vez que a partir de um único sequenciamento, foi possível identificar mais de cem regiões repetitivas do tipo microssatélites. Além disso, uma vez que microssatélites oriundos de sequências do transcriptoma são provavelmente de genes codificadores de proteínas, o uso destes marcadores se torna de importância ainda maior em mapas genéticos (Shin et al., 2012), pois por genômica comparativa (usando genomas de peixes já sequenciados) podemos inferir a posição de cada *locus* estudado.

Com os resultados obtidos a partir do processo de validação dos marcadores moleculares, foi possível selecionar 8 *loci* de microssatélites polimórficos, alguns deles localizados em regiões gênicas que podem ser úteis para características zootécnicas de interesse na aquicultura. Neste caso, foi encontrado um *locus* de microssatélite no gene *Tetraspanin – 3 isoform x1* (C1832), que tem papel na patologia de infecção viral (Shoshana et al., 2005; Martin et al., 2005); e outro no gene *Cytosolic 5 – nucleotidase 3 a-like* (NTC5C3) (C25) (Aksoy et al., 2009), que contribui na produção de hemácias e sua mutação pode causar anemia hemolítica e influência no sistema imunológico. Os microssatélites descritos neste estudo poderão ser importantes em futuras análises de QTL (*quantitative trait loci*) ligados a características do sistema imunológico, uma vez que programas de seleção direcionados para resistência a doenças tem recebido atenção especial na aquicultura (Gjedrem et al., 2012; Odegard et al., 2011). Além disso, também foi caracterizado um *locus* de microssatélite no gene *apolipoprotein e* (C410), associado ao sistema nervoso central e processo de envelhecimento (Wang et al., 2014). Esses marcadores podem fornecer

informações úteis para estudos relacionados ao conhecimento da biologia da pirapitinga, além de servir como modelo para outras espécies nativas.

Em relação à análise de variabilidade genética, de forma geral, o presente estudo revelou um baixo nível de diversidade genética (riqueza de alelos e heterozigosidade) para os oito marcadores microssatélites validados, tanto em populações naturais como em populações cultivadas. Estudos anteriores realizados com pacu (*Piaractus mesopotamicus*) (Calcagnotto e DeSalle, 2009), outra espécie do mesmo gênero, mostraram falta de diferenciação genética em populações da bacia do Paraná-Paraguai, com níveis de H_{obs} semelhantes entre 0,090 e 1,000 e valores de F_{is} positivos (entre 0 e 0,636) mais altos que os negativos, sugerindo deficiência de heterozigotos na maioria dos loci, além de serem consideradas uma população panmítica. A perda de variabilidade genética em população natural pode ser decorrente de alguns fatores ecológicos e evolutivos, como modificações nas características do ambiente, deriva genética e endogamia (Primack e Rodrigues, 2001). No caso de estoques cultivados, este fato pode ser devido ao baixo número de reprodutores da população base inicial, a aquisição de plantéis oriundos de outras pisciculturas, e mesmo a seleção de indivíduos adaptáveis em sistemas de cultivo. Além disso, diferentemente de marcadores neutros, deve-se levar em consideração o fato de que foram utilizados microssatélites em associação com genes, portanto, mais susceptíveis a pressão de seleção e, conseqüentemente, a redução de alelos e heterozigosidade.

Análises da composição e diversidade genética de estoques cultivados de pirapitinga foram realizadas em quatro pisciculturas de diferentes regiões do Brasil, sendo duas provenientes da região Sudeste (Estado de São Paulo: SP1 e SP2) e duas da região Norte (Estado do Tocantins: TO1 e TO2). Por meio da análise hierárquica de variância molecular (AMOVA), foi demonstrado que a maior variação genética está presente dentro das populações (62,5%), do que entre populações ou entre os grupos de populações (Selvagem,

SP e TO). Esta característica genética também foi registrada em estudos realizados em pacu (Iervolino et al., 2010; Calcagnotto e De Salle., 2009) e tambaqui (Aguar et al., 2013), com maior proporção de variação genética representada dentro das populações. Por outro lado, além do padrão de diversidade genética (riqueza de alelos e heterozigosidade) serem similares, os valores de F_{st} analisados por população mostram o baixo nível de estruturação genética do ambiente natural (TO) com as populações cultivadas na região Norte (Estado do Tocantins: TO1 e TO2). Essa semelhança de composição genética pode estar relacionada com uma maior proximidade geográfica à população selvagem do Rio Tocantins, o que facilita a renovação dos estoques de reprodutores oriundos da natureza. Enquanto isso, devido à falta de fontes alternativas de reprodutores, nas populações de pisciculturas da região Sudeste (Estado de São Paulo: SP1 e SP2) ocorrem valores menores de riqueza alélica e heterozigosidade, além de alto nível de estruturação com a população selvagem, resultando em maiores chances de endogamia.

Em geral, uma maior suscetibilidade a doenças pode ocorrer em populações endogâmicas, o que afeta a produção de peixes sem técnicas de seleção e acasalamentos apropriados entre indivíduos (Akvaforsk, 2005). Análise de parentesco é considerada uma ferramenta essencial em programas de melhoramento genético de peixes, pois minimiza taxas de endogamia pelo direcionamento de acasalamentos entre indivíduos não aparentados (Pino-Querido et al., 2010). Portanto, o coeficiente de Wang (r) foi utilizado para avaliar grau de parentesco nos indivíduos das populações cultivadas na região Norte (TO1 e TO2) e região Sudeste (SP1 e SP2). Os resultados demonstraram que as populações da região Sudeste tiveram uma maior proporção de indivíduos aparentados, que não são viáveis para aquicultura, pois podem causar depressão endogâmica, com proles de baixa viabilidade e deformidades esqueléticas (Su et al., 1996 ; Imsland et al., 2001). Esta análise revela a aplicabilidade direta destes marcadores moleculares para as pisciculturas brasileiras.

A geração de dados genéticos para a pirapitinga é um dos aspectos prioritários para ser incorporada dentro do cenário de pesquisa agropecuária no país, frente à sua importância econômica para piscicultura nacional e mundial, a crescente demanda do consumidor por produtos melhorados e a expansão da piscicultura brasileira nos últimos anos. A identificação de microssatélites genes associados por NGS são fundamentais para conhecimento da estrutura genética de populações naturais e cultivadas, e realização de análises de parentesco das espécies não modelo, fornecendo melhor controle sobre o direcionamento das famílias nos acasalamentos em programas de manejo genético. Somente nestas condições será possível realizar um programa de melhoramento genético acelerado para a pirapitinga, como tem sido feito para outras espécies da piscicultura mundial, como salmão, catfish, carpa e tilápia (Liu et al., 2011; Lien et al., 2011; Jin et al., 2012; Guyon et al., 2012).

5. CONCLUSÕES

- aumento significativo dos recursos genéticos disponíveis para pirapitinga (*Piaractus brachypomus*), um peixe não modelo da aquicultura e pouco estudado geneticamente, mas com alto valor econômico em diversos países.
- o sequenciamento do transcriptoma se mostrou eficaz na prospecção de microssatélites gene-associados na pirapitinga.
- os estoques cultivados de pirapitinga do Norte apresentaram maior variabilidade genética do que os lotes do Sudeste.
- alto nível de estruturação genética foi encontrado entre lotes cultivados e a população natural, bem como entre os estoques cultivados do Norte e do Sudeste.
- alta porcentagem de indivíduos aparentados foi encontrada na piscicultura SP1 (95,6%), o que poderá acarretar em depressão endogâmica e prejuízos econômicos.
- os resultados obtidos com a análise dos microssatélites mostram a necessidade e aplicabilidade rotineira destas ferramentas moleculares na piscicultura nacional, especialmente para propor planos eficientes de manejo genético.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Alcântara PF, Oliveira AA, Nobre MISN (1990) Considerações sobre a amostragem da pirapitinga, *Colossoma brachypomum*, Cuvier, no estado do Ceará (Brasil). *Ciências Agronômicas*. **21**: 43-49.
- Amemiya CT, Alföldi J, Lee AP, et al. (2013) The African coelacanth genome provides insights into tetrapod evolution. *Nature*. **496**: 311–316.
- Aguiar J, Scheneider H, Gomes F, Carneiro J, Santos S, Rodrigues LR, Sampaio I (2013) Genetic variation in native and farmed population of tambaqui (*Colossoma macropomum*) un the Brazilian Amazon: regional discrepancies is farming systems. *Anais de Academia Brasileira de Ciências*. **85(4)**: 1439-1447.
- Aparicio S, Chapman J, Stupka E, et al. (2002) Whole-Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes*. *Science*. **297**: 1301-1310.
- Alcântara PF, Oliveira AA, Nobre MISN (1990) Considerações sobre a amostragem da pirapitinga, *Piaractus brachypomum*, Cuvier, no estado do Ceará (Brasil). *Ciências Agronômicas*. **21**: 43-49.
- Araújo-Lima CARM, Ruffino ML, Resende EK. (2003) Migratory fishes of the Brazilian Amazon. In: Carolsfield J, Harvey B, Ross C, Baers A, editors. Migratory fishes of South America: biology, Fisheries and conservation states. *Victoria*: World Bank. 237-291.
- Aksoy L, Zhu MJ, Kalari KR, Moon I, Pellymounter LL, Eckloff BW, Wleben ED, Yee VC, Richar M (2009) Cytosolic 5`Nucleotidase III (NT5C3): Gene sequence variation e functional genomics. *Pharmacogenet Genomics*. **19(8)**: 567-576.
- Akvaforsk A. Selection and Breeding Programs in Aquaculture (2005) Ed. Trygve Gjedrem. Dordrecht: *Springer*.

- Bai J, Li Q, Cong RH, Sun WJ, et al. (2011) Development and characterization of 68 EST-SSR markers in the Pacific oyster, *Crassostrea gigas*. *Journal of the World Aquaculture Society*. **42(3)**: 444–455.
- Brawand D, Wagner CE, Li YI, et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. **513**: 375–381.
- Calgnotto D, Schaefer SA, DeSalle R(2005) Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. *Molecular Phylogenetics and Evolution*. **36**: 135–153
- Calcagnotto D, DeSalle R (2009) Population genetic structuring in pacu (*Piaractus mesopotamicus*) across the Paraná-Paraguay basin: evidence from microsatellites. *Neotropical Ichthyology*. **7(4)**: 607-616.
- Conesa A, Götz S, García-Gómez JM, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21**: 3674-3676.
- Chen S, Zhang G, Shao C, et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature genetics*. **46**: 253-260.
- Chen HP, Li SS, Xie ZZ, Zhang Y, Zhu CH, Deng SP, Li GL, Huang H (2016) The complete mitochondrial genome of the *Piaractus brachypomus* (characiformes: Characidae). *Mitochondrial DNA*. **27(2)**: 1289-1290.
- Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*. **255**: 1-29.

- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature*. **12(7)**: 499-510.
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred: II. Error probabilities. *Genome Research*. **8**: 86–94.
- Ewing B, Hillier L, Wendl MC, et al. (1998) Base-calling of automated sequencer traces using Phred: I. Accuracy assessment. *Genome Research*. **8**: 175–85.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. **131**: 479-491.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. **10**: 564-567.
- Foresti F (2000) Biotechnology and fish culture. *Hydrobiologia*. **420**: 45 -47
- Faircloth BC (2008) Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Research*. **8**: 92-94.
- Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.
- Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220p.
- Fungaro MHP, Vieira MLC (2001) Marcadores moleculares. In: Biotecnologia na agricultura e na agroindústria. Guaíba: Agropecuária. 153-199.
- Flores Nava A (2007) Aquaculture seed resources in Latin America: a regional synthesis. In: Assessment of freshwater fish seed resources for sustainable aquaculture (ed Bondad-Reantaso MG). FAO Fisheries Technical Paper, No. 501, FAO, Rome.

- Fresneda A, Lenis G, Agudelo E, *et al.* (2004) Espermiación inducida y crioconservación de semen de cachama blanca (*Piaractus brachypomus*). *Revista Colombiana Ciencias Pecuárias*. **17**: 46-52.
- Gao Z, Luo W, Liu H, Zeng C, *et al.* (2012) Transcriptome analysis and SSR/SNP markers information of the Blunt Snout Bream (*Megalobrama amblycephala*). *PLoS One*. **7(8)**:e42637.
- Galhardo E (1989) Estudos genéticos do polimorfismo da trnasferrinas em populações selvagens e cultivadas de curimatá, *Prochilodus scrofa* (Pisces, Prochilodontidae) Disserrtação de mestrado. Instituto de Biociências de São Paulo, São Paulo.
- Gross JB, Furterer A, Carlson BM, *et al.* (2013) An integrated transcriptome-wide analysis of cave and surface dwelling *Astyanax mexicanus*. *PLoS One* **8**: e55659.
- Guyon R, Rakotomanga M, Azzouzi N, *et al.* (2012) A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics*. **13**: 222.
- Gjedrem T, Robinson N, Rye M. (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture*. 350-353: 117-129.
- Hashimoto DT, Alves AL, Varela ES, *et al.* (2012) Genética na Piscicultura: Importância da variabilidade genética, marcação e coleta para análise de DNA. 1. ed. Brasília: Embrapa, 32p.
- Hancock RE, Diamond G (2000) The role of cationic antimicrobial peptides in innate host defences. *Trends Microbiology*. **8(9)**: 402-410.
- Hardy OJ, Vakemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618–620

- Helyar SJ, Limborg MT, Bekkevold D, et al. (2012) SNP Discovery Using Next Generation Transcriptomic Sequencing in Atlantic Herring (*Clupea harengus*). *Plos one* **7(8)**: e42089.
- Hernández A (1994) Estado actual del cultivo de *Colossoma* y *Piaractus* en Brasil, Colombia, Panamá, Perú y Venezuela. In: Memorias del VIII Congreso Latinoamericano de Acuicultura y V Seminario Nacional de Acuicultura, Acuicultura y Desarrollo Sostenible. Santafé de Bogotá. Colombia: 9-23.
- Honglang, H. (2007) Freshwater fish seed resources in China. In Assessment of Freshwater Fish Seed Resources for Sustainable Aquaculture (Ed. M. G. BondadReantaso), pp. 185–199. FAO Fisheries Technical Paper No. 501. Rome: FAO.
- Howe K, Clark MD, Torroja CF, et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498-503.
- Hulata G (2001) Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* **111**: 155-173.
- IBAMA (2007) Estatística da Pesca (2007): Brasil – Grandes regiões e unidades da Federação. 113 pp.
- IBGE. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal 2013. 2014; Rio de Janeiro, 41: 1-108. Available in <http://www.ibge.gov.br/home/estatistica/economia/ppm/2013/>.
- Iervolino F, Resende EK, Hilsdorf AWS (2010) The lack genetic differentiation of pacu (*Piaractus mesopotamicus*) Populations in the Upper-Paraguay basin revealed by the mitochondrial DNA D-loop region: Implication for fishery management. *Fisheries Research*. **101**: 27-31

- Imslund AK, Foss A, Nævdal G, Stefansson SO (2001) Selection Or Adaptation: differences in growth performance of juvenile turbot (*Scophthalmus maximus* Rafinesque) from two Close-by localities of Norway. *Sarsia*. **86**: 43-51
- Jaillon O, Aury JM, Brunet F, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*. **431**: 946-957.
- Jégu M (2003) Check list of the freshwater fishes of South and Central America. *Edipucrs*. 182-196.
- Ji P, Liu G, Xu J, Wang X, Li J; Zhao Z, et al. (2012) Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. *PLoS One*. **7 (4)**: e35152.
- Jones FC, Grabherr MG, Chan YF, et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. **484**: 55-61.
- Kasahara M, Naruse K, Sasaki S, et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature*, **447**: 714-719.
- Kincaid HL (1983) Inbreeding in fish populations used for aquaculture. *Aquaculture*. **33(1-4)**: 215 – 227.
- Koljonen ML, Tahtinen J, Saisa M , Koskiniemi J (2002) Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture*. **212**: 69-92.
- Lehoczky I, Magyary I, Hancz C, et al. (2005) Preliminary studies on the genetic variability of six Hungarian common carp strains using microsatellite DNA markers. *Hydrobiologia*. **533**: 223-228.
- Liu S, Zhou Z, Lu J, et al. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*. **12**: 53.

- Liu S, Zhou Z, Lu J, et al. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*. **12**: 53.
- Lien S, Gidskehaug L, Moen T, et al. (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*. **12**:615.
- Long Y, Li Q, Zhou B, et al. (2013) De novo assembly of Mud Loach (*Misgurnus anguillicaudatus*) skin transcriptome to identify putative genes involved in immunity and epidermal mucus secretion. *Plos One*. **8**: e56998.
- Martin F, Roth DM, Jans DA, Pouton CW, Partridge LJ, P. N. Monk PN , Moseley GW. (2005) Tetraspanins in Viral Infections: a Fundamental Role in Viral Biology? *American Society for Microbiology* .**79(17)**: 10839–10851.
- Machado-Schiaffino G, Dopico E, Garcia-Vazquez E (2007) Genetic variation losses in Atlantic salmon stocks created for supportive breeding. *Aquaculture*. **264**: 59-65.
- Macmanes MD (2014) On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*. **5**:1-7.
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annual Review Genomics Human Genetics*. **9**: 387-402.
- Micallef G, Bickerdike R, Reiff C, Fernandes JMO, Bowman AS, Martin SAM. (2012) Exploring the transcriptome of Atlantic salmon (*Salmo salar*) skin, a major defense organ. *Marine Biotechnology*. **14(5)**: 559-569.
- Mirande JM (2010) Phylogeny of the family Characidae (Teleostei: Characiformes): from characters to taxonomy. *Neotropical Ichthyology*. **8(3)**: 385-568.
- Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*. **92**: 255-264.

- MPA (2013a) Boletim estatístico da pesca e aquicultura – Brasil 2011. Ministério da Pesca e Aquicultura.
- MPA (2013b) Censo aquícola nacional - ano 2008. Ministério da Pesca e Aquicultura.
- Oliveira EJ, Pádua JG, Zucchi MI, et al. (2006) Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*. **2**: 294-307.
- Odegard J, Baranski M, Gjerde B, Gjerdem T. (2011) Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquaculture Research*. **42**: 103-114.
- Ortí G, Sivasundar A, Dietz K, et al. (2008) Phylogeny of the Serrasalminae (Characiformes) based on mitochondrial DNA sequences. *Genetics and Molecular Biology*. **31**: 343-351.
- Parkinson J, Blaxter M. (2009) Expressed sequence tags: an overview. *Methods in Molecular Biology*. **533**: 1-12.
- Peakall ROD, Smouse PE. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. **6**: 288-295
- Pino-Querido A, Hermida M, Vilariño M, Bouza C, Martínez P. Statistical properties and performance of pairwise relatedness estimators using turbot (*Scophthalmus maximus* L.) family data. *Aquaculture Research*. 2010; 41: 528-534.
- Primack, RB e Rodrigues E. (2001) *Biologia da conservação*. Editora Rodrigues, Londrina.
- Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology Research*. **19**: 115-131.
- Rozen S, Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*. **132**:365-86.
- Rousset F. GENEPOP'007: (2008) A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*. **8**: 103-106.

- Salem M, Paneru B, Al-Tobasei R, Abdouni F, Thorgaard GH, Rexroad CE, et al. (2015) Transcriptome assembly, gene annotation, and tissue gene expression atlas of the rainbow trout. *PLoS ONE*. **10(3)**: e0121778.
- Sarropoulou E, Nousdili D, Magoulas AGK. (2008) Linking the genomes of nonmodel teleosts through comparative genomics. *Marine Biotechnology*. **10**: 227-233.
- Seeb JE, Carvalho G, Hauser L, et al. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Research*. **11**: 1-8.
- Su GS, Liljedahl LE, Gall GAE. (1996) Effects of inbreeding on growth and reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. **142**: 139-148.
- Suplicy FM (2007) Freshwater fish seed resources in Brazil. In: Bondad-Reantaso MG (ed.), *Assessment of freshwater fish seed resources for sustainable aquaculture*. FAO Fisheries Technical Paper. No. 501. FAO, Rome, 129-143.
- Schlotterer C. (2004). The evolution of molecular markers – just a matter of fashion? *Nature Reviews in Genetics*. **5**: 63-69.
- Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology*. **18**: 233-234.
- Shin SC, Kim SJ, Lee JK, et al. (2012) Transcriptomics and comparative analysis of three Antarctic notothenioid fishes. *PLoS One* **7**: e43762.
- Shoshana L, Shoham T (2005) Tetraspanin web modulates immune – signaling complexes. *Nature*. **5**: 136 -148.
- Tave D (1999) Inbreeding and brood stock management. *Fisheries Technical Paper*. n° 392. Roma, FAO.

- Toledo-Filho SA, Almeida-Toledo LF, Foresti F, Calcagnotto D, Santos SBAF e Bernardino G (1998) Programas genéticos de seleção, hibridação e endocruzamento aplicados à piscicultura. *Cadernos de Ictiogenética* 4, CCS/USP, São Paulo.
- Wang X, Xiamoi S, Lian J, Zhang S (2014) Identification, expression and function of apolipoprotein e in animal fish *Nothobranchis guentheri* : implication for an aging marker. *Biogerontology*. **15**: 233-243.
- Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*. **160**: 1203-1215.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*. **38**: 1358-1370.
- Vera M, Alvarez-Dios JA, Fernandez C, Bouza C, Vilas R, Martinez P (2013) Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *International Journal of Molecular Sciences*. **14**: 5694-5711.
- Venkatesh B, Lee AP, Ravi V, et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature*. **505**:174–179.
- Xu J, Ji P, Zhao Z, et al. (2012) Genome-wide SNP discovery from transcriptome of four common carp strains. *Plos One*. **7**: e48140.
- Xu J, Liu L, Xu Y, Chen C, et al. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Research*. **20**: 497–50
- Zheng W, Xu H, Lam SH, et al. (2013) Transcriptomic analyses of sexual dimorphism of the zebrafish liver and the effect of sex hormones. *PLoS One*. **8**: e53562.