

Construção de um denso mapa genético e associação genômica para
peso corporal e sexo em *Piaractus mesopotamicus*

NATÁLIA JADE MENDES

Botucatu - SP

2019

Construção de um denso mapa genético e associação genômica para peso corporal e sexo em *Piaractus mesopotamicus*

Natália Jade Mendes
Prof. Dr. Fausto Foresti
Dr. Diogo Teruo Hashimoto

Tese apresentada ao Instituto de
Biociências, Campus de Botucatu,
UNESP, para a obtenção do título de
Doutor no Programa de Pós-Graduação
em Ciências Biológicas (Genética).

Botucatu – SP

2019

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

Mendes, Natália Jade.

Construção de um denso mapa genético e associação genômica para peso corporal e sexo em *Piaractus mesopotamicus* / Natália Jade Mendes. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

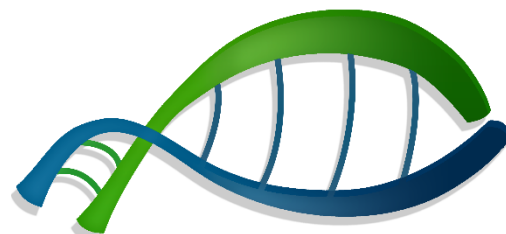
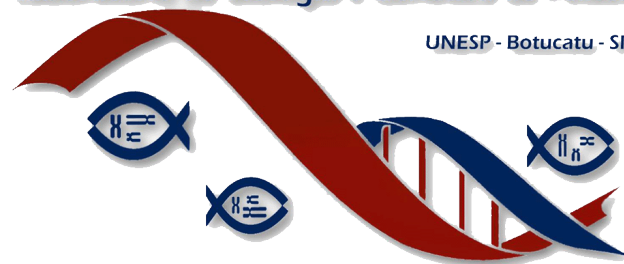
Orientador: Fausto Foresti

Coorientador: Diogo Teruo Hashimoto

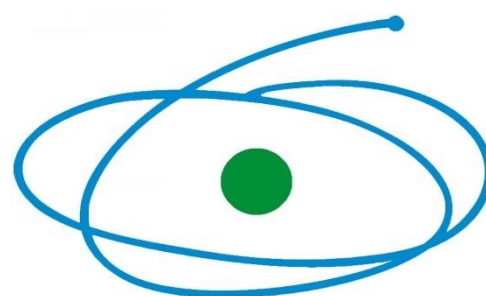
Capes: 20204000

1. Pacu (Peixe). 2. Genômica. 3. Marcadores genéticos.
4. Técnicas de genotipagem.

Palavras-chave: GWAS; Mapa de ligação; Pacu; Radseq.



L a G e A C



C A P E S



Dedico esta tese ao meu marido e a minha filha!

*“O período de maior ganho em conhecimento e experiência
é o período mais difícil da vida de alguém”*

Dalai Lama

AGRADECIMENTOS

Primeiramente eu agradeço a Deus por ter me guiado até aqui.

Agradeço a minha filha Gigi por ter me acompanhado nessa longa etapa em Jaboticabal e ao meu marido por ter me dado todo o suporte durante o doutorado. Eu amo vocês!

A minha família Rute, Félix, Bruno, Hícaro, Vitor, Vitória e em memória ao meu pai Donizete; vocês são o meu porto seguro. Em especial a minha mãe por sempre estar ao meu lado; eu não sei o que seria da minha vida sem você. E ao meu pai que por apenas 2 meses não conseguiu me acompanhar fisicamente até essa conquista, mas que está presente espiritualmente e no meu coração.

Obrigada a todas as pessoas que me ajudaram a cuidar da Giovanna para que eu pudesse me dedicar ao doutorado, especialmente minha mãe, ao meu segundo pai Félix, a minha sogra Cleusa, as minhas tias Érica e Andréia e a minha prima Laysla.

Agradeço ao meu orientador Prof. Dr. Fausto Foresti pela confiança, pelos conselhos e por diversas vezes ter me mostrado o caminho correto. O Sr. é um exemplo de pesquisador e pessoa.

Ao meu coorientador Dr. Diogo Teruo Hashimoto por todo o ensinamento, confiança e pela paciência. Não é fácil ter uma mãe com filho pequeno na equipe.

Ao Dr. Claudio Oliveira pelo exemplo como pesquisador e pela amizade.

Agradeço a todos do Laboratório de Biologia e Genética de Peixes, aos meus amigos que em algum momento estivemos juntos aprendendo e dando risadas no laboratório: Silvana, Fábio (Fio), Bruno Melo, Gabriel (Pink), Luz, Vanessa, Yuldi, Sâmia (Alka), Jefferson (Menudo), Dani, Carol (Papiro), Guilherme (Varvito), Cristiane, Camila (Sova), Ricardo (Tarja).

Também agradeço a todos da equipe Lageac (Vito, Valéria, Milena, Raquel, Jonh, Carol e Rubens). Foram quatro anos de muito aprendizado, em especial à Valéria que por diversas vezes me socorreu com a Gigi.

A toda equipe do Aquaculture Genomics Laboratory que me recebeu de braços abertos. Ao Prof. José Yáñez por ter me recebido em seu laboratório e ao Agustin por toda atenção durante o meu estágio no Chile.

Aos amigos que fiz em Botucatu, em especial às minhas amigas Camila (Judi), Letícia (Loba), Camila (Travs), Bruna (Trep), Thaís (Balboa). Apesar da distância, vocês estiveram presentes na minha vida. Judi, obrigada pelos conselhos, pela amizade, por ter sido o meu apoio emocional durante essa jornada, amo você minha irmã.

Também agradeço a todos os meus amigos de São Carlos que fizeram a minha vida mais feliz e, em especial, a todos do grupo “Etanóis”.

Aos órgãos de fomento Capes, Fapesp e CNPq por financiarem as nossas pesquisas.
E ao CNPq pela bolsa de estudos concedida.

Muito obrigada!

RESUMO

Estudos genéticos com foco no aprimoramento de espécies de peixes Neotropicais são raros e, nesse sentido, uma espécie alvo é o pacu (*Piaractus mesopotamicus*), uma vez que é considerada uma das espécies de peixe de grande interesse econômico, com grande produção e valor comercial para a piscicultura brasileira. O presente trabalho teve como objetivo desenvolver ferramentas genéticas e genômicas através da genotipagem dos marcadores genéticos do tipo SNPs (*Single Nucleotide Polymorphisms*) para desenvolvimento de um mapa de ligação para esta espécie e, dessa maneira identificar regiões correlacionadas com características de peso corporal e sexo. Inicialmente foram formadas 14 famílias, sendo os indivíduos marcados com *pit-tags* e alocados em 3 tanques distintos. Os indivíduos foram pesados e tiveram o sexo identificado; no entanto, apenas 3 famílias foram genotipadas (família1, família2 e família3). Na etapa seguinte, foi realizada a extração do DNA e a elaboração de uma biblioteca genômica para cada família através da metodologia *Restriction-Site Associated DNA-Sequencing (RADseq)*. As bibliotecas foram sequenciadas na plataforma *Hiseq 4000 Illumina*, resultando em uma média de 35.697.700 de sequências brutas de cada família. Um total de 11.240 marcadores SNP foram prospectados e filtrados, sendo que destes 6.437 foram utilizados na construção do primeiro mapa genético da espécie, que possui 2.814,92cM com intervalo médio de 0.44. Esses mesmos marcadores também foram utilizados para desenvolver o primeiro GWAS (Estudo de Associação Ampla do Genoma) prospectado com indivíduos genotipados e não genotipados, totalizando 399 amostras. A análise do GWAS foi realizada por meio da abordagem ssGBLUP e explicou 3,21% para o peso corporal e 2,89% para o sexo. A partir da identificação dos *loci* foi realizado o *Blast* e a busca pelos genes candidatos, sendo identificados genes para o peso corporal envolvidos com fatores de crescimento, regulação e metabolismo do osso, codificação do colágeno, proliferação e crescimento celular. Do mesmo modo foi verificado que os genes candidatos para identificação sexual estavam relacionados à espermatogênese, diferenciação sexual, células germinativas e expressão do hormônio anti-Mülleriano (AMH) em células de Sertoli. Considera-se que as ferramentas genéticas e genômicas desenvolvidas neste trabalho poderão auxiliar em investigações sobre a evolução da espécie, organização do genoma, identificação de QTLs (*loci de característica quantitativa*) e em programas de melhoramento genético para o pacu e espécies aparentadas.

Palavras-chave: Pacu, Mapa de ligação, GWAS, RADseq

ABSTRACT

Genetic studies focusing on the improvement of neotropical fish species are rare and, in this sense, one target species is pacu (*Piaractus mesopotamicus*), as it is considered one of the most produced fish species and commercial value for Brazilian fish farming. The present work aimed to develop genetic and genomic tools through the genotyping of SNPs (Single Nucleotide Polymorphisms) genetic markers for the development of a linkage map for this species, and thus identify regions correlated with body weight and sex. Initially, 14 families were formed, individuals were tagged with Pit-tags and allocated in 3 distinct tanks. The subjects were weighed and had the gender identified; however, only 3 families were genotyped (family1, family2 and family3). In the next step, DNA extraction and the elaboration of a genomic library for each family were performed through the Restriction-Site Associated DNA-Sequencing (RADseq) methodology. The libraries were sequenced on the platform Hiseq 4000 Illumina, resulting in an average of 35.697.700 raw sequences from each family. A total of 11.240 SNP markers were prospected and filtered, of which 6.437 were used to construct the first genetic map of the species, which has 2.814,92cM with an average interval of 0.44. These same markers were also used to develop the first GWAS (Genome Wide Association Study) with genotyped and non-genotyped individuals, totaling 399 samples. The GWAS analysis was performed using the ssGBLUP approach and explained 3,21% for body weight and 2,89% for sex. Based on the loci identification, the Blast was performed and the candidate genes were searched, and genes for body weight involved with bone growth, regulation and metabolism, collagen coding, proliferation and cell growth factors were identified. Similarly, it was found that candidate genes for sexual identification were related to spermatogenesis, sexual differentiation, germ cells and anti-Müllerian hormone (AMH) expression in Sertoli cells. It is considered that the genetic and genomic tools developed in this work can assist in investigations on species evolution, genome organization, QTL identification (quantitative trait loci) and breeding programs for pacu and related species.

Keywords: Pacu, Linkage Map, GWAS, RADseq

SUMÁRIO

INTRODUÇÃO GERAL.....	1
Panorama geral da aquicultura.....	1
Melhoramento genético na piscicultura.....	2
Genômica na piscicultura.....	4
Caracterização da espécie <i>Piaractus mesopotamicus</i>	7
OBJETIVOS.....	9
Objetivo geral.....	9
Objetivos específicos.....	9
MATERIAIS E MÉTODOS.....	9
CAPÍTULO 1 Descoberta de SNPs e construção de um mapa genético de alta densidade através do ddRADseq de <i>Piaractus mesopotamicus</i>.....	10
1.1 INTRODUÇÃO.....	12
1.2 MATERIAIS E MÉTODOS.....	14
1.2.1 Formação dos grupos genéticos, instalações e manejo.....	14
1.2.2 Construção de bibliotecas ddRADseq.....	15
1.2.3 Identificação de filtro de SNPs.....	17
1.2.4 Construção do Mapa Genético.....	19
1.2.5 Mapeamento Comparativo.....	19
1.3 RESULTADOS.....	20
1.3.1 Construção das bibliotecas ddRADseq.....	20
1.3.2 Identificação e filtro de SNPs.....	20
1.3.3 Construção do Mapa Genético.....	21
1.3.4 Mapeamento Comparativo.....	24
1.4 DISCUSSÃO.....	25
1.4.1 Construção do Mapa Genético.....	25
1.5 REFERÊNCIAS.....	27
CAPÍTULO 2 Estudo de Associação Genômica Ampla para o Peso Corporal e Sexo em <i>Piaractus mesopotamicus</i>.....	31

2.1 INTRODUÇÃO.....	33
2.2 MATERIAIS E MÉTODOS.....	35
2.2.1 Formação dos grupos genéticos, instalações e manejo.....	35
2.2.2 Avaliação do peso corporal e do sexo.....	36
2.2.3 Estudo de associação genômica ampla (GWAS) e herdabilidade.....	38
2.2.4 Candidate genes.....	39
2.3 RESULTADOS.....	40
2.3.1 Avaliação do peso corporal e do sexo.....	40
2.3.2 Estudo de associação genômica ampla (GWAS) e herdabilidade.....	41
2.3.3 Candidate genes.....	43
2.4 DISCUSSÃO.....	45
2.4.2 Estudo de associação genômica ampla (GWAS) e herdabilidade.....	45
2.4.3 Candidate genes.....	46
2.5 REFERÊNCIAS.....	50
 CONSIDERAÇÕES FINAIS.....	 56
REFERÊNCIAS GERAIS.....	58
 MATERIAL COMPLEMENTAR.....	 65
APÊNDICE I.....	66
APÊNDICE II.....	76
APÊNDICE III.....	88
APÊNDICE IV.....	97
APÊNDICE V.....	110

INTRODUÇÃO GERAL

Panorama geral da aquicultura

Com a exploração pesqueira praticamente estática desde meados da década de 1980, de acordo com a Organização das Nações Unidas para a Agricultura e Alimentação (FAO, 2018), a oferta de pescado dependerá certamente de desenvolvimento e de uma correta exploração dos recursos aquícolas (Porto-Foresti & Foresti, 2004). Neste sentido, pesquisas devem ser realizadas com o objetivo de aumentar a produção mundial de peixes de maneira sustentável (Longo *et al.*, 2019), ou seja, promover o desenvolvimento da aquicultura com medidas menos impactantes e com o investimento de recursos cada vez menor.

O setor da aquicultura manteve uma taxa de crescimento anual médio de 8,30% em todo o mundo entre 1970 e 2008 (FAO, 2010). No período de 2000 a 2016, o crescimento foi de 5,80%, sendo que o setor continua crescendo quando comparado aos outros setores de produção animal (FAO, 2018). No ano de 2014, a aquicultura continental mundial produziu cerca de 47.1 milhões de toneladas (FAO, 2016), enquanto em 2016 este valor aumentou para 51.4 milhões de toneladas e foi responsável por aproximadamente 30,00% de toda a produção de organismos aquáticos (FAO, 2018). Ainda de acordo com os registros da FAO (2018), os maiores produtores de peixes de águas continentais no ano de 2016 foram a China com 28.2 milhões de toneladas, a Índia com 5.0 milhões de toneladas e a Indonésia com 3.4 milhões de toneladas. Neste cenário, o Brasil aparece entre os 10 maiores produtores, com aproximadamente 507.1 mil toneladas.

No Brasil, a aquicultura é um setor com grande potencial, principalmente devido à produção recorde de grãos, indústria de rações estabelecida e amplo território (8,5 milhões de km²), situado em grande parte sob um clima tropical com boa disponibilidade hídrica e áreas favoráveis para a construção de tanques e açudes (Kubitza, 2015). O início efetivo da aquicultura no Brasil foi estabelecido na década de 1990, quando houve um crescimento acima da média mundial. No período de 1991 a 2004 a aquicultura brasileira cresceu em média 21,10%/ano, enquanto a mundial cresceu cerca de 9,50%/ano (FAO, 2006). Entre o período de 2007 a 2010 a produção de pescado pela aquicultura brasileira teve um avanço de 87,00%, cujos valores totais passaram de 210.644 toneladas para 394.340 toneladas, segundo estatísticas do IBAMA (2007) e do Ministério da Pesca e Aquicultura (MPA, 2012). Em 2014, o IBGE (Instituto Brasileiro de Geografia e Estatística) incluiu a aquicultura em seu censo agropecuário.

Segundo essa instituição, em 2013 o Brasil produziu 392.5 mil toneladas de pescados cultivados.

Em 2018, a piscicultura brasileira continuou crescendo e atingiu um total de 519.3 mil toneladas de peixes (IBGE, 2019). Neste sentido, houve um crescimento de produtos da ordem de 4,40% em relação ao ano de 2017. De acordo com os dados divulgados pela FAO (2016), a aquicultura brasileira terá forte expansão nos próximos anos, sendo que o consumo per capita de pescados no Brasil deverá chegar a 12.7 kg em 2025, ou seja, 32,00% a mais do que os 9.6 kg consumidos entre 2013 e 2015. Os dados apontam também que a produção de pescado em cativeiro no País mais que dobrará em dez anos.

Levando em consideração este cenário, existem basicamente duas alternativas para os aquicultores melhorarem o desempenho produtivo nos sistemas de cultivo. A primeira seria por meio de uma ampliação em suas instalações, o que frequentemente é inviável, uma vez que nem sempre existem áreas ou volume de água disponíveis, além da ampliação frequentemente resultar na elevação dos custos de produção. A segunda seria através da maximização da produtividade, ou seja, pela melhora do rendimento da produção de peixes produzidos por unidade de área cultivada, tanto por meio de melhorias ambientais realizadas na propriedade, que envolveria a aplicação de práticas de manejo adequadas, com equilibrada taxa de arraçamento e controle da qualidade da água, quanto por meio da instalação de um programa de melhoramento genético dos plantéis, com o objetivo de aumentar o ganho genético nos estoques de peixes que estão sendo cultivados (Tave, 1995).

Melhoramento genético na piscicultura

Entre as técnicas utilizadas no Brasil para se obter estoques genéticos mais produtivos na área da aquicultura, a hibridação interespecífica é responsável pela produção de milhares de toneladas de peixes ao ano (IBGE, 2019). Porém, apesar do aparente resultado zootécnico resultante da hibridação entre espécies nativas, existe uma séria preocupação com a possibilidade de escape e introgressão genética destes híbridos sobre as populações selvagens (Hashimoto *et al.*, 2012; 2013), além do risco da implementação incidental de híbridos no plantel de reprodutores, frequentemente devido à dificuldade de se identificar e diferenciar exemplares híbridos daqueles correspondentes às espécies puras, principalmente nos estágios iniciais de vida (Hashimoto *et al.*, 2014). Portanto, para evitar possíveis impactos ambientais e na estrutura das populações naturais, a aplicação de técnicas de seleção e acasalamentos

controlados entre indivíduos de uma mesma espécie usando a variabilidade genética presente nestas populações apresentam-se como alternativas mais sustentáveis (Hiltsdorf & Orfão, 2011).

O Brasil apresenta excelentes recursos naturais e uma rica biodiversidade de peixes com potencial para a implantação de programas de melhoramento genético, sendo que aproximadamente 40 espécies nativas de rios brasileiros são utilizadas nos programas desenvolvidos pela aquicultura nacional (Godinho, 2007). Entretanto, a maior parte dos cultivos utiliza espécies exóticas, principalmente a tilápia (IBGE, 2019), espécie que passa no Brasil por rígidos e longos programas de melhoramento genético desde a década de 2000. Este panorama revela a carência de projetos de melhoramento e inovação genética com enfoque nas espécies de peixes nativos.

Em relação aos objetivos de seleção utilizados pelos programas de melhoramento genético de animais em aquicultura, a maioria destes está ligada à produção, tais como taxa de crescimento, resistência a patógenos e rendimento de filé. Entre tantos fenótipos de interesse zootécnico, características relacionadas ao crescimento, assim como ao peso corporal, constituem uma das primeiras a serem buscadas, pois considera-se que quanto maior for o crescimento dos animais melhor será a utilização dos recursos (Gjedrem *et al.*, 2012) e, dessa forma, o retorno financeiro poderá ser mais rápido.

Diversos estudos foram realizados para avaliar o potencial de crescimento das espécies de peixes de interesse econômico tendo por base ações de melhoramento genético, tendo sido observadas diferenças significativas no peso corporal entre as gerações de várias espécies. Entre estas, a tilápia do Nilo *Oreochromis niloticus* apresentou um ganho genético de 11,9% por geração, cerca de 219 g (Hamzah *et al.*, 2014); a espécie *Cyprinus carpio*, conhecida como carpa comum, teve um ganho de 18,2% por geração, sendo que este valor equivale a 533 g (Ninh *et al.*, 2013), enquanto que na espécie *Oncorhynchus mykiss*, conhecida como truta arco-íris, o ganho foi de 7,0 %, aproximadamente 2.7 kg por geração (Kause *et al.*, 2005).

Além das características relacionadas ao ganho de peso, o conhecimento sobre o sexo dos indivíduos é de grande importância na piscicultura, pois além de auxiliar no controle do sistema de acasalamentos, também a exploração de apenas um sexo é mais vantajosa em algumas espécies (Dong *et al.*, 2019a). Outro benefício do controle na identificação do sexo dos indivíduos seria para a formação de lotes de produção monosexo, que resultaria na contenção de gastos energéticos com a atividade reprodutiva, na obtenção de uniformidade de tamanho na colheita, na redução dos efeitos da maturação sexual, bem como na aparência e na qualidade da carne (Beardmore *et al.*, 2001).

Entretanto, existem espécies que não apresentam um dimorfismo sexual bem definido e as ferramentas moleculares são utilizadas para auxiliar na determinação do sexo. Neste sentido, alguns estudos foram realizados com o intuito de identificar os genes que determinam o sexo em peixes, tendo já resultado na caracterização de vários genes, como exemplo há o gene *amhy* na espécie *Odontesthes hatcheri* (Hattori *et al.*, 2012); do gene *GsdfY* em *Oryzias luzonensis* (Myosho *et al.*, 2012); e do gene *DMY/dmrt1bY* em *Oryzias latipes* (Matsuda *et al.*, 2002, 2003).

Partindo desta questão, a espécie de peixe utilizada no presente trabalho é uma espécie nativa conhecida popularmente como pacu (*Piaractus mesopotamicus*) que possui grande aplicação na piscicultura (Santos *et al.*, 2012). Contudo, sua utilização em programas de melhoramento genético ainda apresenta algumas limitações e necessita de investimentos mais significativos. Considera-se, pois, que seria de grande importância o desenvolvimento de variedades com características voltadas à obtenção de maior peso corporal e com um sistema de determinação sexual bem conhecido, visto que esta espécie não possui características fenotípicas diferenciais relacionadas ao sexo.

Genômica na piscicultura

Com o desenvolvimento da genômica e a identificação de um denso painel de SNPs através dos avanços das tecnologias NGS (*Next Generation Sequencing*), foi possível a identificação de milhares de marcadores moleculares em um curto espaço de tempo, ao mesmo tempo que foi possível a obtenção de produtos com preço cada vez mais reduzido e maior acurácia no sequenciamento (Kumar *et al.*, 2017). Entretanto, apesar do expressivo progresso verificado na área da genômica, este avanço não foi completamente acompanhado pela geração de dados referentes às espécies de peixes Neotropicais, sendo observada uma situação aquém das possibilidades quanto à descrição de marcadores SNPs. Assim, algumas referências encontradas com relação à descrição desses marcadores moleculares referem-se ao estudo com o híbrido *Amphilophus spp* (Recknagel *et al.*, 2013), com o tambaqui *Colossoma macropomum* (da Silva Nunes *et al.*, 2017) e com o pacu (Mastrochirico-Filho *et al.*, 2016).

Entre as diversas metodologias existentes para a identificação dos SNPs, destaca-se o RADseq (*Restriction Site Associated DNA- Sequencing*) (Baird *et al.*, 2008). O RADseq é uma tecnologia de sequenciamento de DNA associada aos sítios de restrição e à fragmentação mecânica, sendo a análise desenvolvida a partir de uma grande quantidade de amostras formando um *pool* que pode ser sequenciadas de uma só vez (Andrews *et al.*, 2016). Porém,

uma variação chamada de ddRADseq (*Double digest RADseq*) é caracterizada pelo uso de dupla restrição enzimática, pode resultar em alta capacidade de multiplexação e apresentar maior confiabilidade no tamanho dos RADtags obtidos (Peterson *et al.*, 2012).

Com a implementação dos marcadores moleculares nos programas de melhoramento genético, as estratégias convencionais de seleção tornaram-se mais eficientes, promovendo a possibilidade de maiores ganhos genéticos nas características de interesse econômico (Sonesson *et al.*, 2012), podendo ser realizado através da identificação dos *loci de característica quantitativa* (QTLs) (Rosa *et al.*, 2013). Existem duas principais estratégias para a descoberta dos QTLs, a tradicional (mapeamento de ligação) e o GWAS (estudo de associação genômica ampla).

A primeira metodologia explora a diferença alélica entre apenas dois parentais e geralmente identifica uma baixa quantidade de QTL devido à baixa taxa de recombinação e a dificuldade em avaliar uma grande quantidade de *loci*. (Zhu *et al.*, 2008). A segunda abordagem indica os SNPs associados com características de interesse através de um denso painel de SNPs e faz a correlação entre marcador e o *loci* com base no desequilíbrio de ligação (LD), dessa forma este processo identifica a região do genoma (QTL) em que esses SNPs estão inseridos (Kijas *et al.*, 2014). A vantagem do GWAS é poder amostrar inúmeros *loci*, identificar diferentes níveis de recombinação (Yu *et al.*, 2008) e testar várias características ao mesmo tempo, dessa forma um grande número de regiões pode ser associado.

O estudo de associação genômica tem sido utilizado para várias características economicamente importantes para a aquicultura em espécies modelos e não-modelos, incluindo características de crescimento (Neto *et al.*, 2019; Yu *et al.*, 2019., Yang *et al.*, 2020, Shi *et al.*, 2020), resistência a doenças (Barría *et al.*, 2019, Wang *et al.*, 2019; Zhou *et al.*, 2019), *loci* e genes candidatos relacionados a determinação sexual (Gabián *et al.*, 2019; Yossa *et al.*, 2019; Wang *et al.*, 2019), entre outros.

Os resultados obtidos do mapeamento QTL e GWAS são aplicadas através do MAS (seleção assistida por marcadores) e GS (seleção genômica). Como exemplo, atualmente, vários QTLs são comercializados pela empresa norueguesa Aquagen de salmonídeos em germoplasma qualificado, livres de diversas doenças, sem a síndrome cardiomiopática e também com melhor pigmentação de filé (Aquagen, 2014, 2015a, 2015b). E ainda, se tratando de salmonídeos, um caso que se destacou na aquicultura foi a utilização do mapa de QTL em salmão do atlântico para seleção da resistência ao IPN – Necrose Pancreática Infeciosa (Houston *et al.*, 2008; Moen *et al.*, 2009). Com a implementação do MAS, a salmonicultura europeia reduziu em 75,00% os

focos da doença e tem garantido atualmente uma economia anual de 24 milhões de libras no Reino Unido (Houston *et al.*, 2008; Moen *et al.*, 2009).

Entretanto, para a identificação de QTLs é necessário o conhecimento prévio sobre a localização dos marcadores moleculares no genoma, tornando essencial a construção de um mapa de ligação (Moen *et al.*, 2004). Os mapas de ligação além de auxiliar no estudo de associações genômicas, também possibilitam a localização de genes, estudos comparativos de sintenia genômica, cobertura e análise completa de genomas, além de gerar informações a respeito da arquitetura genética e pleiotropia (Guo *et al.*, 2007; Li *et al.*, 2016). A função dos mapas genéticos é ordenar os marcadores moleculares em um cromossomo ou em grupos de ligação, sendo que a distância entre eles é estimada a partir da frequência da taxa de recombinação (Xu *et al.*, 2008). Para espécies de peixes nativos atualmente existe apenas o mapa de ligação desenvolvido para a espécie *Colossoma macropomum* (Silva Nunes *et al.*, 2017).

Além da identificação de QTLs, a utilização da genômica pode ter diversas aplicações de interesse zootécnico que são de difícil medição ou nas quais o fenótipo demora para se expressar; entre estes casos podem ser citados a resistência a doenças (Houston *et al.*, 2008), o acompanhamento do crescimento corporal (Gutierrez *et al.*, 2015), as características de formato do corpo (Dong *et al.*, 2019b). A utilização desta metodologia também auxilia o estabelecimento de populações de base para utilização em programas de melhoramento genético, otimiza a diversidade genética, controla a consanguinidade das gerações subsequentes produzidas (Fernández *et al.*, 2014) e prever valores genômicos de reprodução para caracteres de interesse (Yoshida *et al.*, 2018). Pode ser considerada ainda a utilização dos marcadores moleculares em estudos evolutivos, assim como em estudos da genética de populações e filogenia (Bernatchez *et al.*, 2016) e no conhecimento da biologia e estrutura do genoma (Yáñez *et al.*, 2015).

A implementação das ferramentas moleculares em programas de melhoramento genético proporciona expressivos avanços com relação a aumento nos ganhos genéticos. Considera-se, assim, que a aplicação das técnicas genômicas no estudo das espécies nativas pode determinar um desenvolvimento acelerado nos processos de produção, bem como auxiliar nos estudos sobre a evolução e biologia das espécies utilizadas na piscicultura.

Caracterização da espécie *Piaractus mesopotamicus*

A espécie *Piaractus mesopotamicus* (Holmberg, 1887) (Figura 1) componente da família Serrasalminidae (Calcagnotto *et al.*, 2005; Oliveira *et al.*, 2011) é um peixe conhecido popularmente como pacu-caranha, caranha, pacu-guaçu, ou simplesmente pacu, sendo nativa das bacias dos rios Paraná, Paraguai e Uruguai, com especial distribuição nas planícies alagadas da região Centro-Oeste, no Pantanal do Mato Grosso (Petrere Jr, 1989; Resende, 2003). Os exemplares desta espécie possuem corpo ovalado, estreito e robusto, com dorso cinza escuro e ventre amarelado (Britski *et al.*, 2007), sendo que existem relatos de exemplares que atingiram 82,0 cm de comprimento total e peso de 18,5 kg (Moreira *et al.*, 2001). No ambiente natural é comum encontrar exemplares com 7 kg que atingem entre 60 a 80 cm de comprimento total (Britski *et al.*, 2007). A primeira maturidade sexual ocorre aos 3 anos, com duração média até os 8 anos, sendo caracterizados por um período de reprodução que atinge o seu ápice nos meses de novembro a janeiro, época de temperaturas mais altas e maior incidência de chuvas (Bernardino, 1999). Na estação chuvosa, o pacu permanece nas áreas de inundação onde se alimenta principalmente de itens de origem vegetal; porém, durante a seca permanece no leito dos rios, local onde pode haver maior disponibilidade de alimento (Menton, 1989; Urbinati *et al.*, 2013).

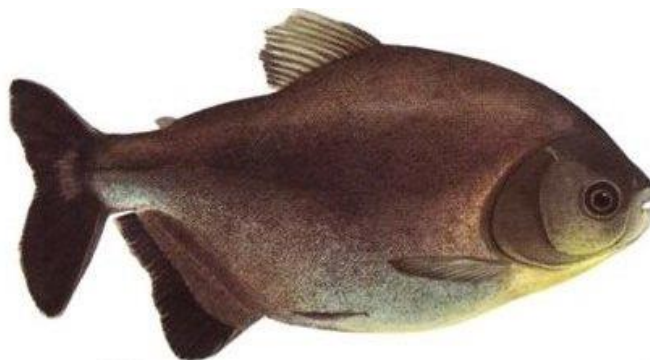


Figura 1. Imagem representativa da espécie *Piaractus mesopotamicus* Holmberg, 1887 (pacu).

Fonte: Britski *et al.*, 2007

O pacu é uma das espécies de peixes mais estudadas no Brasil, principalmente com relação aos aspectos de reprodução, larvicultura, engorda, nutrição e parasitologia (Urbinati *et al.*, 2013). Sua criação com fins comerciais vem crescendo principalmente devido características como uma dieta com menor exigência de proteínas, bom sabor da carne, baixo

custo de manutenção, boa resistência a patógenos, baixa exigência quanto à qualidade da água, alta adaptabilidade ao cultivo em tanques e viveiros, além de ser uma das espécies mais apreciadas na pesca esportiva (Fernandes *et al.*, 2000; Abimorad & Carneiro, 2004; Jomori *et al.*, 2008). Tais características levaram o pacu a ser a quarta espécie nativa mais produzida na aquicultura continental nacional (IBGE, 2017).

De acordo com publicação da FAO (2018), a maior produção nacional do pacu ocorreu em 2010, com 21.245 toneladas. No ano de 2018, de acordo com o IBGE (2019), o pacu e seu híbrido patinga (fêmea *P. mesopotamicus* x macho de *P. brachypomus*), foram relacionados como ocupantes do quarto maior renda de produção de pescados entre as espécies nativas, com aproximadamente 11.600 toneladas. Ainda em relação à piscicultura, a espécie *P. mesopotamicus* também tem considerável importância econômica para a aquicultura de outros países da América do Sul como Colômbia, Peru, Venezuela e Argentina, bem como de países asiáticos como China, Mianmar, Tailândia e Vietnã (Flores Nava, 2007; Honglang, 2007; FAO, 2010).

Em relação à pesca extrativa, o último dado referente à estatística pesqueira continental nacional foi referente ao ano de 2011, no qual o pacu ocupou o sexto lugar no ranking de espécies mais pescadas, representando, juntamente com a pescada, aproximadamente 10,00% do total de capturas (MPA, 2013). No entanto, a partir de 2011, a maior parte das informações obtidas sobre estatística pesqueira é proveniente de dados regionais, aos quais pode ser citado os dados coletados e analisados por meio do Sistema de Controle da Pesca de Mato Grosso do Sul SCPESCA/MS. No ano de 2016, o pacu ocupou o terceiro lugar como a espécie mais pescada (52t - 13,80%), para os setores comercial e esportivo (Catella *et al.*, 2017).

Além da dificuldade enfrentada pela sobrepesca, a espécie também tem sentido os efeitos da fragmentação de seu habitat, causado pela construção de sucessivas barragens ao longo dos cursos dos rios (Agostinho *et al.*, 2005) e, em consequência do impacto sofrido pelas populações naturais, a espécie foi classificada como “quase ameaçada” no ano de 2014, de acordo com o Decreto N° 60.133 (São Paulo, 2014). Por isso, a reprodução desta espécie em cativeiro tem sido uma boa alternativa para suprir a demanda do mercado e reduzir o impacto nos estoques das populações naturais.

Embora sendo o pacu uma espécie bastante estudada e importante na aquicultura brasileira, ainda são escassas as informações relativas à genômica e ao melhoramento genético aplicados no seu cultivo. Considera-se, portanto, que os resultados obtidos no presente estudo como a prospecção de SNPs, desenvolvimento do primeiro mapa genético da espécie e o

GWAS para o peso corporal e identificação do sexo dos indivíduos, certamente se constituirão em ferramentas importantes para um manejo biológico adequado e para o melhoramento genético da espécie.

OBJETIVOS

Objetivo geral

O presente estudo teve como objetivo a prospecção e aplicação de marcadores SNPs para a construção de um mapa genético e estudo de associação genômica ampla (GWAS), para testar a correlação com o peso corporal e determinação sexual na espécie *Piaractus mesopotamicus*.

Objetivos específicos

- caracterização dos marcadores SNPs provenientes das *reads* do ddRADseq para os parentais e dos indivíduos da geração F1;
- desenvolvimento de um mapa genético de SNPs para esta espécie;
- estudo de associação genômica ampla (GWAS) as características de peso corporal e determinação do sexo nos exemplares;
- identificação dos genes candidatos próximos aos QTLs correlacionados com as características de peso corporal e determinação do sexo citadas acima.

MATERIAIS E MÉTODOS

Os materiais e métodos utilizados no presente trabalho serão descritos na parte específica de cada capítulo dos resultados.

Capítulo 1

Descoberta de SNPs e construção de um mapa genético de alta densidade através do ddRADseq de *Piaractus mesopotamicus*

RESUMO

O desenvolvimento da tecnologia de NGS (*next-generation sequencing*) tem permitido a genotipagem de milhares de marcadores genéticos, possibilitando a rápida produção de mapas genéticos para espécies modelo e não-modelo. Em se tratando de espécies não-modelo, o pacu (*Piaractus mesopotamicus*) é uma espécie-alvo interessante por ser considerado um dos peixes com maior produção e valor comercial para a piscicultura nacional. O objetivo do presente estudo foi identificar marcadores moleculares do tipo SNP (polimorfismos de nucleotídeo único) e, com isso, caracterizar um mapa de ligação para esta espécie. Inicialmente foi realizada a extração do DNA de 48 indivíduos de 3 famílias de irmãos completos e, a seguir, utilizando a metodologia *ddRADseq* (*double digest restriction-site associated DNA sequencing*), 3 bibliotecas genômicas foram geradas com a utilização das enzimas de restrição *MluCI* e *SphI*. As amostras foram sequenciadas na plataforma *Hiseq 4000* e obtida uma média de 35.697.700 de sequências em cada família. Na etapa seguinte, a análise bioinformática foi realizada pela aplicação do programa *Stacks* que identificou 18.421 marcadores em 132 indivíduos. O controle de qualidade (QC) foi realizado pelo programa *Plink*, com o qual foi verificado o *Minor Allele Frequency* (MAF) ≥ 0.01 , *individual call rate* > 0.70 e o *SNP call rate* > 0.80 , com a permanência de 11.240 SNPs e 113 indivíduos. Para construção do mapa genético foi utilizado o software *Lepmap* que adicionou os SNPs nos 27 grupos de ligação (LG) desta espécie (2n=54 cromossomos) com um LOD (*minimum logarithm of odds*) de 8.7. O mapa consenso envolveu 6.437 SNPs que abrangem uma distância total de 2.814,92 cM, com intervalo médio 0.44 cM. O mapa de ligação dos machos apresentou 4.649 marcadores, num total de 2.810,41 cM e intervalo médio de 0.60cM, enquanto o mapa feminino foi construído com 4.690 SNPs e consiste de 2.874,89 cM, com um intervalo médio dos marcadores de 0.61 cM. Com a aplicação do programa *Circa*, foi analisada a sintonia existente entre as sequências distribuídas nos 27 LG de *P. mesopotamicus* e nos 25 cromossomos de *Astyanax mexicanus*. Foram identificadas 574 sequências homólogas, sendo que 4 LG possuem poucas RADtags, 22 possuem sintonia direta com 22 grupos correspondentes e apenas o grupo de ligação AM3 de *A. mexicanus* compartilhou as RADtags com os grupos de ligação LG10 e LG20 do pacu. Considera-se, pois, que o mapa genético de alta densidade construído para *P. mesopotamicus* poderá contribuir com programas de melhoramento genético, auxiliar na investigação do processo evolutivo e melhorar o entendimento de aspectos da biologia desta espécie.

Palavras-chave: Genotipagem, Sintonia genômica, Mapa de ligação, peixe Neotropical

1.1 INTRODUÇÃO

Avanços notáveis no sequenciamento do genoma em diferentes organismos ocorreram após o advento da metodologia de NGS (*Next Generation Sequencing*), sendo que técnicas efetivas foram desenvolvidas para obtenção e genotipagem de milhares de marcadores moleculares do tipo SNP (*Single Nucleotide Polymorphisms*) a um custo reduzido (Mardis *et al.*, 2008).

Os marcadores SNPs e como o próprio nome já diz, referem-se ao polimorfismo de nucleotídeos simples ou polimorfismo de nucleotídeo único. Neste caso, ocorre a variação na sequência de DNA que afeta somente uma base. Como características para estudos genéticos, este tipo de marcador possui a vantagem de estar presente e com uma elevada frequência na maioria dos organismos, geralmente estão espalhados uniformemente ao longo do genoma, são de natureza bi-alélica, codominantes, altamente informativos, possui eficiência e automação de genotipagem, além de apresentarem facilidade de isolamento com a aplicação das novas biotecnologias moleculares (Liu *et al.*, 2011; Rasal *et al.*, 2017). Dentre as técnicas para a identificação dos marcadores moleculares, destaca-se o *Double digest RADseq* – ddRADseq (Peterson *et al.*, 2012), esta abordagem utiliza duas enzimas de restrição e possui uma alta capacidade de multiplexação, requer pouco DNA genômico para a construção da biblioteca e apresenta custo reduzido, uma vez que a combinação de muitos indivíduos no *pool* proporciona uma diminuição considerável do valor por amostra, e o DNA fragmentado é ligado a adaptadores que contêm códigos de barras (barcodes) específicos para cada amostra (Peterson *et al.*, 2012); esta técnica é uma variação do RADseq que utiliza apenas uma enzima de restrição.

Em se tratando do *Piaractus mesopotamicus*, na literatura existem 32 SNPs descritos provenientes do transcriptoma, porém torna-se necessário um painel de SNPs mais denso para o desenvolvimento de pesquisas mais amplas e robustas.

A espécie de peixe *Piaractus mesopotamicus*, conhecida popularmente como pacu, está classificada como pertencendo à família Serrasalminidae e é nativa da região Neotropical. De acordo com o IBGE (2019) o pacu é a quarta espécie nativa mais produzida na aquicultura continental nacional e tal posição se deve ao fato de possuir diversas características que são apreciadas pela piscicultura como bom sabor de sua carne, boa resistência a patógenos, rusticidade no manejo, além da alta adaptabilidade ao cultivo em tanques e viveiros (Fernandes *et al.*, 2000; Abimorad & Carneiro, 2004; Jomori *et al.*, 2008). A produção do pacu em cativeiro também se apresenta como uma alternativa para diminuir o impacto decorrente da sobrepesca

para fins comerciais e esportivos (MPA, 2013; Catella *et al.*, 2017), bem como pela diminuição das populações naturais devido à crescente fragmentação de seu habitat causada pela construção de barragens (Agostinho *et al.*, 2005).

Nos últimos anos, a identificação de uma grande quantidade de marcadores moleculares do tipo SNPs têm permitido o desenvolvimento de metodologias mais robustas, assim como o desenvolvimento de mapas de ligação ou mapas genéticos mais densos em diferentes espécies. Mapas genéticos são conjuntos ordenados de marcadores genéticos agrupados em *clusters* ou em grupos de ligação no genoma de uma espécie. Sua construção ocorre a partir de estimativas da frequência de recombinação entre pares de marcadores moleculares. Portanto, a existência de um grande número de marcadores genéticos disponíveis é um pré-requisito fundamental para a aplicação desta metodologia (Xu *et al.*, 2008), sendo que, nos últimos anos este marcador tem sido cada vez mais utilizado. Os mapas genéticos têm ampliado o conhecimento sobre algumas espécies de peixes, permitindo um maior conhecimento sobre a organização do genoma, análises genômicas e genéticas (Maroso *et al.*, 2018) e identificação de QTLs (Zhu *et al.*, 2019a), além de contribuir para estudos de sintenia ou mapeamentos comparativos (Zhu *et al.*, 2019b).

Os mapas de ligação possibilitam a realização de inúmeras pesquisas, pois através das informações fornecidas por ele, como o número de grupos de ligação e a posição dos marcadores moleculares, é possível auxiliar a montagem e análise de genomas, realizar a identificação de genes, o mapeamento de QTLs, e a comparação entre mapas demonstram as similaridades e diferenças entre as espécies e dessa forma, oferece uma melhor compreensão da evolução de cromossomos e genomas (Garcia *et al.*, 2006; Paterson *et al.*, 2000). E assim como foi realizado nesse trabalho, os mapas genéticos auxiliam na identificação de sentenias correspondentes dos loci nos grupos homólogos de diferentes organismos.

Contudo, ainda são escassos os estudos relacionados ao mapeamento genético de peixes do Brasil, principalmente quando se trata de espécies nativas, entre as quais há apenas um mapa genético disponível que é referente a espécie *Colossoma macropomum* – tambaqui (da Silva Nunes *et al.*, 2017). Neste contexto, considerando o cenário ambiental promissor e a expansão da piscicultura brasileira, a identificação em larga escala de SNPs e a construção de um mapa genético realizada neste trabalho para a espécie *P. mesopotamicus*, pode aprimorar estudos genéticos e também auxiliar programas de melhoramento genético, através da desenvolvimento de ferramentas genômicas que poderão ser aplicadas na seleção genômica e seleção assistida por marcadores.

1.2 MATERIAIS E MÉTODOS

1.2.1 Formação dos grupos genéticos, instalações e manejo

Este estudo foi realizado em conformidade com as recomendações do Conselho Nacional de Controle de Experimentação Animal (CONCEA) (Ministério Brasileiro de Ciência, Tecnologia e Inovação) e foi aprovado pelo Comitê de Ética em Uso Animal (CEUA), protocolo no. 22.255 / 15. O presente estudo foi realizado sob autorização no. 33435-1, emitido pelo Instituto Chico Mendes de Conservação da Biodiversidade, Ministério do Meio Ambiente do Brasil (ICMBio).

Inicialmente, foi necessário formar uma população base através da montagem de diferentes famílias para o desenvolvimento de grupos genéticos, e dessa forma dar início a pesquisa.

A reprodução de *P. mesopotamicus* para a formação das famílias foi realizada em novembro de 2015 com a utilização de 3 famílias, sendo os machos provenientes do plantel de matrizes da piscicultura do Centro de Aquicultura da Unesp - CAUNESP, Jaboticabal/SP e as fêmeas da Piscicultura EMA, Cosmorama, SP. Estes animais foram previamente analisados para testar sua pureza genética através da genotipagem de SNPs por Multiplex-PCR (Reação em cadeia da polimerase com uso de primers espécie-específicos), descritos em Hashimoto *et al.* (2011). A indução da desova foi realizada com a utilização de extrato bruto de hipófise (EBH) de carpa dissolvido em solução salina (0,90% NaCl) aplicado em região intramuscular dos exemplares em duas doses nas fêmeas, com intervalo de 12 h (1ª dose de 0,5 mg hipófise kg⁻¹ de peixe; 2ª dose de 5,0 mg de hipófise kg⁻¹ de peixe). No momento da segunda dose foi realizada a aplicação de 5 ml de Ciosin® (0.25 mg/mL cloprostenol, Schering-Plough Saúde Animal Ind. Com. Ltda. Campinas, SP, BR), conforme protocolo descrito por Ceccarelli *et al.* (2005). Nos machos foi utilizada a aplicação de dose única, coincidindo com a aplicação da 2ª das fêmeas, sendo equivalente a 1,5 mg kg⁻¹ de EBH.

A extrusão dos gametas foi realizada com aproximadamente 240 h/graú, e a incubação das larvas foi realizada em incubadoras de 15 litros durante 1 mês. Em seguida, as larvas de cada família (família1, família2 e família3) foram separadas em tanques redes em caixas de 1.000 l. Depois de 7 meses, quando os indivíduos estavam com 6 meses nos tanques, 42 animais da prole de cada família foram coletados, anestesiados com benzocaína (100mg/L), marcados com PITtags (*passive integrated transponder tag*, modelo *full-duplex* FDX-B, 134.2-kHz) e amostras de sangue de cada exemplar foram coletadas e fixadas em etanol absoluto para

posterior extração de DNA. Estes indivíduos foram mantidos por mais oito meses na mesma caixa sob as mesmas condições ambientais e de manejo no Laboratório de Genética em Aquicultura e Conservação – LaGeAC, CAUNESP. Após esse período foram transportados para um tanque escavado de aproximadamente 200 m² existente nas instalações da piscicultura do CAUNESP.

No início do período experimental, enquanto a prole estava alocada nas instalações do Laboratório de Genética em Aquicultura e Conservação – LaGeAC, CAUNESP, a média da temperatura constatada foi de 28.30°C ± 1.55°C, do oxigênio dissolvido foi de 6.50 ± 0.30 mg/L e o pH de 6.70 ± 0.08. Nos tanques escavados, o valor médio da temperatura foi de 24.45°C ± 2.15°C, do oxigênio dissolvido foi de 8.20 mg/L ± 1.05 e do pH foi de 8.90 ± 0.50. Todos os parâmetros foram avaliados utilizando a sonda *Multi Water Quality Checker U 50 serie Horiba*.

Em relação a alimentação, da eclosão dos ovos até um período de 20 dias, os peixes foram alimentados exclusivamente com náuplios de *Artemia* (BioArtemia Ltda, RN, Brasil) quatro vezes ao dia e após este período, também foi adicionada ração em pó à alimentação. Os náuplios de *Artemia* foram retirados da alimentação quando os peixes atingiram dois meses, enquanto a ração em pó foi retirada aos poucos, de acordo com o crescimento dos peixes, que então se tornavam capazes de se alimentarem com ração extrusada, cujo diâmetro dos grânulos também aumentava de acordo com o crescimento dos peixes.

1.2.2 Construção de bibliotecas ddRADseq

Para prospectar e caracterizar os SNPs, a construção das bibliotecas foi realizada através do ddRADseq de 3 famílias de pacu, com base no protocolo originalmente descrito por Peterson *et al.* (2012). Inicialmente foi extraído o DNA genômico de 42 indivíduos da prole de cada família e de seus parentais utilizando o Kit *PureLink™ Genomic DNA*. O DNA genômico dos parentais foi identificado em triplicatas. Todos os indivíduos receberam um *barcode* para sua identificação no *pool*, totalizando 48 marcações individuais de cinco nucleotídeos para cada família.

As quantidades de DNA genômico para as análises foram determinadas no equipamento Qubit® 3.0 (*Thermo Scientific™*) e padronizadas para 10ng/ul. A análise de pureza das amostras de ácido nucleico foi obtida por espectrofotometria no equipamento *NanoDrop™ One* (*Thermo Scientific™*), onde foi obtido um valor de razão de absorbância (A) A260/A280 entre de 1.80 e 2.00; a integridade do DNA das amostras também foi visualizada utilizando gel de agarose. Para a etapa de digestão do DNA foi obtido um volume final de reação de 6.00 µL, em

que foram utilizados 2.50 µL de DNA (totalizando 25 ng), digeridos com 0.10µL (10U/ml) da enzima de restrição *MluCI* (5'...AATT...3' - corte frequente) e 0.10µL (20U/ml) da enzima *SphI* (5'...GCATGC...3' - corte raro), 0.60 µL do *CutSmart 10X Buffer* e 2.80 µL de água. As condições estabelecidas para o tempo de digestão foram de 37°C durante 60 minutos. Essas enzimas criam pontas coesivas em ambas as extremidades das moléculas de DNA digeridas que permitem a ligação aos adaptadores.

Com base em um genoma de peixe já sequenciado, a espécie piranha *Pygocentrus nattereri*, simulações computacionais foram realizadas para avaliar a frequência do corte de cada enzima de restrição utilizando o programa *Nebcutter v2.0* (Vincze *et al.*, 2003). Foi verificado que a distância média entre sítios enzimáticos da enzima *SphI* era de 6.500 pb, enquanto para *MluCI* era de 250 pb, o que permitiu ajustar as molaridades de cada adaptador específico para cada enzima, de acordo com o cálculo fornecido por Peterson *et al.* (2012). Como a enzima *MluCI* gera mais fragmentos na digestão, foi necessário inserir mais adaptadores desta enzima na reação. A reação de ligação foi realizada em um volume final de 10.00 µl, com 0.20 µl da enzima T4 ligase, 1 µl do tampão 10X *T4 DNA ligase Reaction Buffer*, 1.00 µl do adaptador P1 4.04 µM (*MluCI*) e 1.00 µl do adaptador P2 0.07 µM (*SphI*), sendo que cada indivíduo teve um *barcode* diferente no adaptador P2. Este processo de ligação ocorreu com as amostras incubadas a 23°C durante 90 min e a 65°C por 10 min.

Na etapa seguinte foi realizada uma PCR teste para verificar a efetividade da digestão e da ligação dos adaptadores. A PCR foi realizada em um volume final de 12.50 µL contendo 7.90 µL de H₂O, 1.25 µL de 10X *PCR Buffer*, 0.50 µL do *primer* P1 (10.00 µM), 0.50 µL do *primer* P2 (10.00 µM), 0.50 µL dNTPs (10 mM), 0.75 µl MgCl₂ (25 mM), 0.10 µl *Platinum® Taq DNA Polymerase* e 1.00 µL de DNA digerido. As condições estabelecidas para a PCR determinaram um ciclo inicial a 94°C por 2 minutos, seguido por 20 ciclos de 94°C por 30 segundos, 60°C por 30 segundos e 72°C por 45 segundos. As amplificações foram checadas via eletroforese em gel de agarose. O produto excedente da reação de PCR foi descartado, uma vez que não seria mais utilizado.

Para remover os fragmentos curtos resultante da digestão do DNA genômico e da ligação dos adaptadores foi usado o *Kit Agencourt AMPure XP cleanup step* (Beckman Coulter, DE), seguindo o protocolo proposto pelo fabricante. As *beads* magnéticas, além de serem usadas para remover adaptadores excedentes e fragmentos digeridos não ligados, também foram usadas para concentrar o DNA e combinar o *pool* das amostras em um único tubo, deixando-o num volume ideal (50.00 µL) para ser usado na próxima etapa da construção da

biblioteca. O *pool* foi, então, designado para a seleção dos fragmentos de aproximadamente 350 pb (300 pb de DNA + 50 pb dos adaptadores e *barcodes*) através do *E-Gel® Size Select 2,00% Agarose Gel (Invitrogen)* por eletroforese, sendo que este gel de agarose possui uma cuba específica para a eletroforese e o *E-Gel®* é fabricado pela própria empresa. O gel possui espaço para oito amostras (8 colunas), cada coluna possui dois poços, formando poços de aplicação e de extração. O *pool* é aplicado no primeiro poço e durante a eletroforese a corrente elétrica transporta os fragmentos em direção ao segundo. A corrida eletroforética é monitorada e quando o fragmento de interesse entra na linha de extração, a corrida é interrompida e este é recuperado com o auxílio de uma micropipeta. Neste experimento, o *pool* de cada família foi depositado em poços distintos, porém em um mesmo *E-Gel®* para uma maior padronização entre as três famílias.

Em seguida a este procedimento foi realizada a etapa de enriquecimento via amplificação por PCR com a *Phusion® High-Fidelity PCR Kit*. Esta etapa é necessária para incorporar a sequência do *flowcell Illumina* e as regiões das sequências dos *primers* da *Illumina*. Para aumentar o número dos fragmentos de DNA, foram realizadas 20 reações independentes de PCR. As condições de PCR ocorreram em volume final de 20.00 µL, sendo 13.60 µL de H₂O, 0.40 µL dNTPs 10 mM, 4.00 µL 5X *Phusion HF Buffer*, 2.50 µL de primer P1 5.00 µM, 2,5 µL de *Primer* P2 5.00 µM, 0.12 µL *Phusion DNA Polymerase* e 1.00 µL de DNA digerido. As reações ocorreram em uma etapa inicial de 68°C por 1 min; seguida por 18 ciclos a 93°C por 10 s, 52°C por 35 s, 68°C por 1,5 min; e uma etapa final de 68°C por 7 min. Após o término desta etapa as amostras de PCR foram agrupadas, totalizando um volume final de 400.00 µL e foram novamente purificadas por *AMPure XP beads* (Beckman Coulter, DE), de acordo com o protocolo proposto pelo fabricante. O *pool* das sequências amplificadas foi analisado em eletroforese de gel de agarose.

A concentração na biblioteca final foi verificada por fluorometria no equipamento *Qubit® 3.0 (Thermo Scientific™)*. Em seguida, também foi realizada a análise por eletroforese no equipamento Agilent 2100 Bioanalyzer (Agilent Technologies, USA). As amostras foram enviadas para serem sequenciadas na plataforma Hiseq 4000, capaz de sequenciar fragmentos *paired end* de até 150 pb e gerar até 1500Gb de dados.

1.2.3 Identificação e filtro dos SNPs

O *software* utilizado para analisar as leituras de sequências brutas obtidas do sequenciamento *Illumina* foi o *Stacks* v. 2.1 (Catchen *et al.*, 2011; Catchen *et al.*, 2013). Este

programa foi desenvolvido para analisar *reads* geradas por sequenciamento de nova geração (NGS) com dados de leitura curta para organismos modelos e não-modelos (Davey *et al.*, 2013). Além de analisar as sequências brutas, o *Stacks* fornece os dados em diversos formatos para construção de mapas genéticos, genômica populacional e filogeografia, além de ter capacidade de processamento de *RADtags*, identificação e filtragem dos SNPs e exclusão dos erros de sequenciamento. Dentro do programa *Stacks* existem diversos subprogramas que devem ser executados para a análise dos dados. A análise é iniciada com o programa *Process_radtags* (*Stacks*), no qual os barcodes foram atribuídos aos indivíduos, os adaptadores foram retirados, todas as leituras brutas foram trimadas para 129 pb e foi realizado o controle de qualidade através da retirada de sequências de baixa qualidade. Como as sequências são *paired-end* elas são separadas e apenas uma é utilizada nesta etapa.

Em seguida foi usado o *software Ustacks*, com a função de construir os *loci* para cada indivíduo, tendo sido usada uma profundidade mínima de *RADtags* para construir um alelo putativo (-m 3) e considerada a máxima incompatibilidade de nucleotídeos permitida entre os alelos de uma amostra (-M 3). A montagem de um catálogo com os dados de todos os parentais foi realizada através do programa *cstacks* e, desta forma foi possível criar um conjunto de todos os possíveis alelos para todas as famílias. O número máximo de incompatibilidade permitida entre as *Rad-Tags* dos *loci* do catálogo foi três (-n 3). Através do programa *sstacks* foram criados os *loci* putativos, sendo os resultados obtidos pelo programa *ustacks* (progênie) comparado ao catálogo produzido pelo programa *cstacks* (parentais) para identificar os alelos em comum.

O programa *tsv2bam* utilizado permitiu a transposição dos dados que estavam orientados por amostra, para sua orientação por *loci*. E como as seqüências do trabalho em questão foram separadas no início das análises, as duas seqüências (*paired-end*) podem ser associadas nesta etapa. No sexto estágio, o programa *gstacks* foi executado para identificar locais variantes na metapopulação e, dessa forma, genotipar cada amostra. Para finalizar a execução do *Stacks* foi utilizado o programa *populations*, que cria *outputs* em diversos formatos. Neste trabalho o arquivo foi exportado no formato --vcf para posterior análise no *software Plink* v. 1.07 (Purcell, 2009). O *populations* também possui as opções de realizar alguns filtros de qualidade dos SNPs e calcular várias estatísticas de genética de populações.

O *Software Plink* foi utilizado para realizar o controle de qualidade nos SNPs identificados pelo *software Stacks*. Neste processo foram utilizados filtros para exclusão de SNPs que não fossem genotipados em no mínimo 80,00% das amostras; foram excluídos

indivíduos que não tivessem ao menos 70,00% dos SNPs genotipados, bem como aqueles que apresentassem frequência alélica mínima (MAF) inferior a 5,00%. A aplicação destes filtros é indicada por Ziegler *et al.*, (2008) e se baseia na exclusão de possíveis falhas ocorridas durante o processo de genotipagem. Alguns filtros também foram realizados com a aplicação dos softwares *LepMAP* v. 2 e v. 3 (Rastas *et al.*, 2013).

1.2.4 Construção do Mapa Genético

Para a construção do mapa genético foi usado o programa *LepMAP* v. 2 e v. 3 (Rastas *et al.*, 2013). Inicialmente, o módulo *ParentCall2* foi utilizado para “chamar” o genótipo dos parentais, seguido pela etapa “*SeparateChromosomes*” que atribuiu os marcadores aos grupos de ligação utilizando o valor *LOD score* de 8.7, de modo que o número observado de SNPs nos grupos de ligação fosse adequado ao número de cromossomos haplóides para o pacu, 27 cromossomos, além de um número mínimo de 10 marcadores por grupo de ligação. Nesta etapa também foi utilizado o parâmetro *InformativeMask*=13 para o mapa masculino, *InformativeMask*=23 para o mapa feminino e *InformativeMask*=123 para o mapa consenso, pois estes parâmetros são utilizados para identificar marcadores informativos apenas para os machos, apenas para as fêmeas e para ambos os sexos, respectivamente.

A análise seguinte foi realizada com o módulo “*JoinSingles*”, responsável por adicionar os SNPs não atribuídos originalmente a qualquer LG na etapa anterior através de uma pontuação mínima de *LOD score* de 5. Usando a opção “*dataTolerance*=0.01” na etapa *Filtering2* num procedimento seguinte, foi removido qualquer marcador que mostrasse evidência de distorção de segregação Mendeliana (teste χ^2 de distorção de segregação $P < 0.01$).

Finalmente, o módulo “*OrderMarkers*” ordenou os SNPs em cada grupo de ligação. Inúmeras repetições foram realizadas para melhorar os parâmetros de recombinação e diminuir os erros e a ordem mais apropriada foi determinada pelo melhor valor de *likelihood*. Além dos mapas separados por sexo, a opção “*sexAveraged*=1” também foi incluída neste módulo para construir o mapa consenso. Grupos de ligação foram desenhados usando o pacote *LinkageMapView* (Ouellette *et al.*, 2017) do programa computacional *RStudio* v. 1.1.463.

1.2.5 Mapeamento Comparativo

O mapeamento comparativo foi realizado com o download do genoma completo do *Astyanax mexicanus* ([https://www.ncbi.nlm.nih.gov/genome/?term=txid7994\[orgn\]](https://www.ncbi.nlm.nih.gov/genome/?term=txid7994[orgn])) no site do NCBI. As sequências de *P. mesopotamicus* utilizadas no mapa de ligação foram blastadas

contra o genoma do *Astyanax* usando o programa *BLASTn* (“*Basic Local Alignment Search Tool*”) com *e-value* *cutoff* $1e-10$. Para sequências que tiveram vários *blast hits*, apenas aqueles com o menor *e-value* foram utilizadas na análise comparativa. Por fim, foi utilizado o *software Circa 1.2.1* (<http://omgenomics.com/circa/>) para visualizar a sintonia genômica.

1.3 RESULTADOS

1.3.1 Construção de bibliotecas ddRADseq

Com o uso do ddRADseq foram obtidos dados de sequências *single-end*, 35.394.337 *reads* e 10.60 Gbp (Giga bases) para a Família 1; 37.640.043 *reads* e 11.3 Gbp para a Família 2; e 34.058.722 *reads* e 10.2 Gbp para a Família 3. Em relação à qualidade das sequências, foi utilizado o *Phred quality score* (*Qscore*) que é um índice de qualidade de identificação das nucleobases geradas pelo sequenciamento automatizado de DNA (Ewing *et al.*, 1998). Neste trabalho, mais de 91,00% das informações dos grupos genéticos possuem os dados acima do índice Q30, ou seja, a chance de uma base estar correta é 99,90%, em aproximadamente 91,00% dos dados brutos gerados.

1.3.2 Identificação e filtro de SNPs

Durante o processamento das leituras brutas no programa *Process_radtags*, as sequências foram trimadas para 129 pb para a remoção das extremidades com baixa qualidade e foi realizado o primeiro filtro para cada família, mostrado na Tabela1.

Tabela1. Filtros realizados pelo Stacks nas sequencias brutas

	<i>Reads Paired-end Totais</i>	<i>Barcodes Ambiguos</i>	Baixa Qualidade	<i>RAD-Tags Ambiguos</i>	<i>Reads Paired-end Restantes</i>
Família1	70.788.674	179.214	25.714	522.027	70.061.719
Família2	75.280.086	235.490	393.310	642.997	74.008.289
Família3	68.117.444	180.946	23.076	501.144	67.412.278

Os *loci* foram construídos para cada indivíduo e em seguida foi montado um catálogo das amostras dos parentais; assim, foi possível criar um conjunto de todos os alelos. O catálogo formado pelos parentais e pela prole consistiu em 195.168 *loci*, no qual 114 *loci* não tiveram

ou não apresentaram sequências *paired-end* (0,10%); 4.999 *loci* tinham *reads paired-end*, mas não puderam ser montadas em um *contig* (2,60%); restaram 190.055 *loci* (97,40%) e através deles foi possível montar um *contigs paired-end*, sendo o tamanho médio de 164.5 pb. No total de 211.482.286 sequências *paired-end*, 169.788.586 foram excluídas nas primeiras análises e apenas 41.693.700 foram alinhados com sucesso, com média de 219.4 *reads* por *loci*. Na última etapa dos filtros e identificação dos genótipos, do total de 190.055 foram removidos 180.269 dos *loci* identificados, restando 9.786, com tamanho médio de 259.19pb. Desses *loci* foram identificados 18.421 sites variantes em 132 indivíduos, sendo 126 da prole e 6 dos parentais.

Os marcadores também passaram por um filtro de qualidade usando o *software Plink*, pelo qual foram excluídos SNPs não genotipados em no mínimo 80,00% das amostras; o programa também excluiu indivíduos que não tiveram pelo menos 70,00% dos SNPs genotipados e que apresentaram MAF igual ou inferior a 0.05. Com a aplicação desses filtros o número de marcadores foi reduzido para 11.240 SNPs bi-alélicos, com uma profundidade média de sequenciamento de 14x. Foram excluídos 19 indivíduos da análise, restando 113 indivíduos para a análise final.

1.3.3 Construção do Mapa Genético

O mapa genético construído possui alta densidade de SNPs, embora 43,00% dos marcadores (4.803 SNPs) tenham sido descartados. Ainda assim foi utilizado um total de 6.437 SNPs distribuídos em 27 grupos de ligação, identificados em indivíduos de três famílias F1 (Figura1), o que é consistente com o número de cromossomos haplóides de *Piaractus mesopotamicus* ($2n=54$).

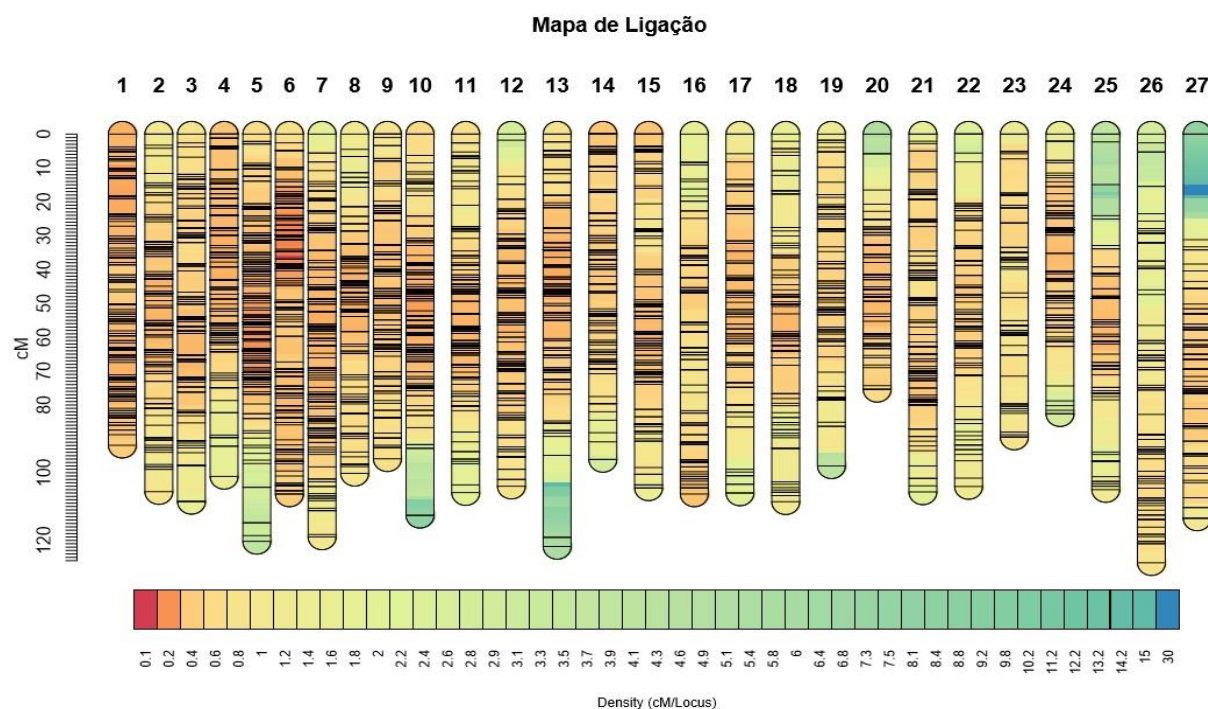


Figura1. Mapa de ligação do *P. mesopotamicus* e densidade dos SNPs nos cromossomos (cM/locus).

A descoberta dos marcadores foi realizada utilizando a pipeline do *software LepMAP* v. 2 e v. 3. Com a aplicação do módulo *SeparateChromosomes* foi adicionado um total de 2.320 marcadores aos 27 grupos de ligação, sendo que o maior valor foi verificado no LG1 com 214 e o menor no LG27 com 28. No módulo *Joingsingle* foram adicionados mais 4.198 marcadores, totalizando 6.518 marcadores distribuídos ao longo dos 27 grupos de ligação, com um range de 90 (LG27) a 492 (LG3). Com a execução do módulo *Filtering* foram eliminados 81 marcadores que não estavam segregando de acordo com a herança mendeliana ($P > 0.01$). E por último foi executado o módulo *Ordermarkers*. Todos os marcadores informativos foram utilizados no mapa consenso, resultando em 6.437 SNPs que abrangem uma distância total de 2.814,92 cM. Foi verificado que o menor grupo de ligação é o LG20 com 75.38 cM e o maior é o LG2 com 126.62 cM, com um intervalo médio entre os marcadores de 0.44 cM (Tabela2).

O mapa de ligação dos machos consistiu em 4.649 marcadores com tamanho variando de 57.93 cM (LG 25) a 136.76 cM (LG 5), um total de 2.810,41 cM e intervalo médio de marcadores de 0.60 (Tabela2). Enquanto o mapa feminino foi construído com 4.690 SNPs e consistia em um total de 2.874,89 cM, com uma variação de 68.68 cM (LG 20) a 144.92 cM (LG8) e um intervalo médio de marcadores de 0.61 (Tabela2).

Tabela2. Resumo do mapa de ligação consenso, das fêmeas e dos machos de *Piaractus mesopotamicus*.

	MACHO			FÊMEA			CONSENSO		
	Distância genética (cM)	Número de marcadores	Intervalo médio entre os SNP (cM)	Distância genética (cM)	Número de marcadores	Intervalo médio entre os SNP (cM)	Distância genética (cM)	Número de marcadores	Intervalo médio entre os SNP (cM)
LG1	124.98	289	0.43	110.26	297	0.37	91.88	383	0.24
LG2	112.51	209	0.54	101.11	221	0.46	105.60	250	0.42
LG3	128.45	177	0.73	114.32	179	0.64	108.50	244	0.44
LG4	95.46	191	0.50	128.33	225	0.57	101.13	300	0.34
LG5	136.76	273	0.50	122.86	247	0.50	120.35	355	0.34
LG6	136.8	330	0.41	134.30	310	0.43	106.32	442	0.24
LG7	124.35	272	0.46	114.83	230	0.50	119.02	341	0.35
LG8	81.45	175	0.47	144.92	177	0.82	100.17	243	0.41
LG9	97.64	173	0.56	128.58	191	0.67	95.91	235	0.41
LG10	92.06	195	0.47	116.22	201	0.58	112.61	283	0.40
LG11	120.45	174	0.69	114.33	164	0.70	105.88	225	0.47
LG12	117.66	183	0.64	90.36	190	0.48	103.96	254	0.41
LG13	111.03	189	0.59	109.53	196	0.56	121.77	280	0.43
LG14	97.98	130	0.75	92.27	154	0.60	96.13	209	0.46
LG15	115.56	207	0.56	110.04	161	0.68	104.54	249	0.42
LG16	85.27	131	0.65	88.6	119	0.74	106.29	176	0.60
LG17	104.48	155	0.67	102.63	182	0.56	106.02	248	0.43
LG18	117.48	149	0.79	103.04	143	0.72	108.60	200	0.54
LG19	110.96	108	1.03	109.77	132	0.83	97.96	163	0.60
LG20	70.41	125	0.56	68.68	107	0.64	75.38	171	0.44
LG21	85.1	148	0.58	100.44	165	0.61	105.69	219	0.48
LG22	89.5	123	0.73	104.80	131	0.80	104.09	168	0.62
LG23	78.56	96	0.82	84.82	117	0.72	89.46	135	0.66
LG24	87.44	126	0.69	84.90	122	0.70	82.61	181	0.46
LG25	57.93	92	0.63	70.73	108	0.65	105.02	167	0.63
LG26	120.63	98	1.23	119.30	107	1.11	126.62	128	0.99
LG27	109.51	131	0.84	104.92	114	0.92	113.41	188	0.60
Total	2810.41	4649	0.60	2874.89	4690	0.61	2814.92	6437	0.44

1.3.4. Mapeamento Comparativo

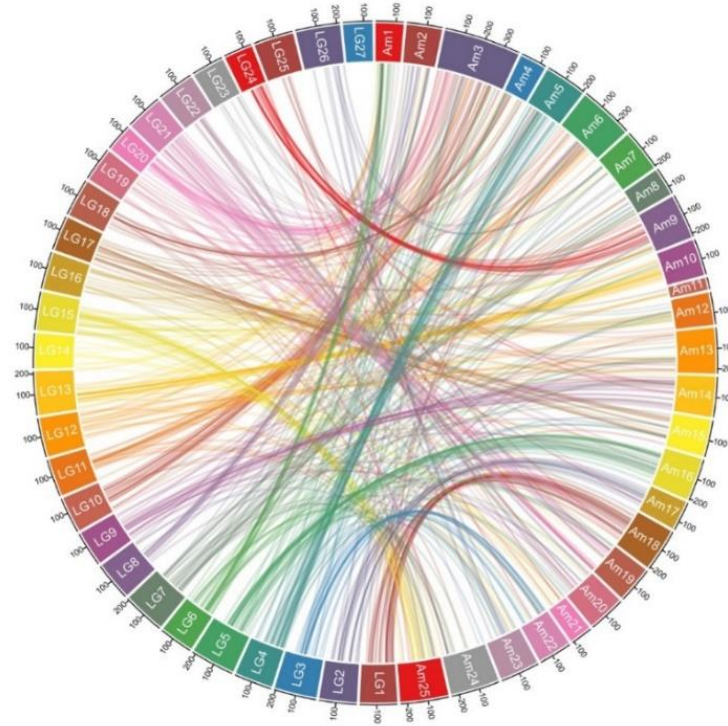


Figura2. Análise sintênica entre grupos de ligação do *Piaractus mesopotamicus* contra os cromossomos do *Astyanax mexicanus*. Cada linha representa uma conexão entre a posição de um *RADtags* do pacu e uma sequência homóloga em *A. mexicanus*. (escala em Mb).

A análise comparativa entre os genomas de *P. mesopotamicus* e de *A. mexicanus* identificou um alto nível de sintenia na maioria dos grupos de ligação. Dos 6.437 SNPs utilizados na construção do mapa genético, 574 marcadores continham sequências que foram alinhados com sucesso no genoma do *A. mexicanus* (Figura2). A análise mostrou ainda a existência de níveis variáveis de conservação genômica nos grupos de ligação.

Entre os 27 grupos de ligação analisados, 22 compartilharam blocos sintenicos, sendo que 17 (LGs 1,2,3,4,5,8,9,11,13,15,16,17,18,21,22,24,26) apresentam mais de 50,00% de alinhamento com um cromossomo correspondente, enquanto 6 grupos de ligação (6,7,10,12,19,20) demonstram entre 30,00% e 50,00% de alinhamento. O LG22 foi o grupo de ligação que apresentou maior sintenia com um mesmo cromossomo, compartilhando 80,00% de suas *RADtags* com o AM20. Por outro lado, 4 grupos de ligação (14,23,25,27) apresentam poucas *RADtags* e estas mostraram correspondência aleatória com os cromossomos. O cromossomo AM3 de *A. mexicanus* compartilhou a maior parte das suas sequências com os grupos de ligação LG10 e LG20 do pacu, mostrando um aparente rearranjo cromossômico.

1.4 DISCUSSÃO

1.4.1 Construção do mapa genético

Neste trabalho foi desenvolvido o primeiro mapa de ligação de alta densidade para a espécie *Piaractus mesopotamicus* que, sendo uma espécie diploide com $2n=54$ cromossomos, seu mapa de ligação consiste de 27 grupos de ligação.

Para o desenvolvimento de um mapa de ligação de alta consistência é necessário utilizar uma grande quantidade de marcadores moleculares. Neste trabalho foram identificados inicialmente 11.240 SNPs e, devido à frequência de recombinação calculada entre os *loci* foi possível utilizar 56,17% dos marcadores para a construção do mapa, resultando na aplicação de 6.437 SNPs e, ainda assim podendo ser considerado um mapa de alta resolução.

A frequência de recombinação foi definida de acordo com o *score LOD* de 8.7 que foi muito semelhante àquela encontrada no trabalho realizado com a espécie *Colossoma macropomum* – tambaqui (da Silva Nunes *et al.*, 2017) que utilizou o valor do *score LOD* de 8.0. Esta espécie é nativa do Brasil e também é classificada na mesma família Serrasalminae. Para obtenção do mapa de ligação relatado nesse trabalho foram identificados 7.192 SNPs, num total de 2.810,9 cM, com um intervalo médio de 0.39 cM em 118 indivíduos. Estes resultados foram muito semelhantes aos aqui descritos para o *Piaractus mesopotamicus*, com uma média total do mapa consenso de 6.437 SNPs, um tamanho de 2.814,9 cM e um intervalo médio entre os marcadores de 0.44 cM em 113 indivíduos. Apesar de ter sido utilizada uma tecnologia distinta, há muitas semelhanças nos dados obtidos para as duas espécies, este fato pode ter ocorrido devido a proximidade taxonômica existente entre elas.

Em se tratando dos mapas de ligação distintos para machos e fêmeas, as quantidades de marcadores informativos para ambos os sexos foram de 72,50% do total de SNPs para o pacu, contrastando com 47,00% e 33,00% para fêmeas e machos, respectivamente, para o tambaqui. Esta superioridade com relação aos pontos nos mapas de ligação também foi identificada em vários trabalhos (Wang *et al.*, 2015; Brown *et al.*, 2016; Sun *et al.*, 2017). Além disso, no mapa unificado também foi possível visualizar melhorias na quantidade de marcadores obtidos, densidade e resolução do mapa quando realizada a comparação com algumas espécies de peixes nas quais também foi utilizado o RADseq e suas variações nas análises, incluindo carpa cabeça grande (*bighead carp*) com 3.121 SNPs, 130 SNPs/LG e 0.75cM (Fu *et al.*, 2016), o peixe mandarim (*mandarin fish*) que apresentou 3.283 SNPs, 137 SNPs/LG e 0.61cM (Sun *et al.*,

2017) e um híbrido de ciclídeo do gênero *Pundamilia* com 1.597 SNPs, 72 SNPs/LG e 1.01cM (Feulner *et al.*, 2018).

Entretanto, quando o mapa genético obtido para o pacu é comparado com o mapa de espécies como salmão do atlântico (Tsai *et al.*, 2016), catfish (Zeng *et al.*, 2017), tilápia (Joshi *et al.*, 2018), espécies que já passaram por rigorosos programas de melhoramento genético e possuem muitas informações genéticas e genômicas, seus mapas de ligação possuem uma grande superioridade em relação ao número de indivíduos analisados e são formados por milhares de marcadores SNPs.

O mapeamento comparativo é uma maneira eficiente de investigação de rearranjos cromossômicos (Kasahara *et al.*, 2007), e dessa forma entender as alterações no genoma de espécies filogeneticamente próximas, localização de QTLs (Araneda *et al.*, 2012), identificação de genes (Guyomard *et al.*, 2012) e inferências sobre função genética (He *et al.*, 2015). Neste estudo, o mapeamento comparativo foi realizado entre a sequência dos marcadores que construíram o mapa genético do pacu (*Piaractus mesopotamicus*) contra o genoma montado de *Astyanax mexicanus*, tendo por base a homologia de sequências. Um total de 574 sequências homólogas entre estas espécies foram identificadas e mostraram-se distribuídos por todos os cromossomos.

Entre os 27 grupos de ligação identificados, 22 grupos compartilharam blocos sintênicos. Porém LG 14, 23 e 25 apresentaram poucas *RADtags*, não sendo identificada a existência de sintênia entre elas. Apenas o grupo de ligação AM3 de *A. mexicanus* compartilha a maior parte das suas *RADtags* com os grupos de ligação LG10 e LG20 do pacu. Essa sintênia pode ter ocorrido devido a um rearranjo cromossômico derivado de um ancestral comum, porém mais pesquisas devem ser realizadas para afirmar este fato.

A realização do presente trabalho possibilitou o desenvolvimento de um denso painel de SNPs e o primeiro mapa de ligação para o pacu (*P. mesopotamicus*) com a utilização de um amplo conjunto de dados em escala genômica. Estas ferramentas fornecem uma forma simples e rápida para a análise de dados, disponibilizando uma ferramenta de importante aplicação e suporte para estudos de melhoramento genético, arquitetura e biologia do genoma (Fernández *et al.*, 2014, Yáñez *et al.*, 2015), bem como de aspectos evolutivos e de conservação deste grupo de organismos (Bernatchez, 2016).

1.5 REFERÊNCIAS

- Abimorad EG & DJ Carneiro. 2004. Métodos de coleta de fezes e determinação dos coeficientes de digestibilidade da fração protéica e da energia de alimentos para o pacu, *Piaractus mesopotamicus* (Holmberg, 1887). *Revista Brasileira de Zootecnia*. 33: 1101-1109.
- Agostinho AA, Thomaz SM, Gomes LC. 2005. Conservação da Biodiversidade em águas continentais do Brasil. *Megadiversidade*. 1: 70-78.
- Araneda C, Díaz NF, Gomez G, López ME, Iturra P. 2012. Comparative mapping reveals quantitative trait loci that affect spawning time in coho salmon (*Oncorhynchus kisutch*). *Genetics and molecular biology*, 35(2), 515-521.
- Bernatchez L. 2016. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, 89(6), 2519-2556.
- Brown JK, Taggart JB, Bekaert M, Wehner S, Palaiokostas C, Setiawan AN, ... Penman, D. J. 2016. Mapping the sex determination locus in the hāpuku (*Polyprion oxygeneios*) using ddRAD sequencing. *BMC genomics*, 17(1), 448.
- Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. *Genes, Genomes, Genetics*, 1, 171–182.
- Catella AC, FL de R. Campos, S. P. Albuquerque. 2017. “Sistema de Controle da Pesca de Mato Grosso do Sul (SCPESCA/MS–23–2016).” *Boletim de Pesquisa e Desenvolvimento Embrapa Pantanal, Corumbá, Brazil*.
- Ceccarelli PS, et al. Piracanjuba, ***Brycon orbignyanus*** (Valenciennes, 1849). In: Baldisserotto, B.; Gomes, L.C. **Espécies nativas para piscicultura no Brasil**. Santa Maria: UFSM, 2005. p.121-147.
- Criscuolo-Urbini E, Kuradomi RY, Urbini EC, Batlouni SR. 2012. The administration of exogenous prostaglandin may improve ovulation in pacu (*Piaractus mesopotamicus*). *Theriogenology*. 78: 2087-2094.
- Da Silva Nunes JDR, Liu S, Pértile F, Perazza CA, Villela PMS, de Almeida-Val, VM F, Hilsdorf AWS, Liu Z, Coutinho LL. 2017. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. *Scientific Reports*. 7: 46112.
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD Sequencing data: implications for genotyping. *Molecular ecology*. 22(11): 3151-3164.

- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome research*, 8(3), 175-185.
- Fernandes JBK, Carneiro DJ, Sakomura NK. 2000. Fontes e níveis de proteína bruta em dietas para alevinos de pacu (*Piaractus mesopotamicus*). *Revista Brasileira de Zootecnia*. 646-653.
- Fernández J, Toro MÁ, Sonesson AK., Villanueva B. 2014. Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet.* 5:414. doi: 10.3389/fgene.2014.00414
- Feulner PG, Schwarzer J, Haesler MP, Meier JI, Seehausen O. 2018. A dense linkage map of Lake Victoria cichlids improved the Pundamilia genome assembly and revealed a major QTL for sex-determination. *G3: Genes, Genomes, Genetics*, 8(7), 2411-2420.
- Fu B, Liu H, Yu X, Tong J. 2016. A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). *Scientific reports*. 6: 28679.
- Garcia AAF, Kido EA, Pinto, LR Meza AN, Silva JAGD, Souza HMD, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, Souza AP. 2006. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theor Appl Genet* 112:298–314
- Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet, E. 2012. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC genetics*, 13(1), 15.
- Hashimoto DT, Mendonça FF, Senhorini JA, Oliveira C, Foresti F, Porto-Foresti F. 2011. Molecular diagnostic methods for identifying Serrasalmid fish (Pacu, Pirapitinga, and Tambaqui) and their hybrids in the Brazilian aquaculture industry. *Aquaculture*, 321(1-2), 49-53.
- He H, Zhu S, Jiang Z, Ji Y, Wang F, Zhao R, Bie T. 2016. Comparative mapping of powdery mildew resistance gene Pm21 and functional characterization of resistance-related genes in wheat. *Theoretical and Applied Genetics*, 129(4), 819-829.
- IBGE. 2017. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal. Rio de Janeiro: IBGE, 2016.
- Jomori RK, Ducatti C, Carneiro DJ, Portella MC. 2008. Stable carbon ($\delta^{13}\text{C}$) and nitrogen ($\delta^{15}\text{N}$) isotopes as natural indicators of live and dry food in *Piaractus mesopotamicus* (Holmberg, 1887) larval tissue. *Aquaculture Research* 39: 370-381.
- Joshi R, Arnyasi M, Lien S, Gjøen HM, Alvarez AT, Kent M. 2018. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (*O. niloticus*). *Frontiers in genetics*, 9, 472.
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, ... Jindo T. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature*, 447(7145), 714.

- Liu H, Zhou Z, Lu J, Sun F, Wang S, Liu H, Jiang Y, Kucuktas H, Kaltenboeck L, Peatman E, Liu Z. 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*. 12: 53.
- MPA (Ministério da Pesca e Aquicultura. 2013. Boletim estatístico da pesca e aquicultura, Brasil 2011.
- Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24, 133–141.
- Maroso F, Casanova A, do Prado FD, Bouza C, Pardo BG, Blanco A, ... Martínez P. 2018. Species identification of two closely exploited flatfish, turbot (*Scophthalmus maximus*) and brill (*Scophthalmus rhombus*), using a ddRADseq genomic approach. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(5), 1253-1260.
- Paterson AH, Bowers JE, Burow MD, Draye X, Elisk CG, Jiang CX, Katsar CS, Lan TH, Lin YR, MingR, et al.. 2000. Comparative genomics of plant chromosomes. *The Plant cell* 12: 1523–40.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*. 7(5): e37135.
- Purcell S. "PLINK, version 1.07. 2009. " URL: <http://pngu.mgh.harvard.edu/~purcell/plink>.
- Ouellette, L. A., Reid, R. W., Blanchard, S. G., & Brouwer, C. R. 2017. LinkageMapView—rendering high-resolution linkage and QTL maps. *Bioinformatics*, 34(2), 306-307.
- Rasal KD, Chakrapani V, Pandey AK, Rasal AR, Sundaray JK, Ninawe A, Jayasankar P. 2017. Status and future perspectives of single nucleotide polymorphisms (SNPs) markers in farmed fishes: Way ahead using next generation sequencing. *Gene Reports*. 6: 81-86.
- Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P. 2013. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics*, 29(24), 3128-3134
- Sun C, Niu Y, Ye X, Dong J, Hu W, Zeng Q, ... Lu M. 2017. Construction of a high-density linkage map and mapping of sex determination and growth-related loci in the mandarin fish (*Siniperca chuatsi*). *BMC genomics*, 18(1), 446.
- Tsai HY, Robledo D, Lowe NR, Bekaert M, Taggart JB, Bron JE, Houston RD. 2016. Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3: Genes, Genomes, Genetics*, 6(7), 2173-2179.
- Vincze T, Posfai J, Roberts RJ. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res*. 31: 3688–3691.
- Wang L, Wan ZY, Bai B, Huang SQ, Chua E, Lee M, Pang HY, Wen YF, Liu P, Liu F Sun F, Lin G, Ye BQ, Yue GH. 2015. Construction of a high-density linkage map and fine mapping of QTL for growth in *Asian seabass*. *Scientific reports*, 5. 16358

- Wang W, Hu Y, Ma Y, Xu L, Guan J, Kong J. 2015. High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex-and growth-related regions detection. *PloS one*, 10(3), e0120410.
- Xu K, Li Q, Kong L, Yu R. 2008. A first-generation genetic map of the Japanese scallop *Patinopecten yessoensis* - based on AFLP and microsatellite markers. *Aquaculture Research*. 40(1): 35-43
- Yáñez JM, Newman S, Houston RD. 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Frontiers in genetics*, 6, 128.
- Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH. 1998. A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics*, 148(2), 839-850.
- Zeng, Q, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S, ... Liu Z. 2017. Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. *Scientific reports*, 7, 40347.
- Ziegler A, König IR, Thompson JR. 2008. Biostatistical aspects of genome-wide association studies. *Biom J*. 50(1):8-28.
- Zhu Z, Sun B, Wei J, Cai W, Huang Z, Chen C, ... Lei J. 2019a. Construction of a high density genetic map of an interspecific cross of *Capsicum chinense* and *Capsicum annuum* and QTL analysis of floral traits. *Scientific reports*, 9(1), 1054.
- Zhu C, Liu H, Pan Z, Chang G, Wang H, Wu N, ... Yu X. 2019b. Construction of a high-density genetic linkage map and QTL mapping for growth traits in *Pseudobagrus ussuriensis*. *Aquaculture*, 734213.

Capítulo 2

Estudo de Associação Genômica Ampla para o Peso Corporal
e Sexo em *Piaractus mesopotamicus*

RESUMO

Os caracteres relacionados ao crescimento estão entre os principais objetivos dos programas de melhoramento genético na aquicultura, pois estão associados aos processos que aceleram o volume da produção e reduzem custos. Entretanto, são raros os estudos genéticos que enfocam o melhoramento de espécies de peixes Neotropicais e, neste sentido, uma espécie de grande interesse é o pacu (*Piaractus mesopotamicus*), visto ser considerada uma das espécies de peixe com grande produção e valor comercial para a piscicultura nacional. O presente trabalho tem como principal objetivo realizar o estudo de associação genômica ampla (GWAS) para identificar QTLs através das associações entre marcadores moleculares com o peso corporal e o sexo. Os dados fenotípicos foram gerados a partir de 14 famílias de irmãos completos, totalizando 399 exemplares. Desses, 113 indivíduos de 3 famílias foram utilizados para a obtenção dos dados genotípicos com um total de 6K SNPs. O GWAS foi utilizado com a abordagem ssGBLUP e uma janela composta por 20 SNPs adjacentes. Para peso corporal, a maior janela foi associada a um valor de 3,21% no grupo de ligação 21 e para o sexo o valor foi de 2,89% na maior janela situada no LG23. A herdabilidade com informações baseadas no pedigree foi de 0.23 para peso corporal e de 0.42 para sexo, respectivamente, a herdabilidade genômica foi estimada com 0.20 para o peso e 0.45 para o sexo. Portanto, a herdabilidade para a característica peso foi considerada moderadamente herdável e para o sexo foi considerada bastante herdável. A busca por genes candidatos foi realizada no *blastn* e a maior parte das sequências homólogas foram compatíveis com a espécie *Pygocentrus nattereri*. Para o caráter crescimento foram identificados 11 genes que possuem participação na regulação do crescimento e metabolismo do corpo, ciclo celular, codificação do colágeno, fatores de transcrição e demandas glicêmicas. Para o sexo, foram identificados 12 genes candidatos, sendo que os genes encontrados exercem influência sobre espermatogênese, estrutura da cauda do espermatozoide, processo de diferenciação sexual, regulação da função reprodutiva das fêmeas e desenvolvimento de células germinativas nos testículos e nos ovários.

Palavras-chave: Peixe Neotropical, QTL, GBLUP, herdabilidade, genes candidatos

2.1 INTRODUÇÃO

O pacu é uma espécie de peixe nativo das bacias dos rios Paraná, Paraguai e Uruguai, e é considerada uma das espécies de peixes com maior produção e valor comercial para a piscicultura brasileira. De acordo com o IBGE (2017), é a quarta espécie nativa mais produzida na aquicultura continental nacional. Porém, a espécie tem sofrido impactos relacionados à sobrepesca e à fragmentação do habitat (Agostinho *et al.*, 2005, Mastrochirico-Filho *et al.*, 2019), tornando a criação em cativeiro uma ótima opção para suprir a necessidade do mercado e reduzir o impacto nos estoques das populações naturais. De acordo com dados reportados pela FAO (2018), a produção em 2017 foi de aproximadamente 13.200 mil toneladas.

Para a realização de produção mais efetiva em aquicultura, o desenvolvimento de programas de seleção genética aparece como uma ferramenta importante na tentativa de melhorar características de interesse zootécnico. Porém, apesar da diversidade de estudos com o pacu (Urbinati *et al.*, 2013), ainda existe grande carência quando se trata de pesquisas relacionadas ao melhoramento genético e desenvolvimento de ferramentas moleculares.

Na literatura existem alguns trabalhos relacionados aos temas de pesquisas genéticas em pacu, como a prospecção de marcadores moleculares (Calcagnotto *et al.*, 2001; Mastrochirico-Filho *et al.*, 2016), expressão gênica (Duran *et al.*, 2015, Mareco *et al.*, 2015), são necessários mais investimentos para o desenvolvimento de variedades de pacu com características positivas para a aquicultura. As tecnologias propostas recentemente como aquelas de sequenciamento de DNA de nova geração (NGS) estão sendo muito utilizadas por apresentarem precisão, baixo custo por par de base sequenciado e maior eficiência para identificar marcadores moleculares em alta densidade, principalmente SNPs (polimorfismo de nucleotídeo único) (Ghosh *et al.*, 2018). Os SNPs são os marcadores moleculares mais utilizados na atualidade e em diversos estudos de melhoramento genético, no qual eles são essenciais em programas que utilizam a identificação de QTLs (*loci* de característica quantitativa) e realização de GS (seleção genômica) (Liu *et al.*, 2018).

Em geral, QTLs são regiões cromossômicas relacionadas à variação de características quantitativas e estão ligados a caracteres de importância econômica (Dong *et al.*, 2019). A identificação dessas regiões ocorre através do desequilíbrio de ligação existente entre o marcador molecular polimórfico e o *loci*, sendo esta associação consequência da proximidade física que existe entre ambos (Ardlie *et al.*, 2002).

O mapeamento de associação genômica, conhecido como estudo de associação genômica ampla (GWAS) é uma técnica empregada na identificação de regiões associadas a características de interesse. No GWAS é utilizado um painel de SNPs distribuídos por todo o genoma, juntamente com os dados fenotípicos e informações sobre pedigree (Zhang *et al.*, 2012). Além disso, a partir de uma única genotipagem é possível verificar vários fenótipos de interesse comercial (Huang & Han, 2014). O GWAS explora eventos histórico e evolutivo de recombinação e utiliza a variação genética proveniente de um vasto número de genótipos (Zhu *et al.*, 2008). Isto confere importantes vantagens para o conhecimento do genoma das espécies, como menor tempo de pesquisa, maior número de alelos identificados e maior resolução do mapeamento (Yu & Buckler, 2006). Vários trabalhos do GWAS já foram publicados para espécies de peixes (Barría *et al.*, 2018, Li *et al.*, 2017, Gonzalez-Pena *et al.*, 2016)

Entre as características fenotípicas de interesse zootécnico, aspectos relacionados ao crescimento dos exemplares se apresentam como tópicos de grande impacto, uma vez que o aumento do uso de estoques produzido seletivamente leva a uma melhor utilização dos recursos limitados, tais como alimentação, água e espaço e tempo de produção (Gjedrem *et al.*, 2012). Portanto, a taxa de crescimento pode representar o fator determinante para a exploração econômica de uma espécie. Desta maneira, muitos trabalhos utilizam o peso como único caráter para determinar a taxa de crescimento dos indivíduos (Neto *et al.*, 2018; Wang *et al.*, 2019).

E em relação ao GWAS para peso, vários estudos realizados podem ser encontrados (Gonzalez-Pena *et al.*, 2016; Neto *et al.*, 2018 e 2019). Num sentido mais amplo esta característica em algumas espécies de peixes pode ser influenciada pela diferença sexual (Mei & Jian-Fang, 2015) e isso ocorre quando os caracteres de crescimento são mais desenvolvidos em um dos gêneros. Porém, a identificação sexual também está associada a outros benefícios, como o aumento do controle reprodutivo, a melhora da estabilidade dos sistemas de acasalamento, diferenças no valor econômico entre os sexos e ainda pode reduzir o impacto na qualidade dos produtos influenciados pelo fenótipo sexual (Budd *et al.*, 2015), podendo resultar em aumento da eficiência e da produção. Neste contexto, a identificação do sexo dos indivíduos também tem importância biológica, principalmente na identificação sexual de espécies que não possuem um dimorfismo sexual bem determinado.

Portanto, a busca de regiões que proporcionam um melhor conhecimento sobre os processos que atuam nos mecanismos de aumento de peso corporal, bem como na determinação sexual dos organismos se constituem como aspectos importantes para ser compreendido, tanto

em um sentido comercial para auxiliar programas de melhoramento genético quanto biológico para o entendimento genético envolvido nestes processos.

2.2 MATERIAIS E MÉTODOS

2.2.1 Formação dos grupos genéticos, instalações e manejo

Este estudo foi realizado em conformidade com as recomendações do Conselho Nacional de Controle de Experimentação Animal (CONCEA) (Ministério Brasileiro de Ciência, Tecnologia e Inovação) e foi aprovado pelo Comitê de Ética em Uso Animal (CEUA), protocolo no. 22.255 / 15. O presente estudo foi realizado sob autorização no. 33435-1, emitido pelo Instituto Chico Mendes de Conservação da Biodiversidade, Ministério do Meio Ambiente do Brasil (ICMBio).

Inicialmente, foi necessário formar uma população base através da montagem de diferentes famílias para o desenvolvimento de grupos genéticos, e dessa forma dar início a pesquisa.

A reprodução de *P. mesopotamicus* para a formação das famílias foi realizada em dezembro de 2015 com a utilização de 4 fêmeas e 14 machos adultos como reprodutores, sendo os indivíduos provenientes do plantel de matrizes da piscicultura do Centro de Aquicultura da Unesp - CAUNESP, Jaboticabal/SP e as fêmeas da Piscicultura EMA, Cosmorama, SP. Estes animais foram previamente analisados para testar sua pureza genética através da genotipagem de SNPs por Multiplex-PCR (Reação em cadeia da polimerase com uso de primers espécie-específicos), descritos em Hashimoto et al. (2011) e os cruzamentos resultaram na formação de 14 famílias de irmãos completos.

A indução da desova foi realizada com a utilização de extrato bruto de hipófise (EBH) de carpa dissolvido em solução salina (0,90% NaCl) aplicado em região intramuscular dos exemplares em duas doses nas fêmeas, com intervalo de 12 h (1ª dose de 0,5 mg hipófise kg⁻¹ de peixe; 2ª dose de 5,0 mg de hipófise kg⁻¹ de peixe). No momento da segunda dose foi realizada a aplicação de 5 ml de Ciosin® (0.25 mg/mL cloprostenol, Schering-Plough Saúde Animal Ind. Com. Ltda. Campinas, SP, BR), conforme protocolo descrito por Ceccarelli *et al.* (2005). Nos machos foi utilizada a aplicação de dose única, coincidindo com a aplicação da 2ª das fêmeas, sendo equivalente a 1,5 mg kg⁻¹ de EBH.

A extrusão dos gametas foi realizada com aproximadamente 240 h/grau, e a incubação das larvas foi realizada em incubadoras de 15 litros durante 1 mês. Em seguida, 100 larvas de

cada família foram mantidas em tanques redes em caixas de 1.000 l. Após 7 meses, quando os indivíduos estavam com 6 meses nos tanques, os grupos genéticos foram anestesiados em banho de imersão com benzocaína (100mg/L), marcados com *PITtags* (*passive integrated transponder tag*, modelo *full-duplex* FDX-B, 134.2-kHz) e amostras de sangue de cada indivíduo foram coletadas e fixadas em etanol absoluto para posterior extração de DNA.

Os indivíduos das 14 famílias foram distribuídos em dois tanques com uma média de 10 animais/família, totalizando cerca de 140 peixes por tanque (tanque1 e tanque2). As 3 famílias utilizadas para obtenção dos dados genômicos (capítulo anterior) foram avaliadas no tanque3, totalizando mais 126 peixes (42 por família). Os exemplares dos 3 tanques foram mantidos por oito meses nas mesmas condições ambientais e de manejo no Laboratório de Genética em Aquicultura e Conservação - LaGeAC - CAUNESP / Brasil. Após esse período foram alocados para 3 tanques escavados de aproximadamente 200 m² (TK1, TK2, TK3) existente nas instalações da piscicultura do CAUNESP.

No início do período experimental, enquanto a prole estava alocada nas instalações do Laboratório de Genética em Aquicultura e Conservação – LaGeAC, CAUNESP, a média da temperatura constatada foi de 28.30°C ± 1.55°C, do oxigênio dissolvido foi de 6.50 ± 0.30 mg/L e o pH de 6.70 ± 0.08. Nos tanques escavados, o valor médio da temperatura foi de 24.45°C ± 2.15°C, do oxigênio dissolvido foi de 8.20 mg/L ± 1.05 e do pH foi de 8.90 ± 0.50. Todos os parâmetros foram avaliados utilizando a sonda *Multi Water Quality Checker U 50 serie Horiba*.

Em relação a alimentação, da eclosão dos ovos até um período de 20 dias, os peixes foram alimentados exclusivamente com náuplios de *Artemia* (BioArtemia Ltda, RN, Brasil) quatro vezes ao dia. Após este período, também foi adicionada ração em pó à alimentação. Os náuplios de *Artemia* foram retirados da alimentação quando os peixes atingiram dois meses, enquanto a ração em pó foi retirada aos poucos, de acordo com o crescimento dos peixes à medida que se tornaram capazes de se alimentarem com ração extrusada, cujo diâmetro dos grânulos também teve aumento de acordo com o crescimento dos peixes.

2.2.2 Avaliação do peso corporal e do sexo

A avaliação foi realizada em fevereiro de 2017, cerca de 730 dias após a reprodução. Apesar de 430 exemplares terem sido distribuídos em 3 tanques, alguns animais morreram e a avaliação foi realizada em 136 peixes no tanque1 (variando de 6 a 10 por família), 137 peixes no tanque2 (variando de 4 a 10 por família), e 126 no tanque3 totalizando 399 peixes. Os

animais foram eutanasiados em benzocaína, foram pesados em balança de precisão (0.01g) e a cavidade celomática de cada indivíduo foi aberta para a exposição das gônadas (Figura1).

a)



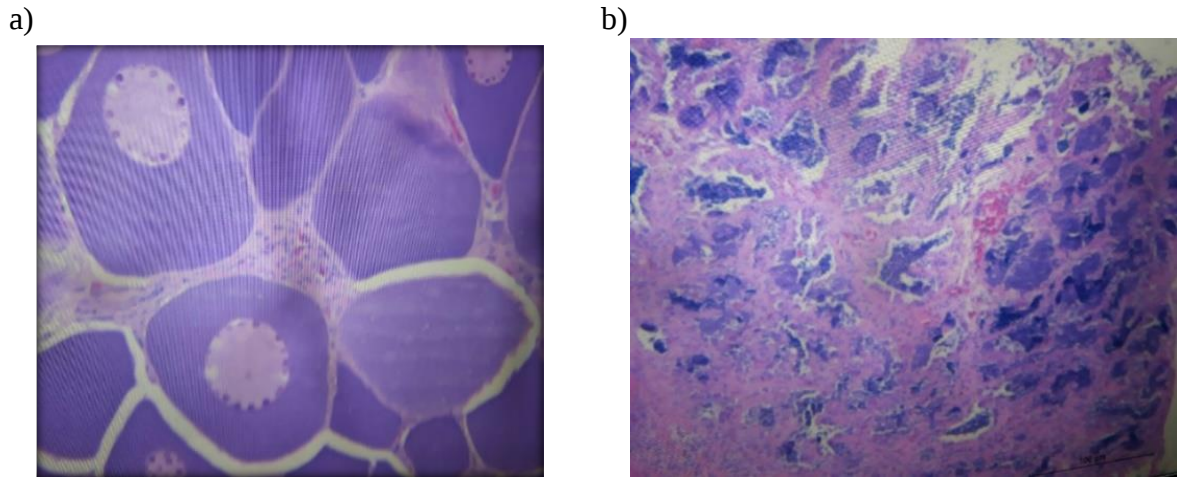
b)



Figura1: Ilustração da atividade de pesagem e de dissecção em *P. Mesopotamicus*. a) Pesagem em balança de precisão. b) Cavidade celomática de um indivíduo macho com exposição dos testículos.

Em muitos exemplares foi possível a visualização macroscópica das gônadas (Figura1), porém quando não era possível a identificação, uma porção do tecido era retirado para a visualização microscópica das células germinativas masculinas e femininas. Para a análise histológica, os fragmentos retirados foram fixados em solução de Karnovsky e processados para inclusão em parafina. As lâminas histológicas foram coradas com Hematoxilina/Eosina (figura2)

O programa computacional R (Team R Core, 2016) foi utilizado para realizar as análises estatísticas. Após a confirmação da homocedasticidade (BoxCox) e da normalidade (teste de *Cramér-von Mises*) dos dados, realizou-se a análise estatística de covariância (ANCOVA) para detectar diferenças significativas ($P < 0.05$) entre as três famílias e o teste de *Tukey* para as comparações entre médias do peso corporal.



Figuras 2: Fotomicrografia evidenciando células germinativas das fêmeas(a) e dos machos (b) de *P. mesopotamicus*. Coloração: HE.

2.2.3 Estudo de associação genômica ampla (GWAS) e herdabilidade

O estudo de associação genômica ampla (GWAS) combina na mesma linhagem de análise, dados de fenótipo e genótipo. Neste estudo foram utilizados 286 indivíduos apenas para a obtenção dos dados fenotípicos e 113 indivíduos na obtenção de dados genotípicos e fenotípicos, totalizando 399 exemplares. A análise foi realizada por meio da abordagem ssGBLUP (*Single step genomic - best linear unbiased prediction*) usando o *software* BLUPF90 (Miszta et al., 2016).

Os dados avaliados foram peso corporal (PC) e sexo (BIN), considerado traço binário com valor 1 para o sexo masculino e 2 para o feminino. A análise foi realizada usando uma janela de 20 SNPs adjacentes para associar as características de interesse com regiões do genoma, sendo que a localização dos SNPs tiveram por base o mapa de ligação (Capítulo anterior). Cada janela representa a soma da variação genética de 20 SNPs adjacentes.

A matriz H usada nesta análise combina informações baseadas na matriz genômica (G) e na matriz de pedigree (A) (Aguilar et al., 2010) segundo a fórmula abaixo:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

Os fenótipos observados das características PC e BIN foram utilizados como variáveis dependentes e independente, respectivamente. Para a característica BIN foi assumido o modelo

de limiar, sendo que este relaciona o fenótipo observado na escala categórica com uma escala subjacente normal contínua.

Todos os animais com genótipo e fenótipo disponíveis, além dos que tinham informação apenas do fenótipo, foram utilizados aplicando-se modelo linear misto para predição dos valores genéticos $y = X\beta + Za + e$, onde valores fenotípicos observados são dados pelo vetor y ; X são matrizes de incidência dos efeitos fixos; β é um vetor de efeitos fixos, para o peso corporal além de tanque, sexo e indivíduo. Para sexo não foram adicionados covariáveis; Z são matrizes de incidência para efeitos genéticos aditivos; o vetor a corresponde a efeitos aleatórios genéticos aditivos considerando a estrutura de covariância estabelecida pela matriz H ; e e é o vetor para resíduos aleatórios

Para visualizar os resultados do GWAS foi utilizado o programa *Manhattan plot* através do pacote *R qqman* (Turner, 2014).

A herdabilidade genômica foi estimada juntamente com a análise do GWAS através do BLUPF90 e a herdabilidade baseada no pedigree foi estimada através do programa ASREML 4.0 package (Gilmour *et al.*, 2009). O resultado das características analisadas foi verificado segundo Bourdon (1997), sendo $h^2 < 0.20$ considerada uma característica pouco herdável; $0.20 \leq h^2 < 0.40$ considerada uma característica moderadamente herdável; e $h^2 > 0.40$ uma característica bastante herdável.

2.2.4 Genes candidatos

A busca por genes candidatos foi realizada com a utilização do programa *Blastn* do NCBI contra as espécies *Pygocentrus nattereri*, *Astyanax mexicanus*, *Danio rerio* e *Oreochromis niloticus*, sendo que *P. nattereli* possui um parentesco mais próximo da espécie *P. mesopotamicus* em análise (Characiformes: Serrasalminidae). Assim sendo, as semelhanças foram maiores com esta espécie. Foram analisadas as 5 maiores janelas de cada característica (PC e BIN) e, dentro de cada janela, foram utilizadas as RADtags com maior associação com peso corporal e sexo, levando-se em consideração os dados do GWAS. Os genes identificados a uma distância máxima de 1Mb do SNP associado foram considerados como candidatos.

A lista de genes realizados pelo *Blastn* consta da Tabela2.

2.3 RESULTADOS

2.3.1 Avaliação do peso corporal e do sexo

Durante o processo de sexagem, em 179 (44,86%) do total de 399 exemplares não foi possível identificar o sexo, uma vez que as células germinativas estavam indiferenciadas.

Tabela1 - Médias e desvio padrão do peso corporal das 14 famílias de pacu *Piaractus mesopotamicus*. A - Famílias e Tanques. B - Sexo e Tanques. Avaliação realizada em fevereiro de 2017.

(A) – Dados de Famílias e Tanques.

	PESO(g)			
	TK1	TK2	TK3	MÉDIA TOTAL
*FAMÍLIA1^{bcd}	866.00±93.93	715.71±248.41	722.89±182.09	736.20±191.31
*FAMÍLIA2^{de}	555.55±193.05	483.00±168.70	613.80±135.90	583.77±159.02
*FAMÍLIA3^{bcd}	658.88±136.58	562.50±144.11	777.71±193.08	724.04±195.46
FAMÍLIA4^{bcd}	700.00±176.14	840.00±486.34		760.00±352.00
FAMÍLIA5^a	1053.33±152.02	908.00±422.74		976.84±332.08
FAMÍLIA6^{ab}	856.00±166.02	1107.00±422.74		981.50±350.30
FAMÍLIA7^{abcd}	716.00±165.96	786.00±259.39		751.00±350.30
FAMÍLIA8^{abcd}	696.25±172.25	741.42±208.76		888.13±400.36
FAMÍLIA9^{abc}	741.42±208.76	891.25±242.40		821.33±239.30
FAMÍLIA10^e	456.25±134.99	380.00±163.16		415.88±155.30
FAMÍLIA11^{cde}	551.51±138.33	620.00±182.51		587.65±166.79
FAMÍLIA12^{cde}	800.00±129.71	565.00±161.44		591.25±154.11
FAMÍLIA13^{de}	561.25±77.36	532.22±258.96		545.88±196.29
FAMÍLIA14^{cde}	630.00±136.38	535.71±180.38		588.75±163.97
TOTAL	691.66±215.67	713.63±368.12	699.73±183.77	701.91±270.53

As letras distintas (a,b,c,d,e) representam os grupos formados no teste de comparação das médias (Tukey, P<0.05).

*Famílias genotipadas somente do TK3

(B) – Dados de Sexo e Tanques.

PESO(g)				
	TK1	TK2	TK3	MÉDIA TOTAL
MACHO	661.53±205.90	575.74±243.36	638.07±122.31	619.52±196.32
FÊMEA	660.00±161.31	663.94±280.39	632.50±145.20	650.86±214.55
TOTAL	691.66±215.67	713.63±368.12	699.73±183.77	632.81±204.83

Após a verificação das médias e do desvio padrão, foi realizada a análise de diagnóstico, sendo que o peso permaneceu dentro da normalidade (teste de *Cramér-von Mises*) e da homocedasticidade/homogeneidade de variância (*BoxCox*).

Posteriormente, as médias foram verificadas através da análise de covariância (ANCOVA) cuja hipótese nula é a de que não existe diferença entre os grupos. Como o *p-valor* foi inferior ao limite de 5,00% apenas para a variável família, a hipótese nula pode ser rejeitada e ser considerado que existe diferença entre elas.

2.3.2 Estudo de associação genômica ampla (GWAS) e herdabilidade

Após o controle de qualidade e a construção do mapa de ligação descrito no capítulo anterior, um total de 6.437 SNPs foi usado para o GWAS. Pelo método ssGBLUP o GWAS explica em porcentagem a variância genética por meio de uma janela de 20 SNPs adjacentes, que são associados ao peso corporal e sexo. Os resultados foram apresentados por gráfico de dispersão através do *Manhattan plot* (Figuras 3 e 4), sendo a variação explicada por cada janela de SNP.

O grupo de ligação 21 está localizado a janela de SNPs mais associada (5295) para o peso corporal, explicando 3,21% da variação genética. Em seguida, as janelas são 1501, 2245, 4255 e 5316 dos grupos de ligação 5, 7, 15 e 21, correspondendo a uma variação genética de 2,37%, 1,78%, 1,58%, 1,57%, respectivamente (Figura 3).

Para o sexo, as janelas e os grupos de ligação mais importantes são 5642 (LG23), 890 (LG4), 5802 (LG24), 2539 (LG8) e 1881 (LG6) e a variação genética explica 2,89%, 2,01%, 1,96%, 1,81%, 1,32%, respectivamente (Figura 4).

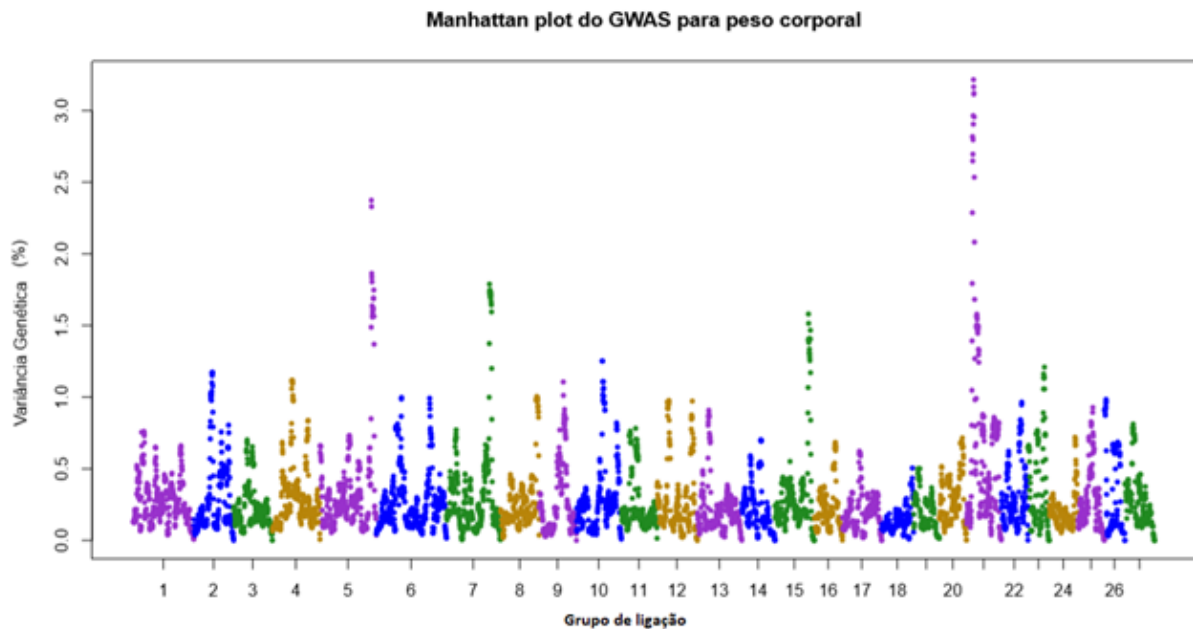


Figura 3. Manhattan plot de variância genética explicada (%) por janela de 20 SNP adjacentes para peso corporal (PC) em *P. mesopotamicus*.

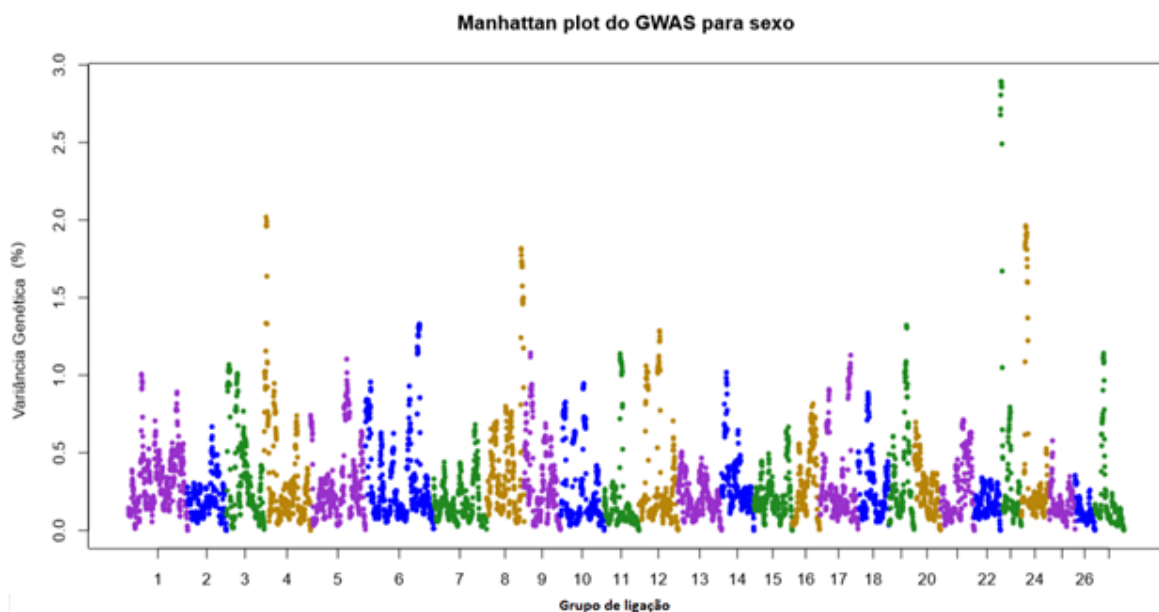


Figura 4. Manhattan plot de variância genética explicada (%) por janela de 20 SNP adjacentes para sexo *P. mesopotamicus*.

A herdabilidade foi considerada de valor alto e moderada para sexo e peso corporal, respectivamente. A h^2 foi estimada em 0.23 e 0.42 para PC e BIN, respectivamente, utilizando uma matriz baseada no pedigree. A h^2 genômica foi estimada com 0.20 para o peso e 0.45 para o sexo utilizando uma matriz G.

2.3.3 Genes candidatos

Apesar do Blastn ter sido realizado com as 4 espécies *Pygocentrus nattereri*, *Astyanax mexicanus*, *Danio rerio* e *Oreochromis niloticus*, praticamente todas as sequências foram similares com a espécie *Pygocentrus nattereri*. Foram encontrados 11 genes candidatos para peso corporal que podem exercer influência sobre o crescimento e destes, o LG5 e 7 possuem a menor quantidade, enquanto o LG21 tem a maior quantidade, apresentando 1 e 4 genes candidatos, respectivamente (Tabela2). Os genes identificados como candidatos exercem influência sobre a regulação do crescimento e metabolismo do corpo (Long *et al.*, 2011), ciclo celular (O'Hurley *et al.*, 2014), codificação do colágeno (Olsen *et al.*, 2000), fatores de transcrição e demandas glicêmicas do corpo (Dagli *et al.*, 2016) (Tabela2).

Para a característica sexo, foram identificados 12 genes candidatos importantes, sendo que a variação nos grupos de ligação foi de 1 a 4, ocorrendo no LG24 e 4, respectivamente (Tabela2). Os genes encontrados exercem influência sobre algumas funções dos testículos, como espermatogênese, estrutura da cauda do espermatozoide, processo de diferenciação sexual (Ma *et al.*, 2018, Agarwal *et al.*, 2018, Wang *et al.*, 2017, Schepers *et al.*, 2003). Além desses genes identificados, também foi verificada associação com genes de influência na regulação da função reprodutiva das fêmeas (Heger *et al.*, 2007) e desenvolvimento de células germinativas nos testículos e nos ovários (Huang *et al.*, 2005; Hong *et al.*, 2007) (Tabela2).

Tabela2. Lista completa de genes localizados nas principais janelas de 20 SNPs adjacentes dos 5 grupos de ligação com maior associados ao peso corporal (PC) e à determinação do sexo (BIN) em *P. mesopotamicus*.

Janela	LG ¹	Genes ^{2,3}
PC		
5295	21	pdlm2, adamts13, pcsk5, agtpbp1, agtpbp1, gkap1, ntrk2, naa35, sorbs3, cep78, cabin1, klhl22, gnrh1, stc1 , kctd9, ankrd39, chmp7, entpd4, usp25, nipbl, nup155, itpr1, noc3l, klhl11, ppp1r16a, uxs1, st6gal2, pdzk1, fam155a, myo16, irs2 , col4a2 , 36 - gypc, ndufa10, slc39a10 , map3k19, mmadh, lypd6b, nmi, ubxn4, tmeff2, lsm12, tmem101, g6pc3, wnk4.
1501	5	lgi2, aco1, sepsecs, acer2, dennd4c, haus6, saxo1, larp7, plin2, col25a1, trmt44, gpr78, slc2a4, gps2, ddit4l, ppp3ca, bank1, ugg1, blvra, erbin, tjp2, vldlr, sh3bp2, parm1, smad2 , sigmar1, il11ra, lpar1, cmpk1, anxa13, imp3, tctex1d1, kdsr.
2245	7	khdrbs3, lynx1, ppt2, clic1, arhgap8, brd1, creld2, alg12, pim3, il17rel, kdm5a, usp15, fbln1, phf21b, nup50, kiaa0930, spin1, nxnl2, ogdh, setd9, pdxp, ash2l, ddx56 , polm, camk2b, mrps24, nudcd3, nxnl2.
4255	15	syt14, sertad4, rdh14, tspan14, trim9, nin, pygl , map4k5, atl1, sav1, cdkl1, dmac2l, srf , tmem254, stk32c, lrcc27, pwwp2b, inpp5a, nkx6-2, tnag-gcc, six2, six3, srbd1, dok7, htt.
5316	21	kbtbd7, rgcc , mycbp2, slain1, scel, stx19, rsph1, cln5, fbxl3, epha3, c3orf38, myo1b, ptgfrn, igsf3, ttf2, vtcn1, trim45, parp4, 4930453n24rik, myo1b, cln5, rsph1, stx19, fbxl3, carf, fam117b, bmpr2 , gpr39, actr3, slc35f5, en1, insig2, dbi, ccde93, ddx18, marco, grb14 .
BIN		
5642	23	nfatc1, cndp2, ggct, klhl7, clptm1l, znrf2, mturn, snrpd1, rplp1, hdac5, metrn1, vwa3a, tbcd, znf516, sall3, atp9b, nfatc1, polr2b, rest, igfbp7, noa1, nfkb1, dok3, fam193b, ddx41 , dgkq, abhd18, cdhr2, pdlim7, rnf44, cltb, smim13, faf2, higd2a, pdgfa, dnaaf5, prkar1b, mvb12a, mri1, mb, cc2d1a, nanos3 , aftph, vps54, c19orf53, khsrp, r3hcc1l, pi4k2a, avpi1, exosc1, grid2.
890	4	dicer1 , syne3, flrt2, vrk1, dnah1 , pabpn1, lrp4, mapk8ip1, slc35c1, tollip, scamp4, ddx49, smim7, ifi30, spata1 , dennd1b, lhx9 , sash1, adgb, ust, slc22a16, adgb, cep57l1, sesn1.
5802	24	col21a1, hmgcll1, gfral, hcrr2, fam83b, mliip, elovl5, arfgef3, fbxo9, klhl31, plpp4, wdr11, ifngr, znf451, apbb1ip, eif2s3, klhl15, cct5, msrb2, armc3, skida1, skida1, ttc27, rasgrp3, matk, cers1, mydgi, ncln, use1, ppp1r2, htr2b, muc20, dact1, kiaa0586, irf2bpl , slc18b1, ahi1, rmnd1.
2539	8	prkar1b, dnaaf5, cdhr2, cltb, faf2, higd2a, unc5a, dok3, dgkq, abhd18, trnaa-cgc, polr2b, rest, igfbp7, sox8 , sstr5, rnf213, septin12 , cunh16orf89, alg1, eef2kmt, grap2, rnf25, thoc6, kat8, bcl2l12, rcn3, nosip, prr12, prrg2, rras, ap2a1, prmt1, pih1d1, aldh16a1, lin7b, strada, snrnp70
1881	6	pcsk2, bfsp1, ppp3r1, wdr92, plek, tnfaip3, perp, lxn, mlf1, mfsd1, gfm1, shox2, mmp28, cul3, ppp5c, gc, mob1b, gsf1, npffr2, adamts3, dmrt3a , dapk1, tmem8b, ggt6, rtkn, mccc2, pex5l, tp63, dpt, mrpl47, nme7, actl6a, pde6d, ptma, spa17 , scn3b, drd2, sorl1, sdhd, dixdc1, pih1d2.

¹ Grupo de ligação;

² Genes candidatos localizados nas janelas de 20 SNPs adjacentes;

³ *Pygocentrus nattereri* usado como principal referência de espécie.

2.4 DISCUSSÃO

2.4.1 Estudo de associação genômica ampla (GWAS) e herdabilidade

Existem diversos mecanismos de determinação sexual em peixes e estes podem ser definidos por fatores genéticos e/ou pelo meio ambiente. Em relação ao pacu, a alta herdabilidade genômica (0.45) e por pedigree (0.42) estimada para o sexo no presente trabalho com *P. mesopotamicus* também foi visualizada para o European sea bass (Vandeputte *et al.*, 2007). Nesta espécie a h^2 encontrada com base no pedigree foi de 0.62 e, de acordo com o autor a espécie possui mecanismo de determinação sexual poligênico e sensível a variações ambientais. De forma semelhante, poderia ser considerado que essa característica no pacu também seja determinada por vários genes, levando em consideração os resultados do GWAS realizados neste estudo. Entretanto, mais pesquisas devem ser realizadas para confirmar a natureza poligênica desta característica e para investigar possíveis influências ambientais, uma vez que não há informações referentes à influência de variações do ambiente na determinação sexual desta espécie.

Em relação ao carácter peso, as análises revelaram resultados indicando valores de herdabilidade considerados moderados. Tal resultado foi coincidente com valores observados em outras espécies de peixe, sendo de 0.31 a 0.34 em *Gadus morhua* (Kristjánsson & Arnason, 2016); de 0.21 a 0.36 em *Cynoglossus semilaevis* (Liu *et al.*, 2016); e de 0.2 a 0.4 em *Cyprinus carpio* (Hu *et al.*, 2017). Esta característica é muito visada pelos programas de melhoramento em aquicultura por apontar para a possibilidade de desenvolvimento de processos que podem acelerar a produção, as quais geralmente se apresentam moderadamente herdáveis (Ashton *et al.*, 2019).

Nas análises de associação genômica ampla (GWAS), foi encontrada uma variação máxima de 3,21% para o carácter peso corporal, este resultado foi muito satisfatório visto que as características de crescimento são reguladas por meio da expressão de vários genes e sofrem influências de fatores não genéticos. Em algumas pesquisas com espécies modelos que utilizaram um painel de SNPs muito denso, foi identificado uma variação genética aditiva para peso corporal igual ou menor a deste trabalho, no estudo da truta arco-íris também para peso corporal realizado por Gonzales-Pena *et al.*, (2016) utilizando 38K e uma janela de 20 SNPs, a variação genética encontrada foi de 1,38% e 0,95% para 10 e 13 meses de pesagem, respectivamente. Nesta mesma espécie e utilizando painel de 57K, a associação encontrada por

Neto *et al.*, (2019) foi de 2,14% e 3,02% para peso corporal, quando os exemplares tinham 15 e 18 meses, respectivamente.

As informações obtidas também poderão servir como base para vários estudos, assim como desenvolvimento de um chip de SNPs e seleção genômica (SG). Na SG a soma dos efeitos dos SNPs poderá auxiliar na predição do mérito genético dos indivíduos candidatos a seleção (HAYES *et al.*, 2007), porém mais estudos e validações precisam ser realizados.

O GWAS também já foi utilizado em vários estudos com peixes teleósteos, sendo que em alguns casos a maior parte da variação fenotípica é explicada pela identificação de apenas um *loci*, como é o caso da espécie *Siniperca chuatsi* (Sun *et al.*, 2017) na qual apenas 1 QTL (LG 23) explica 53,3% da variação e do *Atlantic halibut* (Palaikostas *et al.*, 2013) no LG 13 verificou-se 89,00% da variabilidade. Entretanto, ter essa característica não é uma regra e a identificação sexual pode ter uma herança poligênica (Palaikostas *et al.*, 2013). Nos resultados obtidos no presente trabalho há evidências de que o Pacu se encaixa nesse perfil, pois além da variação genética ter sido distribuída entre vários LGs, o valor de 2,89% de variabilidade obtido na maior janela é semelhante ou se apresenta até mesmo abaixo dos valores encontrados para espécies para as quais acredita-se também serem poligênicas para identificação sexual. Dentre essas espécies pode ser citado o European sea bass (Faggion *et al.*, 2019) na qual a maioria dos marcadores foi localizada em 2 grupos de ligação e a maior associação não atingiu 4,97% por GWAS. Considera-se, pois, que se novos marcadores e mais indivíduos forem analisados, uma associação genômica mais significativa para determinação sexual poderá ser encontrada com relação a este LG.

2.4.2 Genes candidatos

O grupo de ligação 21 apresentou duas janelas muito significativas neste estudo, com a janela 5295 tendo o maior valor de associação genômica com a característica relacionada ao peso dos animais. Nesta janela também foi encontrada a maior quantidade de genes candidatos, bem como o gene *stanniocalcin 1* (**stc1**) que está envolvido com a homeostase do cálcio e o fósforo em peixes ósseos (Zlot *et al.*, 2003). Este gene é expresso em tecidos de diversos órgãos como rim, próstata, tireóide, osso e ovário, estando envolvido no desenvolvimento ósseo, reprodutivo, angiogênico e modulação de resposta inflamatória (Chang *et al.*, 2005).

Nesta janela também foram encontrados genes envolvidos na regulação do crescimento e metabolismo do corpo, o gene *insulin receptor substrate 2* (**irs2**) e o *solute carrier family 39 member 10* (**slc39a10**). O *irs2* codifica o substrato receptor da insulina 2, uma molécula

envolvida como mediadora dos efeitos da insulina. Este gene possui uma função distinta de sinalização dependendo do tecido e em camundongos foi detectada uma redução severa no crescimento muscular esquelético causado pela atuação deficiente dos genes *Irs1* e *Irs2* (Long *et al.*, 2011). Outro gene, o *slc39a10* mostra características estruturais relacionada aos transportadores de zinco e está envolvida no controle de células de crescimento, atuando na transcrição, desenvolvimento e diferenciação (Hoiyo *et al.*, 2011 e MacDonald RS, 2000), além de estimular a proliferação e a mobilidade das células de mamíferos e também no embrião do zebrafish (Taylor *et al.*, 2016).

Com relação ao LG 21 janela 5316, também foi identificada a presença do gene *collagen type IV alpha 2 (col4a2)*, um dos genes responsáveis pela codificação do colágeno, constituinte estrutural da matriz extracelular, essencial para ossos e cartilagens (Olsen *et al.*, 2000). A presença deste gene também já foi identificado em catfish (Geng *et al.*, 2017). E considerando a estrutura do osso, também no grupo LG 21 foi identificada a presença do gene *bone morphogenetic protein receptor, type II (bmpr2)*, situado na janela 5216, que codifica um membro da família de receptores da proteína morfogenética óssea (bmp) da serina transmuticína / treonina quinase. As bmps são uma classe diversificada de moléculas que têm efeitos profundos no crescimento e diferenciação das células e na formação de múltiplos tecidos e órgãos (Cui *et al.*, 2017), tendo sido encontradas em várias espécies de peixes incluindo o Ocean Sunfish (Pan *et al.*, 2016), o zebrafish (Monteiro *et al.*, 2008) e a Tilapia (Fiol *et al.*, 2006).

Os fatores de crescimento relacionados a alguns genes também devem ser mencionados por exercerem influência sobre o ganho de peso dos indivíduos. O *SMAD family member 2 (smad2)* foi identificado no LG5 e está ligado a moléculas reguladoras de desenvolvimento e sinalização, bem como a fatores de transcrição (Liu *et al.*, 2011). O *smad2* também atua na cascata da via de sinalização da miostatina/TGF β (fator de crescimento transformador beta), via que regula negativamente o tamanho do músculo (Velloso 2008). No estudo de Kim *et al.* (2011) foi demonstrado que os alelos e haplótipos do promotor *smad2* estavam significativamente ligados à densidade mineral óssea de alguns locais do corpo humano. E quando se trata de peixes, foi demonstrado que no zebrafish as atividades do *smad2/3* são necessárias para a indução da mesoderme (Jia *et al.*, 2008). No LG21, janela 5216, também foi detectada a presença do gene *growth factor receptor- bound protein 14 (grb14)* que codifica uma proteína de ligação ao receptor do fator de crescimento que interage com os receptores de insulina e dos fatores de crescimento. Esse gene pode desempenhar um importante papel nas

vias de sinalização intracelular que regulam o crescimento e o metabolismo (Holt *et al.*, 2005). Ainda nesta janela também foram identificados genes envolvidos em funções celulares relacionadas com o crescimento como o box helicase 56 (**ddx56**) (Zhang *et al.*, 2011) e a divisão celular do DEAD (Asp-Glu-Ala-Asp), como o *regulator of cell cycle* (**rgcc**) que regula as células e a progressão do ciclo celular (Vlaicu *et al.*, 2016). Deve ser incluído nesta relação o *serum response factor* (**srf**) que além de estimular a proliferação e diferenciação celular, também participa da regulação do ciclo celular, apoptose e crescimento celular (O'Hurley *et al.*, 2014). No LG15 foi identificado o gene *phosphorylase glycogen liver* (**pygl**) que codifica a isozima hepática que atende às demandas glicêmicas do corpo em geral, sendo que mutações dessas isozimas podem causar sérios danos ao corpo, incluindo atraso no crescimento (Dagli *et al.*, 2016).

Em relação à determinação sexual foram identificados vários genes candidatos relacionados com o sistema reprodutivo dos machos, como o *DEAD-box helicase41* (**ddx41**), o *nanos homolog 3* (**nanos3**), e o *si:dkey-205k8.5* (**si:dkey-205k8.5** ou **spata1**) localizados nos LG 23, 23 e 4, respectivamente. Estes genes auxiliam na regulação da espermatogênese (Cheah *et al.*, 2017; Ma *et al.*, 2018; Agarwal *et al.*, 2018). Também o gene *si:dkey-205k8.5*, conhecido em algumas espécies como *spermatogenesis-associated protein 1* (**spata1**) já foi identificado em algumas espécies de peixes como em *zebrafish* e *catfish* (Agarwal *et al.*, 2018). O gene *dynein axonemal heavy chain 1* (**dnah1**) que está presente no LG4 está envolvido com a estrutura da cauda do espermatozoide, sendo que anomalias nesse gene geram baixa fertilidade (Wang *et al.*, 2017). O gene *sept12* (**sepin 12**) é regulado negativamente nos testículos de homens inférteis com defeitos espermatogênicos graves (Kuo *et al.*, 2019) e o gene *sperm autoantigenic protein 17* (**spa17**) codifica uma proteína presente na superfície celular que auxilia a ligação do espermatozoide à zona pelúcida do oócito (Li *et al.*, 2009). Estes genes foram encontrados no LG8 e 6, respectivamente. E em relação a todos os genes candidatos relacionados exclusivamente ao desenvolvimento testicular, o *SRY-box transcription factor 8* (**Sox8**) é o que mais se destaca, sendo que alguns estudos relatam que o Sox8, assim como o Sox9, podem estar envolvidos no processo de diferenciação sexual e expressão do hormônio anti-Mülleriano (AMH) em células de Sertoli (Schepers *et al.*, 2003); este gene foi identificado no LG8. Em se tratando do sistema reprodutivo feminino, no grupo de ligação 24 foi encontrado o gene *interferon regulatory factor 2 binding protein like* (**irf2bpl**) que codifica um fator de transcrição que pode regular a função reprodutiva (Heger *et al.*, 2007).

O gene mais significativo para ambos os sexos é o *double sex and mab-3 related transcription factor 3a* (**dmrt3a**) identificado no LG6. Este gene pertence ao grupo dmrt, que se mostra importante para a diferenciação sexual e está intimamente ligado ao desenvolvimento de ambas os tipos de células germinativas (testículos e ovários) (Huang *et al.*, 2005; Hong *et al.*, 2007). Está presente em algumas espécies de peixes teleósteos como no zebrafish (Li *et al.*, 2008) e no yellow drum (Qiu *et al.*, 2018). Também foram identificados genes ligados às células primordiais, com manifestação nas linhagens germinativas como o *dicer 1 ribonuclease type III* (**dicer1**) que está envolvido com o desenvolvimento de tumores especialmente em células de Sertoli–Leydig (Verrier *et al.*, 2018) e o gene *LIM homeobox 9* (**lhx9**) que codifica uma proteína muito similar a uma proteína que está envolvido com o desenvolvimento gonadal em camundongos (Ottolenghi *et al.*, 2001). Ambos os genes relacionados foram identificados no LG4.

Este trabalho apresentou o primeiro estudo de associação genômica ampla para o *P. mesopotamicus*, através de um denso painel de SNPs que indicou algumas regiões nos grupos de ligação de maior associação com características de peso corporal e determinação sexual. A descoberta dessas regiões pode ajudar no entendimento do controle genético destas características e poderá contribuir para a inclusão das mesmas no processo de seleção dos indivíduos. Este estudo também mostrou evidências de que a determinação sexual se apresenta como uma característica poligênica em *P. mesopotamicus*, sendo, porém, necessário aumentar a quantidades de indivíduos estudados para efetivar a confirmação deste fato. E ainda que os *loci* identificados através dos marcadores moleculares utilizados também abrigam genes candidatos com funções biológicas que podem ser associadas de modo direto ou indireto à expressão das características de crescimento e de determinação sexual.

2.5 REFERÊNCIAS

- Agarwal D, Gireesh-Babu P, Pavan-Kumar A, Koringa P, Joshi CG, Gora A, ... Chaudhari A. 2018. Molecular characterization and expression profiling of 17-beta-hydroxysteroid dehydrogenase 2 and spermatogenesis associated protein 2 genes in endangered catfish, *Clarias magur* (Hamilton, 1822). *Animal biotechnology*, 1-14.
- Agostinho AA, Thomaz SM, Gomes LC. 2005. Conservação da Biodiversidade em águas continentais do Brasil. *Megadiversidade*. 1: 70-78.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of dairy science*, 93(2), 743-752.
- Ardlie Kristin G, Leonid Kruglyak, Mark Seielstad. 2002 "Patterns of linkage disequilibrium in the human genome." *Nature Reviews Genetics* 3.4: 299.
- Ashton DT, Hilario E, Jaksons P, Ritchie PA, Wellenreuther M. 2019. Genetic diversity and heritability of economically important traits in captive Australasian snapper (*Chrysophrys auratus*). *Aquaculture*, 505, 190-198.
- Barría A, Christensen KA, Yoshida GM, Correa K, Jedlicki A, Lhorente JP, ... Yáñez JM. 2018. Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3: Genes, Genomes, Genetics*, g3-200053.
- Budd AM, Banh QQ, Domingos JA, Jerry DR. 2015. Sex Control in Fish: Approaches, Challenges and Opportunities for Aquaculture. *J Mar Sci Eng*, v.3, p.329-355.
- Calcagnotto D, Russello M, DeSalle R. 2001. Isolation and characterization of microsatellite loci in *Piaractus mesopotamicus* and their applicability in other Serrasalminae fish. *Molecular Ecology Notes*, 1(4), 245-247.
- Chang ACM, Cha J, Koentgen F, Reddel RR. 2005. The murine stanniocalcin 1 gene is not essential for growth and development. *Molecular and Cellular Biology*, 25(23), 10604-10610.
- Cheah JJ, Hahn CN, Hiwase DK, Scott HS, Brown AL. 2017. Myeloid neoplasms with germline DDX41 mutation. *International journal of hematology*, 106(2), 163-174.
- Criscuolo-Urbiniati E, Kuradomi RY, Urbiniati EC, Batlouni SR. 2012. The administration of exogenous prostaglandin may improve ovulation in pacu (*Piaractus mesopotamicus*). *Theriogenology*. 78: 2087-2094.
- Cui X, Yang Y, Jia D, Jing Y, Zhang S, Zheng S, ... Dong, K. 2017. Downregulation of bone morphogenetic protein receptor 2 promotes the development of neuroblastoma. *Biochemical and biophysical research communications*, 483(1), 609-616.

- Dagli A, Sentner CP, Weinstein DA. 2016. Glycogen storage disease type III. In *GeneReviews®[Internet]*. University of Washington, Seattle.
- Dong C, Jiang P, Zhang J, Shengjie L, Bai J, Fan J, Xu P. 2019. High-density linkage map and mapping for sex and growth-related traits of largemouth bass (*Micropterus salmoides*). *Frontiers in genetics*, 10, 960.
- Duran BO, Fernandez GJ, Mareco EA, Moraes LN, Salomão RAS, de Paula TG, ... Dal-Pai-Silvca M. 2015. Differential microRNA expression in fast-and slow-twitch skeletal muscle of *Piaractus mesopotamicus* during growth. *PloS one*, 10(11), e0141967.
- Faggion S, Vandeputte M, Chatain B, Gagnaire PA, Allal F. 2019. Population-specific variations of the genetic architecture of sex determination in wild European sea bass *Dicentrarchus labrax* L. *Heredity*, 122(5), 612.
- FAO. 2018. Organização das Nações Unidas para a Agricultura e Alimentação. The state of world fisheries and aquaculture: Rome
- Fiol DF, Chan SY, Kültz, D. 2006. Identification and pathway analysis of immediate hyperosmotic stress responsive molecular mechanisms in tilapia (*Oreochromis mossambicus*) gill. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 1(3), 344-356.
- Geng X, Liu S, Yuan Z, Jiang Y, Zhi, D, Liu Z. 2017. A genome-wide association study reveals that genes with functions for bone development are associated with body conformation in catfish. *Marine biotechnology*, 19(6), 570-578.
- Ghosh M, Sharma N, Singh AK, Gera M, Pulicherla KK, Jeong DK. 2018. Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. *Critical reviews in biotechnology*, 38(8), 1157-1175.
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. 2002. ASReml user guide release 1.0.
- Gjedrem T. 2012. Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture*. 344: 12-22.
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB, ... Leeds TD. 2016. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Frontiers in genetics*, 7, 203.
- Hashimoto DT, Mendonça FF, Senhorini JA, Oliveira C, Foresti F, Porto-Foresti F. 2011. Molecular diagnostic methods for identifying Serrasalmid fish (Pacu, Pirapitinga, and Tambaqui) and their hybrids in the Brazilian aquaculture industry. *Aquaculture*, 321(1-2), 49-53.
- Hayes BJ, Chamberlain AJ, Mcpartlan H, Macleod I, Sethuraman L, Goddard ME. 2007. Accuracy of marker assisted selection with single markers and marker haplotypes in cattle. *Genetical Research*, V. 89, P. 215- 220,

- Heger S, Mastronardi C, Dissen GA, Lomniczi A, Cabrera R, Roth CL, ... Ojeda SR. 2007. Enhanced at puberty 1 (EAP1) is a new transcriptional regulator of the female neuroendocrine reproductive axis. *The Journal of clinical investigation*, 117(8), 2145-2154.
- Hojyo S, Fukada T, Shimoda S, Ohashi W, Bin BH, Koseki H, Hirano T. 2011. The zinc transporter SLC39A14/ZIP14 controls G-protein coupled receptor-mediated signaling required for systemic growth. *PloS one*, 6(3), e18059.
- Holt LJ & Siddle K. 2005 "Grb10 and Grb14: enigmatic regulators of insulin action—and more?." *Biochemical Journal* 388.2: 393-406.
- Hong CS, Park BY, Saint-Jeannet JP. 2007. The function of Dmrt genes in vertebrate development: it is not just about sex. *Developmental biology*, 310(1), 1-9.
- Hu X, Li C, Shang M, Ge Y, Jia Z, Wang S, ... Shi L. 2017. Inheritance of growth traits in Songpu mirror carp (*Cyprinus carpio* L.) cultured in Northeast China. *Aquaculture*, 477, 1-5.
- Huang XY, Guo Y, Shui S, Gao H. Yu et al., 2005 Multiple alternative splicing and differential expression of Dmrt1 during gonad transformation of the rice field eel. *Biol. Reprod.* 73: 1017–1024
- Huang X & Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology*, 65, 531-551.
- IBGE. 2017. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal. Rio de Janeiro: IBGE, 2016.
- Jia S, Ren Z, Li X, Zheng Y, Meng A. 2008. smad2 and smad3 are required for mesendoderm induction by transforming growth factor- β /nodal signals in zebrafish. *Journal of Biological Chemistry*, 283(4), 2418-2426.
- Kim BJ, Hwang JY, Han BG, Lee JY, Lee JY, Park EK, ... Koh JM. 2011. Association of SMAD2 polymorphisms with bone mineral density in postmenopausal Korean women. *Osteoporosis international*, 22(8), 2273-2282.
- Kristjánsson T & Arnason T. 2016 "Heritability of economically important traits in the Atlantic cod *Gadus morhua* L." *Aquaculture research* 47.2: 349-356.
- Kuo PL, Tseng JY, Chen HI, Wu CY, Omar HA, Wang CY, ... Teng YN. 2019. Identification of SEPTIN12 as a novel target of the androgen and estrogen receptors in human testicular cells. *Biochimie*, 158, 1-9.
- Li Q, Zhou X, Guo Y, Shang X, Chen H, Lu H, ... Zhou R. 2008. Nuclear localization, DNA binding and restricted expression in neural and germ cells of zebrafish Dmrt3. *Biology of the Cell*, 100(8), 453-463.
- Li FQ, Han YL, Liu Q, Wu B, Huang WB, Zeng SY. 2009. Overexpression of human sperm protein 17 increases migration and decreases the chemosensitivity of human epithelial ovarian cancer cells. *BMC cancer*, 9(1), 323.

- Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S, ... Liu Z. 2017. Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Molecular Genetics and Genomics*, 1-13.
- Liu F, Yang Y, Li Y, Guo H, Dai H, Gao J, ... Shao C. 2016. Phenotypic and genetic parameter estimation of juvenile growth and bottom color traits in half-smooth tongue sole, *Cynoglossus semilaevis*. *Acta Oceanologica Sinica*, 35(10), 83-87
- Liu S, Vallejo RL, Evenhuis JP, Martin KE, Hamilton A, Gao G, ... Palti Y. 2018. Retrospective evaluation of marker-assisted selection for resistance to bacterial cold water disease in three generations of a commercial rainbow trout breeding population. *Frontiers in genetics*, 9.
- Liu Z, Lin X, Cai Z, Zhang Z, Han C, Jia S, ... Wang Q. 2011. Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *Journal of Biological Chemistry*, 286(32), 28520-28532.
- Long YC, Cheng Z, Copps KD, White MF. 2011. Insulin receptor substrates Irs1 and Irs2 coordinate skeletal muscle growth and metabolism via the Akt and AMPK pathways. *Molecular and cellular biology*, 31(3), 430-441.
- Ma HT, Niu CM, Xia J, Shen XY, Xia MM, Hu YQ, Zheng Y. 2018. Stimulated by retinoic acid gene 8 (Stra8) plays important roles in many stages of spermatogenesis. *Asian journal of andrology*, 20(5), 479.
- MacDonald RS. 2000. The role of zinc in growth and cell proliferation. *J Nutr* 130: 1500S–1508S.
- Mareco EA, Serrana DG, Johnston IA, Dal-Pai-Silva M. 2015. Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). *BMC genomics*, 16(1), 182.
- Mastrochirico-Filho VA, Hata ME, Sato LS, Jorge PH, Foresti F, Rodriguez MV, Martínez P, Porto-Foresti F, Hashimoto DT. 2016. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conservation Genetics Resources*. 8(2): 109-114.
- Mastrochirico-Filho VA, Pazo F, Hata ME, Villanova GV, Foresti, F, Vera M, ... Hashimoto DT. 2019. Assessing Genetic Diversity for a Pre-Breeding Program in *Piaractus mesopotamicus* by SNPs and SSRs. *Genes*, 10(9), 668.
- Mei J & Gui JF. 2015. "Genetic basis and biotechnological manipulation of sexual dimorphism and sex determination in fish." *Science China Life Sciences* 58.2: 124-136.
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. 2016. Manual for BLUPF90 family of programs.
- Monteiro R, Dinther M, Bakkers J, Wilkinson R, Patient R, Dijke P, Mummery C. 2008. Two novel type II receptors mediate BMP signalling and are required to establish left–right asymmetry in zebrafish. *Developmental biology*, 315(1), 55-71.
- Neto RR, Yoshida G, Lhorente JP, Yáñez JM. 2018. Genome-wide Association Study for body weight in rainbow trout (*Oncorhynchus mykiss*).

- Neto RR., Yoshida G, Lhorente JP, Yáñez, JM. 2019. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics*, 294(3), 563-571.
- O'Hurley G, Prencipe M, Lundon D, O'Neill A, Boyce S, O'Grady A, ... Watson RWG. 2014. The analysis of serum response factor expression in bone and soft tissue prostate cancer metastases. *The Prostate*, 74(3), 306-313.
- Olsen BR, Reginato AM, Wang W. 2000. Bone development. *Annual review of cell and developmental biology*, 16(1), 191-220.
- Ottolenghi C, Moreira-Filho C, Mendonça BB, Barbieri M, Fellous M, Berkovitz GD, McElreavey K. 2001. Absence of mutations involving the LIM homeobox domain gene LHX9 in 46, XY gonadal agenesis and dysgenesis. *The Journal of Clinical Endocrinology & Metabolism*, 86(6), 2465-2469.
- Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, ... Migaud H. 2013. Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC genomics*, 14(1), 566.
- Pan H, Yu H, Ravi V, Li C, Lee AP, Lian MM, ... Zhang G. 2016. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *GigaScience*, 5(1), 36.
- Qiu C, Han Z, Li W, Ye K, Xie Y, Wang Z. 2018. A high-density genetic linkage map and QTL mapping for growth and sex of yellow drum (*Nibea albiflora*). *Scientific reports*, 8(1), 17271.
- Schepers G, Wilson M, Wilhelm D, Koopman, P. 2003. SOX8 is expressed during testis differentiation in mice and synergises with SF1 to activate the Amh promoter in vitro. *J. Biol. Chem.* 278,28101 - 28108.
- Sun C, Niu Y, Ye X, Dong J, Hu W, Zeng Q, ... Lu M. 2017. Construction of a high-density linkage map and mapping of sex determination and growth-related loci in the mandarin fish (*Siniperca chuatsi*). *BMC genomics*, 18(1), 446.
- Taylor KM, Muraina IA, Brethour D, Schmitt-Ulms G, Nimmanon T, Ziliotto S, ... Hogstrand C. 2016. Zinc transporter ZIP10 forms a heteromer with ZIP6 which regulates embryonic development and cell migration. *Biochemical Journal*, 473(16), 2531-2544.
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, ... Houston RD. 2015. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC genomics*, 16(1), 969.
- Turner SD. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots
- Urbinati EC, Gonçalves FD, Takahashi LS. 2013. Pacu (*Piaractus mesopotamicus*). In: Espécies nativas para a piscicultura no Brasil (eds Baldisserotto B, Gomes LC). UFSM, Santa Maria.
- Vandeputte M, Dupont-Nivet M, Chavanne H, Chatain B. 2007. A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics*, 176(2), 1049-1057.

- Velloso CP (2008) Regulation of muscle mass by growth hormone and IGF-I. *Brit J Pharmacol* 154:557–568
- Verrier F, Dubois d'Enghien C, Gauthier-Villars M, Bonadona V, Faure-Contier C, Dijoud F, ... Golmard L. 2018. Multiple DICER1-related lesions associated with a germline deep intronic mutation. *Pediatric blood & cancer*, 65(6), e27005
- Vlaicu SI, Tatomir A, Boodhoo D, Ito T, Fosbrink M, Cudrici C, ... Rus V. 2016. RGC-32 is expressed in the human atherosclerotic arterial wall: role in C5b-9-induced cell proliferation and migration. *Experimental and molecular pathology*, 101(2), 221-230.
- Wang Q, Yu Y, Zhang Q, Zhang X, Yuan J, Huang H, ... Li F. 2019. A novel candidate gene associated with body weight in the Pacific white shrimp *Litopenaeus vannamei*. *Frontiers in Genetics*, 10.
- Wang X, Jin H, Han F, Cui Y, Chen J, Yang C, ... Hao C. 2017. Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. *Clinical genetics*, 91(2), 313-321.
- Yu J & Buckler ES. 2006. Genetic association mapping and genome organization of maize. *Current opinion in biotechnology*, 17(2), 155-160.
- Zhang H, Wang Z, Wang S, Li H. 2012. Progress of genome wide association study in domestic animals. *Journal of animal science and biotechnology*, 3(1), 26.
- Zhang Y, Forsys JT, Miceli AP, Gwinn AS, Weber JD. 2011. Identification of DHX33 as a mediator of rRNA synthesis and cell growth. *Molecular and cellular biology*, 31(23), 4676-4691.
- Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *The plant genome*, 1(1), 5-20.
- Zlot C, Ingle G, Hongo J, Yang S, Sheng Z, Schwall R, ... Gerritsen ME. 2003. Stanniocalcin 1 is an autocrine modulator of endothelial angiogenic responses to hepatocyte growth factor. *Journal of Biological Chemistry*, 278(48), 47654-47659.

CONSIDERAÇÕES FINAIS

Este trabalho constituiu um dos primeiros passos no desenvolvimento de ferramentas genômicas para espécies de peixes Neotropicais, com foco especial no pacu, *Piaractus mesopotamicus*, de grande importância econômica para a piscicultura nacional. Foi formado um denso painel de marcadores moleculares, com a caracterização de mais de 6.000 SNPs que poderão servir de suporte para diferentes abordagens biológicas envolvendo programas de melhoramento genético, revelação de mecanismos evolutivos e de conservação desta espécie e de espécies aparentadas. Esta ferramenta genômica permitiu a construção de um rico mapa genético que foi imprescindível para realizar associações entre o marcador e características fenotípicas (GWAS) e, a partir desta, proceder à identificação de QTLs.

Os resultados obtidos com a identificação dos QTLs evidenciaram a natureza poligênica de características relacionadas ao peso corporal dos indivíduos. Em relação ao crescimento esse resultado já era esperado, pois trata-se de uma característica tipicamente controlada por vários genes (Tsai *et al.*, 2015) na maioria dos organismos. Em relação aos mecanismos envolvidos na determinação sexual desta espécie foram identificadas muitas variantes com pequenos efeitos para a formação desta característica, de conformidade com referências existente para outras espécies de peixes (Yoshida *et al.*, 2018). Porém, no caso específico da espécie aqui estudada, trata-se da primeira constatação desta natureza. Contudo, considera-se que este caminho de pesquisa aberto com as análises realizadas deve ser complementado com a realização de mais estudos para comprovar essa questão. Além disso, os *loci* identificados podem ser utilizados como um critério complementar para a identificação de indivíduos candidatos a processos de seleção, aumentando a acurácia nas avaliações genéticas dos indivíduos (Resende *et al.*, 2012).

O painel de SNPs produzido certamente para esta espécie poderá ser utilizado para identificar *loci* relacionados a diversas características de interesse e caso seja detectada uma forte associação entre o marcador e a característica alvo, com um alto valor de associação genética, os indivíduos poderão ser selecionados diretamente pelo genótipo. Tal metodologia possibilitaria, então, a realização do processo de seleção de uma forma mais precoce e acurada, do mesmo modo como foi descrito no trabalho de Houston *et al.* (2008) para a seleção de peixes resistentes à uma doença identificada como necrose pancreática infecciosa (IPN). O mapa de ligação resultante das análises realizadas com estes marcadores nas progênies resultantes dos cruzamentos realizados envolvendo casais de reprodutores desta espécie proporcionou uma

abordagem mais direta do genoma, com possibilidades de aplicação direta nos estoques desta espécie.

Os genes candidatos encontrados nas regiões identificadas neste trabalho também podem explicar e/ou mostrar a sua influência sobre o fenótipo. Entendendo como eles funcionam e como exercem sua influência nas características de interesse zootécnico, certamente serão de grande auxílio nos programas de melhoramento genético.

REFERÊNCIAS GERAIS

- Abimorad EG & DJ Carneiro. 2004. Métodos de coleta de fezes e determinação dos coeficientes de digestibilidade da fração protéica e da energia de alimentos para o pacu, *Piaractus mesopotamicus* (Holmberg, 1887). *Revista Brasileira de Zootecnia*. 33: 1101-1109.
- Agostinho AA, Thomaz SM, Gomes LC. 2005. Conservação da Biodiversidade em águas continentais do Brasil. *Megadiversidade*. 1: 70-78.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*: 17(2): 81-92.
- AQUAGEN. CMS Protection against CMS – Better heart health. p. 7462, 2014. Disponível em: <<https://aquagen.no/en/products/salmon-eggs/product-documentation/resistance-against-cms/>>. Acesso em: 24 abr. 2018.
- AQUAGEN. Strong and even fillet colour. p. 7462, 2015a. Disponível em: <<https://aquagen.no/en/products/salmon-eggs/product-documentation/reliable-and-even-fillet-colour/>>. Acesso em: 24 abr. 2018.
- AQUAGEN. Sterile salmon. p. 7462, 2015b. Disponível em: <<https://aquagen.no/en/products/salmon-eggs/product-documentation/sterile-salmon/>>. Acesso em: 24 abr. 2018.
- Baird NA, Etter PD, Atwood TS, *et al.* 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, 1–7.
- Barria A, Christensen KA, Yoshida GM, Correa K, Jedlicki A, Lhorente JP, ... Yáñez JM. 2018. Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3: Genes, Genomes, Genetics*, g3-200053.
- Beardmore JA, Mair GC, Lewis RI. 2001. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. In *Reproductive Biotechnology in Finfish Aquaculture* (pp. 283-301). Elsevier.
- Bernatchez L. 2016. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, 89(6), 2519-2556.
- Bernardino G & Lima V. 1999. Situação da criação de Colossoma e Piaractus no sudeste do Brasil (1988-1991). SOUZA, RHS Criação de Colossoma e Piaractus no Brasil. *IBAMA (Ed), Brasília*, 173-232.
- Britski HA, Silimon KZS, Lopes BS. 2007. Peixes do Pantanal: manual de identificação. Embrapa, Serviço de Produção de Informação, Brasília.

- Calcagnotto D, Schaefer SA, De Salle R. 2005. Relationships among Characiformes fishes inferred from analysis of nuclear and mitochondrial gene sequences. *Molecular Phylogenetics and Evolution*, 36: 135-153.
- Catella AC, FL de R. Campos, SP Albuquerque. 2017. "Sistema de Controle da Pesca de Mato Grosso do Sul (SCPESCA/MS-23-2016)." *Boletim de Pesquisa e Desenvolvimento Embrapa Pantanal, Corumbá, Brazil*.
- Da Silva Nunes JDR, Liu S, Pértille F, Perazza CA, Villela PMS, de Almeida-Val, VM F, Hilsdorf AWS, Liu Z, Coutinho LL. 2017. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. *Scientific Reports*. 7: 46112.
- Dong, C, Jiang P, Zhang J, Shengjie L, Bai J, Fan J, Xu P. 2019a. High-density linkage map and mapping for sex and growth-related traits of largemouth bass (*Micropterus salmoides*). *Frontiers in genetics*, 10, 960.
- Dong L, Han Z, Fang M, Xiao S, Wang Z. 2019b. Genome-wide association study identifies loci for body shape in the large yellow croaker (*Larimichthys crocea*). *Aquaculture and Fisheries*, 4(1), 3-8.
- Fernandes JBK, Carneiro DJ, Sakomura NK. 2000. Fontes e níveis de proteína bruta em dietas para alevinos de pacu (*Piaractus mesopotamicus*). *Revista Brasileira de Zootecnia*. 646-653.
- 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.
- FAO. 2006. Organização das Nações Unidas para a Agricultura e Alimentação. The state of world fisheries and aquaculture: Rome.
- FAO. 2010. Organização das Nações Unidas para a Agricultura e Alimentação. The state of world fisheries and aquaculture: Rome.
- FAO. 2016. Organização das Nações Unidas para a Agricultura e Alimentação. The state of world fisheries and aquaculture: Rome
- FAO. 2018. Organização das Nações Unidas para a Agricultura e Alimentação. The state of world fisheries and aquaculture: Rome
- Fernández J, Toro MÁ, Sonesson AK., Villanueva B. 2014. Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet*. 5:414. doi: 10.3389/fgene.2014.00414
- Gabián, M, Morán P, Fernández AI, Villanueva B, Chtioui A, Kent MP, ... Saura M. 2019. Identification of genomic regions regulating sex determination in Atlantic salmon using high density SNP data. *BMC genomics*, 20(1), 764.
- Gjedrem, T. 2012. Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture*. 344: 12-22.

- Godinho HP. 2007. Estratégias reprodutivas de peixes aplicadas à aquicultura: bases para o desenvolvimento de tecnologias de produção. *Rev. Bras. Reprod. Anim.* 31: 351-360.
- Gonzalez-Pena D, Gao G, Baranski M, Moen T, Cleveland BM, Kenney PB.... Leeds, T. D. (2016). Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Frontiers in genetics*, 7, 203.
- Gutierrez AP, Yáñez JM, Fukui S, Swift B, Davidson WS. 2015. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One*, 10(3), e0119730.
- Guo W, Cai C, Wang C, Han Z, Song X, Wang K, ... Zhang T. 2007. A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics*, 176(1), 527-541.
- Hamzah A, Ponzoni RW, Nguyen NH, Khaw HL, Yee HY, Nor SAM. 2014. Genetic evaluation of the Genetically Improved Farmed Tilapia (GIFT) strain over ten generations of selection in Malaysia. *Journal of Tropical Agricultural Science*. 37: 411–429.
- Hashimoto DT, Senhorini JA, Foresti F, Porto-Foresti F. 2012. Interspecific fish hybrids in Brazil: management of genetic resources for sustainable use. *Reviews in Aquaculture*. 4(2): 108-118.
- Hashimoto DT, Prado FDD, Senhorini JA, Foresti F, Porto-Foresti F. 2013. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. *Aquaculture Research*. 44(6): 876-884.
- Hashimoto DT, Senhorini JA, Foresti F, Martínez P, Porto-Foresti F. 2014. Genetic identification of F1 and post-F1 Serrasalmid juvenile hybrids in Brazilian aquaculture. *PloS one*. 9(3): e89902.
- Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, ... Strüssmann, C. A. 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*, 109(8), 2955-2959.
- Hilsdorf AWS, Orfão LH. 2011. Aspectos gerais do melhoramento genético em peixes no Brasil. *Rev. Bras. Zootec.* 40: 317-324.
- Honglang H. 2007. Freshwater fish seed resources in China. In: Assessment of freshwater fish seed resources for sustainable aquaculture (Ed: Bondad-Reantaso MG). FAO Fisheries Technical Paper, No. 501, FAO, Rome.
- Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB, ... Bishop SC. 2008. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, 178(2), 1109-1115.
- IBAMA. 2007. Estatística da Pesca 2007: Brasil – Grandes regiões e unidades da Federação.
- IBGE. 2017. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal. Rio de Janeiro: IBGE, 2016.

- IBGE. 2014. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal. Rio de Janeiro: IBGE, 2013.
- IBGE. 2019. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal. Rio de Janeiro: IBGE, 2018.
- Jomori RK, Ducatti C, Carneiro DJ, Portella MC. 2008. Stable carbon ($\delta^{13}\text{C}$) and nitrogen ($\delta^{15}\text{N}$) isotopes as natural indicators of live and dry food in *Piaractus mesopotamicus* (Holmberg, 1887) larval tissue. *Aquaculture Research* 39: 370-381.
- Kause A, Ritola O, Paananen T, Wahlroos H, Meantysaari EA. 2005. Genetic trends in growth, sexual maturity and skeletal deformations, and rate of inbreeding in a breeding programme for rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 247: 177–187.
- Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R., ... International Sheep Genomics Consortium. 2014. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal genetics*, 45(5), 754-757.
- Kubitza F. 2015. Aquicultura no Brasil, conquistas e desafios. Revista. *Panorama da aquicultura*. 25(150): 10-23.
- Kumar G & Martin K. 2017. "Applications of next-generation sequencing in fisheries research: A review." *Fisheries Research* 186: 11-22.
- Li X, Jin X., Wang H, Zhang, X, Lin Z. 2016. Structure, evolution, and comparative genomics of tetraploid cotton based on a high-density genetic linkage map. *DNA Research*, 23(3), 283-293.
- Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S, ... Liu Z. 2017. Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Molecular Genetics and Genomics*, 1-13.
- Longo SB, Clark B, York R., Jorgenson AK. 2019. Aquaculture and the displacement of fisheries captures. *Conservation Biology*.
- Mastrochirico-Filho VA, Hata ME, Sato LS, Jorge PH, Foresti F, Rodriguez MV, Martínez P, Porto-Foresti F, Hashimoto DT. 2016. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conservation Genetics Resources*. 8(2): 109-114.
- MPA (Ministério da Pesca e Aquicultura. 2013. Boletim estatístico da pesca e aquicultura, Brasil 2011.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, ... Hori, H. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(6888), 559.
- Matsuda M. 2003. Sex determination in fish: Lessons from the sex-determining gene of the teleost medaka, *Oryzias latipes*. *Development, growth & differentiation*, 45(5-6), 397-403.
- Menton DJ. 1989. Research considerations into the nutrition of *Colossoma* and *Piaractus* in relation to culture conditions. *Cultivo de Colossoma*. Bogotá: Guadalupe, 75-84.

- Moen T, Hoyheim B, Munck H, Gomez-Raya L. 2004. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Animal genetics*, 35(2), 81-92.
- Moen T, Baranski M, Sonesson AK, Kjøglum S. 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics*, v. 10, p. 1-14.
- Moreira HLM. 2001. *Fundamentos da moderna aquicultura*. Editora da ULBRA.
- Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, ... Sakaizumi, M. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*, 191(1), 163-170.
- Ninh NH, Ponzoni RW, Nguyen NH, Woolliams JA, Taggart JB, McAndrew BJ, Penman DJ. 2013. A comparison of communal and separate rearing of families in selective breeding of common carp (*Cyprinus carpio*): Responses to selection. *Aquaculture*. 408: 152-159.
- Oliveira C, Avelino GS, Abe KT, Mariguela TC, Benine RC, Ortí G, ... Castro RMC. 2011. Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling. *BMC evolutionary biology*, 11(1), 275.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*. 7(5): e37135.
- Petrere Jr M. 1989. River fisheries in Brazil: a review. *Regulated Rivers Res. Manage.* 4: 1-16
- Porto-Foresti F, Foresti F. 2004. Genética e biotecnologia em piscicultura: usos na produção, manejo e conservação dos estoques de peixes. In: Tópicos especiais em piscicultura de água doce tropical intensiva (eds Cyrino JEP, Urbinati EC, Fracalossi DM, Castagnolli N). TecArt: São Paulo.
- Recknagel H, Elmer KR, Meyer A. 2013. A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus spp.*) obtained by massively parallel DNA sequencing (ddRADSeq). *G3: Genes, genomes, genetics*. 3(1): 65-74.
- Resende EK. 2003. Migratory fishes of the Paraguay-Paraná basin excluding the Upper Paraná River. In *Migratory fishes of South America: biology, fisheries and conservation states* (eds Carolsfeld J, Harvey B, Ross C, Baers A), World Bank, Victoria.
- Resende M. (2012). Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP). *Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial*, 1.
- Santos VB, Santos RS, Salomão RAS Silva RM. 2012. Reprodução induzida de pacu (*Piaractus mesopotamicus*) com o uso de diferentes hormônios comerciais. *Pesquisa & Tecnologia*. 9(1): 1-6.

- São Paulo. 2014. Assembleia Legislativa do Estado de São Paulo (ALESP). Decreto Nº 60.133, de 7 de fevereiro de 2014: As espécies da fauna silvestre ameaçadas de extinção, as quase ameaçadas e as deficientes de dados para avaliação no Estado de São Paulo e dá providências correlatas. 2014."
- Shi Y, Zhou Z, Liu B, Kong S, Chen B, Bai H, ... & Xu, P. 2020. Construction of a High-Density Genetic Linkage Map and QTL Mapping for Growth-Related Traits in *Takifugu bimaculatus*. *Marine Biotechnology*, 1-15.
- Sonesson AK, Woolliams JA, Meuwissen TH. 2012. Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution*, 44(1), 27.
- Suárez-Salgado D, Parra-Bracamonte GM, Benavides-González F, Alfaro IOM, Rincón AMS, Moreno-Medina VR, De la Rosa-Reyna XF. 2020. Non-synonymous polymorphisms in candidate gene associated with growth traits in Channel catfish (*Ictalurus punctatus*, Rafinesque, 1818). *Molecular Biology Reports*, 47(1), 87-95.
- Tave D. 1995. Selective breeding programmes or medium-sized fish farms, FAO Fisheries Technical Paper, Rome, 122p.
- Tsai HY, Hamilton A, Guy DR, Tinch AE, Bishop SC, Houston RD. 2015. The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*). *BMC genetics*, 16(1), 51.
- Urbinati EC, Gonçalves FD, Takahashi LS. 2013. Pacu (*Piaractus mesopotamicus*). In: Espécies nativas para a piscicultura no Brasil (eds Baldisserotto B, Gomes LC). UFSM, Santa Maria.
- Xu K, Li Q, Kong L, Yu R. 2008. A first-generation genetic map of the Japanese scallop *Patinopecten yessoensis* - based on AFLP and microsatellite markers. *Aquaculture Research*. 40(1): 35-43
- Wang L, Xie N, Shen Y, Ye B, Yue GH, Feng X. 2019. Constructing high-density genetic maps and developing sexing markers in northern snakehead (*Channa argus*). *Marine Biotechnology*, 21(3), 348-358.
- Wang W, Tan S, Luo J, Shi H, Zhou T, Yang Y, ... Gao D. 2019. GWAS analysis indicated importance of NF-κB signaling pathway in host resistance against motile *Aeromonas septicemia* disease in catfish. *Marine Biotechnology*, 21(3), 335-347.
- Yáñez JM, Newman S, Houston RD. 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Frontiers in genetics*, 6, 128.
- Yoshida GM, Bangera R, Carvalheiro R, Correa K, Figueroa R, Lhorente JP, Yáñez JM. 2018. Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3: Genes, Genomes, Genetics*, 8(2), 719-726.
- Yoshida GM, Carvalheiro R, Lhorente JP, Correa K, Barria A, Figueroa R, Yáñez JM. 2018. Bayesian genome-wide association analyses reveal different genetic architecture of *Piscirickettsia salmonis* resistance in three salmonid species. In *Proceedings of the World Congress on Genefics Applied to Livestock Production* (Vol. 11, p. 721).

- Yossa R, Bardon-Albaret A, Chiasson MA, Liu Q, Duston J, Manning T, Benfey TJ. 2019. Controlling preharvest maturity in farmed Arctic char: A review from the Canadian perspective. *Journal of the World Aquaculture Society*, 50(5), 894-907.
- Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178(1), 539-551.
- Zhou Q, Su Z, Li Y, Liu Y, Wang L, Lu S, ... Wei, M. 2019. Genome-wide association mapping and gene expression analyses reveal genetic mechanisms of disease resistance variations in *Cynoglossus semilaevis*. *Frontiers in Genetics*, 10, 1167.

MATERIAL COMPLEMENTAR

O presente trabalho teve como material complementar o Apêndice I “Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology” DOI 10.7717/peerj.2205. O Apêndice II “Structure and genetic variability of the oceanic whitetip shark, *Carcharhinus longimanus*, determined using mitochondrial DNA” DOI 10.1371/journal.pone.0155623. O Apêndice III “Evolution and conservation of *Characidium* sex chromosome” DOI 10.1038/hdy.2017.43. O Apêndice IV “Genetic Characterization of the Fish *Piaractus brachypomus* by Microsatellites Derived from Transcriptome Sequencing” DOI 10.3389/fgene.2018.00046. E por último o Apêndice V “Genetic Applications in the Conservation of Neotropical Freshwater Fish” DOI: 10.5772/intechopen.73207.

Apêndice I



Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology

Natália J. Mendes¹, Vanessa P. Cruz¹, Fernando Y. Ashikaga¹,
Sâmia M. Camargo^{1,2}, Claudio Oliveira¹, Andrew N. Piercy³,
George H. Burgess⁴, Rui Coelho⁵, Miguel N. Santos⁵,
Fernando F. Mendonça² and Fausto Foresti¹

¹ Laboratório de Biologia e Genética de Peixes, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil

² Laboratório de Genética Pesqueira e Conservação, Institute of Marine Sciences, Universidade Federal de São Paulo, Santos, São Paulo, Brazil

³ Lake Nona Campus, Valencia College, Orlando, Florida, United States

⁴ Florida Museum of Natural History, Florida Program for Shark Research, University of Florida, Gainesville, Florida, United States

⁵ IPMA, Portuguese Institute for the Ocean and Atmosphere, Algarve, Portugal

ABSTRACT

The tiger shark (*Galeocerdo cuvier*) has a global distribution in tropical and warm temperate seas, and it is caught in numerous fisheries worldwide, mainly as bycatch. It is currently assessed as near threatened by the International Union for Conservation of Nature (IUCN) Red List. In this study, we identified nine microsatellite loci through next generation sequencing (454 pyrosequencing) using 29 samples from the western Atlantic. The genetic diversity of these loci were assessed and revealed a total of 48 alleles ranging from 3 to 7 alleles per locus (average of 5.3 alleles). Cross-species amplification was successful at most loci for other species such as *Carcharhinus longimanus*, *C. acronotus* and *Alopias superciliosus*. Given the potential applicability of genetic markers for biological conservation, these data may contribute to the population assessment of this and other species of sharks worldwide.

Submitted 29 February 2016

Accepted 11 June 2016

Published 30 August 2016

Corresponding author

Fernando F. Mendonça,
fernando.mendonca@unifesp.br

Academic editor:
James Reimer

Additional Information and
Declarations can be found on
page 6

DOI 10.7717/peerj.2205

© Copyright
2016 Mendes et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Aquaculture, Fisheries and Fish Science, Conservation Biology, Genetics, Marine Biology, Zoology

Keywords High-throughput sequencing, *Galeocerdo cuvier*, Shark, Microsatellites, Population structure

INTRODUCTION

The tiger shark *Galeocerdo cuvier* (Péron & Lesueur, 1822), is a shark from the order Carcharhiniformes and family Carcharhinidae. It has a worldwide distribution in tropical and temperate seas, and is considered a top predator generally requiring large foraging areas (Heupel et al., 2014). Recent data shows that this species can move long distances and occupies different regions, including coastal areas being therefore more susceptible than pelagic sharks to anthropogenic threats (Heupel et al., 2014).

Caught in many world fisheries as bycatch, the tiger shark is currently classified as "Near Threatened" by the International Union for Conservation of Nature (IUCN).

How to cite this article Mendes et al. (2016), Microsatellite loci in the tiger shark and cross-species amplification using pyrosequencing technology. *PeerJ* 4:e2205; DOI 10.7717/peerj.2205

However, some basic information, such as the characterization of population genetic structure, variability, identification of geographical restrictions to gene flow with possible local populations remains broadly unknown. In light of this, information on their conservation status is difficult to assess. For this reason, molecular markers have been increasingly used in species conservation and management programs, including microsatellite molecular markers (Simple Sequence Repeats-SSR). A range of SSR markers have been developed using the pyrosequencing technique, generating information with millions of base pairs in a single run and in a short period of time.

Specifically for the tiger shark, nine SSR markers were previously developed on specimens from the Hawaiian archipelago (Bernard, Feldheim & Shivji, 2015), but cross-application was not tested for other shark species yet. Thus, the objectives of this study were to identify other microsatellites for the tiger shark in specimens from the Atlantic, and design additional molecular markers that can now be used in this and other shark species for population genetics studies.

MATERIAL AND METHODS

Sampling

In fulfillment of data archiving guidelines (Baker, 2013), primary data have been deposited with Dryad. Samples of tiger shark were collected in landings of the fishing fleet from São Paulo coast (n = 12) and in scientific cruises in the Fernando de Noronha archipelago (n = 6) by researchers from the Biosciences Institute of Botucatu, São Paulo State University, and Marine Sciences Institute of the São Paulo Federal University, in Brazil. Additionally, 11 samples were collected from the east coast of Florida by the Florida Program for Shark Research, University of Florida, USA. For evaluating cross-amplification we used six samples of *Carcharhinus acronotus* collected from São Paulo coast, five samples of *C. longimanus* and five samples of *Alopias superciliosus*, collected in the northeast Atlantic by onboard observers from the Portuguese Institute for the Ocean and Atmosphere (IPMA), Portugal. All sampled tissues were stored in 95% ethanol to ensure the integrity and quality of tissue for molecular analysis.

454 GS-FLX pyrosequencing and microsatellite discovery

The total genomic DNA was extracted from each sample following the protocol described by Ivanova, Dewaard & Hebert (2006). A sample (voucher: TIG03SP) of 100 µg of tiger-shark DNA from São Paulo coast (−25.1164, −47.6082) was sequenced on a Roche 454 GS FLX sequencer with Titanium platform "Genome sequencer 20 System" (Instituto Agrobiotecnológico de Rosário-INDEAR, Argentina), following procedures described in Margulies et al. (2005).

To isolate microsatellites and design primers for population genetics all sequences of the SSR were compiled using Primer3 (Rozen & Skaletsky, 1999) and BatchPrimer3 (You et al., 2008). Primers were designed based on the following criteria: primer size of 20 bp (min = 18, max = 22 bp), ideal annealing temperature of 60 °C (min = 55 °C, max = 63 °C), GC optimum of 60% (min = 40%, max = 80%) and the size of the amplified product ranging from 50–500 bp. The sequences were then grouped and aligned

in the Clustal W software (Thompson, Higgins & Gibson, 1994), identifying duplicated sequences for the same locus.

Novel microsatellite markers

The PCR amplifications to test the synthesized primers were performed in a Thermal Cycler Veriti™ (Applied Biosystems, Foster City, CA, USA) under the following conditions: initial denaturing for 10 min at 95 °C; 30 cycles of 94 °C for 45 s, the primer annealing temperature (TA) was tested from 51 to 57 °C for 50 s; 72 °C for 50 s, and a final extension at 72 °C for 20 min. The total reaction volume was 10 µL and composed of 1.0× PCR Buffer, 0.25 mM MgCl₂, 0.05 mM of each dNTP, 0.5 units of Platinum Taq DNA polymerase (Invitrogen/Carlsbad, CA, USA), 0.10 µM reverse primer, 0.10 µM forward primer, and 30 ng of template DNA.

To verify the effectiveness of the reaction and the amplification of the fragments, 1.5 µL of the PCR product were subjected to electrophoresis on a 1% agarose gel. The amplified products were compared with a 1 Kb plus ladder (Invitrogen), subsequently visualized on a transilluminator and photographed with a digital camera using the Kodak Digital Science software.

Genotyping was done with the M13-tail PCR method of Schuelke (2000). The best loci, that showed high polymorphism and quality of bands, were selected and further analyzed on an ABI 3130xl sequencer (Applied Biosystems, Life Technologies). The allele sizes were determined using ROX 500 (Applied Biosystems) as an internal standard with the software package GeneMapper 3.7 (Applied Biosystems).

We used the software GenAlex analysis 6.1 (Peakall & Smouse, 2006) to convert our data to run in other analysis programs. Arlequin 3.5 (Excoffier & Lischer, 2010) was used to calculate heterozygosity, number of alleles, Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium. The program Cervus v.3.0.7 (Marshall et al., 1998) was used to test for the presence of null alleles and estimate the inbreeding coefficient (Fis) and polymorphism information content (PIC).

RESULTS AND DISCUSSION

From the genomic material generated by the pyrosequencing technology, a total of 71,059 reads with an average size of 367 bp was obtained, consisting of 26,075,405 nucleotides, which accounts for approximately 0.75% of the *G. cuvier* genome, assuming a genome size of 3.44 Gb (estimated from the size of *Rhincodon typus*, Read et al., 2015). For the identification of microsatellite sequences, the online software Batch Primer3 was used, and 615 microsatellite loci were identified. A second filtration was subsequently performed with the software Primer 3.0 which identified 159 microsatellite loci. From these, we selected 30 loci which contained the best scores of each primer pair with a size of 15–20 bp, a GC of 40–50% and little variation in the annealing temperature in the PCR reaction. Of these, 20 pairs of primers were synthesized and tested, nine being polymorphic with one trinucleotide and eight dinucleotide primers (Table 1). The sequences for polymorphic microsatellite markers in this study have been deposited in GenBank (Accession numbers: KT598263–KT598271).

Table 1 General information about the microsatellite analysis.

Analyzes	N of sequences
Number of reads	71.059
Selection of microsatellites (using BatchPrimer3)	615
Secondary selection of microsatellite (using Primer 3.0)	159
Amplification and control of PCR product on agarose gel	30
Microsatellite loci to synthesize with fluorescent dye	20
Polymorphism test with capillary sequencer	10
Microsatellite loci in linkage equilibrium	9

Table 2 Data for microsatellite loci of the cross-amplification in *Carcharhinus longimanus*, *Carcharhinus acronotus* and *Alopias superciliosus*.

<i>C. acronotus</i>			<i>C. longimanus</i>		<i>A. superciliosus</i>	
Loci	Na	Range (bp)	Na	Range (bp)	Na	Range (bp)
TIG_1	2	116–118	2	118–134	2	104–118
TIG_5	3	260–264	2	331–335	2	265–273
TIG_7	2	170–180	2	162–170	3	152–170
TIG_10	2	251–253	1	304	1	307
TIG_12	2	296–364	2	246–296	2	372–418
TIG_15	1	336	3	290–310	2	288–312
TIG_17	3	242–270	2	210–224	1	268
TIG_19	0	0	1	316	2	386–394
TIG_25	1	396	1	358	2	388–398

Note:
Na, number of alleles.

The application of the developed markers resulted in 48 alleles, with a minimum of 3 (TIG_25) to 7 (TIG_1, TIG_7, TIG_12) and average of 5.3 alleles per loci. Transferability tests of the markers in other species showed positive amplification in *C. longimanus*, *Alopias superciliosus* and *C. acronotus*. For the *C. acronotus*, two loci were polymorphic (TIG_17, TIG_5), and for *C. longimanus* and *Alopias superciliosus* only one polymorphic locus were observed in five samples of each species, TIG_15 and TIG_7, respectively (Table 2).

In tiger shark the observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0.16 (TIG_17) to 1.0 (TIG_10) and 0.20 (TIG_25) to 0.72 (TIG_7), respectively (Table 3). The H_o was higher than H_e , suggesting an excess of heterozygotes relative to the model of HWE. Significant differences from HWE after Bonferroni correction ($p < 0.01$) were detected in only two loci (TIG_10 and TIG_17). The deviation in the HWE for locus TIG_17 (0.715) can be explained by a significant value in intrapopulation F_{is} (Kordicheva et al., 2010). This locus was the only one with a positive value for the F_{is} , and may be evidence of a heterozygous deficiency (Holsinger & Weir, 2009), resulting in a decrease in genetic variability.

Table 3 Data for microsatellite loci in the tiger shark, *Galeocerdo cuvier*.

Loci	Primer sequence (5'→3')	MOTIF	T °C	N	Na	Range (bp)	Ho	He	HWE	Fis	PIC	F (Null)
TIG_1	F_CTCTTGACGGTGCTCGATC R_AATGGCAACTTTTCCTGTCC	(AC)10	53	29	7	116–154	0.758	0.642	0.711	–0.184	0.710	–0.194
TIG_5	F_GCCAGCATCCATTCATACAG R_AGAGGGAAGTGGTGTGTGGT	(CT)8	51	26	4	203–257	0.384	0.337	1.000	–0.141	0.589	–0.239
TIG_7	F_CACCAACCTCCCATCAC R_CAGACATTCTCTCCATCC	(AC)15	57	27	7	169–183	0.925	0.726	0.318	–0.280	0.811	–0.101
TIG_10	F CTCAGCAGGTCTGGACAACA R_GGTGGTAGGAACATGGAACG	(GT)10	59	29	5	256–276	1.000	0.655	0.000	–0.539	0.608	–0.245
TIG_12	F_TGCCATGAGTGCTGTTTTC R_TGCCGCTTGTACTGCTAC	(CA)11	53	28	7	364–376	0.535	0.520	0.543	–0.030	0.682	–0.213
TIG_15	F_AACTGCCAAAAGGACAAGA R_GTAAGCCCAACAGACCATCC	(TG)15	55	25	6	231–241	0.520	0.463	0.675	–0.124	0.650	–0.233
TIG_17	F_TGAAGCTAACGAGGGTCTG R_AGCGCAGAAGATCAAGAGGA	(GT)11	57	25	4	268–286	0.160	0.554	0.000	0.715	0.734	–0.138
TIG_19	F_TGCTTGTGTCTGAGGTGAGTG R_TGGAGGTTCAATCCGAGAC	(TG)10	53	27	5	337–353	0.555	0.443	0.677	–0.260	0.627	–0.214
TIG_25	F_CCGTGCTATGTGGATTCT R_CTTGAAGAGAGTGGGCGAAG	(CCT)5	55	27	3	331–349	0.222	0.206	1.000	–0.075	0.511	–0.285

Notes:

T °C, primer annealing temperature; N, number of individuals analyzed; Na, number of alleles; He, expected heterozygosity; Ho, observed heterozygosity; HWE, probability of departure from Hardy-Weinberg equilibrium; Fis, inbreeding coefficient; PIC, polymorphism information content; F (Null), null alleles.

Imbalance values in Hardy-Weinberg equations when considering microsatellite locus may be due to the presence of null alleles (Kordicheva et al., 2010). However, the presence of null allele was not detected in the present study, indicating that the markers developed are of high quality. Further, the PIC was highly informative for all the loci (PIC > 0.5), also indicating a high quality marker (Botstein et al., 1980).

In the present study, the average expected heterozygosity was approximately 0.50 and the average observed heterozygosity was 0.55. The levels of genetic variability seen in this study may be due to population differences resulting from remote sample locations. This is to be expected given that the samples are coming from different oceans and the finding of significant differences in the levels of heterozygosity among different groups would not be unforeseen.

ACKNOWLEDGEMENTS

The authors thank all those who contributed to this study, including the researchers who carried out the sample collection in the Fernando de Noronha Archipelago and the fishermen for facilitating tissue collection during landings on the São Paulo coast. Sampling and data collection from the Portuguese fishery were obtained and supported by the Programa Nacional de Amostragem Biológica (PNAB) within the scope of the EU Data Collection Framework (DCF). The authors are grateful to all fishery observers and longline skippers that helped collect samples for this study.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

R. Coelho is supported by an Investigador-FCT contract from the Portuguese Foundation for Science and Technology (FCT, Fundação para a Ciência e Tecnologia) supported by the EU European Social Fund and the Programa Operacional Potencial Humano (Ref: IF/00253/2014). N.J. Mendes was supported by a grant from Foundation for Research Support of the São Paulo State—FAPESP (Ref: 2013/14555-4). This work was funded by FAPESP (Ref: BIOTA 2011/23787-0). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Natália J. Mendes conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Vanessa P. Cruz conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Fernando Y. Ashikaga conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, reviewed drafts of the paper.
- Sâmia M. Camargo performed the experiments, analyzed the data, reviewed drafts of the paper.
- Claudio Oliveira analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Andrew N. Piercy contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- George H. Burgess contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Rui Coelho analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Miguel N. Santos contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Fernando F. Mendonça conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Fausto Foresti analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Institutional Review Board Ministério do Meio Ambiente–MMA approval number: 50463-1.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

GenBank

BankIt1851998 Seq1 [KT598263](#)
 BankIt1851998 Seq2 [KT598264](#)
 BankIt1851998 Seq3 [KT598265](#)
 BankIt1851998 Seq4 [KT598266](#)
 BankIt1851998 Seq5 [KT598267](#)
 BankIt1851998 Seq6 [KT598268](#)
 BankIt1851998 Seq7 [KT598269](#)
 BankIt1851998 Seq8 [KT598270](#)
 BankIt1851998 Seq9 [KT598271](#).

REFERENCES

- Baker CS. 2013.** Journal of heredity adopts joint data archiving policy. *Journal of Heredity* **104**(1):1 DOI [10.1093/jhered/ess137](#).
- Bernard AM, Feldheim KA, Shivji MS. 2015.** Isolation and characterization of polymorphic microsatellite markers from a globally distributed marine apex predator, the tiger shark (*Galeocerdo cuvier*). *Conservation Genetics Resources* **7**(2):509–511 DOI [10.1007/s12686-014-0408-0](#).
- Botstein D, White RL, Skolnick M, Davis RW. 1980.** Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**(3):314–331.
- 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 DOI [10.1111/j.1755-0998.2010.02847.x](#).
- Heupel M, Knip D, Simpfendorfer C, Dulvy N. 2014.** Sizing up the ecological role of sharks as predators. *Marine Ecology Progress Series* **495**:291–298 DOI [10.3354/meps10597](#).
- Holsinger KE, Weir BS. 2009.** Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics* **10**(9):639–650 DOI [10.1038/nrg2611](#).
- Ivanova NV, Dewaard JR, Hebert PDN. 2006.** An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6**:998–1002 DOI [10.1111/j.1471-8286.2006.01428.x](#).
- Kordicheva SY, Rubtsova GA, Shitova MA, Shaikhaev GO, Afanasiev KI, Zhivotovsky LA. 2010.** Search for null alleles at the microsatellite locus of chum salmon (*Oncorhynchus keta* Walbaum). *Russian Journal of Genetics* **46**(8):1019–1022 DOI [10.1134/S1022795410080168](#).
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR,**

- Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380 DOI 10.1038/nature03959.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7(5):639–655 DOI 10.1046/j.1365-294x.1998.00374.x.
- Peakall ROD, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6(1):288–295 DOI 10.1111/j.1471-8286.2005.01155.x.
- Read TD, Petit RA III, Joseph SJ, Alam MT, Weil R, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb H, Dove ADM. 2015. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. *PeerJ PrePrints* 3:e837v1 DOI 10.7287/peerj.preprints.837v1.
- Rozen S, Skaletsky H. 1999. *Primer3 on the WWW for General Users and for Biologist Programmers*. Totowa: Humana Press Inc.
- Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18(2):233–234 DOI 10.1038/72708.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22):4673–4680 DOI 10.1093/nar/22.22.4673.
- You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hanc D, Lazo GR, Dvorak J, Anderson OD. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9(1):253 DOI 10.1186/1471-2105-9-253.

Apêndice II

RESEARCH ARTICLE

Structure and Genetic Variability of the Oceanic Whitetip Shark, *Carcharhinus longimanus*, Determined Using Mitochondrial DNA

Sâmia M. Camargo^{1,11}, Rui Coelho^{2,3}, Demian Chapman⁴, Lucy Howey-Jordan⁵, Edward J. Brooks⁶, Daniel Fernando^{7,8,9}, Natalia J. Mendes¹, Fabio H. V. Hazin¹⁰, Claudio Oliveira¹, Miguel N. Santos², Fausto Foresti¹, Fernando F. Mendonça^{11*}



OPEN ACCESS

Citation: Camargo SM, Coelho R, Chapman D, Howey-Jordan L, Brooks EJ, Fernando D, et al. (2016) Structure and Genetic Variability of the Oceanic Whitetip Shark, *Carcharhinus longimanus*, Determined Using Mitochondrial DNA. PLoS ONE 11(5): e0155623. doi:10.1371/journal.pone.0155623

Editor: Roberta Cimmanuta, Tuscia University, ITALY

Received: September 2, 2015

Accepted: May 2, 2016

Published: May 17, 2016

Copyright: © 2016 Camargo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The 215 nucleotide sequences that compose the data base is represented by 12 haplotypes containing 724 pairs base. All 12 files are available from the GenBank database (accession numbers KT160318 to KT160329) available in <http://www.ncbi.nlm.nih.gov/genbank/>.

Funding: Funding for north Atlantic sampling was provided by grants from The Moore Bahamas Foundation and Save our Seas Foundation. R. Coelho is supported by an Investigador-FCT contract from the Portuguese Foundation for Science and Technology (FCT, Fundação para a Ciência e

¹ Laboratório de Biologia e Genética de Peixes, Instituto de Biotecnologias de Botucatu, Universidade Estadual Paulista, (UNESP), Botucatu, São Paulo, Brasil, ² Instituto Português do Mar e da Atmosfera (IPMA), Olhão, Portugal, ³ Centro de Ciências do Mar (CCMAR), Universidade Algarve, Faro, Portugal, ⁴ School of Marine and Atmospheric Science, Stony Brook University, Stony Brook, New York, United States of America, ⁵ Microwave Telemetry, Inc., Columbia, Maryland, United States of America, ⁶ Shark Research and Conservation Program, Cape Eleuthera Institute, Eleuthera, The Bahamas, ⁷ The Manta Trust, Catemwood House, Corscombe, Dorchester, United Kingdom, ⁸ Department of Biology and Environmental Science, Linnaeus University, Lund, Sweden, ⁹ Blue Resources, Colombo, Sri Lanka, ¹⁰ Departamento de Pesca e Aquicultura, Universidade Federal Rural de Pernambuco, UFRPE, Recife, Pernambuco, Brasil, ¹¹ Laboratório de Genética Pecuária e Conservação, Instituto do Mar, Universidade Federal de São Paulo (UNIFESP), Santos, São Paulo, Brasil

* fernando.mendonca@unifesp.br

Abstract

Information regarding population structure and genetic connectivity is an important contribution when establishing conservation strategies to manage threatened species. The oceanic whitetip shark, *Carcharhinus longimanus*, is a highly migratory, large-bodied, pelagic shark listed by the IUCN (International Union for Conservation of Nature) Red List as "vulnerable" throughout its range and "critically endangered" in the western north Atlantic. In 2014, the species was protected globally under Appendix II of CITES (Convention on International Trade in Endangered Species), limiting and regulating trade. This study used partial sequences of mitochondrial DNA (mtDNA) control region to determine the population genetic structure of oceanic whitetip sharks across the Atlantic and Indian Oceans. 724 base pairs were obtained from 215 individuals that identified nine polymorphic sites and defined 12 distinct haplotypes. Total nucleotide diversity (π) was 0.0013 and haplotype diversity (h) was 0.5953. The Analysis of Molecular Variance (AMOVA) evidenced moderate levels of population structure ($\Phi_{ST} = 0.1039$) with restricted gene flow between the western and eastern Atlantic Ocean, and a strong relationship between the latter region and the Indian Ocean. Even though the oceanic whitetip is a highly migratory animal the results presented here show that their genetic variability is slightly below average of other pelagic sharks. Additionally, this study recommends that at least two populations in the Atlantic Ocean should be considered distinct (eastern and western Atlantic) and conservation

Tecnologia) supported by the EU European Social Fund and the Programa Operacional Potencial Humano (Ref. IF00253/2014). The Microwaves Telemetry, Inc. provided support in the form of salaries for author L. Howey-Jordan. This work was funded by FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo (Ref. BIOTA 2011/23787-0) and coordinated by FFM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Microwaves Telemetry Inc. provided support in the form of salaries for author Lucy Howey-Jordan, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing Interests: Lucy Howey-Jordan is employed by Microwaves Telemetry Inc. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

efforts should be focused in areas with the greatest genetic diversity by environmental managers.

Introduction

The effects of unsustainable fishing on populations of sharks and rays (elasmobranchs) has been well-documented globally [1], and studies have shown that over the last several decades many large migratory species commonly caught in large scale pelagic marine fisheries are rapidly declining [2, 3]. The *International Union for Conservation of Nature* (IUCN), which quantifies the conservation status of most taxa, currently lists over a thousand species of elasmobranchs [4, 5]. The highly migratory sharks are amongst the species with highest conservation concerns [6].

The oceanic whitetip shark, *Carcharhinus longimanus*, is a pelagic shark with worldwide distribution, found mainly in tropical and subtropical open waters [7, 8]. Historically grouped with the silky shark *C. falciformis* and the blue shark *Prionace glauca*, *C. longimanus* was once considered a highly abundant oceanic shark [7–11]. However, recent population assessments and anecdotal reports suggest that the oceanic whitetip shark is now only occasionally recorded in the Atlantic. A recent study by Coelho et al. [12] covering a wide area of the Atlantic in both hemispheres indicated that the oceanic whitetip shark bycatch in pelagic longline fisheries is less than 1% of the total elasmobranch catches. Furthermore, philopatric behavior, defined as the tendency of individuals to return or stay in their home areas, natal (birth) sites or adopted locales [13], was recently described in this species, highlighting the need for designated protective areas for oceanic whitetip sharks.

The IUCN Red List currently lists *C. longimanus* as "vulnerable" throughout its range and "critically endangered" in the western north Atlantic [14], and in 2014 was listed in Appendix II of CITES (Convention on International Trade in Endangered Species), regulating international trade and requirement of appropriate permits, including documentation validating that catch is legally sourced. A 2012 Pacific-wide stock assessment for oceanic whitetip sharks showed that the populations are currently overfished and stocks have declined to levels below maximum sustainable yield [15]. In the Atlantic and Indian Oceans, Ecological Risk Assessments for pelagic sharks, show that the oceanic whitetip shark has a relatively high vulnerability to pelagic fisheries, due to its low fecundity (average three pups per litter) and high susceptibility [16, 17]. Retaining oceanic whitetip sharks for commercial purposes has been prohibited in tuna RFMOs (Regional Fisheries Management Organizations) of all Oceans, including the ICCAT (International Commission for the Conservation of Atlantic Tunas) Atlantic convention area since 2010 (ICCAT Recommendation 10–07), the IOTC (Indian Ocean Tuna Commission) Indian Ocean convention area since 2013 (IOTC Resolution 13/06), the IATTC (Inter-American Tropical Tuna Commission) in the eastern Pacific since 2012 (Resolution C-11-10), and the WCPFC (Western and Central Pacific Fisheries Commission) in the western and central Pacific since 2013 (CMM 2011–04).

An aspect that is fundamental to sustainable use of marine resources and quantification of productivity of ocean ecosystems is the identification and maintenance of distinct stocks [18]. The conservation of the genetic variability is one of the basic objectives in regulatory programs assisting in the recovery of endangered species [19]. As such, the sequencing of the mitochondrial DNA control region (D-loop) is one of the most commonly applied method to population genetic studies of vertebrates, including sharks [20–26].

Population declines, such as that observed in the oceanic whitetip shark over the last 50 years, raises questions about the maintenance of genetic variability and the putative loss of

evolutionary lineages. These issues are more relevant in species with population structure where the distribution of genetic characteristics is geographically limited. Therefore, this study aimed to describe the genetic variability and population structure of *C. longimanus* from the Atlantic and Indian Oceans by using control region of mitochondrial DNA with an intention to provide a framework for future assessments and tracking studies.

Materials and Methods

Sampling

Samples of *C. longimanus* were collected by fishery observers on Portuguese, Brazilian and American commercial pelagic longline vessels in several regions of the Atlantic and Indian Oceans. Tissue samples, including small muscle or fin fragments ($<1\text{cm}^2$), were collected opportunistically during the normal fishing operations of the vessels, and were subsequently frozen in 95% ethanol. Tissue samples from the north Atlantic were collected in The Bahamas as part of an ongoing study of oceanic whitetip movement and life history. All samples were collected under the Cape Eleuthera Institute research permit (MAF/FIS/17 and MAF/FIS/34) and in accordance with Stony Brook University and Cape Eleuthera Institute regulations developed within the guidelines of the Association for the Study of Animal Behaviour and the Animal Behavior Society [27]. Sampling from the Portuguese fishery was conducted by the Portuguese Institute for the Ocean and Atmosphere (IPMA), within the scope of the European Data Collection Framework (PNAB/DCF). This sampling was conducted under an ICCAT permit for biological sampling of shark species in the Atlantic (ICCAT Recommendation 13/10) and an IOTC permit for biological sampling of oceanic whitetip sharks in the Indian Ocean (IOTC Resolution 13/06). No specific permissions were required for internationally transporting samples of *C. longimanus* before the species effectively became part of CITES in September 2014, all samples used in this study were collected before that date.

A total of 215 specimens were sampled for this study. Specifically, 206 were sampled from various regions of the Atlantic Ocean, including 28 from the Northwest Atlantic, 51 from the Western Equatorial Atlantic, 54 from the Northeast Tropical Atlantic, 50 from the Meso-Atlantic, 17 from the Southeast Atlantic and six from the Southwest Atlantic. In addition, nine samples were collected from the Indian Ocean (Fig 1).

DNA extraction and sequencing

Genomic DNA was extracted using the NucleoSpin® Tissue XS Kit (Macherey & Nagel, Düren, Germany). Partial sequences of the control region of the mitochondrial DNA were attained according to Mendonça et al. [28]. Individual reactions were performed with approximately 30 ng DNA template, 3.2 pmol primer, 1 μl terminator mix, and 5 μl Better Buffer (The Gel Co.) in a total volume of 15 μl . PCR sequencing profiles consisted of an initial denaturation step of 4 min at 96°C, followed by 30 cycles of 30s at 96°C, 15s at 50°C, and 4 min at 60°C. Cycle sequencing was performed with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing was completed on an automated sequencer, ABI 3130, manufactured by Applied Biosystems. The consensus sequences were assembled and edited using Geneious 4.8.5 [29] and aligned with Muscle algorithm implemented within Geneious 4.8.5. The generated haplotypes were deposited in GenBank under accession numbers KT160318 to KT160329.

Population genetics analyses

The relative nucleotide composition, number of polymorphic sites, haplotype diversity (h), nucleotide diversity (π), and number of pairwise nucleotide differences among populations

were calculated using ARLEQUIN 3.5.1.3 software [30]. To estimate levels of genetic divergence among populations of *C. longimanus* index Φ_{ST} was calculated using the Analysis of Molecular Variance (AMOVA) [31] under the nucleotide evolution model of Tamura-Nei [32]. The Φ_{ST} estimates were tested with 1,000 non-parametric bootstrap pseudo-replicates using the ARLEQUIN 3.5.1.3 software, and the p-values were adjusted for simultaneous pairwise comparisons using the sequential Bonferroni correction [33]. The AMOVA test considered the possibilities of the maximum and minimum number of populations using hypothetical groups of population structure. Groupings that were most and least inclusive were tested, first by geographic proximity of the collection points, considering maritime currents; and secondly, by using the geographical distribution and correlation between the North and South, and East and West Atlantic. A minimum-spanning haplotype network was estimated using the Network 4.6.1.1 program [34]. The mismatch analysis in ARLEQUIN 3.5.1.3 included a raggedness index to determine goodness-of-fit to a unimodal distribution [35].

Results

Among the 215 specimens of *C. longimanus* analyzed, we obtained a total matrix with 724 bp with 9 polymorphic sites, yielding 12 haplotypes (Table 1).

In the overall analyses, we found low relative levels of haplotype diversity $h = 0.5953$ and nucleotide diversity $\pi = 0.0013$, with the greatest diversities found in the Southwest Atlantic ($h = 0.8000$, $\pi = 0.00184$) and in the Meso-Atlantic ($h = 0.7571$, $\pi = 0.00192$). Haplotype numbers three (H3 = 131 individuals) and one (H1 = 32 individuals) were the most common and were found in all sampled regions, representing 75.8% of the analyzed oceanic whitetip shark specimens. Numbers of polymorphic sites, haplotypes, nucleotide diversity and haplotypic diversity for different samples are shown in Tables 2 and 3.

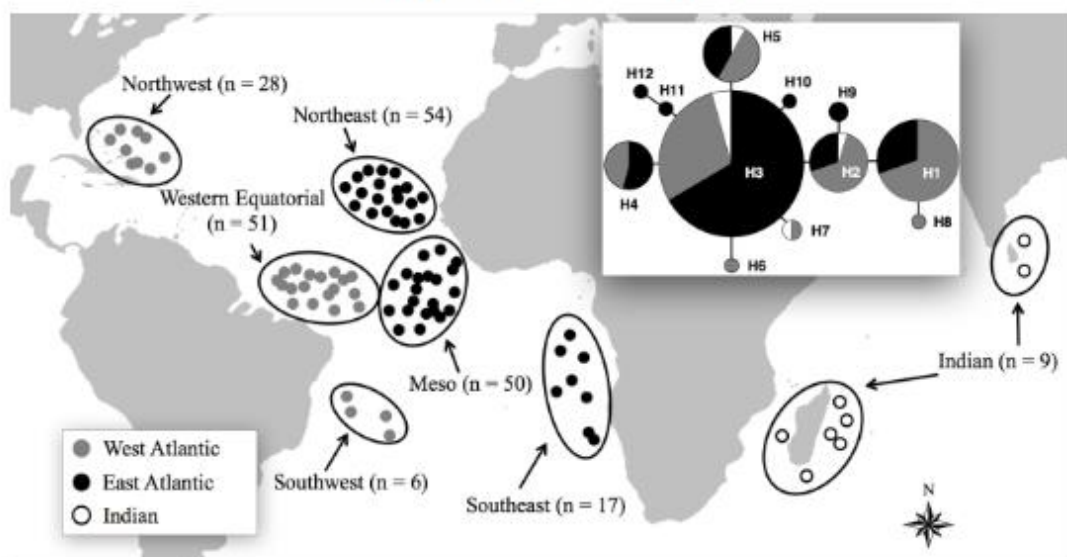


Fig 1. Geographic distribution of samples of *Carcharhinus longimanus* with the network haplotypes analyzed and compiled from the sequences of the mitochondrial DNA control region.

doi:10.1371/journal.pone.0155623.g001

Table 1. Polymorphisms found in the twelve haplotypes of *Carcharhinus longimanus* along several regions of the Atlantic and Indian Oceans.

Haplotypes	Polymorphic sites								
	1	2	2	2	2	2	3	4	5
	9	1	3	5	7	9	9	0	2
	2	0	0	0	6	4	9	0	9
1	G	C	T	A	T	G	T	C	T
2	-	-	-	-	-	-	C	-	-
3	-	-	C	-	-	-	C	-	-
4	-	-	C	-	-	A	C	-	-
5	-	-	C	-	C	-	C	-	-
6	-	-	C	G	-	-	C	-	-
7	-	-	C	-	-	-	C	-	C
8	-	-	-	-	C	-	-	-	-
9	-	-	-	-	-	A	C	-	-
10	A	-	C	-	-	-	C	-	-
11	-	T	C	-	-	-	C	T	-
12	-	T	C	-	C	-	-	T	-

doi:10.1371/journal.pone.0155623.t001

The AMOVA test was adopted following hypothetical groups of population structure. While examining different scenarios of structuring, a greater difference ($\Phi_{ST} = 0.1039$, $P < 0.001$) was observed when populations were grouped in two geographic regions: Eastern (Northeast Atlantic, Southeast Atlantic and Meso-Atlantic) and Western (Northwest Atlantic, Western Equatorial Atlantic and Southwest Atlantic) Atlantic Ocean. In this case, the indices related to genetic variability were $h = 0.716$ and $\pi = 0.00165$ in West Atlantic, and $h = 0.471$ and $\pi = 0.00104$ in the East Atlantic. Due to the low sampling rate in the Indian Ocean, the AMOVA did not detect a statistically significant structure when left as a distinct group. However, in the F_{ST}

Table 2. Geographical distribution of haplotypes of *Carcharhinus longimanus* delimited by the different areas of the Atlantic and Indian Oceans.

	West Atlantic				East Atlantic				Total (n = 130)
	Northwest Atlantic (n = 28)	West. Eq. Atlantic (n = 51)	Southwest Atlantic (n = 6)	Total (n = 85)	Meso-Atlantic (n = 50)	Northeast Atlantic (n = 34)	Southeast Atlantic (n = 17)	Indian Ocean (n = 9)	
h1	8	13	1	22	8	1	1	-	10
h2	4	7	-	11	3	1	1	1	5
h3	13	22	3	38	22	50	15	6	91
h4	1	3	1	5	0	-	-	-	6
h5	1	4	1	6	7	1	-	1	9
h6	1	-	-	1	-	-	-	-	-
h7	-	1	-	1	-	-	-	1	1
h8	-	1	-	1	-	-	-	-	1
h9	-	-	-	-	1	1	-	-	2
h10	-	-	-	-	1	-	-	-	1
h11	-	-	-	-	1	-	-	-	1
h12	-	-	-	-	1	-	-	-	1

n, haplotype.

doi:10.1371/journal.pone.0155623.t002

Table 3. Population statistics of *Carcharhinus longimanus*—n, number of individuals; P, polymorphic sites; Nh, number of haplotypes; h, haplotype diversity; π , nucleotide diversity.

	n	P	Nh	h	π
Northwest Atlantic	28	5	6	0.7037	0.00177
Western Equatorial Atlantic	51	5	7	0.7431	0.00170
Southwest Atlantic	6	4	4	0.8000	0.00184
West Atlantic	85	6	8	0.716	0.00165
Meso-Atlantic	50	7	9	0.7567	0.00192
Northeast Atlantic	54	4	5	0.1440	0.00034
Southeast Atlantic	17	2	3	0.2279	0.00047
Indian Ocean	9	3	4	0.5833	0.00092
East Atlantic	130	8	10	0.471	0.00104

doi:10.1371/journal.pone.0155623.t003

pairwise analysis an absence of structure was observed between the Indian Ocean and the East Atlantic groups (Table 4).

A statistical parsimony haplotype network was constructed using the Network 4.611 program. More than 60% analyzed sequences shared a single haplotype (H3), which was found in all sampled regions (Fig 1).

Discussion

Population genetics of the oceanic whitetip shark

The results show low levels of genetic diversity for the oceanic whitetip shark in the Atlantic and also in the Indian Ocean, although the latter has been evaluated with a small number of individuals and in a relatively small area. Population genetic studies with marine vertebrates, such as bony fishes, cetaceans and other elasmobranchs (including pelagic sharks) also tend to report low values of nucleotide and haplotypic diversity in the Atlantic Ocean [36–39]. Although low levels of nucleotide diversity are not considered standard for elasmobranchs, similar results to those obtained for oceanic whitetip sharks have been found in other highly migratory pelagic species including basking shark (*Cetorhinus maximus*), crocodile shark (*Pseudocarcharias kamoharui*), the whale shark (*Rhincodon typus*), pelagic thresher shark (*Alopias pelagicus*) and the white shark (*Carcharodon carcharias*) [25, 26, 38, 40–44]. The combination of low haplotypic diversity with a single haplotype shared by most of the individuals is also found in the nurse shark (*Ginglymostoma cirratum*) [39], white shark (*Carcharodon carcharias*) [43] and sicklefin lemon shark (*Negaprion acutidens*) [44].

Table 4. Differentiation Index (F_{ST}) between pairs of sampled regions in the Atlantic and Indian Oceans. Pairwise F_{ST} numbers are below the diagonal line, and the significance of the p-values are represented above the diagonal (+ represents a statistically significant difference with p-value < 0.01).

	Northwest Atlantic	Western Eq. Atlantic	Southwest Atlantic	Mezo-Atlantic	Northeast Atlantic	Southeast Atlantic	Indian Ocean
Northwest Atlantic	0	-	-	-	+	+	-
Western Eq. Atlantic	-0.0228	0	-	-	+	+	-
Southwest Atlantic	-0.0005	-0.0224	0	-	-	-	-
Mezo-Atlantic	0.0309	0.0201	-0.0943	0	+	-	-
Northeast Atlantic	0.2716	0.2085	0.1352	0.0965	0	-	-
Southeast Atlantic	0.1170	0.0949	0.0081	0.0290	-0.0137	0	-
Indian Ocean	0.1092	0.0814	-0.0623	0.0019	0.0237	-0.0224	0

doi:10.1371/journal.pone.0155623.t004

Considering two distinct groups for the oceanic whitetip shark in the Atlantic Ocean and based on the hypothetical population groupings and genetic structuring simulations, restrictions to gene flow with moderate genetic divergence were observed, with one population occurring in the West and another in the East Atlantic. Population structure between the eastern and western portions of an oceanic basin was observed by Cardenosa et al [41] for the thresher shark with strong differentiation in the Pacific. In the latest study on population genetic structure of the white shark, presented by O'Leary et al. [43], a similar structure was observed to the one in our study, being the $\Phi_{ST} = 0.10$ with sequences of mtDNA control region and $F_{ST} = 0.1057$ with microsatellite marker, between the northwest Atlantic and southern Africa.

It would not be surprising to observe panmictic populations given the migration potential of this species. However, ocean currents and their interactions produce the oceanic barriers, organismal dispersal potential, behavior and organismal environmental tolerances which all contribute to patterns of diversification in the oceans [45, 46]. Scientific findings indicate that the western Atlantic is an area of both origin and accumulation of biodiversity [47]. This is supported by population growth rates of *C. longimanus* as rates of haplotype and nucleotide diversity are approximately 35% higher in western Atlantic. The haplotype network in star contraction generated from the haplotype H3 suggests an expansion event, possibly from the individuals of the eastern Atlantic, where the highest rates were observed in diversity with higher frequency on the H3.

Currently, very little is known about the movements and habitat use of oceanic whitetips in the Atlantic. The most comprehensive movement study of this species used Pop-up satellite archival tags applied to 11 mature females and one mature male, near Cat Island in the central Bahamas. In that study, the maximum individual displacement from the tagging site dispersed 290–1940 km after 30–245 days at liberty, with individuals moving to several different destinations. The tag affixed to the male shark never reported. All of the tagged oceanic whitetips remained within 500 km of their tagging location for the first 30 days of their tracks and none of them moved away from the northwestern Atlantic [27].

The shortest route between the western and eastern Atlantic is between Brazil and Guinea-Bissau, requiring an oceanic crossing measuring ~2400 km. Although this distance does not seem to be a barrier to potential oceanic whitetip migration, such a trans-Atlantic route does not seem to be an aspect of the female behavior of the species. In those regions the oceanic whitetip shark populations in the western and east Atlantic are genetically differentiated from one another at mitochondrial loci, at least by female lineages. Philopatry is one of the factors that could also influence the reduction of gene flow across the Atlantic, by the migration of females to their original birth place [27]. Philopatry is well documented in many other sharks reviewed in Chapman et al. [48] and trans-Atlantic structure may have developed in oceanic whitetips because females remain within or return to give birth on one side of the basin or the other. A survey of biparentally inherited genetic markers is needed to determine whether or not there is male-mediated gene flow across the Atlantic.

The haplotype network denotes genetic differences among individuals without showing correlations between the different branches and their geographic distributions, with a characteristic pattern of an unstructured population. Although dispersion of fishes between the Indian and Atlantic Oceans are rarely observed, mtDNA data show that some species of tropical and sub-tropical fish found in the Indian Ocean may cross the barrier of the Benguela current [49]. The water transport between the Indian and Atlantic Oceans is summarized in [50]. A strong genetic connectivity between the Indian Ocean group and the Eastern Atlantic groups through the pairwise F_{ST} , despite low statistical significance, highlight the dispersal capability of *C. longimanus*. Considering the movement of individuals between Indian and eastern Atlantic, the absence of significant genetic structure can indicate the existence of only one genetic stock of

oceanic whitetip sharks around the African continent, both in the eastern Atlantic and western Indian Ocean. The warm Agulhas current passes the Cape of Good Hope and is incorporated into the cool Benguela current. The Agulhas current is one of the strongest currents in the world, and it may help the transport of oceanic whitetip sharks from the southwest Indian Ocean to the Southeast Atlantic. However, the question on whether the flow of these sharks can also occur in the reverse order remains unknown.

More genetic studies and their importance in the evolutionary history of elasmobranchs with oceanic distribution, as well as for the management of commercially exploited shark species, are needed. Phylogeographic studies presenting extensive sampling and appropriate geographic scales including habitats with different ecological characteristics (such as oceanic and continental, tropical or subtropical) have a greater chance of detecting evolutionarily significant units, which would facilitate the development and application of appropriate conservation measures [51].

Conservation of the oceanic whitetip shark

Though shark populations are often impacted by overfishing, pelagic species, including *C. longimanus*, ubiquitous in all tropical oceans, with its susceptibility to incidental longline capture, make it prone to population declines, prompting the IUCN to designate it "vulnerable". Considering the important relationship between the sustainability of a natural population and their genetic diversity, there is an urgent need to uncover information about global levels of genetic diversity and geographical distribution as well as patterns of gene flow and the detection of significant evolutionary units.

Here we present evidence that two populations of oceanic whitetip sharks exist in the Atlantic, specifically one in the western and another in the eastern regions of the basin. Despite their highly migratory nature, barriers to the gene flow of oceanic whitetips in the Atlantic result in two genetically distinct and demographically independent populations. This structure should be incorporated into assessments and monitoring of this species. The factors that restrict gene flow in the Atlantic may be absent between the east Atlantic and parts of the Indian Ocean; as there appears to be connectivity between those two regions. Isolated populations, especially if reduced by fishing or other sources of mortality, are vulnerable to inbreeding as reproducing adults have an elevated probability of mating with relatives [52]. Nonrandom matings generate deviations in allele frequencies with heterozygous deficiencies. Therefore, resource managers, tasked with the conservation of this species, should recognize that intrinsic reproductive barriers need to be considered concurrently with immigration rates, for sustaining at-risk populations. In addition to the low levels of genetic variability found for whitetip sharks throughout the study area, an important difference was observed between the two populations, with the haplotype diversity of the eastern population 34.2% less than the western population. This difference was even more striking for the nucleotide diversity with 36.9% less in the eastern population. These low genetic variability rates found may represent a dramatic risk to the adaptive potential of the species leading to a weaker ability to respond to environmental changes, and consequently, could promote extinction of some lineages in the future. Describing differences in indices of genetic diversity allows for managers to prioritize conservation efforts between populations with the highest diversity index, maintaining evolutionary potential, or populations with the lowest genetic variability, in an effort to prevent further inbreeding and reduced reproductive success. In order to prevent further population declines, we suggest that global genetic diversity should be preserved for all haplotypes through multinational cooperation, and particularly for the two distinct populations of oceanic whitetips identified in the Atlantic.

Acknowledgments

Sampling from the Portuguese fishery in the Atlantic and Indian Oceans was conducted by the Portuguese Institute for the Ocean and Atmosphere (IPMA), Portugal, within the scope of the European Data Collection Framework (PNAB/DCF). The authors thank all collaborators who contributed with samples for this study, particularly Annabelle Brooks, Sean Williams, Debra Abercrombie, Lance Jordan and Mark Bond. Funding for north Atlantic sampling was provided by grants from The Moore Bahamas Foundation and Save our Seas Foundation. Thanks are also due to the technicians from IPMA for collecting and carefully handling samples during the Portuguese pelagic longline fishery observer program. R. Coelho is supported by an Investigador-FCT contract from the Portuguese Foundation for Science and Technology (FCT, *Fundação para a Ciência e Tecnologia*) supported by the EU European Social Fund and the *Programa Operacional Potencial Humano* (Ref: IF/00253/2014). The Microwave Telemetry, Inc. provided support in the form of salaries for author L. Howey-Jordan. This work was funded by FAPESP—Fundação de Amparo à Pesquisa do Estado de São Paulo (Ref: BIOTA 2011/23787-0) and coordinated by FFM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: SMC RC CO FF FFM. Performed the experiments: SMC NJM. Analyzed the data: SMC RC FFM. Contributed reagents/materials/analysis tools: SMC RC DC EJB DF FHVH CO MNS FF FFM. Wrote the paper: SMC RC DC LHJ EJB NJM FHVH CO MNS FF FFM.

References

1. Stevens JD, Bonfil R, Dulvy NK, Walker PA. The effects of fishing on sharks, rays, and chimaerids (chondrichthyan), and the implications for marine ecosystems. *ICES Journal of Marine Science: Journal du Conseil*. 2000; 57(3):476–94.
2. Baum JK, Myers RA, Kehler DG, Worm B, Harley SJ, Doherty PA. Collapse and conservation of shark populations in the Northwest Atlantic. *Science*. 2003; 299(5605):389–92. PMID: 12532010
3. Burgess GH, Beerkircher LR, Cailliet GM, Carlson JK, Cortés E, Goldman KJ, et al. Is the collapse of shark populations in the Northwest Atlantic Ocean and Gulf of Mexico real? *Fisheries*. 2005; 30(10):19–26.
4. Camhi M, Fowler S, Musick J, Bräutigam A, S. F. Sharks and their relatives. 1998. 20 p.
5. IUCN. The IUCN Red List of Threatened Species 2014 [cited Version 2014.3].
6. Dulvy NK, Baum JK, Clarke S, Compagno LJ, Cortés E, Domingo A, et al. You can swim but you can't hide: the global status and conservation of oceanic pelagic sharks and rays. *Aquatic Conservation: Marine and Freshwater Ecosystems*. 2008; 18(5):459.
7. Backus RH, Springer S, Arnold EL. A contribution to the natural history of the white-tip shark, *Pterolamiops longimanus* (Poey). *Deep Sea Research* (1953). 1956; 3(3):178–88.
8. Bonfil R. Overview of world elasmobranch fisheries: Food & Agriculture Org.; 1994.
9. Castro JL, Woodley CM, Brudek RL. A preliminary evaluation of the status of shark species: Food & Agriculture Org.; 1999.
10. Compagno LJV. Species Catalogue. Sharks of the World. A, Annotated and Illustrated Catalogue of Shark Species Known to Date. Part 2. Carcharhiniformes. 1984. 251–635 p.
11. Hoerig J, Gruber S, Pratt H, Taniuchi T. Elasmobranchs as living resources: advances in the biology, ecology, systematics, and the status of the fisheries. Elasmobranchs as living resources: advances in the biology, ecology, systematics, and the status of the fisheries. 1990.
12. Coelho R, Fernandez-Carvalho J, Lino PG, Santos MN. An overview of the hooking mortality of elasmobranchs caught in a swordfish pelagic longline fishery in the Atlantic Ocean. *Aquatic Living Resources*. 2012; 25(04):311–9.
13. Mayr E. Animal species and evolution. Animal species and their evolution. 1963.

14. Baum J, Medina E, M MJS. *Carcharhinus longimanus*. IUCN Red List of Threatened Species. 2006; IUCN 2010 (Version 2010.1).
15. Rice J, Harley S. Stock assessment of oceanic whitetip sharks in the western and central Pacific Ocean. Scientific Committee Eighth Regular Session of WCPFC, Busan Republic of Korea. 2012.
16. Cortés E, Domingo A, Miller P, Forselledo R, Mas F, Arocha F, et al. K. Yokawa. 2012. Expanded ecological risk assessment of pelagic sharks caught in Atlantic pelagic longline fisheries. Report of the Standing Committee on Research and Statistics (SCRS), ICCAT, Madrid, Spain, October 1–5, 2012. SCRS/2012/167.
17. Munua H, Cohelo R, Neves M, Artzabalaga H, Yokawa K, Romanov E, et al. Preliminary Ecological Risk Assessment (ERA) for shark species caught in fisheries managed by the Indian Ocean Tuna Commission (IOTC). IOTC Technical Paper. 2012; 1:1–26.
18. Carvalho GR, Hauser L. Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology Fisheries 1994; 4:326–50.
19. Lindenmayer D, Lacy R. A simulation study of the impacts of population subdivision on the mountain brushtail possum *Trichosurus caninus* Ogilby (Phalangeridae: Marsupialia) in south-eastern Australia. I. Demographic stability and population persistence. Biological Conservation. 1995; 73(2):119–29.
20. Clarke CR, Karl SA, Horn RL, Bernard AM, Lea JS, Hazin FH, et al. Global mitochondrial DNA phylogeography and population structure of the silky shark, *Carcharhinus falciformis*. Marine Biology. 2015; 162(5):945–55.
21. Mendonça FF, Oliveira C, Gadig OBF, Foresti F. Diversity and genetic population structure of the Brazilian sharpnose shark *Rhizoprionodon lalandi*. Aquatic Conservation-Marine and Freshwater Ecosystems. 2013; 23(6):850–7.
22. Karl SA, Castro ALF, Lopez JA, Charvet P, Burgess GH. Phylogeography and conservation of the bull shark (*Carcharhinus leucas*) inferred from mitochondrial and microsatellite DNA. Conservation Genetics. 2011; 12(2):371–82.
23. Portnoy DS, McDowell JR, Heist EJ, Musick JA, Graves JE. World phylogeography and male-mediated gene flow in the sandbar shark, *Carcharhinus plumbeus*. Molecular Ecology. 2010; 19(10):1994–2010. doi: 10.1111/j.1365-294X.2010.04626.x PMID: 20406387
24. Overden JR, Kashiwagi T, Broderick D, Giles J, Salini J. The extent of population genetic subdivision differs among four co-distributed shark species in the Indo-Australian archipelago. BMC Evolutionary Biology. 2009; 9.
25. Hoelzel AR, Shivji MS, Magnussen J, Francis MP. Low worldwide genetic diversity in the basking shark (*Cetorhinus maximus*). Biol Lett. 2006; 2(4):639–42. PMID: 17148309
26. Ferrelle BLD, Mendonça FF, Coelho R, de Oliveira PGV, Hazin FHV, Romanov EV, et al. High Connectivity of the Crocodile Shark between the Atlantic and Southwest Indian Oceans: Highlights for Conservation. Plos One. 2013; 10(2).
27. Howey-Jordan LA, Brooks EJ, Abercrombie DL, Jordan LK, Brooks A, Williams S, et al. Complex movements, philopatry and expanded depth range of a severely threatened pelagic shark, the oceanic whitetip (*Carcharhinus longimanus*) in the western North Atlantic. PLoS One. 2013; 8(2):e56588. doi: 10.1371/journal.pone.0056588 PMID: 23437180
28. Mendonça FF, Oliveira C, Gadig OBF, Foresti F. Populations analysis of the Brazilian Sharpnose Shark *Rhizoprionodon lalandi* (Chondrichthyes: Carcharhinidae) on the São Paulo coast, Southern Brazil: inferences from mt DNA sequences. Neotropical Ichthyology. 2009; 7(2):213–6.
29. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12):1647–9. doi: 10.1093/bioinformatics/bts199 PMID: 22543367
30. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources. 2010; 10(3):564–7. doi: 10.1111/j.1755-0998.2010.02847.x PMID: 21565059
31. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992; 131(2):479–91. PMID: 1644282
32. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution. 1993; 10(3):512–26. PMID: 8336541
33. Rice WR. Analyzing tables of statistical tests. Evolution. 1989:223–5.
34. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution. 1999; 16(1):37–48. PMID: 10331250

35. Harpending H. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human biology*. 1994;391–600. PMID: [8088750](#)
36. Bradman H, Grewe P, Appleton B. Direct comparison of mitochondrial markers for the analysis of swordfish population structure. *Fisheries Research*. 2011; 109(1):35–9.
37. Mesnick SL, Taylor BL, Archer FI, Martien KK, Trevino SE, Hancock-Hanser BL, et al. Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. *Molecular Ecology Resources*. 2011; 11(51):278–98.
38. Phillips NM, Chaplin JA, Morgan DL, Peverell SC. Population genetic structure and genetic diversity of three critically endangered *Pristis* sawfishes in Australian waters. *Marine Biology*. 2011; 158(4):903–15.
39. Karl SA, Castro AL, Garia RC. Population genetics of the nurse shark (*Ginglymostoma cirratum*) in the western Atlantic. *Marine Biology*. 2012; 159(3):489–98.
40. Schmidt JV, Schmidt CL, Ozer F, Ernst RE, Feldheim KA, Ashley MV, et al. Low genetic differentiation across three major ocean populations of the whale shark, *Rhincodon typus*. *PLoS One*. 2009; 4(4): e4988. doi: [10.1371/journal.pone.0004988](#) PMID: [19352489](#)
41. Cardenaosa D, Hyde J, Caballero S. Genetic Diversity and Population Structure of the Pelagic Thresher Shark (*Alopias pelagicus*) in the Pacific Ocean: Evidence for Two Evolutionarily Significant Units. 2014.
42. Castro A, Stewart B, Wilson S, Hueter R, Meekan M, Motta P, et al. Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*. 2007; 16(24):3183–92. PMID: [18092992](#)
43. O'Leary SJ, Feldheim KA, Fields AT, Natanson LJ, Wintner S, Hussey N, et al. Genetic Diversity of White Sharks, *Carcharodon carcharias*, in the Northwest Atlantic and Southern Africa. *J Hered*. 2015.
44. Schultz JK, Feldheim KA, Gruber SH, Ashley MV, McGovern TM, Bowen BW. Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Mol Ecol*. 2008; 17(24):5336–48. doi: [10.1111/j.1365-254X.2008.04000.x](#) PMID: [19121001](#)
45. Joyeux JC, Floeter SR, Ferreira CEL, Gasparini JL. Biogeography of tropical reef fishes: the South Atlantic puzzle. *Journal of Biogeography*. 2001; 28(7):831–41.
46. Hellberg ME. Gene flow and isolation among populations of marine animals. 2009.
47. Rocha LA, Bowen BW. Speciation in coral-reef fishes. *Journal of Fish Biology*. 2008; 72(5):1101–21.
48. Chapman DD, Feldheim KA, Papastamatiou YP, Hueter RE. There and Back Again: A Review of Residency and Return Migrations in Sharks, with Implications for Population Structure and Management. *Annual review of marine science*. 2015; 7:547–70. doi: [10.1146/annurev-marine-010814-015730](#) PMID: [25251267](#)
49. Rocha LA, Robertson DR, Rocha CR, Van Tassell JL, Craig MT, Bowen BW. Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Mol Ecol*. 2005; 14(13):3921–8. PMID: [16262848](#)
50. Tomczak M, Godfrey JS. Regional oceanography: an Introduction: Elsevier; 2013.
51. Rocha L, Craig M, Bowen B. Phylogeography and the conservation of coral reef fishes. *Coral Reefs*. 2007; 26(3):301–12.
52. Wright S. Evolution and the genetics of populations, volume 3: experimental results and evolutionary deductions: University of Chicago press; 1984.

Apêndice III

ORIGINAL ARTICLE

Evolution and conservation of *Characidium* sex chromosomes

R Utsunomia^{1,8}, PC Scacchetti^{1,8}, M Hermida², R Fernández-Cebrián³, X Taboada⁴, C Fernández², M Bekaert⁵, NJ Mendes¹, D Robledo^{4,6}, JE Mank⁷, JB Taggart⁵, C Oliveira¹, F Foresti¹ and P Martínez²

Fish species exhibit substantial variation in the degree of genetic differentiation between sex chromosome pairs, and therefore offer the opportunity to study the full range of sex chromosome evolution. We used restriction-site associated DNA sequencing (RAD-seq) to study the sex chromosomes of *Characidium gomesi*, a species with conspicuous heteromorphic ZW/ZZ sex chromosomes. We screened 9863 single-nucleotide polymorphisms (SNPs), corresponding to ~1 marker/100 kb distributed across the genome for sex-linked variation. With this data set, we identified 26 female-specific RAD loci, putatively located on the W chromosome, as well as 148 sex-associated SNPs showing significant differentiation (average $F_{ST}=0.144$) between males and females, and therefore in regions of more recent divergence between the Z and W chromosomes. In addition, we detected 25 RAD loci showing extreme heterozygote deficiency in females but which were in Hardy–Weinberg equilibrium in males, consistent with degeneration of the W chromosome and therefore female hemizyosity. We validated seven female-specific and two sex-associated markers in a larger sample of *C. gomesi*, of which three localised to the W chromosome, thereby providing useful markers for sexing wild samples. Validated markers were evaluated in other populations and species of the genus *Characidium*, this exploration suggesting a rapid turnover of W-specific repetitive elements. Together, our analyses point to a complex origin for the sex chromosome of *C. gomesi* and highlight the utility of RAD-seq for studying the composition and evolution of sex chromosomes systems in wild populations.

Heredity (2017) 119, 237–244; doi:10.1038/hdy.2017.43; published online 26 July 2017

INTRODUCTION

Sex chromosomes are usually defined as a pair of chromosomes which carry the sex determination (SD) locus, and they typically occur as either male (XX/XY) or female (ZW/ZZ) heterogamety in diploid organisms (Martínez *et al.*, 2014). Sex chromosome divergence is often assumed to result from selection against recombination between the X and Y or Z and W chromosomes in order to maintain the association between a sex-determining allele and a sexually antagonistic allele at a locus in close proximity (Bull, 1983). The loss of recombination on the sex-limited W or Y can result in degenerative processes, such as the accumulation of recessive deleterious mutations and repetitive DNA elements (Ellegren, 2011). If allowed to progress for a sufficient period of time, this process can produce major differences in size and gene content, referred to as sex chromosome heteromorphy (Ohno, 1967; Bull, 1983).

In contrast to mammals and birds, which show highly heteromorphic sex chromosomes that are conserved within each clade, fish exhibit a wide variety of non-orthologous sex chromosomes that have

emerged independently throughout their evolution. This is reflected both by the rapid rate of sex chromosome turnover as well as significantly reduced heteromorphism compared to mammals and birds (Mank and Avise, 2009; Bachtrog *et al.*, 2014). This rapid turnover has produced many examples of different sex chromosome systems between congeneric species, and even different systems within species (Martínez *et al.*, 2014).

Neotropical fish in particular show a remarkable diversity of SD mechanisms, and this diversity offers a natural laboratory to explore and test evolutionary hypotheses about the origin and evolution of sex chromosomes. Different sex chromosome systems have been described by classical and molecular cytogenetics, including conspicuous sex chromosome heteromorphisms (Oliveira *et al.*, 2009), however, aside from recent work in *Pocilia reticulata* (Wright *et al.*, 2017), sequencing-based studies on sex chromosome evolution have not yet been conducted in Neotropical species.

Characidium, a genus within the order Characiformes, shows a wide Neotropical distribution. All cytogenetic studies on this genus have

¹Departamento de Morfologia, Instituto de Biociências, UNESP, Botucatu, Brazil; ²Faculdade de Veterinária, Departamento de Zootecnia, Xenética e Antropologia Física, Universidade de Santiago de Compostela, Campus Terra, Lugo, Spain; ³GENEAQUA S.L., Rúa Carligas e Flores, Lugo, Spain; ⁴Faculdade de Biologia, Departamento de Zootecnia, Xenética e Antropologia Física, Universidade de Santiago de Compostela, Santiago de Compostela, Spain; ⁵Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling, UK; ⁶The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, Scotland and ⁷Department of Genetics, Evolution and Environment, The Darwin Building, London, UK

Correspondence: Professor F Foresti, Departamento de Morfologia, Instituto de Biociências, UNESP, Botucatu 13618-689, Brazil.
E-mail: fforesti@ibb.unesp.br

or Professor P Martínez, Faculdade de Veterinária, Departamento de Zootecnia, Xenética e Antropologia Física, Universidade de Santiago de Compostela, Campus Terra, 27002 Lugo, Spain.

E-mail: paulino.martinez@usc.es

⁸These two authors contributed equally to this work.

Received 30 March 2017; revised 23 June 2017; accepted 26 June 2017; published online 26 July 2017

thus far revealed a conserved diploid chromosome number ($2n = 50$), and most analyses have revealed visibly heteromorphic female heterogametic sex chromosomes (Scacchetti *et al.*, 2015), suggesting that the ZW/ZZ chromosome system of *Characidium* originated once in the ancestor of the genus (Panosato-Alves *et al.*, 2014). Following this putative single origin, the sex chromosomes in *Characidium* diversified among different species and populations, as there is significant inter- and intraspecific cytogenetic variation in size and heteromorphism (Scacchetti *et al.*, 2015).

In order to characterise the degree of divergence of the sex chromosomes, as well as variation across populations and related species, we used restriction-site associated DNA sequencing (RAD-seq; Baird *et al.*, 2008), which permits the simultaneous discovery and robust scoring of large numbers of single-nucleotide polymorphisms (SNPs) across many individuals. This approach now provides marker densities appropriate for detailed studies of sex chromosomes characterisation and its evolution in fish (Martinez *et al.*, 2014; Pan *et al.*, 2016) and RAD-seq has been applied for the identification of SD regions through linkage analysis in model and aquaculture (Robledo *et al.*, 2017) as well as wild (Wilson *et al.*, 2014; Böhne *et al.*, 2016) species. Results presented here provide new information into the differentiation of the W chromosome and its variation across populations and species within the genus *Characidium*.

MATERIALS AND METHODS

Biological material

Sexually mature adults of *C. gomesi* were collected in the Água da Madalena tributary, Paranapanema River Basin Brazil (PR onwards; Figure 1). Gonad inspection under light microscope and cytogenetic analyses (Figure 2) were used to sex individuals and to check for convergence between histological and cytogenetic data. In order to minimise false positives, a total of 21 females and 18 males of *C. gomesi* from the PR population were used to construct RAD-seq libraries.

DNA from additional sexed specimens from the PR population (14 females and 7 males) and from another distant population (10 males and 10 females, Alumbari River, Tietê River Basin, Brazil; TR onwards; Figure 1) were analysed to validate the identified sex-associated markers and to check for their intraspecific conservation. Furthermore, specimens from two other *Characidium* species, *C. zebra* (5 males and 5 females), a basal *Characidium* species which lacks sex chromosome heteromorphism, and *C. pterostictum* (5 males and 5 females), a distant relative of *C. gomesi* with significant sex chromosome heteromorphism (Panosato-Alves *et al.*, 2014), were used for testing trans-specific conservation of sex-associated markers identified in *C. gomesi*.

All samples were collected in accordance with Brazilian environmental protection legislation (collection permission MMA/IBAMA/SISBIO—number 3245), and the procedures for sampling, maintenance and analysis of the specimens were performed in compliance with the Brazilian College of Animal Experimentation (COBEA) procedures and was approved (protocol 595) by the Bioscience Institute/UNESP Ethics Committee on use of animals (CEUA).

RAD sequencing and SNP genotyping

Genomic DNA from 21 females and 18 males of *C. gomesi* was extracted using the NucleoSpin Tissue Kit (Macherey-Nagel) and treated with RNase to remove residual RNA from the samples. DNA quantity and quality were evaluated by fluorescence (Qubit) and agarose gels prior to library construction.

In order to help achieve an even representation of sequenced individuals, two RAD libraries were prepared, one with 20 individuals, the second with 19 individuals, both with approximately equal numbers of each sex. The RAD library preparation protocol, including the design of RAD-specific P1 and P2 paired-end adaptors and library amplification PCR primer sequences, followed Houston *et al.* (2012). In brief, each sample ($n = 39$; 200 ng DNA) was individually digested with 1.6 U SbfI high fidelity restriction enzyme (New England Biolabs; NEB) in 1 × Reaction Buffer 4 (NEB) at 37 °C for 45 min. The reactions (10 µl final volumes) were then heat inactivated at 65 °C for 20 min. Individual specific P1 adaptors, each with a unique 5 or 7 base barcode (Supplementary Table S1) were ligated to the SbfI digested DNA, at 22 °C for 90 min, by adding 0.5 µl 100 nM P1 adaptor, 0.12 µl 100 mM rATP (Promega), 0.2 µl 10 × Reaction Buffer 2 (NEB), 0.1 µl T4 ligase (NEB, 2 U ml⁻¹) and reaction volumes made up to 12 µl volume with nuclease-free water. Following ligation, the samples were heat inactivated at 65 °C for 20 min, cooled to room

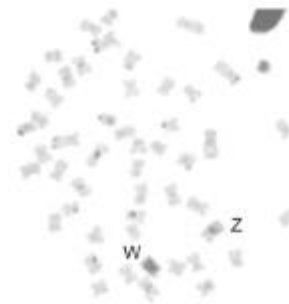


Figure 2 Mitotic plate of *Characidium gomesi* female after C-banding, showing a strong C-band positive pattern on W chromosome.

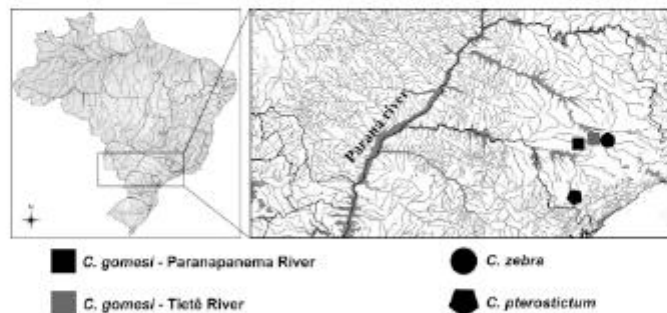


Figure 1 Geographic map of sampling sites for *Characidium* species and populations used in this study.

temperature, then combined into one or other of two library pools. Shearing (Covaris S2 sonication) and initial size selection (c. 200–300 bp) by agarose gel separation of both library pools was followed by gel purification, end repair, dA overhang addition, P2 paired-end adaptor ligation and library amplification. An equimolar combination of two P2 adaptors with 5 and 6 base barcodes (1 µl of 10 µM P2 adaptor mix per library) was used to identify each library. A total of 150 µl of each amplified library (16 PCR cycles) was prepared and size selected (c. 320–650 bp) by gel electrophoresis. Following a final gel elution step into 20 µl EB buffer (MinElute Gel Purification Kit, Qiagen, Hilden, Germany), the libraries were quantified by fluorimetry.

Equimolar amounts of both libraries were combined and sequenced in one lane of an Illumina Genome Analyzer II (100 base paired-ends (PE) reads) at the Wellcome Trust Centre for Human Genetics Sequencing Platform. Raw reads retrieved from the sequencing platform were processed using Stacks v1.08 (Catchen et al., 2011). First, the *process_radtags* module was used to demultiplex raw reads of each individual, discarding reads with uncalled bases, missing restriction site, ambiguous barcodes or average quality score below 20. Barcodes were also removed and all Read 1 sequences (that is, those starting at the RE site) were 3' trimmed to 93 bp. Next, *denovo_map.pl* was used to align these processed reads into exactly-matching stacks, and to score SNPs at each locus using a maximum likelihood framework. The main parameters were as follows: minimum stack depth ($M=3$), maximum nucleotide mismatches allowed within stacks ($m=2$), and mismatches between sample tags when building the catalogue ($n=1$). Third, two data sets were extracted from the SNP data: (i) the *population* module was used to generate an unfiltered set of all RAD loci, from which lists of RAD loci present in all females but not in males and vice versa were manually identified; (ii) the *export_snp.pl* and *populations* modules were used to select a set of highly consistent, robust SNPs. For the latter, RAD-loci with a minimum depth of 10 reads, containing only one SNP with two variants, and genotyped at least in ~75% of the 28 samples were selected. Finally, the 3' end read sequences (Illumina P2 reads) of the RAD loci selected in (i) and (ii) were retrieved and collated using the *sort_rad.pl* module. As the 3' end of RAD-tags are generated by random shearing it is possible to assemble multiple reads from the same RAD-tag into longer more informative contigs (Etter et al., 2011). CAP3 software was employed for this (Huang and Madan, 1999) using the suggested parameters in the CAP3 manual for short reads assembly.

Identification, annotation and population parameters of sex-linked markers in *C. gomesi*

Two different types of sex-linked markers were identified in the PR samples used for RAD-seq. First, female-specific RAD loci were defined as sequences only identified in females from the unfiltered outputs data set. Second, sex-associated markers were defined as SNPs with significant genotypic association with sex, starting from the filtered SNP data set. In light of the likely occurrence of false positives when analysing very large numbers of markers, two statistical approaches were used to identify confident sex-associated SNPs ($P<0.05$). First, we calculated exact G tests using default parameters of GENEPOP (Rousset, 2008). Second, we calculated a logistic regression based genome-wide association strategy using the fast score test for association between a trait and genetic polymorphism implemented in GenABEL (Aulchenko et al., 2007), with sex coded as a binary trait. The common set of markers identified with both approaches was considered as consistently associated with sex, while those markers identified by only one of the two analytical methods were considered suggestive, but not included in the sex-associated SNP list.

RAD loci identified with both approaches were annotated using BLASTn and BLASTx homology searches (Altschul et al., 1990) against NCBI's nr/nt database (<http://www.ncbi.nlm.nih.gov/blast>) using both the 5' end sequence and the 3' end contig sequence of each RAD locus. To assign predictive genomic locations for the sex-linked RAD loci of *C. gomesi*, comparisons were made to the genomes of the blind cave fish (*Asynx mexicanus*, Characidae; NCBI BioProject accession PRJNA89115), the closest related species with an assembled genome within Characiformes, and zebrafish (*Danio rerio*, Cyprinidae; Ensembl GRCh10), the fish species with the most extensive genomic resources and best quality genome assembly within the superorder Ostariophysi. Searches for homology (E-value $<10^{-7}$) against these reference genomes were carried out using both 5' and 3' end sequences of each RAD locus. Gene

mining around the RAD locus in blind cave fish and zebrafish genomes (2 Mb windows: RAD locus position ± 1 Mb) was performed with BioMart (www.ensembl.org), to identify relevant genes related to gonad differentiation. In addition, all sex-linked RAD loci were screened with RepeatMasker to identify putative interspersed repeats and low complexity regions (www.repeatmasker.org; Smit et al., 2015).

In the case of sex-associated SNPs, the relative coefficient of genetic differentiation (F_{ST}) was used to measure the extent of genetic differentiation between male and female groups using GENEPOP (Rousset, 2008; permutation test; $P<0.05$). Theoretically, the maximum F_{ST} expected in a ZW/ZZ system assuming fixation of one allelic variant in the Z and another in the W chromosomes is $F_{ST}=0.5$ (Kirkpatrick and Guerrero, 2014), and we used this threshold to evaluate the magnitude of observed F_{ST} .

Deviation from Hardy-Weinberg (HW) expectations (exact tests; $P<0.05$) and estimation of the F_{IS} fixation index were obtained for all markers, both computed by GENEPOP. In a ZW/ZZ chromosome pair, excess heterozygosity in the putative non-recombining region might occur due to genetic differentiation between the Z and W chromosomes. However, heterozygote deficiency may arise due to null alleles caused by degeneration of the W chromosome (Mark and Avise, 2009). Furthermore, in highly evolved SD chromosome systems, the heterogametic sex may be hemizygous for Z-linked markers, thus we would not expect heterozygous females at those loci where the W has sufficiently degraded. We also used linkage disequilibrium (LD) to confirm sets of loci putatively linked to the sex chromosome pair, again using GENEPOP (exact test; $P<0.05$). To accommodate multiple test issues, we compared the proportion of significant LD deviations of a putative sex-linked SNP set versus the average genome LD, estimated using 20 samples of 100 SNPs each randomly chosen among the 9863 identified SNPs. A frequency distribution of the proportion of LD using the 20 random SNP samples was constructed and the confidence intervals obtained. LD between sex-linked loci should be higher than between an average random SNP sample, and further, LD would likely increase at those regions where recombination between Z and W chromosomes is restricted.

Validation of sex-linked markers

Female-specific RAD sequences were individually validated by PCR screening on an extended panel of 35 females and 25 males from the PR population. Primers were designed using Primer3 (Untergasser et al., 2012) and PCR conditions were those provided by the programme (Supplementary Table S2). Sex-associated SNPs showing high F_{ST} were selected for their validation in the same PR population. Validation was hampered by the short length of the 5' PE (93 bases) and the lack of the species reference genome. Therefore, we scanned the 3' end of assembled contigs with sex-associated SNPs, a region more suitable for primer design (usually >300 bases), as this end is more likely to be in LD with the sex-associated SNPs identified. Primers for the selected SNPs were designed (Supplementary Table S3) and samples genotyped on the matrix assisted laser desorption/ionization-time of flight mass spectrometry platform (Sequenom, San Diego, CA, USA) at the University of Santiago de Compostela.

In order to investigate the inter-population and inter-specific conservation of sex-associated markers, we also amplified our PCR primers on 20 individuals (10 males and 10 females) from the TR *C. gomesi* population (Figure 1), as well as *C. pterostictus* and *C. zebra*, with 5 males and 5 females in each species, using the protocol described above.

Assessment of W-linked markers: microdissection and fluorescent *in situ* hybridisation (FISH)

A *C. gomesi* W-specific chromosome library was constructed through microdissection of the W chromosome and amplified using the GenomePlex Single Cell Whole Genome Amplification Kit (WGA4, Sigma-Aldrich). This library was then used as the DNA template to verify W-linkage using the screening methodologies outlined above. In addition, the amplified female-specific PCR products were labelled with digoxigenin-11-dUTP (Roche) to verify their location and distribution in *C. gomesi* genome (or W chromosome) through subsequent FISH experiments.

FISH analysis was performed as described in Scacchetti et al. (2015). In brief, slides were incubated with RNase (50 µg ml⁻¹) for 1 h at 37 °C, and the

chromosomal DNA was denatured in 70% formamide/2×SSC for 5 min at 70 °C. For each slide, 30 µl of hybridisation solution (containing 200 ng of labelled probe, 50% formamide, 2×SSC and 10% dextran sulphate) was denatured for 10 min at 95 °C, then dropped onto the slides and allowed to hybridise at 37 °C in a moist chamber for 36 h. Post-hybridisation, all slides were washed in 0.2×SSC/15% formamide for 20 min at 42 °C, followed by a second wash in 0.1×SSC for 15 min at 60 °C and a final wash at room temperature in 4×SSC/0.5% Tween for 10 min. Probe detection was carried out with anti-digoxigenin-rhodamine (Roche), and the chromosomes were counterstained with DAPI (4',6-diamidino-2-phenylindole, Vector Laboratories) and visualised by optical photomicroscopy (Olympus BX61). Images were captured using Image Pro Plus 6.0 software (Media Cybernetics).

RESULTS

RAD sequencing, SNP calling and genotyping

After demultiplexing and filtering for quality control, we recovered 247 743 386 PE reads, 87% of the total raw read count (NCBI BioProject PRJNA391395; Accession numbers: SRR5738929–SRR5738965). Two females were removed from further analyses due to the very low number of reads, leaving us with 19 females and 18 males for further analysis with an average of 6 664 512 filtered reads per sample (Supplementary Table S1).

We generated a catalogue of 360 754 unique RAD loci with the Stacks pipeline, of which 89 572 were polymorphic. The number of unique RAD-tags in each sample ranged from 62 945 to 83 463 (Supplementary Table S1). After applying the final filtering steps, we retained 9863 putative biallelic SNPs with a minimum depth of 10 reads and genotyped in at least 28 individuals (Dryad doi: 10.5061/dryad.tr3d8), with similar average number of reads for both sexes (females: 60.8, s.d. = 29.6; males: 53.1, s.d. = 26.1).

Sex-linked genetic markers: annotation and characterisation

RAD-tags in PR samples of *C. gomesi* were analysed for two types of sex-linkage. First, we identified 26 female-specific RAD loci sequences in the full panel of 19 females, which are consistent with W-linkage in regions that are sufficiently distinct from the Z chromosome (Supplementary Table S4a). Interestingly, the average number of reads per individual for the female-specific RAD loci (31.8, s.d. = 18.1) was very similar to that observed for the whole RAD locus data set (38.8, s.d. = 29.8), suggesting that our W-specific set represents a limited proportion of the *C. gomesi* genome.

Seven of the female-specific RAD loci showed sequence similarity with protein related genes, including hydroxysteroid 17-beta dehydrogenase 3 (*hsd17b3*), a steroidogenic factor related to gonad differentiation (Mindnich et al., 2005; Supplementary Table S4b). Around half of the female-specific RAD loci were related to various types of repetitive elements as defined by a range of criteria, with two annotated as transposable elements and seven with ≥ 5 hits in the reference genomes, especially that of the blind cave fish (average = 33.6 hits, s.d. = 34.4). In addition, four were annotated as pol-like proteins, suggesting a relationship to retro-elements and two showed very long microsatellite tracts (> 100 bp). Only two of the female-specific RAD loci rendered a unique hit in the blind cave fish genome and one could be anchored on the genetic map to linkage group (LG) 20. In addition, the *hsd17b3* gene was located in LG6 and LG8 of blind cave fish and zebrafish genomes, respectively. As expected in a female heterogametic species, we observed no male-specific RAD sequences.

We scanned the 9863 SNPs for evidence of significant allelic differentiation (F_{ST}) between males and females ($P < 0.05$; Supplementary Table S5), which would be indicative of regions where the Z and W chromosomes have started to diverge but which still retain significant homology. The 75% genotyping threshold used to

retain RAD-loci (14 for females and males) ensured a minimum number of individuals per sex in order to control the number of false positives in our sample (Brelsford et al., 2017). The average sample size used for SNP genotyping was close to the total number, the mean number of females and males analysed per locus being 18.3 and 17.0, respectively. We identified 148 consistent sex-associated markers common to both statistical approaches used (exact G tests and genome-wide association; $P < 0.05$; Supplementary Table S6a). F_{ST} values of this set ranged between 0.090 and 0.299 with a mean of 0.144 (s.d. = 0.046) (Supplementary Table S6b). Ten loci showed an $F_{ST} > 0.250$ and SNP with the highest $F_{ST} = 0.299$ suggests notable divergence between the Z and W, given the maximum expected $F_{ST} = 0.500$. The F_{ST} frequency distribution of this sex-associated SNP set and a 500 random non-associated reference set (Figure 3) showed distinct modes, and the small overlap between the distributions confirm the robustness of our statistical approach. In fact, the six non-associated SNPs overlapping at the left end of the sex-associated distribution were suggestive, namely they were identified with only one approach, genome-wide association or exact G tests.

We searched public databases to identify candidate genes related to sex differentiation for the 148 sex-linked SNPs (Supplementary Table S6c). A total of 116 sequences were annotated (78.4%), 16 of which (13.8%) were transposable elements and five included long microsatellite tracts (> 100 bp). Two RAD loci were annotated to genes involved in gonad differentiation and reproduction, including *necrin-2*, a gene that codifies for a junction molecule crucial for spermatogenesis (Zhang and Lui, 2014) and also suggested to play a role in the follicular development of the mouse ovary (Kawagishi et al., 2005), and *tgb2*, a gene that plays a pivotal role in gonad development (Ergin et al., 2008). *Necrin-2* and *tgb2* were located at LG23 and LG10 of blind cave fish and at LG19 and LG16 of zebrafish, respectively.

BLASTn searches of the 148 sex-associated RAD loci to the blind cave fish and the zebrafish genomes were undertaken to potentially identify regions of synteny (E-value < 10^{-5} ; Supplementary Table S6c). A total of 24 sex-associated SNPs mapped to unique positions of the *A. mexicanus* genome, and 5 to the *D. rerio* genome (Supplementary Table S6c), which is consistent with the closer phylogenetic relatedness between *A. mexicanus* and *C. gomesi*.

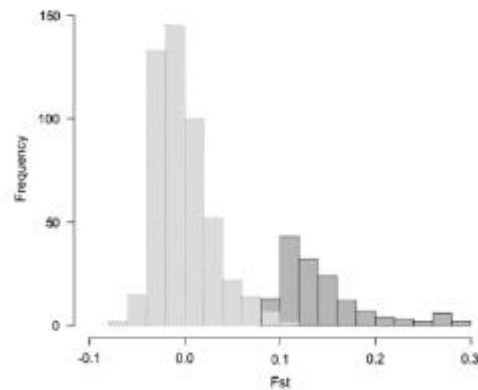


Figure 3 Distribution of genetic differentiation (F_{ST}) between male and female subpopulations for the 148 sex-associated SNPs (dark grey) and a random sample of 500 SNPs not associated with sex (light grey).

Nineteen out of 24 mapped loci were located in the genetic map of the blind cave fish (Carlson *et al.*, 2015), with LG8, LG10, LG15, LG16 and LG22 each containing two hits. Most mapped loci showed F_{ST} values around the mean (0.144; range: 0.094–0.278), except L37075_37 on LG22 ($F_{ST}=0.278$). Gene mining within 1 Mb on either side of the mapped RAD loci identified several genes involved in gonad differentiation, particularly in the blind cave fish genome. Two notable genes were found at LG8 include sperm adhesion molecule 1 gene (*spaml*), which is involved in sperm penetration through the cumulus matrix (Kimura *et al.*, 2009), and stimulated by retinoic acid 6 gene (*strab6*), which is involved in vitellogenesis (Levi *et al.*, 2012). In addition, a gene that functions in female differentiation, b-catenin1 (*ctnnb1*; Chassot *et al.*, 2014), was tightly linked to another RAD locus mapping at LG10, whereas two genes related to the steroid hormone mediated signalling pathway, that is, nuclear receptor subfamily 4 group A member 1 (*nr4a1*; Abdou *et al.*, 2013) and retinoic acid receptor, gamma b (*nr3b*), closely mapped in LG22. Other important genes related to gonad development, bone morphogenetic protein 15 (*bmp15*; Han *et al.*, 2015) and cytochrome P450, family 26, subfamily b, polypeptide 1 (*cyp26b1*; Saba *et al.*, 2014) were found in LG16, while the oestrogen receptor 2b gene (*estrb*; Delalande *et al.*, 2015) in zebrafish LG13.

Our analysis is somewhat limited by the relatively small number of individuals screened (37) and the large number of loci analysed (9863). Nevertheless, we observed several important indicators of differentiation between the Z and W chromosomes. First, we observed a significant deficit of heterozygotes within the 148 sex-associated SNP data set compared to the genomic average ($P<0.05$; 19.6% vs 4.2%, respectively; Supplementary Table S6b), suggesting a higher proportion of null alleles in the sex chromosome pair. This would be expected if the W and Z chromosomes have diverged significantly from one another. Furthermore, we detected 25 loci with high heterozygosity ($H_E \geq 0.4$) showing extreme heterozygote deficiency in females ($F_{IS}=1$) but in HW equilibrium in males (Supplementary Table S7). Only two markers showing these characteristics were detected in males, which suggest a low proportion of false positives within the 25 SNPs identified in females. These loci might be located in regions where the W has significantly diverged, for which females are hemizygous.

The proportion of LD departures ($P<0.05$) in the 148 sex-associated (8.4% pair-wise LD) and in the 25 female-limited (17.7%) SNP sets was much higher than that observed in the genome background (20 random 100 SNP samples over the 9863 loci; mean = $1.725\% \pm 0.062\%$; 95% CI: 1.595–1.855), strongly supporting their linkage. Moreover, pair-wise LD between the 148 and 25 SNP sets was also much higher than the background (9.2%), supporting their linkage to the ZW pair.

Validation of the female-specific and sex-associated markers

The 26 female-specific RAD loci were validated by PCR on the expanded sample of 35 females and 25 males of *C. gomesi* from the PR population. Two RAD loci showed high sequence similarity, and a single primer pair was designed for their amplification (Supplementary Table S2). Seven out of 26 loci showed different amplification patterns between the sexes (Figure 4). Most of these showed a single prominent band of the expected size in females, while no band or a different banding pattern was observed in males. The remaining 19 RAD loci showed a similar banding pattern in both sexes. Nevertheless, this fact does not preclude diagnostic differences between males and females for these RAD loci, possibly reflecting diagnostic polymorphism within the restriction enzyme site associated with the specific tag.

A total of 15 of the 148 initially identified sex-associated RAD loci were selected for validation in the expanded PR sample (Supplementary Table S3). Thirteen primer sets of the 15 selected RAD loci were successfully genotyped in the PR population (Table 1 and Supplementary Table S8). Nearly half of these SNPs showed lower polymorphism than their 5' counterparts ($H_E < 0.2$), diminishing statistical power to detect differences between male and female subpopulations. Of the seven remaining loci ($H_E > 0.25$), two showed significantly greater differentiation between males and females than the 5' SNP counterpart ($F_{ST} > 0.290$), supporting sex-linkage. Interestingly, 33.3% positive F_{IS} values were detected in females but not in males ($P<0.05$), supporting the presence of null alleles linked to the W chromosome.

Trans-population and trans-specific conservation of female-specific and sex-associated markers

Four out of seven validated female-specific markers were consistent in both the PR and TR populations. The remaining three loci showed a

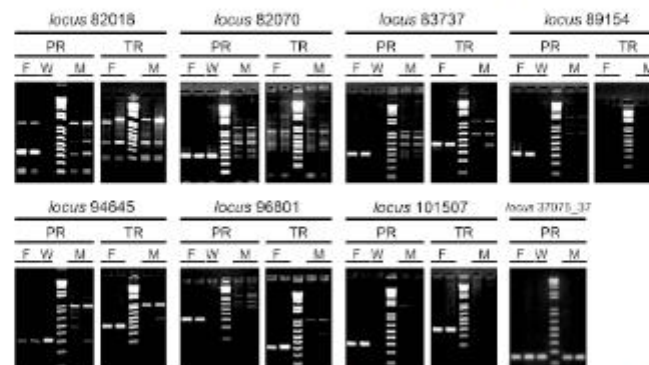


Figure 4 Banding patterns of validated female-specific markers identified in *Characidium gomesi* in Paranapanema (PR) and Tiele (TR) populations. Amplification on the W chromosome DNA library is also shown.

Table 1 Genetic diversity and differentiation between male and female subpopulations for the validated SNPs detected in *Characidium gomesi*

SNP code	Females				Males				F_{ST}	$F_{ST} 5'$ end
	K	H _O	H _E	F _{IS}	K	H _O	H _E	F _{IS}		
L17552_28	2	0.185	0.171	-0.083	2	0.059	0.059	0	-0.023	0.150
L23581_62	2	0	0.073	1.000*	1	0	0	0	-0.016	0.192
L26007_70	2	0.074	0.307	0.763***	2	0.111	0.203	0.46	-0.019	0.299
L29522_74	2	0.522	0.468	-0.128	2	0.321	0.353	0.118	0.298***	0.226
L30946_43	2	0.182	0.512	0.650**	2	0.333	0.434	0.233	0.035	0.164
L37075_37	2	0.208	0.51	0.597**	2	0.200	0.186	-0.077	0.292***	0.278
L38879_69	2	0.16	0.15	-0.068	2	0.333	0.286	-0.172	0.015	0.173
L53050_55	1	0	0	NA	1	0	0	NA	NA	0.183
L55368_62	2	0.963	0.509	-0.926	2	1	0.514	-1	0	0.236
L59864_80	2	0.320	0.327	0.020	2	0.167	0.246	0.329	-0.015	0.177
L7660_75	2	0.522	0.502	-0.039	2	0.588	0.499	-0.185	-0.022	0.280
L8018_33	2	0.185	0.171	-0.083	2	0.056	0.056	0	0.012	0.263
L8459_59	2	0.037	0.037	0	2	0.056	0.056	0	-0.022	0.130

Abbreviations: F_{IS}, inbreeding coefficient; F_{ST}, relative genetic differentiation coefficient; H_O, observed heterozygosity; H_E, expected heterozygosity; K, number of alleles; NA, not applicable; SNP, single-nucleotide polymorphism. *P<0.05; **P<0.01; ***P<0.001. In grey background loci with significant and high F_{ST} in the 5' end tested SNP.

similar banding pattern in males and females (Figure 4). Notably, none of the primer sets for female-specific markers produced female-specific PCR products in *C. pterostictum* and *C. zebra*, and assays of the 13 validated sex-associated markers in the PR population failed or were monomorphic in these two species.

Location of sex validated markers on the *C. gomesi* genome

Two female-specific markers and one sex-associated marker produced PCR products of the expected size when amplified on the microdissected W chromosome library (Figure 4), confirming their location on this chromosome. The seven PCR amplified female-specific markers were used as probes for FISH on *C. gomesi* metaphase plates to ascertain their location, but no specific signals of hybridisation were detected. In keeping with their known transposable elements sequence composition, three of these generated weak scattered signals throughout the karyotype of the species, suggesting multiple locations (Supplementary Figure S1).

DISCUSSION

Teleosts show a high turnover of sex-determining mechanisms (Martinez *et al.*, 2014), including sex chromosomes. This rapid origin makes fish a useful clade for studying the early stages of sex chromosome evolution. In this study, we characterised the sex chromosomes at both population and species level in the genus *Characidium* using a reduced representation genome sequencing strategy (RAD-seq). Our study is based on 9863 filtered SNPs, equivalent to roughly 1 SNP per 100 kb and 1 RAD locus per 3 kb, based on the reported genome size of the species of roughly 1 Gb (Carvalho *et al.*, 1998).

Clear heteromorphism between the Z and W chromosomes has been reported for both size and C-banding pattern in *C. gomesi* (Maistro *et al.*, 1998), and this is consistent with the proportion of sex-linked RAD sequences and SNPs we identified. The W chromosome of *C. gomesi* shows C-positive banding in mitotic plates suggesting a condensation estate (Figure 2; Maistro *et al.*, 1998) theoretically related to the accumulation of repetitive sequences (Charlesworth, 1991). Thus, the genomes of species with heteromorphic sex chromosomes, like *C. gomesi*, are expected to have more sex-specific restriction sites (and hence sex-specific RAD loci) than species with homomorphic sex

chromosomes (Gamble *et al.*, 2015). Consistent with the female heterogametic sex chromosome system reported in this species, we recovered female-specific RAD loci, but not male-specific loci, and roughly half of the female-specific RAD loci showed typical features of repetitive elements.

An increasing number of studies using RAD-seq methods have been carried out in a range of organisms either to identify the SD region by identifying sex-associated molecular markers at the population or species level. In some cases, linkage mapping has been used to identify the sex-determining locus. The first approach can be applied to wild populations based on LD between polymorphic loci and the SD region, while the second method requires a mapping family panel (Gamble and Zarkower, 2014; Fowler and Buonaccorsi, 2016; Gamble, 2016; Brelsford *et al.*, 2017; Robledo *et al.*, 2017).

Inevitably, with small numbers of sexed individuals being screened for a very large number of SNP markers, false positive associations between sex and markers are likely to arise. While limited numbers of individuals were available for our study (19 females and 18 males), analyses from similar studies in other species suggest that data from 12 to 14 random individuals of each sex should be sufficient for relatively robust sex-linked analysis (Lambert *et al.*, 2016; Brelsford *et al.*, 2017). We validated approximately one third of the female-specific RAD loci in a broader sample of the PR population, thereby proving the usefulness of RAD-seq to develop sex-specific markers in non-model species (Gamble and Zarkower, 2014; Fowler and Buonaccorsi, 2016). Moreover, two of these loci were amplified by PCR from the W-chromosome library, confirming W-linkage. The failure of several other sex-specific marker assays to amplify products from the W chromosome library may reflect methodological limitations of W-DNA library construction due to the gross process of microdissection of mitotic chromosomes and their subsequent random amplification.

Interestingly, some of the validated markers were population-specific, suggesting that a fraction of W-specific sequences may evolve rapidly, and populations located in a rather small geographic area may display different W-specific RAD-tags repertoires. Sex chromosomes often display higher rates of molecular evolution compared to autosomes (Bachtrog, 2005; Filatov, 2005; Berlin *et al.*, 2006; Shikano *et al.*, 2011), and this indicates that differentiation between

species should be higher for sex-linked regions. Consistent with that, we found that none of the *C. gomesi* female-specific loci amplified in the two other species evaluated (*C. zebra* and *C. pterostictum*). This finding could be due to several factors, including the rapid differentiation of the W chromosome, as reported recently in molecular cytogenetic studies of other Characiforms (Yano et al., 2016); the phylogenetic distance among the analysed species (Pansonato-Alves et al., 2014); or even possible transitions of sex-determining systems between *C. gomesi* and *C. zebra*.

It is worth noting that we detected a relatively small proportion of female-specific RAD loci (0.0007%), which is a much smaller fraction than expected based on the fact that the W chromosome constitutes 3% of the *C. gomesi* karyotype in metaphase plates. This is most likely due to a small repertoire of repetitive elements that lack SbfI sites on the W chromosome.

We identified 148 consistent sex-associated SNPs in *C. gomesi* in our PR population samples, and the highest F_{ST} value we observed between males and females suggests close proximity of this marker to the putative SD gene. A subset of these markers was additionally validated, and some confirmed high differentiation. Furthermore, the RAD locus with the highest differentiation was shown to be W-linked, and LD evidence indicates that most of these markers are located on the same LG.

Annotation of the sex-linked RAD loci identified single-copy genes associated with sex differentiation. Among the most relevant, the hydroxysteroid 17- β dehydrogenase 3 (*hsd17b3*), a steroidogenic marker that catalyses the conversion of androstenedione to testosterone, was detected in the female-specific RAD loci. This gene is almost exclusively expressed in the testis of mammals, but it shows high ovarian expression in zebrafish (Mindnich et al., 2005). Among the sex-associated RAD loci, we identified *noxin-2*, a gene involved in gonadal differentiation (Kawagishi et al., 2005; Zhang and Lui, 2014), and *tgfb2*, an important mediator of growth and differentiation involved in germ cell and gonadal development (Ergin et al., 2008; Özgüden-Akkoç and Özer, 2012). *Tgfb2* belongs to the Tgfb- β superfamily whose members have been associated with important ovarian and testicular functions (Drummond, 2005; Fan et al., 2012) and different components of the TGF- β signalling pathway have been found to be sex-determining genes in several fish species (reviewed by Martínez et al., 2014).

We used a comparative mapping strategy interrogating the reference genomes of other fish species to gain insight on the genomic organisation of the *C. gomesi* SD region and, specifically, to identify putative SD candidates. The assembled genome and important genomic resources of zebrafish, a species of the superorder Ostariophysi, and the draft genome of blind cave fish, with fewer genomic resources than zebrafish, but within the family Characidae, were used for this purpose. As expected, a much higher number of significant hits were retrieved in our study from the blind cave fish genome than the zebrafish. All three species have a $2n=50$ karyotype (Sola and Gornung, 2001; Kavalco and De Almeida-Toledo, 2007), but there have been important genomic reorganisations between the blind cave fish and zebrafish based on synteny, and only two chromosomes of their karyotype show a consistent macrosyntentic pattern (Carlson et al., 2015). LG10 and LG22 show features, including number of hits, candidate genes and highest F_{ST} , suggesting that they may contain important regions orthologous to the *C. gomesi* sex chromosomes. However, our data indicate a complex origin, involving reorganisations of different chromosomes.

Degeneration of the W chromosome causes heterozygote deficiency in females, as the loss of restriction sites on the W chromosome would

give rise to null alleles. In regions where the W has only begun to diverge from the Z, we expect a significantly higher frequency of W-specific null alleles, manifesting as significant deviations from the null hypothesis ($F_{IS}=0$), as observed in our data (19.6% deviations for sex-associated loci vs 4.2% across the whole data set). In regions where the W chromosome is highly diverged, this situation would be extreme ($F_{IS}=1$) and females will be hemizygous for the Z chromosome, and we observed 25 SNPs with a pattern of female hemizygosity. This suggests heterogeneity in recombination suppression on the *C. gomesi* sex chromosomes, with an older stratum (25 markers), where recombination was suppressed earlier and the W chromosome is more degraded, and a relatively younger stratum (148 markers) where the W has only recently started to diverge. This is consistent with sex chromosomes in other fish species, including *Gasterosteus aculeatus* (Schulteis et al., 2015) and *P. reticulata* (Wright et al., 2017), both of which show sex chromosome strata of differing ages. Importantly, we observed strong evidence of elevated LD among our female-specific SNPs, consistent with the loss of recombination on the W chromosome.

Taken together, our results indicate significant genetic differentiation between the Z and W sex chromosomes in the genus *Characidium*, illustrating the utility of RAD-seq for sex chromosome characterisation in wild species. Both female-specific markers, expected in a ZW/ZZ system, as well as two types of sex-associated markers, suggesting different steps in the process of chromosome differentiation, were identified. The validation of sex-related markers in other populations and species of the genus *Characidium* suggested a quick evolution of sex chromosome associated sequences, as previously reported in other vertebrates. The existence of the genome of the blind cave fish, a closely related species included in the family Characidae, enabled the identification of SD candidate genes and suggested a complex evolution of the sex chromosome pair. From a practical perspective, the identified markers will be valuable for developing a suitable molecular tool for non-invasive sex identification in future population dynamics or ecological studies.

DATA ACCESSIBILITY

Raw data: NCBI BioProject PRJNA391395. Raw sequences: SRR5738929-SRR5738965. Filtered RADtags Dryad doi:10.5061/dryad.tr3d8.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Renato Devidé for help with obtaining samples, and Érica A. Serrano-Freitas and José C. Pansonato-Alves for providing tissue and cell suspensions. This study was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the European Research Council (grant agreements 26023 and 680951) and research funding from the Xunta de Galicia local government (GRC2014/010). MB and JBT were partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (Grant reference HR09011) and contributing institutions.

- Abdou HS, Villeneuve G, Tremblay JJ (2013). The calcium signaling pathway regulates Leydig cell 3 β H steroidogenesis through a transcriptional cascade involving the nuclear receptor NR4A1 and the 3 β H steroidogenic acute regulatory protein. *Endocrinology* **154**: 511–520.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Aulchenko VS, Ripke S, Isaacs A, van Duijn CM (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**: 1294–1296.
- Bachrog D (2005). Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. *Genome Res* **15**: 1393–1401.
- Bachrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL et al. (2014). Sex determination: why so many ways of doing it? *PLoS Biol* **12**: e1001899.
- Baird NA, Elter PD, Alwood TS, Curney MC, Shiver AL, Lewis ZA et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**: e3376.
- Berlin S, Brandström M, Backström N, Axelsson E, Smith NG, Ellegren H (2006). Substitution rate heterogeneity and the male mutation bias. *J Mol Evol* **62**: 226–233.
- Böhne A, Wilson CA, Postlethwait JH, Salzburger W (2016). Variations on a theme: genomics of sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* **17**: 883.
- Bretford A, Lwanczyk G, Semler R, Rausch A, Perin N (2017). Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. *Mol Ecol Resour* **17**: 752–759.
- Bull JJ (1983). *Evolution of Sex Determining Mechanisms*. Benjamin Cummings Press: Menlo Park, CA, USA.
- Carlson BM, Onisko SW, Gross JB (2015). A high-density linkage map for *Aspianus mexicanus* using genotyping-by-sequencing technology. *G3 (Bethesda)* **5**: 241–251.
- Canalho ML, Oliveira C, Navarette MC, Froehlich O, Foresti F (1998). Nuclear DNA content determination in Characiformes fish (Teleostei, Ostariophysi) from the Neotropical region. *Genet Mol Biol* **25**: 49–55.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1**: 171–182.
- Charlesworth B (1991). The evolution of sex chromosomes. *Science* **251**: 1030–1033.
- Chenot AA, Gillet I, Chaboussier MC (2014). R-spondin1, WNT4, and the CTNNB1 signaling pathway: strict control over ovarian differentiation. *Reproduction* **148**: R97–110.
- Delalande C, Goupil AS, Lamyre JJ, Le Gac F (2015). Differential expression patterns of three aromatase genes and of four estrogen receptors genes in the testes of trout (*Oncorhynchus mykiss*). *Mol Reprod Dev* **82**: 694–708.
- Drummond AE (2005). TGF β signaling in the development of ovarian function. *Cell Tissue Res* **322**: 107–115.
- Ellegren H (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet* **12**: 157–166.
- Erkin K, Güroym E, Bayraktar Koca Y, Bayraktar H, Seyrek K (2008). Immunohistochemical detection of insulin-like growth factor-I, transforming growth factor- β 2, basic fibroblast growth factor and epidermal growth factor-receptor expression in developing rat ovary. *Cytokine* **43**: 209–214.
- Elter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011). Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One* **6**: e18561.
- Fan YS, Hu YJ, Yang WX (2012). TGF- β superfamily: how does it regulate testis development. *Mol Biol Rep* **39**: 4727–4741.
- Filatov DA (2005). Substitution rates in a new *Silene latifolia* sex-linked gene, *Sil1*. *Mol Biol Evol* **22**: 402–408.
- Fowler BLS, Buonocorsi VP (2016). Genomic characterisation of sex-identification markers in *Sebastes caninus* and *Sebastes chrysomelas* rockfishes. *Mol Ecol* **25**: 2165–2175.
- Gamble T (2016). Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Mol Ecol* **25**: 2114–2116.
- Gamble T, Zarkower D (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol Ecol Resour* **14**: 902–913.
- Gamble T, Corneil J, Ezaz T, Lynch J, Scantlebury DP, Zarkower D (2015). Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among *Drosophila* sex-determining systems. *Mol Biol Evol* **32**: 1296–1309.
- Han H, Lei Q, Zhou Y, Gao J, Liu W, Li F et al. (2015). Association between BMP15 gene polymorphism and reproduction traits and its tissues expression characteristics in chicken. *PLoS ONE* **10**: e0143298.
- Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A et al. (2012). Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**: 244.
- Huang X, Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res* **9**: 868–877.
- Kawato KF, De Almeida-Toledo LF (2007). Molecular cytogenetics of blind Mexican tetra and comments on the karyotypic characteristics of genus *Aspianus* (Teleostei, Characidae). *Zebrafish* **4**: 103–111.
- Kawaguchi R, Tahara M, Morishige K, Sakata M, Tasaka K, Ikeda W et al. (2005). Expression of nectin-2 in mouse granulosa cells. *Eur J Obstet Gynecol Reprod Biol* **121**: 71–76.
- Kimura M, Kim E, Kang W, Yamashita M, Saigo M, Yamazaki T et al. (2009). Functional roles of mouse sperm hyaluronidases, HYAL5 and SPAM1, in fertilisation. *Biol Reprod* **81**: 939–947.
- Kirkpatrick M, Guerrero RF (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics* **197**: 531–541.
- Lambert MR, Skelly DK, Ezaz T (2016). Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DARTseq provide early insight into sex chromosome evolution. *BMC Genomics* **17**: 844.
- Levi L, Ziv T, Admon A, Levavi-Sivan B, Lubzens E (2012). Insight into molecular pathways of retinal metabolism, associated with vitellogenesis in zebrafish. *Am J Physiol Endocrinol Metab* **302**: E626–E644.
- Maistro EL, Mata EF, Oliveira C, Foresti F (1998). Unusual occurrence of a ZZ/ZX sex-chromosome system and supernumerary chromosomes in *Characidium cf. fasciatum* (Pisces, Characiformes, Characidae). *Genetica* **104**: 1–7.
- Mank JE, Avise JC (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex Dev* **3**: 60–67.
- Martinez P, Vilas AM, Sánchez L, Diaz N, Ribas L, Piferrer F (2014). Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front Genet* **5**: 340.
- Mindrich R, Haller F, Halbach F, Moeller G, Hrabě de Angelis M, Adamski J (2005). Androgen metabolism via 17 β -hydroxysteroid dehydrogenase type 3 in mammalian and non-mammalian vertebrates: comparison of the human and the zebrafish enzyme. *J Mol Endocrinol* **35**: 305–316.
- Ohno S (1967). *Sex Chromosomes and Sex Linked Genes*. Springer-Verlag: Berlin.
- Oliveira C, Foresti F, Hilsdorf AWS (2009). Genetics of Neotropical fish: from chromosomes to populations. *Fish Physiol Biochem* **35**: 81–100.
- Ogden-Akkoç CG, Ozer A (2012). Immunohistochemical localization of transforming growth factor β 1 and β 2 in mouse testes during postnatal development. *Biotech Histochem* **87**: 154–159.
- Pan Q, Anderson J, Bertho S, Herpin A, Wilson C, Postlethwait JH et al. (2016). Vertebrate sex-determining genes play musical chairs. *C R Biol* **339**: 258–262.
- Paronato-Alves JC, Semano EA, Uzunova R, Camacho JPM, Costa-Silva GJ, Vicari MR et al. (2014). Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus *Characidium*. *PLoS ONE* **9**: e107169.
- Robledo D, Palackostas C, Bagnoli L, Martinez P, Houston R (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev Aquac*; doi:10.1111/raq.12193.
- Rousset F (2008). Genepop007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* **8**: 103–106.
- Saba R, Wu Q, Saga Y (2014). CYP26B1 promotes male germ cell differentiation by suppressing STRAS-dependent meiotic and STRAS-independent mitotic pathways. *Dev Biol* **389**: 173–181.
- Scacchetti PC, Uzunova R, Paronato-Alves JC, Vicari MR, Artoni RF, Oliveira C et al. (2015). Chromosomal mapping of repetitive DNAs in *Characidium* (Teleostei, Characiformes): genomic organization and diversification of ZW sex chromosomes. *Cytogenet Genome Res* **146**: 136–143.
- Schulteis R, Villanelli HM, Leder EH (2015). Spatial dynamics of evolving dosage compensation in a young sex chromosome system. *Genome Biol Evol* **7**: 581–590.
- Shikano T, Natri HM, Shimada Y, Menis J (2011). High degree of sex chromosome differentiation in stickleback fishes. *BMC Genomics* **12**: 474.
- Smith AFA, Hultley R, Green P (2015). RepeatMasker Open-4.0. 2013–2015. Available at: <http://www.repeatmasker.org>.
- Sola L, Gornung E (2001). Classical and molecular cytogenetics of the zebrafish, *Danio rerio* (Cyprinidae, Cypriniformes): an overview. *Genetica* **111**: 397–412.
- Untergrasser A, Oulcliche I, Köressar T, Ye J, Fanclothe SC, Remm M et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.
- Wilson CA, High SK, McCluskey BM, Amores A, Yan YL, Titus TA et al. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics* **198**: 1291–1308.
- Wright AE, Canali I, Bloch NI, Oostra V, Sandkam B, Buechel SD et al. (2017). Convergent recombination suppression suggests a role of sexual selection in the guppy sex chromosome formation. *Nat Commun* **8**: 14251.
- Yano CF, Berillo LAC, Ezaz T, Trifunov V, Semler A, Liehr T et al. (2016). Highly conserved Z and molecularly divergent W chromosomes in the fish genus *Triportheus* (Characiformes, Triportheidae). *Heredity* **118**: 276–283.
- Zhang X, Liu WY (2014). Dysregulation of nectin-2 in the testicular cells: an explanation of cadmium-induced male infertility. *Biochim Biophys Acta* **1839**: 873–884.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)

Apêndice IV



Genetic Characterization of the Fish *Piaractus brachypomus* by Microsatellites Derived from Transcriptome Sequencing

Paulo H. Jorge¹, Vito A. Mastrochirico-Filho¹, Milene E. Hata², Natália J. Mendes¹, Raquel B. Ariede¹, Milena Vieira de Freitas¹, Manuel Vera², Fábio Porto-Foresti³ and Diogo T. Hashimoto^{1*}

¹ Aquaculture Center of Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo State University, Jaboticabal, Brazil, ² Veterinary Faculty, University of Santiago de Compostela, Lugo, Spain, ³ School of Sciences, São Paulo State University, Bauri, Brazil

OPEN ACCESS

Edited by:

Rodrigo A. Torres,
Universidade Federal de Pernambuco,
Brazil

Reviewed by:

Pedro Manoel Galetti Jr.,
Federal University of São Carlos, Brazil
Maria Raquel Moura Coimbra,
Federal Rural University of
Pernambuco, Brazil

*Correspondence:

Diogo T. Hashimoto
diogo@caunesp.unesp.br

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 01 September 2017

Accepted: 31 January 2018

Published: 22 February 2018

Citation:

Jorge PH, Mastrochirico-Filho VA,
Hata ME, Mendes NJ, Ariede RB,
Freitas MV, Vera M, Porto-Foresti F
and Hashimoto DT (2018) Genetic
Characterization of the Fish *Piaractus
brachypomus* by Microsatellites
Derived from Transcriptome
Sequencing. Front. Genet. 9:46.
doi: 10.3389/fgene.2018.00046

The pirapitinga, *Piaractus brachypomus* (Characiformes, Serrasalminidae), is a fish from the Amazon basin and is considered to be one of the main native species used in aquaculture production in South America. The objectives of this study were: (1) to perform liver transcriptome sequencing of pirapitinga through NGS and then validate a set of microsatellite markers for this species; and (2) to use polymorphic microsatellites for analysis of genetic variability in farmed stocks. The transcriptome sequencing was carried out through the Roche/454 technology, which resulted in 3,696 non-redundant contigs. Of this total, 2,568 contigs had similarity in the non-redundant (nr) protein database (Genbank) and 2,075 sequences were characterized in the categories of Gene Ontology (GO). After the validation process of 30 microsatellite loci, eight markers showed polymorphism. The analysis of these polymorphic markers in farmed stocks revealed that fish farms from North Brazil had a higher genetic diversity than fish farms from Southeast Brazil. AMOVA demonstrated that the highest proportion of variation was presented within the populations. However, when comparing different groups (1: Wild; 2: North fish farms; 3: Southeast fish farms), a considerable variation between the groups was observed. The F_{ST} values showed the occurrence of genetic structure among the broodstocks from different regions of Brazil. The transcriptome sequencing in pirapitinga provided important genetic resources for biological studies in this non-model species, and microsatellite data can be used as the framework for the genetic management of breeding stocks in Brazil, which might provide a basis for a genetic pre-breeding programme.

Keywords: aquaculture, genetic structure, NGS, Pirapitinga, Serrasalminidae

INTRODUCTION

The pirapitinga (*Piaractus brachypomus*) is a native fish from the Amazon and Orinoco Rivers and can reach up to 20 kg of weight (Alcantara et al., 1990). This species is used for fish farming, is valued for its meat and has fast growth performance (Fresneda et al., 2004). In Brazil, pirapitinga farming represents the third largest fish production operation (about 10,000 tons) among the native

fish species (MPA, 2013a). Furthermore, this species has been widely used for the production of interspecific hybrids, particularly the tambatinga (female tambaqui *Colossoma macropomum* × male pirapitinga *P. brachypomus*), and patinga (female pacu *Piaractus mesopotamicus* × male pirapitinga *P. brachypomus*; IBGE, 2016). The aquaculture production of pirapitinga in Brazil is concentrated mainly in the Midwest and North (87%), followed by the Southeast (9%), Northeast (3%) and South (1%) (MPA, 2013b). This species also has economic importance for aquaculture in other countries in South America (Colombia, Peru, and Venezuela) and in Asia (China, Myanmar, Thailand, and Vietnam; Flores Nava, 2007; Honglang, 2007; Lin et al., 2015).

However, despite this representation of aquaculture production, few scientific studies have focused on understanding the biology of pirapitinga, especially of genetic traits. So, the generation of genetic resources for this species is fundamental to advancing studies of breeding and genetic management, as occurred in model species used in aquaculture, such as salmon, catfish, carp, and tilapia (Lien et al., 2011; Liu et al., 2011; Guyon et al., 2012; Ji et al., 2012).

Model fish species, such as zebrafish *Danio rerio*, have been described with more than 26,000 genes (Howe et al., 2013). However, few genes and their metabolic pathways have been characterized for non-model species without reference genomes, as is the case for pirapitinga. In the field of genetics and molecular biology, Next-Generation Sequencing (NGS) technologies are causing a revolution, allowing the sequencing of genome and transcriptome of any organism, quickly and at low cost (Seeb et al., 2011). RNA-seq (transcriptome sequencing) is considered one of the most used strategies of NGS technology for the transcripts analysis (Qian et al., 2014), wherein all the messenger RNA (mRNA) of a specific tissue or set of tissues are used as a source for sequencing. Moreover, RNA-seq is an effective tool for discovery of molecular markers, particularly for the prospection of gene-associated microsatellites (Teacher et al., 2012; Xu et al., 2013).

Due to the usefulness of revealing the genetic variation among individuals (Liu and Cordes, 2004), microsatellite markers have proven to be efficient for genetic characterization of wild populations and breeding stocks of farmed fish (Koljonen et al., 2002; Lehoczy et al., 2005), such as to prevent inbreeding (Ponzoni et al., 2008), to identify and preserve live gene banks (Machado-Schiaffino et al., 2007), to detect genetic structure (Do Prado et al., 2018), to direct matings during the formation of the population base of breeding programmes (Fernández et al., 2014), and to perform marker assisted selection (MAS) for economic traits (Houston et al., 2010). However, these markers are not available for pirapitinga, one of most important species for the aquaculture in South America.

For the aquaculture of pirapitinga, analysis of genetic variability in farmed stocks still needs to be performed, which will allow three hypotheses to be tested: (1) farmed stocks of pirapitinga have lower genetic diversity in relation to wild stocks; (2) farmed stocks of pirapitinga in Brazil are genetically structured; and (3) gene-linked microsatellites can be associated to economic traits of pirapitinga, such as growth and disease

resistance. These analyses will support the creation of a breeding programme to increase the productivity of pirapitinga, by directed matings which lead to the formation of families, avoiding the problems of bottlenecks and inbreeding in the base population (Fernández et al., 2014), and by the identification of quantitative trait loci (QTL), which will assist the selection of superior genotypes by MAS (Houston et al., 2010).

Thus, the objective of the present study was to characterize genetic resources for the proper management of this non-model species in aquaculture, through transcriptome characterization and genetic variability analysis of stocks using microsatellite markers.

MATERIALS AND METHODS

Ethics Statement

This study was carried out in strict accordance with the animal welfare guidelines of the National Council for Control of Animal Experimentation (Brazilian Ministry for Science, Technology, and Innovation). The present study was performed under authorization N° 33435-1, issued through ICMBio (Chico Mendes Institute for the Conservation of Biodiversity, Brazilian Ministry for Environment). No animal was housed or cared for in the laboratory. Fish were euthanized by benzocaine anesthetic overdose for collection of liver tissue for transcriptome sequencing. For microsatellite validation and genetic variability analysis, fin fragments were collected from each fish under benzocaine anesthesia and all efforts were made to minimize suffering.

Samples for Transcriptome Sequencing

To perform the transcriptome sequencing, samples of liver tissue were taken from 10 individual fish from three different Brazilian fish farms and one wild population: Aquaculture Center of São Paulo State University, CAUNESP, Jaboticabal, SP ($n = 3$); Projeto Peixe fish farm, Sales Oliveira, SP ($n = 1$); Fazenda São Paulo fish farm, Brejinho de Nazaré, TO ($n = 5$); and Tocantins River, Lajeado, TO ($n = 1$). Individuals from different origins were used in order to achieve the highest genetic variability in microsatellite discovery analysis. Liver samples were selected for transcriptome studies because the liver plays a critical role in coordinating various physiological processes, including digestion, metabolism, detoxification, and endocrine system immune response (Martin et al., 2010).

Samples for Genetic Variability Analysis

Analyses of microsatellite validation were performed in 22 individual pirapitinga collected from the Tocantins River (TO) from Lajeado City, Tocantins State, Brazil. We then used the microsatellite markers to study the genetic variability in samples collected from four commercial fish farms: TO1 ($n = 25$) and TO2 ($n = 26$), from Tocantins State (North Brazil); and SP1 ($n = 36$) and SP2 ($n = 20$), from São Paulo State (Southeast Brazil). To maintain the confidentiality of these fish farms, the names of and information on the fish farms have been preserved.

Analysis of Genetic Purity in Pirapitinga Individuals

According Hashimoto et al. (2014), interspecific hybrids have been detected in broodstocks of Brazilian fish farms. The pirapitinga can be crossed with tambaqui *C. macropomum* or pacu *P. mesopotamicus*, resulting in viable and fertile hybrids (Hashimoto et al., 2012, 2014). Therefore, in the present study, special attention was given to analyze pure pirapitinga, and not interspecific hybrids. The analysis of genetic purity in all animals herein studied was performed using the mitochondrial genes, Cytochrome C Oxidase subunit I (*mt-co1*) and Cytochrome b (*mt-cyb*); and the nuclear genes, α -Tropomyosin (*tpm1*) and Recombination Activating Gene 2 (*rag2*), according to the protocols and methods of Hashimoto et al. (2011). Fish identified as interspecific hybrids were excluded from further analysis in this study.

cDNA Library Construction and Roche 454 Platform Sequencing

Samples of ~100 mg of liver fixed in RNAlater were extracted with RNeasy Mini Kit (Qiagen). Each sample was quantified by spectrophotometry using NanoDrop ND-1000 equipment and the quality (integrity) was checked by 2100 Bioanalyzer equipment. It succeeded the preparation of an equimolar pool of total RNA samples (from 10 individuals) to mRNA enrichment with μ MACS mRNA Isolation Kit (Miltenyi Biotec).

A non-normalized cDNA library was prepared using cDNA Synthesis System Kit with random primer GS Rapid Library Prep Kit and GS Rapid Library MID Adaptors Kit (Roche). The High Sensitivity DNA LabChip Kit (Agilent Technologies) with 2100 Bioanalyzer was used for quality analysis of the cDNA library. The concentration of sample (molecules/ μ L) was obtained by QuantiFluor™-ST fluorimeter (Promega). Titration of emPCR (emulsion PCR) was performed with the GS FLX Titanium SV em PCR Kit (Lib-L) (Roche), according to the emPCR Amplification Method Manual—Lib-L SV, GS FLX+ Series, to identify the optimal number of DNA molecules per bead (cpb = copies per bead). After emPCR titration, the emPCR was performed with GS FLX Titanium LV emPCR Kit (Lib-L) (Roche), according to the emPCR Amplification Method Manual—Lib-L LV, GS FLX+ Series. The transcriptome sequencing was conducted using the Roche/454 technology (GS FLX Titanium Sequencing Kit XL+) from HELIXXA company (Campinas, SP, Brazil), which has been used for transcriptome analysis of non-model fish species (Renaut et al., 2010).

Bioinformatic Analysis

Filtering of the initial quality of the 454 sequences in *stf* format was performed using the Roche Newbler programme. Sequence analysis was performed using the high-throughput sequencing module of CLC Genomics Workbench (version 7.5.1; CLC bio, Aarhus, Denmark). The raw reads were cleaned by trimming low quality sequences with quality scores of <20. Terminal nucleotides (five nucleotides at each extremity 5' and 3'), ambiguous nucleotides, adapter sequences and reads <15 base pairs (bp) were discarded. For *de novo* assembly, contigs <200 bp were also discarded and the default local alignment settings were used to rank potential matches (mismatch cost of 2, insertion cost

of 3, deletion cost of 3). The highest scoring matches that shared $\geq 50\%$ of their length with $\geq 80\%$ of similarity were included in the alignment. The assembled transcripts were subjected to cd-hit-est programme with an identity threshold of 90% to remove redundancy (Li and Godzik, 2006; Duan et al., 2012). In order to remove any mitochondrial and ribosomal contamination, sequences were compared against pacu mitochondrial genome and zebrafish ribosomal RNA RefSeqs (NCBI database) using CLC Genomic Workbench (version 8.0.3; CLC Bio, Aarhus, Denmark).

Functional annotation of the unique consensus sequences was performed by homology searches against the National Center for Biotechnology Information (NCBI) non-redundant protein database (nr) (cutoff E-value of $1E-3$) using BLAST2GO software (Conesa et al., 2005) to obtain the putative gene identity. All BLASTx hits were filtered for redundancy in protein accessions. The gene ontology (GO) terms were assigned to each unique gene based on the GO terms annotated to the corresponding homologs in the NCBI database (e-value cutoff $1E-6$). The transcripts were further annotated in InterPro, Enzyme code (EC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways analysis through the Bi-directional Best Hits (BBH) method.

Microsatellites were identified in the contigs using msatcommander software (Faircloth, 2008). Primers flanking the microsatellite loci were designed with Primer3plus software (Rozen and Skaletsky, 2000). The six possible reading frames of the consensus sequence of each functionally annotated contig containing microsatellite were compared against the NCBI protein database using BLASTx (e-value $1E-10$) in order to find Open Reading Frame (ORF) regions. These approaches allowed us to locate microsatellites in coding sequences (CDS) or untranslated regions (5'UTR and 3'UTR) through graphical sequence viewer Tablet (Milne et al., 2013).

Microsatellite Genotyping and Validation

DNA was extracted from fin fragments using the Wizard Genomic DNA Purification Kit (Promega), according to the manufacturer's protocol. Microsatellite validation was performed in 30 loci, selected according to the motif and functional annotation of the contigs. Amplifications were performed by polymerase chain reaction (PCR) in a total volume of 25 μ l containing 100 μ M of each dNTP (dATP, dTTP, dGTP, and dCTP), 1.5 mM MgCl₂, 1X Taq DNA buffer (20 mM Tris-HCl, pH 8.4, and 50 mM KCl), 0.1 μ M of each primer, 0.5 units of Taq Polymerase (Invitrogen) and 10–50 ng of genomic DNA. The reactions were performed in a thermocycler (ProFlex™ PCR System, Life Technologies) following initial denaturing for 10 min at 95°C; 35 cycles of 30 s at 95°C, 30 s at 55–60°C (adjusted for each primer set), 20 s at 72°C; and a final extension at 72°C for 20 min.

Microsatellites that showed polymorphism in 6% polyacrylamide gels were analyzed in a 3130xl sequencer (Life Technologies) to get better accuracy of allele determination. The sequencing strategy adopted in this study was according to protocols described by Schuelke (2000), using the CAGtag primer (5'-CAGTCGGGGCGTCATCA-3'; Shirk et al., 2013) labeled with the fluorochromes HEX or FAM. The genotyping

PCR was performed with the following reagents: 100 μ M of each dNTP, 1.5 mM $MgCl_2$, 1X *Taq* DNA buffer, 0.1 μ M of each primer (F and R), 0.01 μ M of the CAGtag primer, 0.5 units of *Taq* Polymerase (Invitrogen), and 10–50 ng of genomic DNA. The cycling programme for amplification consisted of nine cycles at 95°C for 30 s, 55–60°C for 30 s (adjusted for each primer set), 72°C for 20 s; then, 30 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 20 s. During the first nine cycles, the annealing temperature of 55–60°C allows incorporation of the primers (F and R) from the microsatellite loci. Then, in the following 30 cycles, the temperature of 50°C facilitates the annealing of the fluorescent dye-labeled CAGtag primer. PCR products were analyzed by capillary electrophoresis with a 3130xl genetic analyzer, using the DS-30 matrix, with the GeneScan 500 ROX dye Size Standard (Thermo). The programme GeneMapper 3.7 (Applied Biosystems) was used to determine the allele sizes.

Microsatellite Diversity and Population Analysis

For statistical analysis, we initially used GenAlex analysis 6.1 software (Peakall and Smouse, 2012) to convert the arrays into specific formats for each programme. The observed (H_o) and expected (H_e) heterozygosity, Hardy-Weinberg Equilibrium (HWE) and Analysis of Molecular Variance (AMOVA) (Excoffier et al., 1992) were calculated using the Arlequin 3.5 programme (Excoffier and Lischer, 2010). The levels of significance for the HWE test were adjusted with the Bonferroni correction (Rice, 1989). The inbreeding coefficient (F_{IS}) was performed using Genepop 4.0.11 (Rousset, 2008), based on Weir and Cockerham (1984) estimates. The fixation index (F_{ST}) was calculated using FSTAT 9.3.2 software (Goudet, 1995). Wright (1965) threshold values were adopted, F_{ST} = little genetic differentiation (0–0.05); moderate genetic differentiation (0.05–0.25); high level of genetic differentiation (> 0.25). The programme Cervus v.3.0.7 (Marshall et al., 1998) was applied to verify the presence of null alleles. Linkage disequilibrium (LD) was estimated using Arlequin v.3.5.2.2. The levels of significance were adjusted to multiple tests using the Bonferroni correction.

After LD analysis, level of admixture among population samples was inferred by estimating the optimum number of clusters (K), as suggested by Evanno et al. (2005), using the programme STRUCTURE version 2.3.4 (Pritchard et al., 2000) without prior information about population. Primarily, we determined the distribution of ΔK , an *ad hoc* statistic based on the rate of change in the log probability of data between successive K values. The range of clusters (K) was predefined from 1 to 5. The analysis was performed in 25 replicated runs using 200,000 iterations after a burn-in period of 50,000 runs. The K value most likely to explain the population structure is the modal value of this ΔK . The outputs of STRUCTURE analysis were visualized through the STRUCTURE HARVESTER programme (Earl, 2012).

Analysis for population bottlenecks was tested using BOTTLENECK (Cornuet and Luikart, 1996; Piry et al., 1999), by using the mutation-drift equilibrium assuming the two-phase model (TPM) with 70% stepwise mutation model (SMM)

and 30% infinite allele model (IAM). Deviations between the observed and expected frequency distributions were tested using the Wilcoxon's signed rank test. BOTTLENECK was run for 10,000 iterations.

RESULTS

Transcriptome Sequencing

The results of liver transcriptome sequencing in pirapitinga yielded a total of 192,373 reads, which were deposited in the Short Read Archive (SRA) of NCBI under the accession number SRR6303971. The raw reads presented an average length of 395.5 bp, comprising a total of ~76 Mbp. After the trimming process, the average length of the reads was of 362.1 bp, resulting in a total of ~69 Mbp (192,077 reads; Table 1). As *P. brachyomus* is considered a non-model organism, and therefore without reference genome, *de novo* assembly strategy was performed for transcriptome analysis, which yielded 3,696 non-redundant contigs as a result of 174,272 overlapping reads (63,460,229 bp). The size characteristics of the contigs are presented in Table 1. A total of 17,805 remaining reads (6,084,530 bp) was considered as singletons, and therefore they were not used for subsequent analysis.

Non-redundant sequences were annotated by BLASTx algorithm against the NCBI databases: non-redundant protein (*nr*), protein RefSeq of *zebrafish* and *fugu*. A total of 2,568 unique protein accessions (69.4% of transcripts) had significant similarity in the *nr* database. In relation to the protein RefSeq of *zebrafish* and *fugu*, we found similar numbers of annotated genes, which were of 2,498 (67.6%) and 2,419 (65.4%), respectively. No sequence showed homology with known pirapitinga protein sequences deposited in NCBI database, because the available sequences database is still limited to mostly mitochondrial sequences.

Of the 2,568 contigs with correspondence in the *nr* database, 2,075 (80.8%) were annotated in the categories of Gene Ontology (GO). A total of 1,831 assignments to Biological Process (88.2%) were found, followed by 1,757 to Molecular Function (84.6%) and 1,378 to Cellular Component (66.4%). In relation to the GO subcategories, the most abundant terms were related to: metabolic process, cellular process, and single-organism process of the Biological Process category; binding, catalytic activity, and

TABLE 1 | Data of *de novo* assembly from liver transcriptome of pirapitinga *Pirapetichthys brachyomus*.

Matched reads for assembly	174,272
Total nucleotides of matched reads	63,460,229
Number of contigs	3,797
Total of contig nucleotides	2,898,680
Minimum contig length (bp)	202
Maximum contig length (bp)	7,812
Average contig length (bp)	790
N75 (bp)	583
N50 (bp)	861
N25 (bp)	1,383

transporter activity of the Molecular Function; cell, organelle, and membrane of the Cellular Component (Figure 1). In the present study, genes assigned to the immune system, growth and reproduction were found, and therefore these data will serve as support for future studies on the aquaculture of pirapitinga.

The transcripts characterization in the KEGG database demonstrated that 1,122 sequences were identified in 106 metabolic pathways. Genes involved in the biosynthesis of antibiotics, purine metabolism and glycolysis/gluconeogenesis were able to be highlighted (Figure 2).

Microsatellite Diversity and Population Analysis

The search for short sequence repeats (SSR) in the 3,696 contigs resulted in the discovery of 130 microsatellite markers distributed in 95 contigs. In total, 75 pairs of primers were designed adjacent to the microsatellite loci, including the following sequence repeats: 56 di, 13 tri, 4 tetra, and 2 pentanucleotide. Among the dinucleotide motifs, the main repeats were the types AC (48.28%), AG (39.65%), AT (10.35%), and CG (1.72%). In relation to the trinucleotide motifs, we identified seven types (AGC, AGG, ATC, AAT, ACG, CCG, and AAG). The tetranucleotide (ATCT, AAAG, AATG, and AAAT) and pentanucleotide (ACTAT and ATAGT) sequences were described with the presence of four and two types of motifs. In relation to the gene position, 26.76% of the microsatellite markers were found in the 3'UTR (untranslated region), 19.71% in the 5'UTR, and 29.58% in the cds (coding sequence).

In the process of microsatellite validation, 30 markers were evaluated in 22 samples of pirapitinga collected from the wild. Of these markers, eight microsatellite loci showed polymorphism (GenBank accession numbers MG595996–MG596003), revealed by the presence of different fragment sizes (Table 2). The number of alleles was low, which ranged from 2 (loci C25, C64, C410, and C1832) to 5 (C1376) and mean of 2.750 ± 0.366 . The expected (H_e) and observed (H_o) heterozygosity in the wild population had an average of 0.466 ± 0.061 and 0.355 ± 0.076 , respectively. Most of the loci showed positive values for F_{IS} , except the locus C64. Three microsatellite loci (C13, C25, and C1716) showed significant deviation from the Hardy–Weinberg Equilibrium (HWE) after Bonferroni correction (adjusted $p = 0.00625$).

The results of genetic variability in farmed stocks revealed that North fish farms TO1 and TO2 had higher diversity than the wild population, demonstrated by number of alleles (mean of 4.500 ± 0.423 and 3.375 ± 0.596 , respectively) and average values of H_e (0.589 ± 0.033 and 0.488 ± 0.044 , respectively) and H_o (0.520 ± 0.060 and 0.447 ± 0.040 , respectively; Table 3). The Southeast fish farms SP1 and SP2 showed the lowest genetic variability when compared to other populations, with lower allele number (mean of 2.250 ± 0.313 and 3.125 ± 0.581 , respectively), and average of H_e (0.226 ± 0.077 and 0.278 ± 0.085 , respectively; $p < 0.05$) and H_o (0.259 ± 0.103 and 0.251 ± 0.079 , respectively; Table 3). Most of the microsatellite loci were characterized with positive values of F_{IS} , except for SP1 and SP2. The mean value of

F_{IS} and null alleles was positive in most populations, with the exception of SP1 (-0.071 ± 0.076 and -0.011 ± 0.080). The majority of the markers were in concordance to HWE, after Bonferroni correction, with the exception of C25 (TO1, SP1, and SP2), C64 (SP1 and SP2) and C1376 (SP1) (Table 3). Linkage disequilibrium was found between the microsatellites C410 and C1005 ($p < 0.00625$). Although molecular markers on linkage disequilibrium were not applied in genetic variability studies, this information can be useful in future analysis of genetic mapping.

In bottleneck analyses, evidence for recent reductions in population size (bottleneck) using TPM was not found, except for the wild population of Tocantins River ($p = 0.027$).

In the evaluation of the level of admixture among stocks by STRUCTURE, the model-based clustering analyses detected $K = 3$, allowing the identification of 3 main clusters between the populations: Group 1 (SP1 and SP2), Group 2 (TO1 and TO2), and Group 3 (wild) (Figure 3).

The global F_{ST} was 0.379, which showed high genetic differentiation among the populations ($p < 0.05$). Pairwise F_{ST} detected a higher genetic differentiation between the wild population and all farmed stocks, particularly when compared to SP1 ($F_{ST} = 0.538$, $p < 0.05$) and SP2 ($F_{ST} = 0.537$, $p < 0.05$). Additionally, high genetic structure was found between the populations from North and Southeast Brazil, as observed between TO2 with SP1 ($F_{ST} = 0.463$, $p < 0.05$) and TO1 with SP2 ($F_{ST} = 0.380$, $p < 0.05$; Table 4). Moreover, values of pairwise F_{ST} after stock clustering detected a higher genetic differentiation when comparing Group 1/Group 2 ($F_{ST} = 0.379$, $p < 0.05$), Group 1/Group 3 ($F_{ST} = 0.549$, $p < 0.05$), and Group 2/Group 3 ($F_{ST} = 0.144$, $p < 0.05$).

The results of AMOVA showed that the majority of genetic variation (29.11%, $F_{CT} = 0.291$, $p < 0.001$) occurred between groups (according to STRUCTURE clustering), while the variation among individuals within populations was only 8.21% ($F_{IS} = 0.126$, $p < 0.001$) and among populations within groups presented 6.06% of genetic variation ($F_{SC} = 0.085$, $p < 0.001$).

DISCUSSION

Transcriptome Sequencing

Currently, genetic resources for pirapitinga *P. brachypomus* are limited only to sequences of the mitochondrial genome (Chen et al., 2016). Thus, one of the main results of this study was the data generated through transcriptome sequencing, because little knowledge was available about the genes of this species. The efficiency of the Roche/454 sequencing system in the functional genomics analysis of pirapitinga can be observed because of the 3,696 transcripts that were generated in this study. According to Seeb et al. (2011), genome reduction strategies for NGS sequencing (e.g., transcriptome sequencing) are more viable when the objective is to prospect molecular markers and genetic information for use in aquaculture, in a low cost and fast way. Roche/454 sequencing technology is one of the main methods used in NGS transcriptome of non-model fish (Renaut et al., 2010;

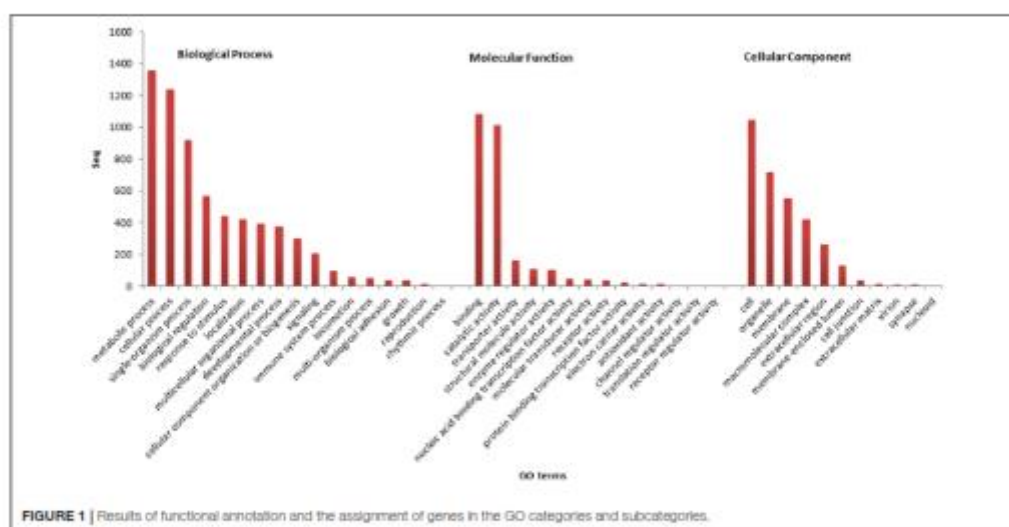


FIGURE 1 | Results of functional annotation and the assignment of genes in the GO categories and subcategories.

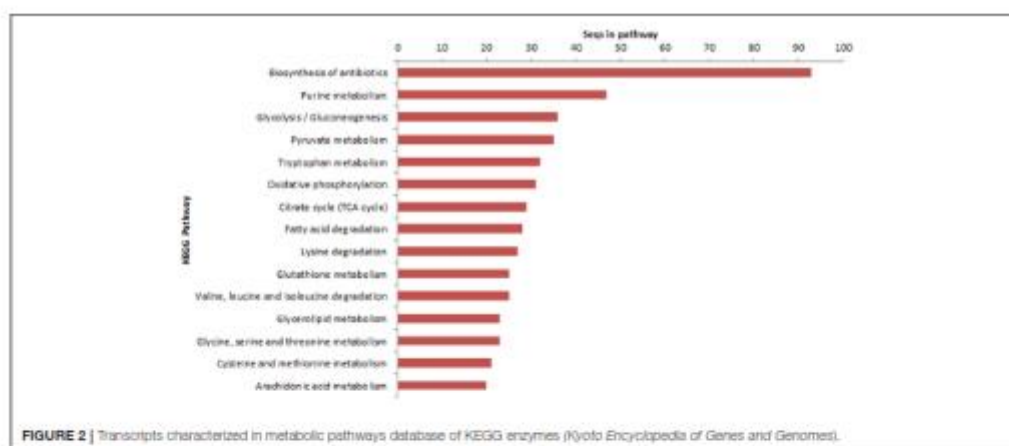


FIGURE 2 | Transcripts characterized in metabolic pathways database of KEGG enzymes (Kyoto Encyclopedia of Genes and Genomes).

Shin et al., 2012; Calduch-Giner et al., 2013; Mutz et al., 2013).

The results of functional annotation showed that the sequences of pirapitinga had a high proportion of annotated genes when compared to the database of *zebrafish* and *fugu* proteins. The gene annotation allowed identification of genomic regions responsible for ontogenetic development processes, biological regulation, the immune system, and regions involved in processes of growth and reproduction. Consequently, the present data can be used as the basis of further biological studies of other areas of aquaculture

or for future breeding programmes. In addition, through transcriptome sequencing, the discovery of gene-associated microsatellites can be considered to be the main result which can be applied to pirapitinga aquaculture, as already demonstrated in previous studies of fish (Renaut et al., 2010; Helyar et al., 2012; Shin et al., 2012). The use of gene-associated markers becomes even more important in the construction of genetic maps (Shin et al., 2012) because, by comparative genomics using fish genome references already sequenced, it is possible to presume the location of each studied locus.

TABLE 2 | Characterization of the genetic diversity of eight polymorphic microsatellites in the wild population of pirapitinga (*Piaractus brachipomus*).

Locus	Sequence description	Gene position	Motifs	Primers 5'–3'	T _A °C	Size range	N _a	H _o	H _e	P(HWE)	F _{IS}	F(Nul)
C13	dihydroxyvitamin d24-like	2'UTR	(AGC) ₈	F: TCTCTTCAAGCCTCTCTGC R: ATGCTGCAGCTCTCTCTGT	60°C	143–148	3	0.545	0.669	0.000	0.188	0.075
C25	cytosolic 5-nucleotidase 3a-like isoform x1	3'UTR	(AT) ₁₁	F: CTTTGTCTGCTTTGGGTCGT R: CTAGAAGAATGTGCAAAATTGAAA	60°C	117–120	2	0.000	0.169	0.001	1.000	0.887
C64	sodium-coupled neutral amino acid transporter	5'UTR	(AAG) ₇	F: CAAAGCAAACCTCAAAAGGAAAA R: TGGGAACGTTTAGCATCTCA	55°C	143–151	2	0.545	0.474	0.650	–0.156	–0.082
C410	apolipoprotein e	2'UTR	(AG) ₆	F: CGCACAGGTCTAAAGGCACT R: CTCACACAGTGAAGAAAGCA	60°C	125–127	2	0.273	0.359	0.271	0.245	0.125
C1005	apolipoprotein e	3'UTR	(AG) ₇	F: AGTTGTGTGACCAATGACAG R: CTTGTTCCTCCACACAGT	60°C	137–141	3	0.318	0.369	0.225	0.140	0.080
C1376	fructose-bisphosphatase 1-like	5'UTR	(AC) ₁₀	F: GTGTATACATGGCAGGCGTTT R: CAAGTGAGACCAATCCAAGG	60°C	157–175	5	0.500	0.608	0.185	0.180	0.073
C1716	vacuolar atp synthase 16 kda proteolipid subunit	2'UTR	(GT) ₇	F: AACCGAAGAGAGGGGAGTGT R: GCATTTACAAAGGGACGAC	60°C	155–171	3	0.545	0.659	0.000	0.175	0.059
C1832	–	–	(AC) ₆	F: GGTGCTATGTCTAGAGGCC R: AGGAAGCATGACCACTGTG	60°C	159–169	2	0.111	0.529	0.036	0.800	ND

The wild population analyzed corresponds to 22 individuals collected on the Tocantins River. T_A, annealing temperature (°C); N_a, number of alleles per locus; H_o, observed heterozygosity; H_e, expected heterozygosity; P (HWE), p-value from Hardy-Weinberg equilibrium and F_{IS}, inbreeding coefficient. F(Nul), Null alleles; ND, not performed.

Moreover, some examples have demonstrated that gene-linked microsatellite markers can be correlated with interesting productive traits, especially for growth performance. In the fish *Sparus aurata*, a dinucleotide microsatellite in the 5' UTR of the growth hormone gene (GH) is linked with faster growth rate, especially the alleles 250 and 254, which can be used for breeding management and genetic selection for this trait (Almuly et al., 2005). In other fish species, such as *Oreochromis niloticus* and *Lates calcarifer* (Yue et al., 2001; Yue and Orban, 2002), microsatellites have also been reported for genes of interest (prolactin, GH and igf2) and, therefore, they can be used in marker-assisted selection (MAS) programmes. In the present study, eight polymorphic microsatellite loci were validated, some of them located in gene regions that may be useful for productive characteristics in aquaculture. In this case, a microsatellite locus was found in the gene *Tetraspanin-3 isoform x1* (C1832), which plays a role in viral infection pathology (Martin et al., 2005; Shoshana and Shoham, 2005). There is another microsatellite in the gene *Cytosolic 5-nucleotidase 3a-like* (NTC5C3) (C25), which contributes in the production of red blood cells and its mutation can cause hemolytic anemia and influence on the immune system (Aksoy et al., 2009). Thus, the microsatellites described in this study will be also important in future analysis of (QTL) linked to traits of disease resistance, which has received special attention in aquaculture species, such as turbot (*Scophthalmus maximus*), rainbow trout (*Oncorhynchus mykiss*), salmon (*Salmo salar*), Nile tilapia (*O. niloticus*), and cod (*Gadus morhua*), investigating the resistance to pathogens (Pardo et al., 2008; Ødegård et al., 2010, 2011;

Yáñez et al., 2014; Evenhuis et al., 2015). Furthermore, one microsatellite locus was also detected in the gene *apolipoprotein e* (C410), which is associated with the central nervous system and the senescence process (Wang et al., 2014). These markers can provide useful information for studies of the biology of the pirapitinga, besides serving as a framework for other native species.

Population Analysis

The validation of eight microsatellites showed a low level of genetic diversity in these loci, both in wild and farmed stocks. In the wild, the observed heterozygosity (H_o) ranged from 0.000 to 0.545 and an average of 2.750 alleles per locus. These values confirm the low genetic variability when compared with related species, such as pacu *P. mesopotamicus* (H_o range from 0.068 to 0.911 and average of 8.5 alleles per locus), and tambaqui *C. macropomum* (H_o range from 0.430 to 0.880 and average of 12.8 alleles per locus; Calcagnotto and DeSalle, 2009; Fazzi-Gomes et al., 2017). In contrast to neutral markers (microsatellites in noncoding regions), gene-associated microsatellites might be more susceptible to selection pressure and, therefore, they have low values of gene diversity.

Analysis of the genetic diversity in pirapitinga farmed stocks showed significant differences between fish farms in different regions of Brazil, two from the Southeast (São Paulo State: SP1 and SP2) and two from the North (Tocantins State: TO1 and TO2). In general, farmed stocks were expected to have low genetic variability as a result of genetic decline, genetic

TABLE 3 | Values of genetic diversity of eight microsatellite loci of *Piaractus brachipomus*.

Stocks		Loci							
		C13	C25	C64	C410	C1005	C1376	C1716	C1832
Wild	N	22	22	22	22	22	22	22	9
	Na	3	2	2	2	3	5	3	2
	H _o	0.545	0.000	0.545	0.273	0.318	0.500	0.545	0.111
	H _e	0.669	0.169	0.474	0.359	0.369	0.608	0.659	0.529
	P(HWE)	0.000	0.001	0.650	0.271	0.225	0.185	0.000	0.036
	F _{IS}	0.188	1.000	-0.156	0.245	0.140	0.180	0.175	0.800
	F(Null)	0.075	0.887	-0.082	0.125	0.090	0.073	0.059	ND
TO1	N	25	25	25	25	25	25	25	25
	Na	4	5	3	3	4	6	6	5
	H _o	0.600	0.160	0.560	0.520	0.560	0.440	0.760	0.560
	H _e	0.536	0.704	0.485	0.537	0.562	0.609	0.782	0.580
	P(HWE)	0.650	0.001	0.365	0.267	0.421	0.010	0.545	0.388
	F _{IS}	-0.121	0.776	-0.133	0.034	0.004	0.282	0.028	0.035
	F(Null)	-0.067	0.620	-0.065	-0.015	-0.034	0.164	0.011	-0.001
TO2	N	26	21	26	26	26	26	26	26
	Na	3	2	2	2	2	6	5	5
	H _o	0.500	0.380	0.346	0.384	0.384	0.538	0.461	0.653
	H _e	0.528	0.315	0.382	0.506	0.506	0.632	0.515	0.638
	P(HWE)	0.840	1.000	0.626	0.256	0.256	0.020	0.030	0.670
	F _{IS}	0.055	-0.212	0.096	0.244	0.244	0.151	0.107	-0.253
	F(Null)	0.018	-0.103	0.040	0.127	0.127	0.055	0.067	-0.027
SP1	N	14	14	14	14	14	14	14	14
	Na	2	1	1	3	2	3	3	3
	H _o	0.143	-	-	0.287	0.500	0.857	0.143	0.143
	H _e	0.137	-	-	0.264	0.494	0.634	0.140	0.203
	P(HWE)	1.000	-	-	1.000	1.000	0.000	1.000	0.109
	F _{IS}	-0.040	-	-	-0.083	-0.011	-0.368	-0.019	0.306
	F(Null)	-0.028	-	-	-0.069	-0.023	-0.215	-0.027	0.272
SP2	N	19	19	19	19	19	16	18	19
	Na	2	2	1	3	3	5	3	6
	H _o	0.368	0.000	-	0.157	0.157	0.625	0.222	0.473
	H _e	0.308	0.193	-	0.152	0.152	0.790	0.207	0.486
	P(HWE)	1.000	0.002	-	1.000	1.000	0.112	1.000	0.387
	F _{IS}	-0.200	1.000	-	-0.038	-0.385	0.214	-0.070	0.027
	F(Null)	-0.099	0.916	-	-0.032	-0.032	0.093	-0.051	0.024

Wild, population from the Tocantins River; TO1 and TO2, fish farms from Tocantins; SP1 and SP2, fish farms from São Paulo; Na, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; P (HWE), P-value from Hardy-Weinberg equilibrium; F_{IS}, inbreeding coefficient; F(Null), Null alleles; ND, not performed.

drift, selection and inbreeding (Theodorou and Couvet, 2015). However, the results of this study showed higher genetic variability in breeding stocks from North fish farms in relation to the wild stocks ($p < 0.05$; higher values of allelic frequency and heterozygosity), which was also observed in studies with other related species (Barroso et al., 2005; Panarari-Antunes et al., 2011). The basis of this result could be considered from three different perspectives: (1) North fish farms had originated from different wild stocks resulting in high level of genetic variability;

(2) problems of sample size bias, such as few microsatellite loci and individuals analyzed; (3) evidence for recent genetic bottlenecks in the wild population. Some studies of fish have reported bottlenecks in natural populations, particularly due to habitat loss and fragmentation by human disturbance (Brauer et al., 2016). In the case of pirapitinga, the fragmentation of the Tocantins River by hydroelectric dams in the 80's and 90's (e.g., Tucuruí and Luiz Eduardo Magalhães dams, where wild fish were collected for this study) could be responsible for

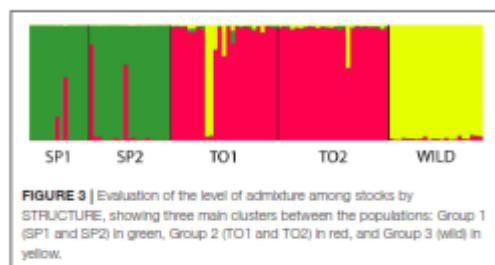


FIGURE 3 | Evaluation of the level of admixture among stocks by STRUCTURE, showing three main clusters between the populations: Group 1 (SP1 and SP2) in green, Group 2 (TO1 and TO2) in red, and Group 3 (WILD) in yellow.

TABLE 4 | Analysis of pairwise F_{ST} based on eight microsatellite loci between populations of *Pirapitinga brachycephala*.

	SP1	SP2	TO1	TO2	wild
SP1	–				
SP2	0.18438	–			
TO1	0.39201	0.38080	–		
TO2	0.40373	0.42999	0.08345	–	
wild	0.53856	0.53782	0.16086	0.17915	–

Wild, population from the Tocantins River; TO1 and TO2, fish farms from Tocantins; SP1 and SP2, fish farms from São Paulo. All results of F_{ST} were significant statistically $p < 0.05$.

a population reduction and subsequent genetic variation loss detected by our microsatellite analysis. There are considerable numbers of hydropower dams in the basin, which can affect the reproduction, migratory routes, and egg and larvae drift of fish (Agostinho et al., 2008). Alteration of the migratory flow consequently leads to a decrease in or interruption of the gene flow, reducing the population size, which makes the fish more susceptible to the effects of genetic drift (Hatanaka and Galetti, 2003), which results in genetic structure for some fish species (Calcagnotto and DeSalle, 2009; Do Prado et al., 2018).

STRUCTURE and pairwise F_{ST} analyses suggested a high genetic structure between the stocks herein analyzed, particularly as result of the fixation of specific alleles in some loci, which resulted in three clusters (Figure 3). There are three hypothetical explanations for these genetic patterns: (1) differentiation of wild population in relation to farmed stocks, which could be due to the selection of the fittest individuals for farming systems or low number of founders for the establishment of the farmed broodstocks; (2) lower genetic structure in North/wild than Southeast/wild, which suggests that North fish farms had frequent broodstock renovation from the wild; (3) fish farms were genetically clustered due to the geographic distribution, i.e., the degree of genetic similarity is higher when one fish farm is closer to the other, indicating interchange of individuals between nearby fish farms, common origin of the farmed broodstocks, or fixation/selection of specific alleles for different climatic conditions that are found in Brazil (North and South). However, these genetic patterns should be also evaluated using neutral

markers (microsatellites in noncoding regions) and through techniques of higher genome coverage (SNP, single-nucleotide polymorphism).

Through AMOVA analysis, the main genetic variation was found to be present within populations (64.8%). This genetic pattern has also been reported in studies carried out with pacu (Calcagnotto and DeSalle, 2009; Iervolino et al., 2010) and tambaqui (Aguiar et al., 2013). Moreover, highly significant genetic variation was associated with differences between groups (Wild, SP, and TO), which represented 29.11% of genetic variation, in contrast to low differences among populations within groups (6.06%).

In general, our study of genetic characterization in pirapitinga farmed stocks provides important insights which can lead to better management of this species in aquaculture. Our results are fundamental to beginning a breeding programme, since the genetic structure should be taken into consideration when composing an initial base population, where matings between farmed individuals from North and Southeast Brazil are shown to result in higher genetic variability in the families. Moreover, the data suggested levels of genetic diversity which were higher in farmed stocks than in wild fish, discarding the occurrence of inbreeding. In general, lack of knowledge on genetic variability of stocks can result in inbreeding and fixation of deleterious genes, reduced growth rates, disease resistance problems and reduced ability to adapt to new environments (Arkush et al., 2002; Gallardo et al., 2004; Neira et al., 2006; Hillen et al., 2017). Therefore, besides the identification of QTL to assist in the selection of superior genotypes by MAS, studies of microsatellites are important for genetic monitoring, supporting pirapitinga aquaculture and increasing its productivity.

FINAL CONSIDERATIONS

The prospection of genetic data for pirapitinga is one of the priority issues for aquaculture, since this species is of high economic importance in national and global fish farming. The identification of gene-associated microsatellites by NGS is fundamental to understanding the genetic structure of wild and farmed populations, providing support for further management programmes and genetic pre-breeding programmes. Moreover, the microsatellites described herein are interesting targets used to find QTL markers, specifically related to the immune system of pirapitinga.

AUTHOR CONTRIBUTIONS

PJ: Acquisition, analysis and interpretation of data, draft of the work, final approval of the version; VM-F, RA, and MdF: Draft of the work, development of intellectual content, final approval of the version; MH: Analysis and interpretation of data, draft of the work, final approval of the version; NM: Analysis and interpretation of data, draft of the work, final approval of the version; MV: Analysis and interpretation of data, draft

of the work, final approval of the version; FP-F: Acquisition, Interpretation of data, development of intellectual content, final approval of the version; DH: Draft of the work, development of intellectual content, writing of the manuscript, final approval of the version.

REFERENCES

- Agostinho, A. A., Pelicice, F. M., and Gomes, L. C. (2008). Dams and the fish fauna of the Neotropical region: impacts and management related to diversity and fisheries. *Braz. J. Biol.* 68, 1119–1132. doi: 10.1590/S1519-69842008000500019
- Aguar, J., Schneider, H., Gomes, F., Carneiro, J., Santos, S., Rodrigues, L. R., et al. (2013). Genetic variation in native and farmed populations of *Tambaqui* (*Colossoma macropomum*) in the Brazilian Amazon: regional discrepancies in farming systems. *Am. Acad. Bras. Ciênc.* 85, 1439–1447. doi: 10.1590/0001-376520130007
- Akopy, P., Zhu, M. J., Kalari, K. R., Moon, I., Pellemounter, L. L., Eckloff, B. W., et al. (2009). Cytosolic 5'-nucleotidase III (NT5C3): gene sequence variation & functional genomics. *Pharmacogenet. Genomics* 19, 567–576. doi: 10.1097/PPG.0b013e3181652832c14b8
- Alcantara, P. F., Oliveira, A. A., and Nobre, M. L. S. N. (1990). Considerações sobre a amostragem da pirapitinga, *Colossoma brachypomum*, Cuvier, no estado do Ceará (Brasil). *Ciênc. Agem.* 21, 43–49.
- Almaly, K., Poleg-Darin, Y., Gershkov, S., Gershkov, G., Rapoport, B., Soller, M., et al. (2005). Characterization of the 5' flanking region of the growth hormone gene of the marine teleost, gilthead sea bream *Sparus aurata*: analysis of a polymorphic microsatellite in the proximal promoter. *Fish. Sci.* 71, 479–490. doi: 10.1111/j.1444-2906.2005.00991.x
- Arkuh, K. D., Giam, A. R., Mendonça, H. L., McBride, A. M., Marty, G. D., and Hedrick, P. W. (2002). Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. *Can. J. Fish. Aquat. Sci.* 59, 966–975. doi: 10.1139/f02-066
- Barroso, B. M., Hildorf, A. W. S., Moreira, H. L. M., Cabello, P. M., and Traub-Csisko, Y. M. (2005). Genetic diversity of wild and cultured populations of *Brycon opulens* (Cuvier, 1819) (Characiformes, Characidae, Bryconinae) using microsatellites. *Aquaculture* 247, 51–65. doi: 10.1016/j.aquaculture.2005.02.004
- Brauer, C. J., Hammer, M. P., and Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Mol. Ecol.* 25, 5093–5113. doi: 10.1111/mec.13830
- Calcagno, D., and DeSalle, R. (2009). Population genetic structuring in pacu (*Piaractus mesopomus*) across the Paraná-Paraguay basin: evidence from microsatellites. *Neotrop. Ichthyol.* 7, 607–616. doi: 10.1590/S1679-62252009000400008
- Caldwell-Gibson, J. A., Bermejo-Nogales, A., Benedito-Palos, L., Estensoro, I., Ballester-Lorenzo, G., Sijth-Bobadilla, A., et al. (2013). Deep sequencing for de novo construction of a marine fish (*Sparus aurata*) transcriptome database with a large coverage of protein-coding transcripts. *BMC Genomics* 14, 178. doi: 10.1186/1471-2164-14-178
- Chen, H. P., Li, S. S., Xu, Z. Z., Zhang, Y., Zhu, C. H., Deng, S. P., et al. (2016). The complete mitochondrial genome of the *Piaractus brachipomus* (Characiformes: Characidae). *Mitochondrial DNA* 27, 1289–1290. doi: 10.3109/19401736.2014.945560
- Comesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Cornuet, J. M., and Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144, 2001–2014.
- Do Prado, P. D., Fernandez-Cabrán, R., Foresti, F., Oliveira, C., Martínez, P., and Porto-Forrest, F. (2018). Genetic structure and evidence of anthropogenic effects on wild populations of two Neotropical catfishes: baselines for conservation. *J. Fish. Biol.* 92, 55–72. doi: 10.1111/jfb.13486
- Duan, J., Xia, C., Zhao, G., Jia, J., and Kong, X. (2012). Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *BMC Genomics* 13, 392. doi: 10.1186/1471-2164-13-392
- Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12666-011-9548-7
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Evenhuis, J. P., Leida, Y. D., Marancik, D. P., LaPatra, S. E., and Wiens, G. D. (2015). Rainbow trout (*Oncorhynchus mykiss*) resistance to columnaris disease is heritable and favorably correlated with bacterial cold water disease resistance. *J. Anim. Sci.* 93, 1546–1554. doi: 10.2527/jas.2014-8566
- Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- Faircloth, B. C. (2008). Maelstrom: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Res.* 8, 92–94. doi: 10.1111/j.1471-8286.2007.01864.x
- Fazri-Gomes, P., Guerrero, S., Pallotta, G. D. A., Melo, N. P. A. C., Santos, S., and Hamay, I. (2017). High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. *Genet. Mol. Biol.* 40, 142–146. doi: 10.1590/1678-4665-gmb-2015-0222
- Fernández, J., Toro, M. A., Sonesson, A. K., and Villanueva, B. (2014). Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Front. Genet.* 5, 414. doi: 10.3389/fgen.2014.00414
- Flores Nava, A. (2007). "Aquaculture seed resources in Latin America: a regional synthesis," in *Assessment of Freshwater Fish Seed Resources for Sustainable Aquaculture*, ed M.G. Bonlad-Reantao (Rome: FAO Fisheries Technical Paper, No. 501, FAO), 91–102.
- Freunde, A., Lemis, G., Agudelo, E., and Olvera Angel, M. (2004). Espermación inducida y criopreservación de semen de cachama blanca (*Piaractus brachipomus*). *Rev. Colomb. Ciênc.* 17, 46–52.
- Gallardo, J. A., Garcia, X., Llorentia, J. P., and Neira, B. (2004). Inbreeding and inbreeding depression of female reproductive traits in two populations of Coho salmon selected using BLUP predictors of breeding values. *Aquaculture* 234, 111–122. doi: 10.1016/j.aquaculture.2004.01.008
- Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *J. Hered.* 86, 485–486. doi: 10.1093/oxfordjournals.jhered.a111627
- Guyon, R., Rakotoniranga, M., Azzouzi, N., Coutanceau, J. P., Bonillo, C., D'Costa, H., et al. (2012). A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* 13, 222. doi: 10.1186/1471-2164-13-222
- Hashimoto, D. Y., Mendonça, F. P., Senhorini, J. A., Oliveira, C., Foresti, F., and Porto-Forrest, F. (2011). Molecular diagnostic methods for identifying Serrasalmid fish (*Pacu*, *Pirapitinga*, and *Tambaqui*) and their hybrids in the Brazilian aquaculture industry. *Aquaculture* 321, 49–53. doi: 10.1016/j.aquaculture.2011.08.018
- Hashimoto, D. Y., Senhorini, J. A., Foresti, F., Martínez, P., and Porto-Forrest, F. (2014). Genetic identification of F1 and post-F1 Serrasalmid juvenile hybrids in Brazilian aquaculture. *PLoS ONE* 9, e89902. doi: 10.1371/journal.pone.0089902
- Hashimoto, D. Y., Senhorini, J. A., Foresti, F., and Porto-Forrest, F. (2012). Interspecific fish hybrids in Brazil: management of genetic resources for sustainable use. *Rev. Aquacult.* 4, 108–118. doi: 10.1111/j.1753-5131.2012.01067.x

- Hatanaka, T., and Galetti, P. M. Jr. (2003). RAPD markers indicate the occurrence of structured populations in a migratory freshwater fish species. *Genet. Mol. Biol.* 26, 19–25. doi: 10.1590/S1415-47572003000100004
- Helyar, S. J., Limborg, M. Y., Bekkevold, D., Nabbucci, M., van Houli, J., Maes, G. E., et al. (2012). SNP discovery using next generation transcriptomic sequencing in Atlantic herring (*Clupea harengus*). *PLoS ONE* 7:e42089. doi: 10.1371/journal.pone.0042089
- Hillen, J. E. J., Crocic, T., Vandeputte, M., Herlen, K., Hellemans, B., Maroso, F., et al. (2017). Estimates of genetic variability and inbreeding in experimentally selected populations of European sea bass. *Aquaculture* 479, 742–749. doi: 10.1016/j.aquaculture.2017.07.012
- Hongling, H. (2007). "Freshwater fish seed resources in China," in *Assessment of Freshwater Fish Seed Resources for Sustainable Aquaculture*, ed. M. G. Boodad-Kanantao (FAO Fisheries Technical Paper No. 501, Rome: FAO), 185–199.
- Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Mota-Velasco, J. C., Ghayas, A. A., et al. (2010). The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* 105, 318–327. doi: 10.1038/hdy.2009.171
- Hovse, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., et al. (2015). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–505. doi: 10.1038/nature12111
- IBGE (2016). Produção da Pecuária Municipal 2016. *Rio de Janeiro* 44, 1–51.
- Iervolino, P., Rueda, E. K., and Milder, A. W. S. (2010). The lack of genetic differentiation of paco (*Piaractus mesopotamicus*) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: implications for fishery management. *Fish. Res.* 101, 27–31. doi: 10.1016/j.fishres.2009.09.003
- 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:e35152. doi: 10.1371/journal.pone.0035152
- Keljonen, M. L., Tahtinen, J., Saita, M., and Koskinen, 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. doi: 10.1016/S0044-8486(01)00080-0
- Labrecque, L., Magyari, I., Hancz, C., and Weiss, S. (2005). Preliminary studies on the genetic variability of six Hungarian common carp strains using microsatellite DNA markers. *Hydrobiologia* 533, 223–228. doi: 10.1007/s10750-004-2490-x
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and compare large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Lien, S., Gidskehaug, L., Moen, T., Høyem, B. J., Berg, P. R., Daviduin, W. S., et al. (2011). A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homologies and striking differences in sex-specific recombination patterns. *BMC Genomics* 12:615. doi: 10.1186/1471-2164-12-615
- Lin, Y., Gao, Z., and Zhan, A. (2015). Introduction and use of non-native species for aquaculture in China: status, risk and management solutions. *Rev. Aquacult.* 7, 28–58. doi: 10.1111/raq.12052
- Liu, S., Zhou, Z., Lu, J., Sun, P., Wang, S., Liu, H., 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. doi: 10.1186/1471-2164-12-53
- Liu, Z. J., and Cordes, J. P. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37. doi: 10.1016/j.aquaculture.2004.05.027
- Machado-Schiaffino, G., Dopico, E., and Garcia-Vazquez, E. (2007). Genetic variation losses in Atlantic salmon stocks created for supportive breeding. *Aquaculture* 264, 59–65. doi: 10.1016/j.aquaculture.2006.12.028
- Marshall, T., Slate, J., Kruuk, L., and Pemberton, J. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7, 639–655. doi: 10.1046/j.1365-294x.1998.00374.x
- Martin, F., Roth, D. M., Jans, D. A., Pouton, C. W., Partridge, L. J., Monk, P. N., et al. (2005). Tetraspanins in viral infections: a fundamental role in viral biology? *J. Virol.* 79, 10839–10851. doi: 10.1128/JVI.79.17.10839-10851.2005
- Martin, S. A. M., Douglas, A., Houlihan, D. P., and Secombes, C. J. (2010). Starvation alters the liver transcriptome of the innate immune response in Atlantic salmon (*Salmo salar*). *BMC Genomics* 11:418. doi: 10.1186/1471-2164-11-418
- Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* 14, 193–202. doi: 10.1093/bib/bbs013
- MPA (2013a). *Boletim Estatístico da Pesca e Aquicultura—Brasil 2011*. Ministério da Pesca e Aquicultura.
- MPA (2013b). *Centro Aquícola Nacional—Ano 2008*. Ministério da Pesca e Aquicultura.
- Mutz, K. O., Heikenhinker, A., Lönne, M., Walter, J. G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* 24, 22–30. doi: 10.1016/j.copbio.2012.09.004
- Neto, R., Diaz, N. P., Gall, G. A., Gallardo, J. A., Llorente, J. P., and Manterola, R. (2006). Genetic improvement in Coho salmon (*Oncorhynchus kisutch*): I: selection response and inbreeding depression on harvest weight. *Aquaculture* 257, 9–17. doi: 10.1016/j.aquaculture.2006.03.002
- Odegaard, J., Baranski, M., Gjerde, B., and Gjerdem, T. (2011). Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquacult. Res.* 42, 103–114. doi: 10.1111/j.1365-2109.2010.02669.x
- Odegaard, J., Sommer, A. L., and Praebel, A. K. (2010). Heritability of resistance to viral nervous necrosis in Atlantic cod (*Gadus morhua*). *Aquaculture* 300, 59–64. doi: 10.1016/j.aquaculture.2010.01.006
- Panarari-Antunes, R. S., Prioli, A. J., Prioli, S. M. A. P., Galdino, A. S., Junior, J., Ferreira, H., et al. (2011). Genetic variability of brycon orthogyanus (Valenciennes, 1850) (*Characiformes: Characidae*) in cultivated and natural populations of the Upper Paraná River, and implications for the conservation of the species. *Braz. Arch. Biol. Technol.* 54, 839–848. doi: 10.1590/S1516-89132011000400025
- Pardo, B. G., Fernández, C., Millán, A., Bouza, C., Vázquez-López, A., Vera, M., et al. (2008). Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. *BMC Vet. Res.* 4:37. doi: 10.1186/1746-6148-4-37
- Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Piry, S., Luikart, G., and Cornuet, J. M. (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.* 90, 502–503. doi: 10.1093/hered/90.4.502
- Ponroni, B. W., Nguyen, N. H., Khaw, H. L., and Ninh, N. H. (2008). Accounting for genotype by environment interaction in economic appraisal of genetic improvement programs in common carp *Cyprinus carpio*. *Aquaculture* 285, 47–55. doi: 10.1016/j.aquaculture.2008.08.012
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Qian, X., Ba, Y., Zhuang, Q., and Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *OMICS* 18, 98–110. doi: 10.1089/omi.2013.0110
- Renaut, S., Nolte, A. W., and Bernatchez, L. (2010). Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Ecol. Res.* 10, 115–131. doi: 10.1111/j.1365-294X.2009.04477.x
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* 43, 223–225. doi: 10.1111/j.1558-5646.1989.tb04220.x
- Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for windows and linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x
- Rosen, S., and Skolitsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386. doi: 10.1385/1-59259-192-2:365
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234. doi: 10.1038/72708
- Seeb, J. E., Carvalho, G., Hauser, L., Nalsh, K., Roberts, S., and Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol. Ecol. Res.* 11, 1–8. doi: 10.1111/j.1755-0998.2010.02979.x

- Shin, S. C., Kim, S. J., Lee, J. K., Ahn, D. H., Kim, M. G., Lee, H., et al. (2012). Transcriptomics and comparative analysis of three Antarctic notothenioid fishes. *PLoS ONE* 7:e43762. doi: 10.1371/journal.pone.0043762
- Shirk, R. Y., Glenn, T. C., Chang, S., and Hamrick, J. L. (2013). Development and characterization of microsatellite Primers in *Geranium carolinianum* (Geraniaceae) with 454 sequencing. *Appl. Plant Sci.* 1:1300006. doi: 10.3732/app.1300006
- Shoshana, L., and Shoham, T. (2005). The tetraspanin web modulates immune – signaling complexes. *Nat. Rev. Immunol.* 5, 136–148. doi: 10.1038/nri1548
- Teicher, A., Kähkönen, K., and Merilä, J. (2012). Development of 63 new transcriptome-derived microsatellites for the Atlantic herring (*Clupea harengus*). *Conserv. Genet. Resour.* 4, 71–74. doi: 10.1007/s12686-011-9477-5
- Theodorou, K., and Couvet, D. (2015). The efficiency of close inbreeding to reduce genetic adaptation to captivity. *Heredity* 114:38. doi: 10.1038/hdy.2014.63
- Wang, X., Shang, X., Luan, J., and Zhang, S. (2014). Identification, expression and function of apolipoprotein v in animal fish *Notobranchia guentheri*: implication for an aging marker. *Biogerontology* 15, 233–243. doi: 10.1007/s10522-014-9493-4
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19, 395–420. doi: 10.1111/j.1558-5646.1965.tb01731.x
- Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., et al. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Res.* 20, 497–509. doi: 10.1093/dnares/det026
- Yáñez, J. M., Houston, R. D., and Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415. doi: 10.3389/fgene.2014.00415
- Yue, G. H., and Orban, L. (2002). Microsatellites from genes show polymorphism in two related *Oreochromis* species. *Mol. Ecol. Notes* 2, 99–100. doi: 10.1046/j.1471-8286.2002.00159.x
- Yue, G., Li, Y., and Orban, L. (2001). Characterization of microsatellites in the *HGF-2* and *GII* genes of Asian seabass (*Lateolabrax niloticus*). *Mol. Biotechnol.* 3, 1–5. doi: 10.1007/s101260000043

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Jorge, Mastroianni-Filho, Hake, Mendes, Arielle, Freitas, Vem, Porto-Foresti, and Hashimoto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Apêndice V

Genetic Applications in the Conservation of Neotropical Freshwater Fish

Vito Antonio Mastrochirico Filho, Milena V. Freitas,
Raquel B. Ariede, Lieschen V.G. Lira,
Natália J. Mendes and Diogo T. Hashimoto

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.73207>

Abstract

Neotropical fish correspond to approximately 30% of all fish species worldwide. The diversity of fish species found in Neotropical basins reflects variations in life-history strategies and exhibition of particular morphological, physiological and ecological attributes. These attributes are mainly related to different forms of feeding, life maintenance and reproduction. Today, fish populations are being threatened by anthropogenic actions that are having a visible impact on the natural state of continental aquatic ecosystems. The main causes are overfishing, non-native species introduction, reservoir-dam systems, mining, pollution and deforestation. The biology and population dynamics of the species are still unclear due to lack of research. Genetic tools can be useful resources for the conservation of Neotropical fish species in several ways. Molecular genetic markers are considered powerful tools to identify cryptic and hybrid fish and also allow the evaluation of the genetic variability and structure of populations of Neotropical ichthyofauna. Several analyses of molecular markers have been performed on Neotropical fish, including allozyme analysis, restriction fragment length polymorphisms in regions of DNA (RFLP), randomly amplified polymorphic DNA (AFLP), randomly amplified polymorphic DNA (RAPD), microsatellites, single nucleotide polymorphisms (SNPs) and mitochondrial DNA (mtDNA) markers. In order to analyse a high number of markers, next generation sequencing has allowed researchers to generate a large amount of genomic information that can be applied to the conservation of Neotropical fish.

Keywords: molecular markers, genetic conservation, Neotropical ichthyofauna, overfishing, dam