UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CENTRO DE AQUICULTURA DA UNESP

Análise de parentesco e variabilidade genética de pacu (*Piaractus mesopotamicus*) por meio de marcadores SNPs: subsídios para o melhoramento genético

Vito Antonio Mastrochirico Filho

Jaboticabal, São Paulo
2016

UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CENTRO DE AQUICULTURA DA UNESP

Análise de parentesco e variabilidade genética de pacu (*Piaractus mesopotamicus*) por meio de marcadores SNPs: subsídios para o melhoramento genético

Vito Antonio Mastrochirico Filho

Orientador: Prof. Dr. Diogo Teruo Hashimoto

Dissertação apresentada ao Programa de Pós-graduação em Aquicultura do Centro de Aquicultura da UNESP-CAUNESP, como parte dos requisitos para obtenção do título de Mestre.

Jaboticabal, São Paulo
2016

Mastrochirico-Filho, Vito Antonio
M423a Análise de parentesco e variabilidade genética de pacu (*Piaractus mesopotamicus*) por meio de marcadores SNPs: subsídios para o melhoramento genético / Vito Antonio Mastrochirico Filho. – – Jaboticabal, 2016
vi, 85 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Centro de Aquicultura, 2016
Orientador: Diogo Teruo Hashimoto
Banca examinadora: Fernanda de Alexandre Sebastião, Danielly Veloso Blanck
Bibliografia

1. RNA-seq. 2. Polimorfismo de base única. 3. Aquicultura. 4. Diversidade genética. 5. Relação de parentesco. I. Título. II. Jaboticabal-Centro de Aquicultura.

CDU 639.3.03

Para Walkyria, minha mãe

e em memória de

Davina e Priscila, meus anjos

Pouco conhecimento faz com que as pessoas se sintam orgulhosas.

Muito conhecimento, que se sintam humildes. É assim que as espigas

sem grãos erguem desdenhosamente a cabeça para o Céu, enquanto

que as cheias as baixam para a terra, sua mãe.

Leonardo da Vinci

AGRADECIMENTOS

Agradeço primeiramente a Deus, Jesus, e Nossa Senhora Aparecida por me conceder a oportunidade, saúde, paz e otimismo para a realização deste trabalho.

Ao Prof. Dr. Diogo Teruo Hashimoto, por toda a orientação, dedicação, apoio, paciência e confiança para realização deste trabalho. Além de ser um exemplo a ser seguido como ser humano e como profissional.

Aos meus amigos e companheiros de LaGeAC: Luquita, Paulo Jorges, Paolita, Milene, Bruna, Natália, Raquel e Milena , por toda a ajuda, incentivo, pelas risadas e momentos que transmitiram garra e vontade de continuar sempre em frente.

À minha família, à minha amada mãe Walkyria que com seu amor, carinho, e compreensão foi indispensável para a realização deste trabalho. À Nina, Willy e Bó por tantos momentos de alegria, e amor incondicional.

Ao meu anjo da guarda, ou quem me acompanha, me guia e me protege!!

Muito Obrigado!!

Apoio Financeiro

A FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo, pelo auxílio da bolsa concebida (Proc. FAPESP 2014/12412-4, 01/11/2014 a 29/02/2016) e auxílio à pesquisa (Proc. FAPESP 2014/03772-7).

Ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo auxílio concebido (Proc. CNPq 130262/2014-5), como bolsa inicial, durante a vigência de março a outubro de 2014; e ao auxílio à pesquisa (Proc. CNPq 446779/2014-8).

Sumário

Resumo Geral

Pacu (*Piaractus mesopotamicus*) é uma espécie de peixe Neotropical amplamente distribuída nas bacias dos rios Paraná e Paraguai, e uma das espécies de peixe neotropicais de maior valor para a aquicultura. Uma melhor compreensão do genoma do pacu é necessária para o manejo genético na conservação dos estoques naturais e cultivados. O principal objetivo foi identificar SNPs (*Single Nucleotide Polymorphisms*) gene-associados no transcriptoma de fígado do pacu e, em seguida, aplicar em análises de variabilidade genética e de parentesco visando um manejo adequado desta importante espécie não modelo na aquicultura. O sequenciamento do transcriptoma foi realizado por meio da plataforma Roche/454, que resultou na formação de 4.110 *contigs* não redundantes. Destes, 2.051 genes foram identificados e funcionalmente anotados a fim de revelar genes relacionados às características econômicas interessantes para a aquicultura. Foram encontrados 464 SNPs localizados em 5'UTR (10.0%), 3'UTR (17.2%) e em regiões CDS (71,1%), e classificados como sinônimos (70,6%) e não sinônimos (29,4%). Foram genotipados 32 SNPs por meio da técnica Sequenom MassARRAY, dos quais alguns estavam relacionados com sistema imune. A variabilidade genética foi estimada em populações de indivíduos selvagens (Rio Paraná) e de indivíduos cultivados em sete pisciculturas do estado de São Paulo (FF1, FF2, FF3, FF4, FF5, FF6 e FF7). Não foram observadas diferenças significativas entre heterozigosidade observada (H_{obs}) e esperada (H_{exp}) para cada população. Análises de diferenciação genética mostraram baixo nível de estruturação genética entre as populações (F_{st} = 0.064, AMOVA = 93,59% da variação dentro de populações, P<0,05). Análises de parentesco mostraram que a maioria das estações de piscicultura possuíam pelo menos 40% de indivíduos aparentados, com risco de endogamia e necessidade de realização de um programa de acasalamentos direcionados. Nossos resultados proporcionaram importantes recursos genéticos para o pacu, com aplicabilidade para a aquicultura.

Palavras-chave: RNA-Seq, polimorfismo de base única, aquicultura, diversidade genética, relação de parentesco

Abstract

Pacu (*Piaractus mesopotamicus*) is a Neotropical freshwater fish widely distributed in Parana, Paraguay Basin. Wild populations of pacu are threatened by overfishing and it is one of the fish species of highest commercial value for aquaculture. An understanding of the pacu genome is appropriate to genetic management in the conservation of wild and cultivated stocks. The main objective was identify gene-associated SNPs in liver transcriptome of pacu. We used SNPs (*Single Nucleotide Polymorphisms*) to perform genetic variability and kinship analysis for suitable management of this important non-model species in aquaculture. Transcriptome sequencing was done with the Roche/454 technology and yielded 4,110 non-redundant contigs. Of these, 2,051 genes were identified and functionally annotated to reveal genes correlated to economical traits in aquaculture. We found 464 SNPs in 5'UTR (10.0%), 3'UTR (17.2%) and CDS (71,1%), classified in synonymous (70,6%) and non-synonymous (29,4%). We genotyped 32 feasible SNPs through Sequenom MassARRAY platform and we obtained some SNPs related to immune system. Genetic diversity was estimated in wild individuals (Parana river) and in seven farm fish populations (FF1, FF2, FF3, FF4, FF5, FF6 and FF7). There were no significant differences between observed heterozygosity (H_{obs}) and expected (H_{exp}) for each population; and also between observed heterozygosity, expected heterozygosity and minimum allele frequency (*MAF*), when the population averages were compared ($P < 0.05$). In addition, genetic differentiation analyzes showed low genetic structure of wild and cultivated populations of pacu (F_{st} = 0.064; AMOVA = 93.59% of the variation within populations; $P<0,05$). Kinship analysis showed most hatchery stations had at least 40% of related individuals, at risk of inbreeding and the need to perform a directed mating program. Our results showed unprecedented genomic resources for pacu.

Keywords: RNA-Seq, single base polymorphism, aquaculture, genetic diversity, parental analysis

Introdução Geral

A demanda global por proteínas derivadas de pescado tem aumentado constantemente ao longo das últimas décadas, e deverá continuar devido ao crescimento da população, urbanização e uma crescente preferência por alimentos saudáveis. No contexto da estagnação da pesca, a aquicultura terá que cumprir a maior parte do futuro aumento na demanda por pescado de maneira correta e sustentável. Esta importância da aquicultura ocorre em um momento em que o mundo se tornou mais consciente sobre questões ambientais, e consumidores passaram a exigir produtos mais seguros (Bacher, 2015).

De acordo com as últimas estatísticas disponíveis, em 2013, a produção mundial de organismos aquáticos atingiu recordes históricos (97,2 milhões de toneladas) com sistemas de cultivo de diferentes intensidades e tecnologias. A produção de peixes pela aquicultura apresentou rápido crescimento e já representa 43,1% da produção total de pescado. Entretanto, grande parte do crescimento da produção aquícola foi devido à expansão de políticas de incentivo de produção em países em desenvolvimento, como na América Latina, com um forte foco em espécies orientadas para a exportação (FAO, 2014; 2015).

No Brasil, a aquicultura passou a ser reconhecida como uma atividade importante para a produção de alimentos de alto valor nutricional somente a partir do ano 2000. Com isso, a atividade passou então a ser alvo de investimentos privados e ações do próprio Governo Federal, com o objetivo de produzir espécies de peixes em escala industrial. Como resultado, a aquicultura brasileira registrou um expressivo avanço nos últimos anos, com um incremento de sua produção de 44%, entre os anos de 2007 a 2009. No ano de 2009, o Brasil atingiu uma produção de 415 mil toneladas, o que representou 33% da produção de pescado do país. Em 2010, houve um incremento de mais 15,3% sobre o ano anterior, com um somatório anual acima de 479 mil toneladas (MPA, 2008). De acordo com o censo de produção agropecuária realizado no ano de 2013 (IBGE, 2014), a produção aquícola brasileira

representou cerca de 488 mil toneladas, sendo 80% correspondentes à produção de peixes.

Entretanto, com a exploração pesqueira em declínio e com a aquicultura ainda insuficiente para atender a população mundial, a atual produção de pescado não atende a demanda crescente de produtos proteicos de origem animal. Portanto, pesquisas devem ser realizadas com o objetivo de aumentar a produção mundial de peixes de maneira sustentável, ou seja, promover o desenvolvimento da aquicultura, com medidas menos impactantes e com o uso cada vez menor de recursos.

Programas de melhoramento genético são cruciais para o desenvolvimento sustentável da aquicultura, pois além de possibilitarem um incremento do desempenho produtivo das espécies cultivadas, podem reduzir os custos de produção e garantir melhor qualidade dos produtos com a utilização mais eficiente de alimento, água, e áreas de cultivo disponíveis (Gjedren and Baranski, 2009). Contudo, menos que 10% da produção aquícola é baseada em estoques geneticamente melhorados, apesar do fato de que os ganhos genéticos anuais registrados são substancialmente maiores do que o desempenho produtivo de organismos normalmente cultivados (Gjedren *et al*., 2012).

Características economicamente importantes como crescimento (Ayllon *et al*., 2015; Tsai *et al.,* 2015), atributos de carcaça (Neira *et al.,* 2016), adaptação ambiental (Guan *et al.,* 2016; Sae-Lim e Bijima, 2016), conversão alimentar (Martens *et al*., 2014) e aumento da eficiência imunológica dos organismos a diversas doenças (Gonen *et al.,* 2015; Correa *et al*., 2015; Evenhuis *et al*., 2015) são o principal foco de pesquisas científicas relacionadas ao aumento da produtividade na aquicultura por meio de programas de melhoramento genético (Li e Ponzoni, 2015). Olesen *et al*. (2000) argumentaram que o desenvolvimento de uma produção sustentável deve levar em consideração, além de melhorias de características zootécnicas (como o crescimento e qualidade de carcaça), preocupações ambientais, sociais e éticas, como a capacidade de adaptação dos organismos (bem estar animal), e resistência a particulares infecções e parasitas sem a necessidade de aplicação de antibióticos no ambiente.

Além da seleção assistida por marcadores gene-associados à características importantes para o desenvolvimento da produtividade dos organismos cultivados, técnicas de seleção e acasalamento em programas de melhoramento genético são consideradas um meio de garantir que populações cultivadas permaneçam viáveis e produtivas (Lind *et al*., 2012). Vários fatores são importantes para que os programas de melhoramento por seleção conduzam a ganhos genéticos expressivos e duradouros. Dentre eles, é fundamental realizar um adequado programa de pré-melhoramento por meio da avaliação da variabilidade genética de populações-base, que evitará problemas decorrentes do estreitamento da base genética de certas espécies ou estoques (Ponzoni, 2006). O estreitamento da base genética ocorre devido às alterações nas frequências alélicas, com consequente redução da variabilidade genética no processo de seleção de organismos geneticamente superiores, o que pode resultar no aumento da probabilidade de acasalamento de indivíduos aparentados e surgimento de genes deletérios, com consequente redução do potencial de linhagem da produção piscícola, redução nas taxas de crescimento e sobrevivência, além do aparecimento de deformidades morfológicas (Kincaid, 1983).

Marcadores moleculares de DNA são considerados ferramentas úteis na detecção de polimorfismos genéticos entre indivíduos, populações e espécies, e têm revolucionado o poder analítico de estudos relacionados à diversidade genética de populações naturais e cultivadas (Avise, 1994; Ferguson *et al*., 1995; Liu e Cordes, 2004). Entretanto, para que as informações de marcadores moleculares sejam úteis aos programas de melhoramento genético, é essencial a identificação de marcadores associados a características economicamente importantes na aquicultura (Yue, 2014).

Os marcadores SNPs (*Single Nuceotide Polymorphisms*) são polimorfismos causados por mutações pontuais, e têm sido amplamente utilizados na aquicultura por serem considerados os marcadores mais promissores da atualidade (Liu and Cordes, 2004). Os polimorfismos são caracterizados por diferenças nas sequências nucleotídicas que são originadas pela substituição de uma única base nitrogenada, e resultam em diferentes alelos pertencentes a um lócus específico (Pierce, 2009). SNPs estão entre os

tipos de variação genética mais abundante e amplamente distribuídos pelo genoma, e constitui o polimorfismo mais adaptável à automação de genotipagem, além de revelarem polimorfismos ocultos não detectados em outros marcadores (Liu, 2011).

Uma vantagem dos SNPs, da mesma forma que os marcadores microssatélites, é o fato de apresentarem a herança do tipo codominância (Liu e Cordes, 2004), que permite caracterizar uma população quanto às frequências alélicas e genotípicas. Entretanto, o nível de polimorfismo de SNPs (*loci* normalmente bi-alélicos) não é tão elevado como nos marcadores microssatélites (*loci* multi-alélicos), mas esta desvantagem é equilibrada pela sua ampla cobertura no genoma (Vignal *et al.,* 2002). A alta adaptação de SNPs em análises do genoma em grande escala, comparados aos marcadores microssatélites que possuem nível de automação limitado e processos de genotipagem lentos e trabalhosos, faz com que a escolha de marcadores SNPs seja uma opção mais viável em estudos genômicos direcionados para a aquicultura (Liu, 2011).

Estas características demonstram que marcadores SNPs são ideais para diversos estudos biológicos, pois possibilitam análises genômicas complexas, com alto rendimento e elevada cobertura, o que tem causado uma revolução em estudos de variabilidade genética (Hauser *et al.,* 2011; Vera *et al.,* 2013; Zhang *et al.,* 2015; Liu *et al.,* 2016) e de seleção assistida por marcadores SNPs associados a características de interesse para a aquicultura, como crescimento (Gutierrez *et al.,* 2015; Tsai *et al.,* 2015), adaptação e domesticação (Sun *et al.,* 2014) e resistência a doenças (Correa *et al*., 2015; Geng *et al.,* 2015; Shao *et al*., 2015).

Para o uso rotineiro de SNPs, é fundamental plataformas de genotipagem que viabilizem analisar um elevado número de marcadores e amostras, de forma rápida e econômica. Em síntese, cada plataforma utiliza uma química de detecção específica, o que geram diferenças no custo de genotipagem, preço de equipamento, número de marcadores, expertise para o uso, volume de amostras, tempo de análise e automação. Para a genotipagem dos SNPs em baixo rendimento, *loci* candidatos foram testados em organismos

cultivados utilizando diferentes metodologias como *Snapshot* em mexilhões (Nie *et al*., 2015) e *Taqman* em truta arco-íris (Hansen *et al*., 2011).

A utilização de técnicas de genotipagem de alto a médio rendimento ainda está em início para peixes. Muitas tecnologias de genotipagem de alto rendimento estão disponíveis no momento. O método *Sequenom MassARRAY* oferece várias características atrativas para os utilizadores que desejam um método eficaz e econômico de genotipagem de SNP, e vêm sendo cada vez mais utilizadas em peixes (Willians *et al*., 2010; Salem *et al*., 2012). O ensaio é baseado primeiramente em uma reação de PCR para o *locus* de interesse, seguido por uma reação de extensão em que um *primer* é anelado adjacente ao polimorfismo existente para que, por meio de espectrometria de massas, a massa do *primer* extendido seja determinada (Gabriel e Ziaugra, 2004).

Historicamente, numerosas abordagens para a descoberta de SNPs foram descritas, principalmente a partir da comparação das sequências de *loci* específicos. A realização de sequenciamento direto (Sanger) de genes candidatos era considerada a estratégia mais simples, apesar de dispendiosa, para busca de SNPs. Em uma escala maior, a melhor alternativa para a caracterização de SNPs se fundamentava na comparação de sequências de fragmentos clonados, particularmente de projetos de *EST* (*Expressed Sequence Tags*) utilizando diferentes tipos de tecidos (Vignal *et al.*, 2002). Entretanto, além dos custos elevados, era necessário muito esforço laboratorial, tempo e expertise para este tipo de análise.

Para suprir esta demanda, as tecnologias de sequenciamento de nova geração (NGS) estão permitindo uma mudança de paradigma no cenário da Genética e Biologia Molecular, pois possibilitam à descoberta de milhares de SNPs por meio de um maior rendimento das leituras de sequências (Nielsen *et al.,* 2011). Neste caso, sequenciadores de nova geração como as plataformas Roche/454 e Illumina/HiSeq são particularmente adaptadas para produzir uma elevada cobertura de sequências com precisão (Mardis, 2008). Com a utilização de tecnologias NGS, a ausência de um genoma de referência é uma das maiores dificuldades para descobrir SNPs em organismos não modelos. A geração primária de sequências genéticas de um determinado indivíduo é chamada de sequenciamento *de novo* (Meng e Yu, 2011). Nestes casos, a

partir de um projeto de sequenciamento, as leituras individuais (reads) podem ser montadas em sequências de consenso denominadas *contigs*, que poderão servir como um genoma de pseudorreferência (Waldbieser, 2011). Além disso, para descoberta de SNPs em organismos sem genoma de referência, uma etapa de redução de genoma para adquirir os conjuntos de *contigs* redundantes é outra estratégia que deve ser realizada (Ekblom e Galindo, 2011).

O sequenciamento do transcriptoma por meio de NGS (*RNA-seq*), especificamente o RNA mensageiro (RNAm), é uma das abordagens mais comuns de redução do genoma de organismos não modelos (Seeb *et al*., 2011). Neste caso, todo o RNAm de um tecido específico ou um conjunto de tecidos é utilizado como fonte para o sequenciamento, ou seja, é possível obter toda a informação de transcritos funcionais do genoma de uma espécie alvo, de forma rápida e com baixo custo. Uma das principais vantagens é que os SNPs derivados do transcriptoma estão associados a genes com informação funcional, evitando íntrons e regiões extragênicas que podem dificultar a análise de dados (Wang e Liu, 2011).

O *RNA-seq* ainda permite o uso de marcadores para mapeamento genético e subsequente uso em estudos comparativos (Sarropoulou *et al*., 2008), que é importante para espécies da aquicultura sem o genoma completo, pois permite estudar funcionalmente os SNPs gene-associados, uma vez que estes podem apresentar diferentes implicações dependendo da exata localização genômica. Além disso, também é possível realizar a anotação funcional dos *contigs* para verificar quais as classes funcionais e vias metabólicas em que estão inseridos, resultando em melhores inferências sobre os efeitos que SNPs podem causar (Vera *et al*., 2013).

A análise de transcriptoma tem demonstrado ser eficaz para identificação de SNPs em peixes, principalmente em espécies não modelo (Montes *et al*., 2013; Vidotto *et al*., 2013; Wang *et al*., 2014; Wang *et al*., 2015; Zhang *et al*., 2015). Porém, não há na literatura qualquer dado de transcriptoma com identificação de SNPs para espécies de peixes nativos economicamente importantes para a aquicultura Neotropical. Com os avanços de tecnologias de NGS, somado com as vantagens do sequenciamento de

transcriptoma, é fundamental um projeto que vise integrar estas ferramentas moleculares com a aplicação de marcadores SNPs em espécies com potencial de produção na aquicultura brasileira.

Pacu (*Piaractus mesopotamicus*) apresenta uma ampla distribuição nas bacias dos rios Paraná, Paraguai e Prata, sendo que sua maior distribuição ocorre nas planícies alagadas do Pantanal (Resende, 2003). É uma das mais importantes espécies Neotropicais cultivadas no Brasil (IBGE, 2014) e até mesmo em outras partes do mundo (Honglang, 2007; FAO, 2010). Porém, os estoques das populações selvagens de pacu têm diminuído nas últimas décadas devido principalmente à sobrepesca e às grandes alterações em seu habitat (Resende, 2003).

Estudos genéticos direcionados à espécie são praticamente ausentes, sendo limitados a estudos de variabilidade genética por meio de sequências mitocondriais (Iervolino *et al*., 2010) e por marcadores microssatélites (Calcagnotto *et al*., 2001; Calcagnotto e DeSalle, 2009) em ambiente selvagem. Adicionalmente, existe somente um registro de estudo genético relacionado ao transcriptoma de pacu, que buscou um melhor entendimento do papel do metabolismo muscular em características de crescimento, e que possui aplicabilidade no desenvolvimento de subsídios para programas de melhoramento genético (Mareco *et al*., 2015). Portanto, devido ao declínio das populações naturais e importância econômica na aquicultura, um melhor entendimento do genoma do pacu é necessário para o desenvolvimento de um manejo genético apropriado tanto para conservação dos estoques selvagens, quanto para aumento de produtividade na aquicultura.

O objetivo principal do estudo foi analisar o grau de parentesco e variabilidade genética de estoques de pacu (*Piaractus mesopotamicus*) utilizando SNPs obtidos por sequenciamento de transcriptoma, por meio de NGS. Além disso, a prospecção de SNPs associados a genes com características úteis para melhorar o desempenho da produção do pacu foi uma análise inicial que servirá como base para um futuro estudo de desenvolvimento de *QTL (Quantitative Trait Loci*), que poderão ser úteis em programas de seleção assistida por marcadores. Nossos dados de transcriptoma de fígado representam um estudo pioneiro de prospecção de

SNPs em espécies de peixes Neotropicais e, portanto, podem servir também como base para o desenvolvimento de programas de melhoramento genético para outras espécies de interesse econômico, mas que ainda não possuem um genoma de referência.

Referências Bibliográficas

Avise, J.C. *Molecular Markers, Natural History and Evolution*; Chapman and Hall: New York, USA, 1994.

Ayllon, F.; Kjaerner-Semb, E.; Furmanek, T.; Wennevik, V.; Solberg, M.F.; et al. The vgll3 locus controls age at maturity in wild and domesticated Atlantic Salmon (*Salmo salar* L.) males. *Plos Genetics* 2015, 11(11), e1005628.

Bacher, K. *Perceptions and Misconceptions of Aquaculture: A Global Overview*. *Globefish Research Programme*, vol.120; FAO: Rome, Italy, 2015; 35 pp.

Calcagnotto, D.; DeSalle, R. Population genetic structuring in pacu (*Piaractus mesopotamicus*) across the Paraná-Paraguay basin: evidence from microsatellites. *Neotropical Ichthyology* 2009, 7(4), 607-616.

Calcagnotto, D.; Russello, M.; DeSalle, R. Isolation and characterization of microsatellite loci in *Piaractus mesopotamicus* and their applicability in other Serrasalminae fish. *Molecular Ecology* Notes 2001, 1, 245-247.

Correa, K.; Lhorente, J.P.; Lopez, M.E; Bassini, L.; Naswa, S.; et al. Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* 2015, 16, 854.

Ekblom, R.; Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 2011, 107(1), 1-15.

Evenhuis, J.P.; Leeds, T.D.; Marancik, D.P.; LaPatra, S.E.; Wiens, G.D. Rainbow trout (*Oncorhynchus mykiss*) resistance to columnaris disease is heritable and favorably correlated with bacterial cold water disease resistance. *Journal of Animal Science*. 2015, 93(4), 1546-1554.

FAO. *The State of World Fisheries and Aquaculture 2010*. FAO: Rome, Italy, 2010; 197 pp.

FAO. *The State of World Fisheries and Aquaculture 2014*. FAO: Rome, Italy, 2014; 223 pp.

FAO. *Global Aquaculture Production database updated to 2013 – Summary information*. FAO: Rome, Italy, 2015.

Ferguson, A.; Taggart, J.B.; Prodohl, P.A.; McMeel, O.; Thompson, C.; et al. The application of molecular markers to the study and conservation of fish populations with special reference to Salmo. *Journal of Fish Biology* 1995, 47(sA), 103-126.

Gabriel, S.; Ziaugra, L. SNP genotyping using Sequenom MassARRAY 7K Platform. *Current Protocols in Human Genetics* 2004, 42:2.12, 2.12.1–2.12.16.

Geng, X.; Sha, J.; Liu, S.K.; Bao, L.S.; Zhang, J.R.; et al. A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* 2015, 16, 196.

Gonen, S.; Baranski, M.; Thorland, I.; Norris, A.; Grove, H.; et al. Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar)*. *Heredity* 2015, 115(5), 405-414.

Gutierrez, A.P.; Yanez, J. M.; Fukui, S.; et al. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic Salmon (Salmo salar). *Plos One* 2015, 10(3), UNSP e0119730.

Gjedren, T.; Baranski, M. *Selective Breeding in Aquaculture: An Introduction*. Springer Science & Business Media: 2009. 221 pp.

Gjedren, T.; Robinson, N.; Rye, M. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 2012, 350-353, 117-129

Guan, J.T.; Hu, Y.L.; Wang, M.S.; Wang, W.J.; Kong, J.; et al. Estimating genetic parameters and genotype-by-environment interactions in body traits of turbot in two different rearing environments. *Aquaculture* 2016, 450, 321-327.

Hauser, L.; Baird, M.; Hilborn, R.; Seeb, L.W.; Seeb, J. E. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources* 2011, 11(1), 150-161.

Hansen, M.H.H; Young, S.; Jorgensen, H.B.H.; Pascal, C.; Henryon, M, et al Assembling a dual purpose TaqMan-based of single-nucleotide polymorphism markers in rainbow trout and steelhead (*Oncorhynchus mykiss*) for association mapping and population genetics analysis. *Molecular Ecology Resources* 2011, 11 (1), 67-70.

Honglang, H. Freshwater fish seed resources in China. In: *Assessment of freshwater fish seed resources for sustainable aquaculture*, FAO Fisheries Technical Paper No 501 Bondad-Reantaso, M.G., Eds.; FAO: Rome, Italy, 2007, 628 pp.

Iervolino, F.; Resende, E. K.; Hilsdorf, A.W.S. The lack of genetic differentiation of pacu (*Piaractus mesopotamicus*) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: Implications for fishery management. *Fisheries Research* 2010, 101, 27-31.

IBGE. Instituto Brasileiro de Geografia e Estatística. *Produção da Pecuária Municipal 2013, Vol 41*. Rio de Janeiro, Brasil, 2014; 1-108.

Kincaid, H.L. Inbreeding in fish populations used for aquaculture. *Aquaculture* 1983, 33, 215-227.

Liu, Z. J.; Cordes, J. F. DNA marker Technologies and their applications in aquaculture genetics. *Aquaculture* 2004, 238, 1-37.

Liu, Z. *Next Generation Sequencing and Whole Genome Selection in Aquaculture*; Wiley-Blackwell: Iowa, USA, 2011; 221 pp.

Lind, C.E.; Ponzoni, R.W.; Nguyen, N.H.; Khaw, H.L. Selective breeding in fish and conservation of genetic resources for aquaculture. *Reproduction in Domestic Animals* 2012, 47(4), 255-263.

Li, Y.; Ponzoni, R.W. Some aspects of design and analysis of selection programmes in aquaculture species. *Journal of Animal Breeding and Genetics* 2015, 132, 169-175.

Liu, S.; Palti, Y.; Gao, G.; Rexroad III, C.E. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture* 2016, 452, 178-182.

Mardis, E.R. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 2008, 9, 387-402.

Mareco, E. A.; Serrana, D.G.; Johnston, I.A.; Dal-Pai-Silva, M. Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). *BMC Genomics* 2015, 16, 182.

Martens, M.T.; Wall, A.J.; Pyle, G.G.; Wasylenko, B.A.; Dew, W.A. Growth and feeding efficiency of wild and aquaculture genotypes of rainbow trout (*Oncorhynchus mykiss*) common to Lake Huron, Canada. *Journal of Great Lakes Research* 2014, 40(2), 377-384.

Meng, Q.; Yu, J. Next Generation DNA Sequencing Technologies and Applications. In: *Next Generation Sequencing and Whole Genome Selection in Aquaculture*; Liu, Z, Ed.; Wiley-Blackwell: Oxford, UK, 2011; 221 pp.

Montes, I.; Conklin, D.; Albaina, A.; Creer, S.; Carvalho, G.R.; et al. SNP Discovery in European Anchovy (*Engraulis encrasicolus*, L) by high-throughput transcriptome and genome sequencing. *Plos One* 2013, 8(8), e70051.

MPA – Ministério da Pesca e Aquicultura. *Censo Aquícola Nacional: Ano 2008*. Brasília, Brasil, 2013; 336pp.

Nielsen, R.; Paul, J.S.; Albrechtsen, A.; Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* 2011, 12(6), 443-451.

Nie, Q.; Yue, X.; Liu, B. Development of Vibrio spp. infection resistance related SNP markers using multiplex SNaPshot genotyping method in the clam *Meretrix meretrix*. *Fish & Shellfish Immunology* 2015, 43(2), 469-476.

Neira, R.; Garcia, X.; Lhorente, J.P.; Filp, M.; Yanez, J.M. et al. Evaluation of the growth and carcass quality of diallel crosses of four strains of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 2016, 451, 213-222.

Olesen, I.; Groen, A.F.; Gjerde, B. Definition of animal breeding goals for sustainable production systems. *Journal of Animal Science* 2000, 78, 570-582.

Pierce, B.A. *Genetics: a conceptual approach*; W.H. Freeman and Co: New York and Basingstoke, 2009; 774 pp.

Ponzoni, R.W. Genetic improvement effective dissemination: Keys to prosperous and sustainable aquaculture industries. In: *Development of Aquatic Animal Genetic Improvement and Dissemination Programs: current status and action plans*; Ponzoni, R.W.; Acosta, B.O.; Ponniah, A.G., Eds.; WorldFish Center: Penang, Malásia, 2006; 114 pp.

Resende, E.K. Migratory fishes of the Paraguay-Paraná basin excluding the Upper Paraná River. In: *Migratory fishes of South America: biology, fisheries and conservation states*; Carolsfeld, J.; Harvey, B.; Ross, C.; Baers, A., Eds.; World Bank: Victoria, Canada, 2003; pp.99-156.

Sae-Lim, P.; Bijima, P. Comparison of designs for estimating genetic parameters and obtaining response to selection for social interaction traits in aquaculture. *Aquaculture* 2016, 451, 330-339.

Salem, M.; Vallejo, R.L.; Leeds, T.D.; Palti, Y.; Liu, S.; et al. RNA-Seq identifies SNP markers for growth traits in rainbow trout. *Plos One* 2012, 7(5), e36264.

Sarropoulou, E.; Nousdili, D.; Magoulas, A.; Kotoulas, G. Linking the genomes of nonmodel teleosts through comparative genomics. *Marine Biotechnology* 2008, 10(3), 227-233.

Seeb, J.E.; Carvalho, G.; Hauser, L.; Naish, K.; Roberts, L.W.; et al. Single-nucleotide polymorphism (SNP) Discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*. 2011, 11(1), 1-8.

Sun, L.Y.; Liu, S.K.; Wang, R.J.; Jiang, Y.L.; Zhang, Y.; et al. Identification and analysis of genome-wide SNPs provide insight into signatres of selection and domestication in channel catfish (*Ictalurus punctatus*). *Plos one* 2014, 9(10), e109666.

Shao, C.W.; Niu, Y.C.; Rastas, P.; Liu, Y.; Xie, Z.Y.; et al. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. *DNA Research* 2015, 22(2), 161-170.

Tsai, H.; Hamilton, A.; Tinch, A.E.; Guy, D.R.; Gharbi, K.; et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 2015, 16, 969.

Vidotto, M.; Grapputo, A.; Boscari, E.; Barbisan, F.; Coppe, A.; et al. Transcriptome sequencing and de novo annotation of the critically endangered Adriatic sturgeon. *BMC Genomics* 2013, 14, 407.

Vera, M.; Alvarez-Dios, J.; Fernandez, C.; Bouza, C.; Vilas, R.; et al. Development and validation of Single Nucleotide Polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *International Journal of Molecular Sciences* 2013, 14, 5694-5711.

Vignal, A.; Milan, D.; SanCristobal, M.; Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 2002, 34(3), 275-306.

Waldbieser, G.C. SNP discovery through de novo deep sequencing using the next generation of DNA sequencers. In: *Next generation sequencing and whole genome selection in aquaculture*; Liu, Z. Ed.; Wiley-Blackwell: Oxford, UK, 2011; 221 pp.

Wang, S.; Liu, Z. SNP Discovery through EST Data Mining. In: *Next generation sequencing and whole genome selection in aquaculture*; Liu, Z. Ed.; Wiley-Blackwell: Oxford, UK, 2011; 221 pp.

Wang, P.; Xiao, S.; Han, Z.; Wang, Z. SNP discovery in large yellow croaker (Larimichthys crocea) using Roche 454 pyrosequencing sequencing platform. *Conservation Genetic Resources* 2015, 7(4), 777-779.

Wang, W.; Yi, Q.; Ma, L.; Zhou, X.; Zhao, H.; et al. Sequencing and characterization of the transcriptome of half-smooth tongue sole (*Cynoglossus semilaevis*). *BMC Genomics* 2014, 15, 470.

Willians, L.M.; Ma, X.; Boyko, A.R.; Bustamante, C.D.; Oleksiak, M.F. SNP identification, verification, and utility for population genetics in a non-model genus. *BMC Genetics* 2010, 11, 32.

Yue, G.H. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries* 2014, 15, 376-396.

Zhang, H.W.; Yin, S.W.; Zhang, L.J.; Hou, X.Y.; Wang, Y.Y. Development and validation of single nucleotide polymorphism markers in *Odontobutis potamophila* from transcriptomic sequencing. *Genetics and Molecular Research* 2015, 14(1), 2080-2085.

Capítulo 1

Liver Transcriptome and SNP Discovery in the Fish *Piaractus mesopotamicus*

Resumo

Pacu (*Piaractus mesopotamicus*) é uma ameaçada espécie de peixe Neotropical, e uma das espécies de peixe de maior valor comercial para a aquicultura. Um melhor entendimento do genoma do pacu é apropriado para um manejo genético aplicado em programas de conservação e em estoques cultivados. Análises de variabilidade genética através de marcadores moleculares é uma técnica essencial para o melhor manejo destes estoques. O principal objetivo foi identificar e validar SNPs gene-associados no transcriptoma de fígado do pacu, para um adequado manejo desta importante espécie não modelo na aquicultura. O sequenciamento do transcriptoma foi realizado por meio da plataforma Roche/454, que resultou na formação de 4.110 *contigs* não redundantes. Destes, 2.051 genes foram identificados e funcionalmente anotados em categorias GO (*Gene Ontology*), *Interpro, enzyme codes* e *KEGG*, que revelaram genes relacionados à características econômicas interessantes para a aquicultura. Nós encontramos 464 SNPs classificados em regiões 5'UTR (10.0%), 3'UTR (17.2%) e em regiões CDS (71,1%), e identificados como sinônimos (70,6%) e não sinônimos (29,4%). Foram genotipados 32 SNPs através da técnica *Sequenom MassARRAY* e obtivemos alguns SNPs relacionados ao sistema imunológico. Análise de variabilidade genética foi estimada em uma população originada do rio Paraná que mostrou que os valores de heterozigosidade observada e esperada variaram entre 0.059 e 0.706 e 0.058 e 0.507, respectivamente, e não apresentaram diferença significativa ($P<0,05$), com todos os loci em equilíbrio de Hardy-Weinberg ($P>0.05$). Nossos resultados forneceram importantes e pioneiros recursos genéticos para o pacu, através da identificação e utilização de marcadores SNPs para verificar a estrutura de populações e aplicabilidade em futuros programas de melhoramento genético para espécies nativas Neotropicais.

Palavras-chave: RNA-Seq, NGS, montagem *de novo*, genotipagem de SNPs, diversidade genética, Neotropical

Abstract

Pacu (*Piaractus mesopotamicus*) is a threatened Neotropical freshwater fish, and one of the fish species of highest commercial value for aquaculture. An understanding of pacu genome is appropriate to genetic management applied in conservation programs and cultivated stocks. Genetic variability analysis by molecular markers is an essential technique to genetic management of these stocks. The main objective was to identify and validate gene-associated SNPs in liver transcriptome of pacu, for suitable management of this important non-model species in aquaculture. Transcriptome sequencing was done with the Roche/454 technology and yielded 4,110 non-redundant contigs. Of these, 2,051 genes were identified and functionally annotated in *GO (Gene Ontology)* terms, *Interpro*, *enzyme codes* and *KEGG* analysis (Kyoto Encyclopedia of Genes and Genomes), which revealed genes correlated to economical traits in aquaculture. There were found 464 SNPs classified into 5'UTR (10.0%), 3'UTR (17.2%) and CDS (71,1%) regions, and identified as synonymous (70,6%) and non-synonymous (29,4%). 32 SNPs were genotyped and validated by *Sequenom MassARRAY* technique. Some of SNPs were related to immune system. Genetic diversity analysis was estimated in Parana river population which showed observed and expected heterozygosity ranged from 0.059 to 0.706 and 0.058 to 0.507, respectively, with no significant difference ($P<0,05$) and all loci in Hardy-Weinberg equilibrium ($P>0.05$). Our results showed unprecedented and pioneer genomic resources for pacu, through the identification and applicability of SNPs to verify population structure with use in conservation purposes and aquaculture industry.

Keywords: RNA-Seq, NGS, *de novo* assembly, SNP genotyping, genetic diversity, Neotropical

1. Introduction

Pacu (*Piaractus mesopotamicus*) is a freshwater fish of South America waters where it is widely distributed in floodplain areas of the La Plata Basin. Wild populations of pacu are threatened by overfishing (Resende, 2003), particularly in Brazil (in São Paulo State, according to the Decree nº 56.031, SSP, 2010), since this species is considered of high commercial value, with large-scale catches by the industrial and recreational fisheries (MPA, 2013). Furthermore, this fish is important for aquaculture and represents one of the most cultivated species in Brazil, with annual production estimated at about 13.600 tons (IBGE, 2013), and in aquaculture from other countries in South America (Colombia, Peru, Venezuela and Argentina) and in Asian countries (China, Myanmar, Thailand and Vietnan) (Flores Nava, 2007; Honglang, 2007; FAO, 2010).

However, there are still no breeding programs in order to develop a profitable aquaculture for pacu. Genetic studies directed to this species are insufficient, since few genetic resources are available and limited mostly in microsatellites and mitochondrial sequences (Calcagnotto *et al*., 2001; Calcagnotto and DeSalle, 2009; Iervolino *et al*., 2010). Therefore, due to environmental concerns and economic importance to aquaculture production, a better understanding of pacu genome is necessary to enable appropriate genetic management for conservation of wild stocks and to increase productivity in aquaculture.

The advance in genetic studies had reached a plateau until the development of the next generation sequencing (NGS) technologies and simultaneous maturation of bioinformatic tools, which has now made possible to investigate the functional complexity of transcriptome through capture and annotation of genes involved in many biological processes of economic aquaculture species such as carp (Ji *et al.,* 2012), japanese flounder (Huang *et al*., 2015), rainbow trout (Salem *et al*., 2015) and Atlantic salmon (Micallef *et al*., 2012).

Recent advances in massively parallel cDNA sequencing, or RNA-seq, provide a cost-effective way to obtain large amounts of transcriptome data from

many tissue types (Shin *et al*. 2012; Cui *et al.,* 2014; Gomes *et al*., 2014) through the production of high accurately *EST* (*Expressed Sequence Tags*) sequences (Mardis, 2008). Thus, whole mRNA sequences (cDNA library) from a specific tissue or set of tissues can be aligned to a reference genome (or reference transcripts), or assembled *de novo*, which is the most common approaches to reduce genome of non-model organisms (Zhao *et al*., 2011). In this case, all the *EST* sequences are used as a source for pseudo-reference genome which allows obtaining all the functional transcripts information of a target species (Wang *et* al., 2009; Lanes *et* al., 2013; Gallardo-Escárate *et al*., 2014). Model species, such as zebrafish *Danio rerio*, may present more than 26.000 genes (Howe *et al*., 2013). However, few genes and their metabolic pathways have been characterized for non-model species. In this sense, through NGS technologies, *RNA-seq* is considered a perfect strategy to generate a better understanding of non-models functional genome, avoiding introns and extragenic regions that can interfere data analysis (Parkinson and Blaxter, 2009).

The *RNA-seq* is also an effective tool for the discovery of molecular markers for genetic mapping and subsequent use in comparative studies (Teacher *et al.,* 2012; Xu *et al*., 2012), particularly SNPs (*Single Nucleotide Polymorphisms*) prospection in non-model species (Seeb *et al.,* 2011; Montes *et al*., 2013). SNPs are polymorphisms caused by point mutations and the major focus for the development of molecular markers, since they are the most abundant type of sequence polymorphism and suitable for high-throughput genotyping (Liu and Cordes, 2004). Their frequency in non-model species has been estimated at ~1 SNP in 200-500 bases for non-coding DNA and ~1 SNP in 500-1000 bases for coding DNA (Brumfield *et al*., 2003). *RNA-seq* allows gene-associated SNPs studies, once these can have different implications depending on the exact genomic location and effects on the metabolic pathways that are inserted (Liu *et al*., 2011). Identification of gene-associated SNPs in high-throughput sequencing transcriptome of a non-model species, can generate genetic resources for application in variated areas of aquaculture, such as biological studies related to metabolic pathways and immunity (Hubert *et al*., 2010; Núñez-Acuña and Gallardo-Escárate, 2013), as well as being as

useful tools to evaluate genetic variability, providing greater control over family representation and inbreeding (Fernández *et al*., 2014; Vandeputte and Haffray, 2014).

Currently, SNP markers are not available for pacu, one of the most important warm water species for the development of aquaculture in South America. Thus, the objective of this study is 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 SNP markers.

2. Material and Methods

2.1 Ethic Statement

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and it was approved by the Animal Use Ethics Committee (CEUA) nº 22.255/15. 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). Fish were euthanized by benzocaine anesthetic overdose to collect liver tissue for transcriptome sequencing. For SNP validation and genetic variability analysis, fin fragments were collected from each fish under benzocaine anesthesia and all efforts were made to minimize suffering.

2.2 Samples for Transcriptome sequencing

To perform the transcriptome sequencing, samples were collected from liver tissue of total of ten individuals (five adults and five juveniles) from three different Brazilian fish farms (National Research and Conservation Center of Freshwater Fish, CEPTA/ICMbio, Pirassununga, São Paulo State; Aquaculture Center of São Paulo State University, CAUNESP, Jaboticabal, São Paulo State; Projeto Peixe fish farm, Sales de Oliveira, São Paulo State) and in a wild

population (Sapucaí-Mirim river). We used individuals from different origins in order to achieve the highest genetic variability in the SNP discovery analysis.

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

2.3 cDNA Library Construction and Roche 454 Platform Sequencing

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). We used the High Sensitivity DNA LabChip Kit (Agilent Technologies) with 2100 Bioanalyzer for quality analysis of the cDNA library. The concentration of sample molecules/uL was obtained by QuantiFluor^{TM} –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 – Libl 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 – LibL LV, GS FLX+ Series.

The transcriptome sequencing was done with the Roche/454 technology (GS FLX Titanium Sequencing Kit XL +) in HELIXXA company (Campinas, SP, Brazil).

The image processing and filtration was carried out by standard pipeline to shotgun libraries. To this end, the GS Run Processor v2.6 and GS Run Reporter v2.6 packages (Roche) were used.

2.4 *De novo* Assembly of Expressed Short Reads

A filter on the initial quality of the 454 sequences in .sff format was performed by Roche Newbler program. Sequence analysis was performed using 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 less than 20. Terminal nucleotides (five nucleotides at each extremity 5' and 3'), ambiguous nucleotides, adapter sequences, reads less than 15 bp were discarded. For *de novo* assembly, contigs less than 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 ≥ 50% of their length with ≥ 80% of similarity were included in the alignment. The assembled transcripts were subjected to cd-hit-est program 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).

2.5 Functional Annotation Analysis

Functional annotation of the unique consensus sequences was performed by homology searches against the NCBI (*National Center for Biotechnology Information*) 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 KEGG (*Kyoto Encyclopedia of Genes and*

Genomes) metabolic pathways analysis through Bi-directional Best Hit method (BBH).

2.6 SNP Markers Identification and Classification

Assembled contigs were screened for putative SNPs using the software CLC Genomics Workbench (version 7.5.1; CLC bio; Aarhus; Denmark), under the following criteria; both central and average surrounding base quality score of ≥20. In order to identify quality SNPs; minimum coverage (read depth) of 20 reads; minimum variant count of four and minimum frequency of the allele less frequent of 20% were counted as quality putative SNP. In addition, no extra SNPs or indels within 15-bp flanking regions were considered; SNPs located in repetitive regions were also discarded. For practical application in SNPs genotyping assays, only bi-allelic SNPs were considered in this study. SNPs located in contigs with Blast hits to the zebrafish Refseq protein database (NCBI) were identified along the zebrafish chromosomes through NCBI MapViewer. The six possible reading frames of the consensus sequence of each functionally annotated contig containing SNP were compared against the NCBI protein database using BLASTx (e-value 1e-10) in order to find ORF (*Open Reading Frame*) regions. These approaches allowed us to locate SNPs in coding sequences (CDS) or untranslated regions (5'UTR and 3'UTR) through graphical sequence viewer TABLET and classify ORF regions on which SNPs are inserted. In addition, for SNPs in CDS regions, the resulting amino acid sequences of both variants were translated to determine whether SNP variants were synonymous or non-synonymous.

2.7 SNPs Genotyping and Validation

DNA was extracted from fin fragments of 34 wild individuals collected in the Paraná River, using the Wizard Genomic DNA Purification Kit (Promega), according the manufacturer's protocol. The DNA concentration was quantified using the Qubit dsDNA BR Assay kit (Life Technologies) and measured on the Qubit 2.0 Fluorometer (Invitrogen). Top fifty SNPs were initially used for

validation and genotyping analysis with the MassARRAY platform (Sequenom, San Diego, CA, USA), in CeGen (Genotyping National Center, Santiago de Compostela, Spain). The technique consists of an initial locus specific polymerase chain reaction (PCR) in multiplex, followed by single-base extension using mass-modified dideoxynucleotide terminators of an oligonucleotide primer that anneals immediately upstream of the polymorphic site (SNP) of interest (Vera *et al*., 2013).

Assays were performed for putative SNPs located in different non-redundant contigs and combined in 2 multiplex reactions (33 + 17 SNPs). SNP multiplexes were designed *in silico* by CeGen and tested on a panel of wild pacu samples. SNPs were classified on manual inspection as “failed assays” (in case that the majority of genotypes could not be scored and/or the samples did not cluster well according to genotype), and feasible SNPs (markers with proper and reliable genotypes), these being either monomorphic or polymorphic (Vera *et al*., 2013).

2.8 SNP Diversity and Population Analysis

Genetic diversity parameters were estimated in 34 wild individuals from Paraná River, in relation to feasible SNPs. Observed (H_{obs}) and expected heterozygosity (H_{exp}) were calculated using Cervus 3.0.7 (Marshall *et al*., 1998). Inbreeding coefficient (F_{is}), minimum allele frequency (*MAF*) parameters and conformance to Hardy-Weinberg equilibrium (HW) and genotypic disequilibrium (*LD*) were performed using Genepop 4.0.11 (Smith *et al*., 2005; Rousset, 2008). Significant differences between H_{obs} and H_{exp} means were obtained by Student’s t-test. F_{is} parameters were estimated using Weir & Cockerham approach (Weir and Cockerhsam, 1984). Conformance to Hardy-Weinberg equilibrium (HWE) was checked using the complete enumeration method (Louis and Dempster, 1987) because only two alleles were identified at each locus. Bonferroni correction was performed when multiple tests were realized (Rice, 1989).

3. Results

3.1 Transcriptome Sequencing and *De novo* Assembly

The results of pacu transcriptome sequencing yielded a total of 212,813 reads which were then deposited in Short Read Archive (SRA) of NCBI under the accession number SRA312243. The raw reads presented an average length of 402 bp (base pairs), comprising a total of ~86 Mbp. After the trimming process, the average length of the reads was of 367.69 bp, resulting in a total of ~78 Mbp (Table 1). These trimmed reads were considered as high-quality sequences, with a quality (Q score) mean value of 33.

As *Piaractus mesopotamicus* is considered a non-model organism and, therefore without reference genome, *de novo* assembly strategy was performed for transcriptome analysis, which yielded in 4,110 non-redundant *contigs* as a result of 193,247 reads overlapped (71,581,413 bp). The average, minimum and maximum lengths of *contigs* were: 800, 203 and 5,727 bp, respectively with 3,373,792 nucleotides of coverage. Of the total, 50% of the assembled bases are contained in contigs of 871 bp or larger. A total of 19,298 remaining reads (6,568,796 bases) were considered as singletons, and therefore were not used for subsequent analysis (Table 2).

3.2 Functional Annotation Analysis

Non-redundant sequences were annotated by BLASTx algorithm against the NCBI non-redundant protein database (nr). A total of 2,051 unique protein accessions (49.9% of transcripts) had significant hits in the nr protein database, being denominated as top Blast hits sequences which presented lower e-values and higher bit-score parameters suggesting profitable alignments. Meanwhile, almost all the analyzed sequences (99.6%) showed similarity values ≥ 50%, with most of values ≥ 80% (87.3%) against NCBI nr database.

The reference fish species with the most significant hits to *Piaractus mesopotamicus* sequences were the blind cavefish (*Astyanax mexicanus*), followed by zebrafish (*Danio rerio*), rainbow trout (*Oncorhynchus mykiss*) and catfish (*Ictalurus punctatus*) (Fig 1).

Through the analysis in *Interpro* member databases, which provides comprehensive information about protein families, domains and functional sites, we found 1,665 annotated sequences with *Interpro* accession numbers, accounting for ~40% of total transcripts. The unique genes with matches in public protein databases were annotated in *Gene Ontology (GO)* categories, which provide a dynamically controlled vocabulary and hierarchical relationships to represent information regarding biological process, molecular function and cellular component categories (Ashburner *et al*., 2000). Among the 2,051 unique genes, 1,773 (86,4%) were annotated with 10,737 *GO* terms. Regarding the most frequent *GO* assignments in this transcriptome study, we found 1,484 annotated with biological process (GO: 0008150), followed by 1,528 of molecular function (GO: 0003674) and 1,130 of cellular component (GO: 0005575). The similarity of the contigs to different *GO* terms was plotted in hierarchical level (level 2) (Fig 2). For biological process, genes involved in cellular process (GO:0009987; 1130 genes), metabolic process (GO:0008152; 1118 genes) and single-organism process (GO:0044699; 795 genes) were highly represented; for molecular function, the *GO* terms of binding (GO:0005488; 1074 genes) and catalytic activity (GO:0003824; 778 genes) were the most representatives; while in cellular component, the *GO* terms of cell (GO:0005623; 875 genes), organelle (GO:0044464; 674 genes) and membrane (GO:0005622; 842 genes) were the most frequent (Fig 2). According to *GO* term mapping results, *enzyme codes (ECs)* numbers are available to 468 transcripts. Meanwhile, we have still identified 619 enzyme code numbers from *KEGG* mapping results which revealed a total of 1,023 transcripts (~25% of total transcripts) mapped to 111 different enzyme pathways. The top enriched pathways are shown in Fig 3 and they are mainly involved in purine metabolism (KO00230), thiamine metabolism (KO00730), biosynthesis of antibiotics (KO01130), aminobenzoate degradation (KO00627) and drug metabolism (KO00983).

3.3 SNP Identification and Classification

The major putative SNP parameters are showed in Table 3. In total, 802 putative SNPs in 229 contigs were found, with an average SNP depth coverage of 54.8x. We detected 568 transitions (70.8%) and 234 transversions (29.2%), where the transition ratios were 37.5% (A/G) and 33.3% (C/T) and transversion ratios were 9.0% (G/T), 7.6% (A/C), 6.5% (A/T) and 6.1% (C/G). This corresponds to a transition:transversion ratio of 2.43:1.

The majority of contigs (~75%) has fewer than 16 SNPs per contig (Fig 4A). Additionally, most of the contigs with SNPs exhibited lengths fewer than 2,000 bp, with average lengths of 1,471.4 bp (Fig 4B). The average number of SNPs per kilobase was 2.4 and about 10% of the total contigs with putative SNPs had high polymorphism rates, with ~1 SNP every 100 bp.

After the filter steps, we selected and classified 464 SNPs to several categories according to their locations, including CDS non-synonymous, CDS synonymous, 5'UTR and 3'UTR. For this, *ORFs* were identified and the informative strand, reading frame, and stop codon at the each contig were recorded using homology with the highest homologous annotated sequence in NCBI database. In total, 8 (1.7%) SNPs could not be positioned and identified, 46 (10.0%) were located in 5'UTR, 80 (17.2%) in 3'UTR and 330 (71.1%) in CDS region (Fig 5A), of which 233 SNPs (70.6%) were classified as synonymous and 97 (29.4%) as non-synonymous (amino acid change) (Fig 5B).

3.4 SNP Genotyping and Validation

Fifty SNPs were selected (only one SNP per contig) with good coverage and quality assumptions to perform validation. Of the total, 32 feasible and polymorphic SNPs were successfully genotyped. Primers for isolation of the SNP locus, and extension primers that contains the SNPs with their different masses, were specified in Table 5.

3.5 SNP Diversity and Population Analysis

Statistical parameters to describe the genetic populations of the Paraná River (PR - wild) were obtained through the validation process of 32 feasible polymorphic SNPs (Table 5). Minimum allele frequency values (*MAF*) in the polymorphic SNPs ranged from 0.029 (C271_399) to 0.485 (C348_245, C585_507) with an average value of 0.275 ± 0.132. Expected (H_{exp}) and observed heterozygosity (H_{obs}) were estimated. H_{exp} presented an average of 0.370 ± 0.130; whereas H_{obs} showed average of 0.385 ± 0.171 and did not present significant differences ($p > 0.05$).

In some loci, the H_{obs} were higher than H_{exp} in PR (C1459_108; C260_818). Moreover, we also observed lower values of H_{obs} in relation to H_{exp}, such as C128_1801 and C585_507 loci in PR. These results interfere with the accordance of Hardy-Weinberg equilibrium (*p-value* > 0.05) affording some marginal p (HW) values due heterozygote excess or deficiency. However, all SNP markers were at Hardy-Weinberg equilibrium after Bonferroni correction (*p* = 0.0015). Additionally, when samples were tested for all *loci*, PR population was in accordance with Hardy-Weinberg expectations with *p-value* reaching 0.293. Significant linkage disequilibrium (*LD*) was detected in 5 of 1488, even with the Bonferroni correction (p = 0,0015). Of the 32 nonsignificant values of inbreeding coefficients (F_{is}), 21 were negative (heterozygote excess) and 11 were positive (heterozygote deficiency) and presented an overall loci value of -0.038 (*p-value* > 0,0002).

4 Discussion

In this study, we designed transcriptome sequencing of pacu by NGS (RNA-seq) in order to reduce the genome of this non-model organism and to obtain information of the its functional transcripts. Despite being a highly valued Neotropical fish in the fishing and aquaculture industries, pacu still has few genetic studies that could be very useful in a better understanding of genetic structure of its natural populations and provide benefits for future breeding programs. Therefore, due to these concerns, transcriptome studies with SNP

mining has been widely used in fish (Montes *et al.,* 2013; Vera *et al*., 2013; Wang *et al.,* 2014).

As a result of transcriptome sequencing, transcripts offered additional genetic resources for genomic studies of the species. For this, these sequences were considered as high-quality sequences, with a quality (Q Phred) score) mean value of 33. The quality of the sequencing can be demonstrated in recent transcriptome analyses in fish which showed inferior Q Phred parameters, such as transcriptome sequencing of sturgeon (Vidotto *et al*., 2013) and large yellow croaker (Xiao *et al*., 2015). In relation to sequence alignment in the *de novo* assembly, the average *contig* length resultant was 800 base pairs with N50 size of the 871 bp. When compared with other *de novo* transcriptomes in fish, results were similar to those found in red cusk-eel (*Genypterus chilensis*) with a N50 value of 846 bp (Aedo *et al*., 2014). While for fast eskeletal muscle transcriptome of the gilthead sea brean (*Sparus aurata*), smaller values of N50 (679 bp) and average *contig* size (450 bp) were found (Garcia de la serrana *et al*., 2012). Therefore, *de novo* assembly for this study was performed within the transcriptome patterns for fishes, presented good quality and reliability in the sequencing.

We designed a BLAST-based assessment of transcriptome sequences in order to design functional annotation by identity with functionally characterized homologous genes. Thus, the reference species that got more homologous genes with pacu sequences was blind cavefish (*Astyanax mexicanus*), the unique species of Top Blast Hits that belongs to the same order of pacu. Additionally, no sequence showed homology with known pacu protein sequences deposited in NCBI database, because of the available sequences database are still limited mostly in microsatellites and mitochondrial sequences (Calcagnotto *et al*., 2001; Calcagnotto and DeSalle, 2009; Iervolino *et al*., 2010). Consequently, the data of the present study increase the knowledge about pacu genes through NGS technologies.

Liver samples were selected for transcriptome studies because it plays a critical role in coordinating various physiological processes, including digestion, metabolism, detoxification, and endocrine system immune response (Martin *et al*., 2010). The liver may still be considered a very interesting model for the

study of interactions between environmental factors and hepatic functions (Dutta *et al*., 1996), which means that not only transcripts related to morphological characteristics will be obtained, but also transcripts related to environmental interactions such as disease resistance.

The proportion of *GO* terms in this study were similar to that found in other transcriptome studies in half-smooth tongue sole, deep-sea black scabbardfish and red cusk-eel (Aedo *et* al., 2014; Stefanni *et al*., 2014; Wang *et al*., 2014) and revealed genes correlated to economical important traits to production of this species. Among them, transcripts related to the immune system (GO0006955) were well represented (17 gene descriptions) as MHC (major histocompatibility complex) (Li *et al*., 2011), complement components C7 and C8 (Aybar *et al*., 2009) and chemokines (Kuroda *et al*., 2003) transcripts. Although these data of transcriptome generate transcripts related to immune system, they did not present SNPs and will be used in future SNP mining for pacu.

This study allowed identification of 802 putative SNPs in 229 *contigs* of pacu. The average SNP depth coverage was 54.8x which was higher when compared to SNP prospection studies in others fish (Renaut *et al.,*2010; Liu *et al*., 2011; Vera *et al*., 2013), allowing the search for true SNPs with improved accuracy (Yu and Sun, 2013). We detected more transitions (70.8%) than transversions (29.2%) and mutation ratios were similar with other SNP mining by trancriptome analysis in fishes, with few differences in abundance order between nucleotide mutations (Renaut *et al.,* 2010; Liu *et al*., 2011; Vera *et al*., 2013). However, we had a transition:transversion rate of 2.43:1, which represented a higher proportion of transitions over transversions than in most research studies of EST-SNPs in fishes, for example in turbot (Ts:Tv = 1.35) (Vera *et al*., 2013) and in lake whitefish (Ts:Tv = 1.65) (Renaut *et al.,* 2010) but similar to the found in European anchovy (*Engraulis encrasicolus*) (Montes *et al*., 2013).

High number of SNPs concentrated in a small area was discarded because it could represent SNPs false positives from error sequencing, paralogous sequence variants or pseudogenes. However the major concern was to identify high quality SNPs and thus make possible minimum chances of

false positives SNPs presence (Liu *et al*., 2011). In order to select SNPs with high confidence, 464 putative SNPs are selected and classified in CDS non-synonymous, CDS synonymous, 5'UTR and 3'UTR.

Non-synonymous SNPs are candidates for functional changes in the correspondent proteins leading to phenotypic changes (Vera *et al*., 2013). The relationship between synonymous and non-synonymous was almost 2:1 because evolutionary constraints should preferentially eliminate non-synonymous variation since it could be associated with deleterious mutation (Hubert *et al*., 2010). SNPs are also more common in 3'UTR than 5'UTR due the approximately the double length of 3'UTR (Pardo *et al*., 2008) even though their presence in untranslated regions are equally important because they play an important role in gene expression regulating (Mignone *et al*., 2002). Finally, the high amount of SNPs located in CDS demonstrates the utility of NGS for SNP detection with gene association.

In relation of 32 validated SNPs, non-synonymous feasible SNPs were located in coding regions of proteins (Table 4). These SNPs were located in regions that may be useful for aquaculture and improving the knowledge of species biology, as coding regions of proteins related to immune system genes. In this case, we found a non-synonymum SNP in vitronectin-like gene (C1013_445). Vitronectin have the potential to contribute to homeostasis in the event of tissue trauma or infection (Bayne *et al*., 2001). We also found SNPs located in untranslated regions (UTR) that could interfere in regulation of the expression of immune genes, such as SNP presents in coagulation factor V genes (C238_1041) related to the process of protection against infectious agents after an injury; or SNPs in untranslated region of gluthatione S-transferase (C627_936), related to detoxification of microcystins produced by cyanobacteria (Best *et al*., 2002). This set of polymorphic SNPs (Table 4) may provide useful information for further pacu breeding programs.

The 32 validated SNPs, as well as characterized to evaluate possible application in aquaculture, were also used to verify the genetic variability in wild population of the Parana river, where the diversity of population was checked. The average observed heterozygosity (0.385 ± 0.171) was slightly higher than average expected heterozygosity (0.370 ± 0.130). However, expected (H_{exp})

and observed (H_{obs}) heterozygosity means did not presented significant differences (p >0.05). The values of heterozygosity found in the present study were lower than that presented in other genetic variability studies of pacu through microsatellite markers (H_{exp} = 0.578; H_{obs} = 0.564) (Calcagnotto and DeSalle, 2009), which could be attributed to highest number of alleles in microsatellites.

The little difference between H_{obs} and H_{exp} means also resulted in accordance with the Hardy Weinberg equilibrium, except in four loci that present heterozygote (C260_818;C1459_108) or homozygote excess (C128_1801; C585_507). However, loci characterized with homozygotes excess showed highest null allele frequencies (0.547 in locus C128_1801 and 0.214 in locus C585_507). The Hardy-Weinberg equilibrium was also verified when the F_{is} values were analyzed. The results showed predominance of negative F_{is} values with nonsignificant over all loci value of -0,038, indicating a population in equilibrium, although there were more loci with negative values.

The present study offers genetic resources for pacu (*Piaractus mesopotamicus*), a non-model warmwater species used in aquaculture of several countries, which has high market value, but little studied genetically. Genetic researches of economic traits in pacu are limited to muscle transcriptome study aiming a better understanding of the metabolism involved in growth characteristics (Mareco *et al*., 2015). Beyond the interest in increased growth rates in economically important fish species, selective breeding for disease resistance has received special attention in aquaculture breeding programs worldwide (Pardo *et al*., 2008; Odegard *et al*., 2011). Furthermore, our pioneer SNPs set for Neotropical fishes showed the applicability of these genomic markers in pre-breeding programs and management studies of wild populations. The SNPs obtained by liver transcriptome sequencing will be useful for the construction of a genetic linkage map in pacu, allowing further studies with economically relevant traits for the industry. Moreover, our results may be applied as a basis for the development of breeding programs to other Neotropical species of economic interest.

5. References

Aedo JE, Maldonado J, Estrada JM, Fuentes EN, Silva H, Gallardo-Escarate C, et al. Sequencing and de novo assembly of the red cusk-eel (Genypterus chilensis) transcriptome. Marine Genomics. 2014; 18(B): 105-107.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nature Genetics. 2000; 25: 25-29.

Aybar L, Shin D, Smith SL. Molecular characterization of the alpha subunit of complement component C8 (GcC8α) in the nurse shark (*Ginglymostoma cirratum*). Fish & Shellfish Immunology. 2009; 27(3):397-406.

Bayne CJ, Gerwick L, Fujiki K, Nakao M, Yano T. Immune-relevant (including acute phase) genes identified in the livers of rainbow trout, *Oncorhynchus mykiss*, by means of suppression substractive hybridization. Developmental and Comparative Immunology. 2001; 25(3): 205-217.

Best JH, Pflugmacher S, Wiegand C, Eddy FB, Metcalf JS, Codd GA. Effects of enteric bacterial and cyanobacterial lipopolysaccharides, and of microcystin-LR, on glutathione S-transferase activities in zebra fish (*Danio rerio*). Aquatic Toxicology. 2002; 60:223-231.

Brumfield R T, Beerli P, Nickerson D A, Edwards S V. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution. 2003; 18(5):249-256.

Calcagnotto D, Russello M, DeSalle R. Isolation and characterization of microsatellite loci in Piaractus mesopotamicus and their applicability in other Serrasalminae fish. Molecular Ecology Notes. 2001; 1: 245-247.

Calcagnotto D, DeSalle R. Population genetic structuring in pacu (Piaractus mesopotamicus) across the Paraná-Paraguay basin: evidence from microsatellites. Neotropical Ichthyology. 2009; 7(4): 607-616.

Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21: 3674-3676.

Cui J, Liu S, Zhang B, Wang H, Sun H, Song S, et al. Transcriptome analysis of the gill and swimbladder of *Takifugu rubripes* by RNA-Seq. PLoS ONE. 2014; 9(1): e85505.

Duan J, Xia C, Zhao G, Jia J, Kong X. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics. 2012; 13:392.

Dutta HM, Datta Munshi JS. Fish morphology: horizon of new research. Science Publishers: Lebanon, USA, 1996; 300 pp.

FAO. The State of World Fisheries and Aquaculture 2010. FAO: Rome, Italy, 2010; 197 pp.

Fernández J, Toro MÁ, Sonesson AK, Villanueva B. Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. Frontiers in Genetics. 2014; 5:414.

Flores Nava, A. Aquaculture seed resources in Latin America: a regional synthesis. In: Assessment of freshwater fish seed resources for sustainable aquaculture. FAO Fisheries Technical Paper No 501; Bondad-Reantaso, M.G., Ed.; FAO: Rome, Italy, 2007; 628 pp.

Garcia de la serrana D, Estévez A, Andree K, Johnston IA. Fast skeletal muscle transcriptome of the Gilthead sea bream (Sparus aurata) determined by next generation sequencing. BMC Genomics 2012; 13:181.

Gallardo-Escárate C, Valenzuela-Muñoz V, Nuñes-Acuña G. RNA-seq analysis using *de novo* transcriptome assembly as a reference for the salmon louse *Caligus rogercresseyi*. PloS ONE. 2014; 9(4): e92239.

Gomes AS, Alves RN, Stueber K, Thorne MAS, Smáradóttir H, Reinhard R, et al. Transcriptome of the Atlantic halibut (*Hippoglossus hippoglossus*). Marine Genomics. 2014; 18(B): 101-103.

Honglang, H. Freshwater fish seed resources in China. In: Assessment of freshwater fish seed resources for sustainable aquaculture, FAO Fisheries Technical Paper No 501; Bondad-Reantaso, M.G., Ed.; FAO: Rome, Italy, 2007; 628 pp.

Hubert S, Higgins B, Borza T, Bowman S. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). BMC Genomics. 2010; 11:191.

Huang L, Li G, Mo Z, Xiao P, Li J, Huang J. De novo assembly of the japanese flounder (*Paralichthys olivaceus*) spleen transcriptome to identify putative genes involved in immunity. PLoS ONE. 2015; 10(6): e0131146.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013; 496: 498-503

IBGE. Instituto Brasileiro de Geografia e Estatística. Produção da Pecuária Municipal 2013*, Vol 41*. Rio de Janeiro, Brasil, 2014; 1-108.

Iervolino F, Resende EK, Hilsdorf AWS. The lack of genetic differentiation of pacu (Piaractus mesopotamicus) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: Implications for fishery management. Fisheries Research. 2010; 101: 27-31.

Ji P, Liu G, Xu J, Wang X, Li J, et al. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. PLoS ONE*.* 2012; 7(4): e35152.

Kuroda N, Uinuk-ool TS, Sato A, Samonte IE, Figueroa F, Mayer WE, et al. Identification of chemokines and a chemokine receptor in cichlid fish, shark, and lamprey. Immunogenetics. 2003; 54:884-895.

Lanes CFC, Bizuayehu TT, Fernandes JMO, Kiron V, Babiak I. Transcriptome of Atlantic cod (*Gadus morhua* L.) early embryos form farmed and wild broodstocks. Marine Biotechnology. 2013; 15: 677-694.

Li W, Godzik A. Cd-hit: a fast program for clustering and compare large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22(13): 1658-1659.

Li C, Zhang Q, Yu Y, Li S, Zhong Q, Sun Y, et al. Sequence polymorphism of two major histocompatibility (MH) class II B genes and their association with Vibrio anguillarum infection in half-smooth tongue sole (Cynoglossus semilaevis). Chinese Journal of Oceanology and Limnology. 2011; 29(6): 1275-1286.

Liu ZJ, Cordes JF. DNA marker technologies and their applications in aquaculture genetics. Aquaculture. 2004; 238: 1-37.

Liu S, Zhou Z, Lu J, Sun F, Wang S et al. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. BMC Genomics. 2011; 12:53.

Louis EJ, Dempster ER. An exact test for Hardy-Weinberg and multiple alleles. Biometrics. 1987; 43: 805-811.

Mardis ER. Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics. 2008; 9: 387-402.

Mareco E A, Serrana D G, Johnston I A, Dal-Pai-Silva M. Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). BMC Genomics. 2015; 16:182.

Martin SAM, Douglas A, Houlihan DF, Secombes CJ. Starvation alters the liver transcriptome of the innate immune response in Atlantic salmon (Salmo salar). BMC Genomics. 2010; 11:418 (35)

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7: 639-655. doi: 10.1046/j.1365-294x.1998.00374.x

Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biology. 2002; 3(3): reviews0004.1-0004.10.

Micallef G, Bickerdike R, Reiff C, Fernandes JMO, Bowman AS, Martin SAM. Exploring the transcriptome of Atlantic salmon (*Salmo salar*) skin, a major defense organ. Marine Biotechnology. 2012; 14(5): 559-569.

Montes I, Conklin D, Albaina A, Creer S, Carvalho GR, et al. SNP Discovery in European Anchovy (*Engraulis encrasicolus*, L) by high-throughput transcriptome and genome sequencing. PLoS ONE. 2013; 8(8): e70051.

MPA. Ministério da Pesca e Aquicultura. Boletim estatístico da pesca e aquicultura - 2011; Brasília, Brasil, 2013; 60 pp.

Núñez-Acuña G, Gallardo-Escárate C. Identification of immune-related SNPs in the transcriptome of *Mytilus chilensis* through high-throughput sequencing. Fish & Shellfish Immunology. 2013; 35: 1899-1905.

Odegard J, Baranski M, Gjerde B, Gjedrem T. Methodology for genetic evaluation of disease resistence in aquaculture species: challenges and future prospects. Aquaculture Research. 2011; 42: 103-114.

Pardo B G, Fernández C, Millan A, Bouza C, Vázquez-López A, Vera M, et al. Expressed sequence tags (ESTs) from imune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens. BMC Veterinary Research. 2008; 4:37.

Parkinson J, Blaxter M. Expressed Sequence Tags: An overview. Methods in Molecular Biology. 2009; 533: 1-12.

Resende, E.K. Migratory fishes of the Paraguay-Paraná basin excluding the Upper Paraná River. In: Migratory fishes of South America: biology, fisheries and conservation states; Carolsfeld, J.; Harvey, B.; Ross, C.; Baers, A., Eds.; World Bank: Victoria, Canada, 2003; pp.99-156.

Renaut S, Nolte A W, Bernatchez L. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Molecular Ecology 2010, 19(1):115-131.

Rice WR. Analyzing tables of statistical tests. Evolution. 1989; 43: 223-225.

Rousset F. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources. 2008; 8: 103-106.

Salem M, Paneru B, Al-Tobasei R, Abdouni F, Thorgaard GH, Rexroad CE, et al. Transcriptome assembly, gene annotation, and tissue gene expression atlas of the rainbow trout. PLoS ONE. 2015; 10(3): e0121778.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources. 2011; 11(s1): 1-8.

Stefanni S, Bettencourt R, Pinheiro M, De Moro G, Bongiorni L, Pallavicini A. Transcriptome of the deep-sea black scabbardfish, Aphanopus carbo (Perciformes: Trichiuridae): tissue-specific expression patterns and candidate genes associated to depth adaptation. International Journal of Genomics. 2014; 2014, ID:267482.

Smith CT, Templing WD, Seeb JE, Seeb LW. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian chinook salmon caught in Yukon river fisheries. North American Journal of Fisheries Management. 2005; 25(3): 944-953.

Shin SC, Kim SJ, Lee JK, Ahn DH, Kim MG, Lee H, et al. Transcriptomics and comparative analysis of three Antarctic notothenioid fishes. PLoS ONE. 2012; 7(8): e43762.

Teacher A G F, Kähkönen K, Merilä. Development of 61 new transcriptome-derived microsatellites for the Atlantic herring (*Clupea harengus*). Conservation Genetics Resources. 2012; 4(1):71-74.

Vandeputte M, Haffray P. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. Frontiers in Genetics. 2014; 5: 432.

Vera M, Alvarez-Dios JA, Fernandez C, Bouza C, Vilas R, Martinez P. Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. International Journal of Molecular Sciences. 2013; 14: 5694-5711.

Vidotto M, Grapputo A, Boscari E, Barbisan F, Coppe A, et al. Transcriptome sequencing and de novo annotation of the critically endangered Adriatic sturgeon. BMC Genomics 2013; 14: 407

Wang W, Yi Q, Ma L, Zhou X, Zhao H, Wang X, et al. Sequencing and characterization of the transcriptome of half-smooth tongue sole (*Cynoglossus semilaevis*). BMC Genomics. 2014; 15:470.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009; 10(1): 57-63.

Weir BS, Cockerhsam CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38: 1358-1370.

Xiao S, Han Z, Wang P, Han F, Liu Y, et al. Functional marker detection and analysis on a comprehensive transcriptome of large yellow croaker by Next Generation Sequencing. Plos One 2015; 10(4): e0124432.

Xu J, Ji P, Zhao Z, Zhang Y, Feng J, Wang J, et al. Genome-wide SNP discovery from transcriptome of four common carp strains. PLoS ONE. 2012; 7(10): e48140.

Yu X, Sun S. Comparing a few SNP calling algorithms using low coverage sequencing data. BMC Bioinformatics. 2013; 14: 274.

Zhao Q, Wang Y, Kong Y, Luo D, Li X. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics. 2011; 12(14):S2.

Table List

Table 1. Raw reads trimming of pacu *Piaractus mesopotamicus* liver transcriptome. bp (base pairs)

Sequencing data	Value
Number of raw reads	212,813
Average length of raw reads (bp)	402
Total nucleotides of raw reads	85,550,940
Number of trimmed reads	212,545
Average length of trimmed reads (bp)	367.69
Total nucleotides of trimmed reads	78,150,209

Table 2. Statistics of pacu *Piaractus mesopotamicus* transcriptome *de novo* assembly. bp (base pairs)

Transcriptome features	Value
Matched reads for assembly	193,247
Total nucleotides of matched reads	71,581,413
Contig number	4,110
Total of contig nucleotides	3,373,792
Maximum contig length (bp)	5,727
Minimum contig length (bp)	203
Average contig length (bp)	800
N25 (bp)	1,472
N50 (bp)	871
N75 (bp)	595
Singletons	19,298

Table 3. Putative SNPs identification parameters, through *de novo* assembly of pacu (*Piaractus mesopotamicus*).

SNP features		Value
Contig with SNPs		229
Contig average length (min-max) (bp)		1461.8 (209-5,150)
SNP amount		802
SNP per kilobase		2.4
Average coverage of reads (min-max)		54,8x (15-1,619x)
Transitions	A-G	301 (37.5%)
	C-T	267 (33.3%)
Transversions	G-T	72 (9.0%)
	A-C	61 (7.6%)
	A-T	52 (6.5%)
	C-G	49 (6.1%)
Transition:transversion ratio		2.43

Table 4. Predicted position, gene description, SNP location within genes, GO terms and location in *Danio rerio* chromosomes of the 32 feasible SNPs of pacu (*Piaractus mesopotamicus*)

SNP_ID	Gene Description	SNP location/effect	GO term	Chromosome *Danio rerio*
C4_231	unnamed protein product	Synonymous	translation (GO:0006412)	4
C5_660	Uncharacterized protein LOC103025530	3'UTR	not detected	16
C30_132	Ribosomal protein S9	Synonymous	protein binding (GO:0005515)	16
C41_428	Secreted phosphoprotein 24	Synonymous	bone remodeling (GO:0046849)	1
C43_831	Protein AMBP precursor	Non-synonymous	negative regulation of endopeptidase activity (GO:0010951)	10
C83_761	Glutathione peroxidase 3 precursor	3'UTR	hydrogen peroxide catabolic process (GO:0042744)	14
C87_1726	Heparin cofactor 2-like	Non-synonymous	negative regulation of endopeptidase activity (GO:0010951)	8
C128_1801	Proteoglycan 4-like isoform X2	Synonymous	immune response (GO:0006955)	20
C147_351	Vitelline membrane outer layer protein 1 homolog isoform X1	Synonymous	not detected	21
C178_243	Haptoglobin	Synonymous	not detected	7
C191_480	40S ribosomal protein S2	Synonymous	RNA binding (GO:0003723)	3
C213_629	Prostaglandin-H2 D-isomerase-like	3'UTR	not detected	not found
C238_1041	Coagulation factor V-like	3'UTR	response to bacterium (GO:0009617)	9
C239_1594	Basigin isoform X1	3'UTR	protein binding (GO:0005515)	22
C240_1549	Heme oxygenase-like	3'UTR	phospholipid catabolic process (GO:0009395)	3
C260_818	Complement C5-like	Synonymous	endopeptidase inhibitor activity (GO:0004866)	5
C271_399	Ferritin, heavy subunit-like	5'UTR	iron ion transport (GO:0006826)	7
C348_245	Ribosomal protein S7	Synonymous	embryo development (GO:0009790)	20
C379_275	UPF0762 protein C6orf58 homolog	5'UTR	not detected	20
C391_875	Fibronectin	Synonymous	not detected	1
C417_302	Alpha-1-antitrypsin homolog isoform X6	5'UTR	regulation of proteolysis (GO:0030162)	20
C437_455	40S ribosomal protein S18	3'UTR	multicellular organismal development (GO:0007275)	19
C455_315	15-hydroxyprostaglandin dehydrogenase [NAD(+)]isoform X1	5'UTR	prostaglandin metabolic process (GO:0006693)	1
C458_2209	Polyadenylate-binding protein 1-like	3'UTR	nucleotide binding (GO:0000166)	19
C470_159	Peptidyl-prolyl cis-trans isomerase-like isoform X2	5'UTR	regulation of cell cycle (GO:0000074)	8
C564_1273	S-adenosylmethionine synthase isoform type-1-like	Synonymous	methionine metabolic process (GO:0006555)	13
C579_153	60S ribosomal protein L14-like	Synonymous	not detected	19
C585_507	novel protein similar to H.sapiens HPN, hepsin	Synonymous	proteolysis (GO:0006508)	16
C627_936	Glutathione S-transferase A-like	3'UTR	metabolic process (GO:0008152)	19
C857_201	Inter-alpha-trypsin inhibitor heavy chain H3-like isoform X3	Non-synonymous	not detected	11
C1013_445	Vitronectin-like	Non-synonymous	immune response (GO:0006955)	21
C1459_108	Unnamed protein product, partial	Non-synonymous	not detected	not found

Table 5. Variants and diversity values of the 32 technically feasible SNPs of pacu (*Piaractus mesopotamicus*). Paraná River (PR) population corresponds to 34 individuals. *MAF*: minimum allele frequency; HW: Hardy Weinberg p-value; H_{obs}: observed heterozigosity; H_{exp}: expected heterozigosity; Fis: inbreeding coefficient. Allele (wild/rare)

SNP_ID	PCR Primers (F and R) and extension primer (E)	Allele	*MAF*	(HW)	H_{obs}	H_{exp}	Fis
C4_231	F:GCAGCCAGATGCCAAAGTTC	T/C	0.324	1.000	0.471	0.444	-0.0602
	R:AGGCTTCTGGAGAGATTGTG						
	E:TGGCCTGGTTCATGAGAAGTCTCC						
C5_660	F:TTCGACAACTGCCATGATGC	C/T	0.338	0.254	0.559	0.454	-0.234
	R:AAAGCCTGTAGTTCAGTGTG						
	E:GTTCAGTGTGAAGCTCT						
C30_132	F:TCAGCACCACCAGCACCTG	G/A	0.176	0.249	0.235	0.295	0.205
	R:ACATCGACTTCTCCCTGCG						
	E:CCGTGTGAAGAGGAAGAA						
C41_428	F:AAATGCAATCAGGCCCAGAG	G/A	0.235	0.647	0.412	0.365	-0.130
	R:GACGATTTGGATCATAGTGC						
	E:ATCATAGTGCCATGCCAT						
C43_831	F:TATAACTCCTCCCTCATGGC	G/A	0.309	0.686	0.382	0.433	0.119
	R:AGACACTCCTTCTCTGTCAC						
	E:TGTTCTGGTTGCCCA						
C83_761	F:GACACAGAACAGGATTAGTC	G/A	0.191	0.570	0.265	0.314	0.159
	R:CATCCGTCTGATCAGTCAAC						
	E:CAAACACTAATCGACCCCA						
C87_1726	F:AAGATGGCGCCAAAGCTG	C/T	0.177	1.000	0.294	0.295	0.0030
	R:GTCCAGAGGGATAGAGTC						
	E:ACGGCCAGGTGCTCG						
C128_1801	F:ATGGACACTGGATTCCCAAG	T/C	0.118	0.0015	0.059	0.211	0.724
	R:GTACTGAGGCATGGACAAAG						
	E:TTTGACCACTCAGTCC						
C147_351	F:TTCCTACTCCAGACGTTACC	G/A	0.162	0.562	0.324	0.275	-0.179
	R:GAATTAGTCTGCACTGTGTC						
	E:AACACACGCGCAGGCAAGTT						
C178_243	F:TTGGATTCAAGGAAGGCGAG	T/C	0.162	0.562	0.324	0.275	-0.179
	R:GAGAACATGTTCTGCACAGG						

	E:CCCAGCAATTTCCAGGA						
C191_480	F:CAAGCTGTCCATCATTCCTG	C/T	0.353	1.000	0.471	0.464	-0.0154
	R:ACCAGTCACCTTGCAGGGTA						
	E:GGGGTTTGCCGATCTTGTT						
C213_629	F:CATGTCCACAAACTGTCCTG	C/T	0.471	0.165	0.647	0.506	-0.285
	R:AGTTCGTGATTCCACCTCAG						
	E:AGAGTTCTGCAAGCTGGACG						
C238_1041	F:AAACTGAACCAGTCTGGAG	A/T	0.044	1.000	0.088	0.086	-0.0313
	R:TCATTTTGGACACCCTCAC						
	E:TGGACACCCTCACTTGTAATG						
C239_1594	F:TCCCAAAGAACATAAAAGC	C/A	0.339	0.706	0.500	0.454	-0.102
	R:GTTTGGCAGCATGGGATTTG						
	E:CTTGATGTCTAACTGCAGTG						
C240_1549	F:TGCGAAGAGACACATTTCCC	T/C	0.206	0.599	0.294	0.332	0.115
	R:GTATCTTATTGCTATGGCT						
	E:GCTATGGCTTATGAACAG						
C260_818	F:CACTTTGAACAGAGAGGCAC	A/G	0.471	0.0363	0.706	0.506	-0.404
	R:AGACGAGTTCTACTGTGTGG						
	E:TGTGGGCTTCCGAAT						
C271_399	F:ATATTAGGCAAGCGGCTAAG	T/A	0.029	1.000	0.059	0.058	-0.0154
	R:CATGGCCGAATACCTGTTTG						
	E:AGCCCCTTTAACCCCC						
C348_245	F:ATCATCTTTGTGCCTGTGCC	A/G	0.485	1.000	0.500	0.507	0.0141
	R:TGCCGCTGAATTTCTTCTCC						
	E:TCGCGCACAAGCCGCAC						
C379_275	F:AGCTGGTTTATGTGGGTCTG	T/A	0.265	1.000	0.412	0.395	-0.0429
	R:TGGTACACTTGTCCACCATC						
	E:CCACCATCATGACTATGC						
C391_875	F:CGAAAACGGTCAGATGATG	G/A	0.426	0.177	0.618	0.496	-0.249
	R:ATGTGGCTCGCATTTGAAC						
	E:TCCCTTGCCATTGCC						
C417_302	F:AGAGTATCCTCTTCATGGGC	G/A	0.191	1.000	0.324	0.314	-0.0312
	R:CATATTGCTTGCTGTGTGGG						
	E:GTTGGATGCCTCTATGC						
C437_455	F:TCAGCACACTAAGACCACTG	G/A	0.132	1.000	0.265	0.233	-0.138

	R:AGCAGGGAGGGCGATTACTT						
	E:AGGGCGATTACTTCTTCTT						
C455_315	F:CATGAAAATGTTTTACAGG	C/T	0.368	1.000	0.500	0.472	-0.0605
	R:ATTGAATCCCATGGCTGTTG						
	E:AGCTTATAATGAATACTGTTTAAGA						
C458_2209	F:TCTCGGTTTTTTCCGCTCG	C/T	0.044	1.000	0.088	0.086	-0.0313
	R:ACGTACAGGGAGGCCATC						
	E:CCGGGATTCATCTCCGAATT						
C470_159	F:CCCAAAACCCTGCGTCTTTC	A/G	0.353	1.000	0.471	0.464	-0.0154
	R:ATGGGCTAGGAGTAAAACCG						
	E:CTGTTTGAACCAGGATTTTC						
C564_1273	F:ACTGCTTCAGATCGTCAAC	C/T	0.273	0.383	0.485	0.403	-0.208
	R:TTGGCCTCTTCAGCTTCAG						
	E:AATGACACCAGGCCGGAG						
C579_153	F:CAAGAAGATCGAAGCCAGAC	C/T	0.235	0.152	0.471	0.365	-0.294
	R:TCTTGGCCTTCATGACCTTG						
	E:AAGTCGTTCATTTTGGC						
C585_507	F:TCTGTGTAAAGGAGAGCGAG	C/T	0.485	0.0432	0.324	0.507	0.365
	R:AGCACTTCAACAATCTGTCC						
	E:GTGCTGATCTTCTTGCC						
C627_936	F:TCTGAGGGTGAGAACTATG	A/T	0.397	0.285	0.382	0.486	0.216
	R:GTAATAATAACTACAGATAC						
	E:AATAACTACAGATACAATAAGAATTAG						
C857_201	F:AAGAGTGGACTGCTGGCTTG	A/G	0.279	0.233	0.324	0.409	0.211
	R:GAGACACAGTGAAGTCAGAG						
	E:CCCCCTCTGCAGGACAAAC						
C1013_445	F:ATACAACAACAGCCGCTCCC	C/T	0.426	0.289	0.382	0.496	0.233
	R:TTTAGCTGCAGGAAAGCATC						
	E:GGAAAGCATCAAACGAG						
C1459_108	F:ACCATTGAACATCAGGCCAC	T/G	0.344	0.0045	0.688	0.458	-0.512
	R:CAGCTGTGTCTATGAGAACC						
	E:CCTTCAGATTGAAAACTTTCGAG						

Figure List

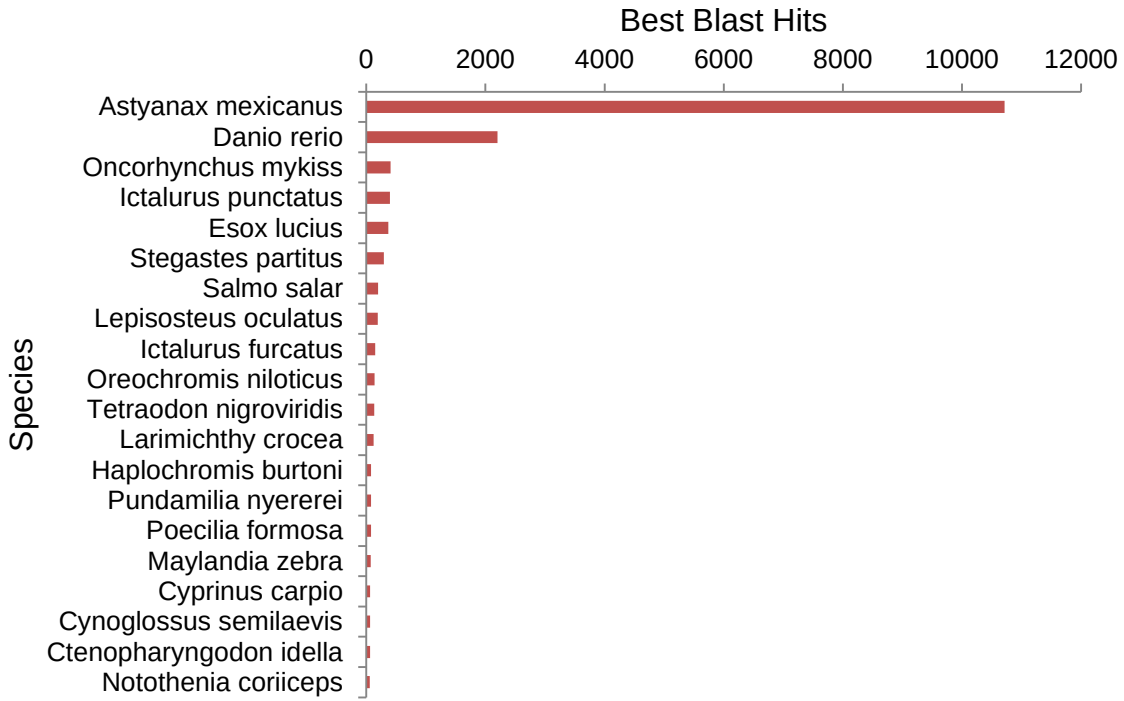

Fig 1. Best hit species distribution in nr database blast (NCBI) for *Piaractus mesopotamicus* functional annotation.

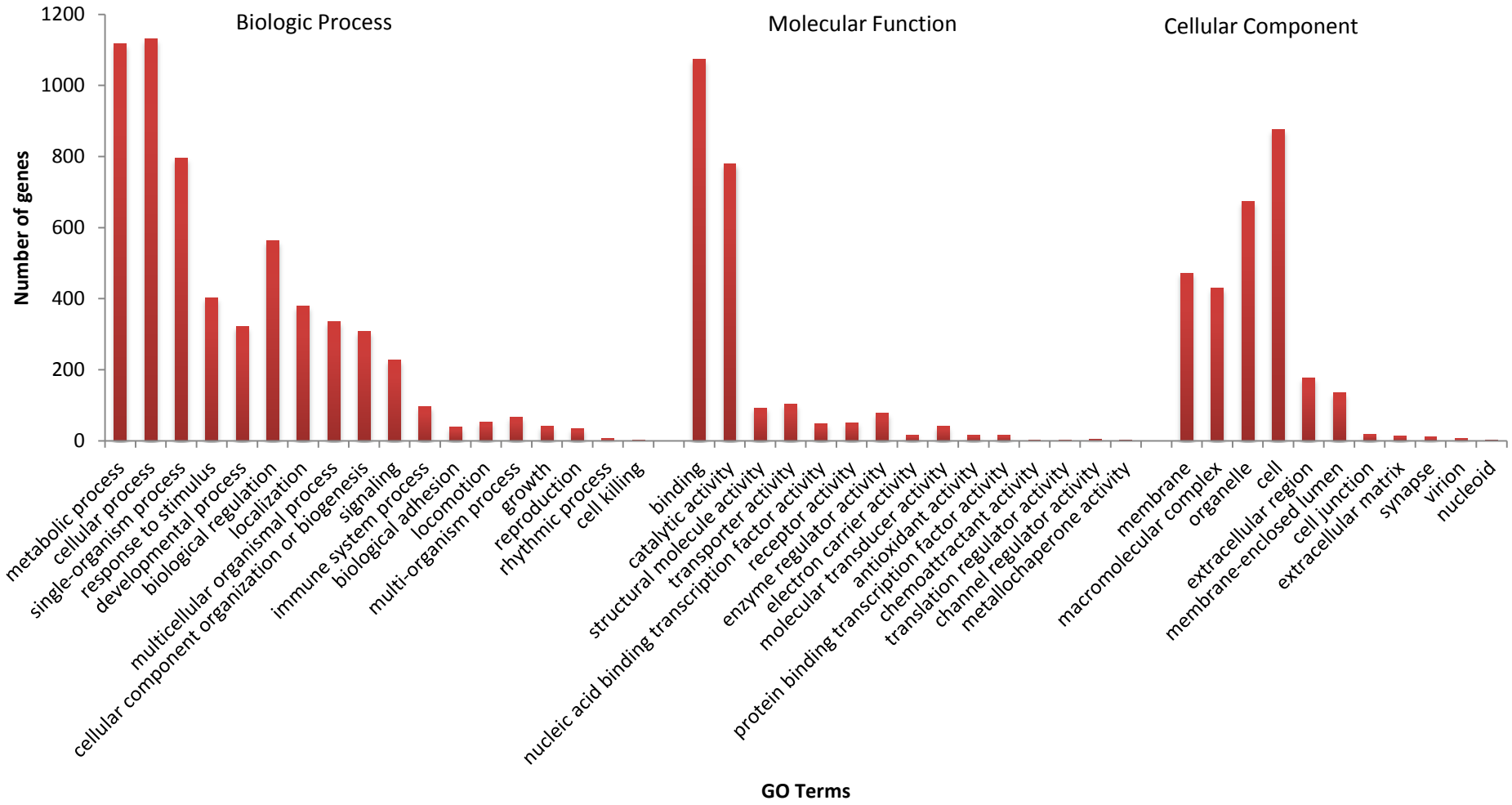

Fig 2. Gene ontology categories of *Piaractus mesopotamicus* sequences.

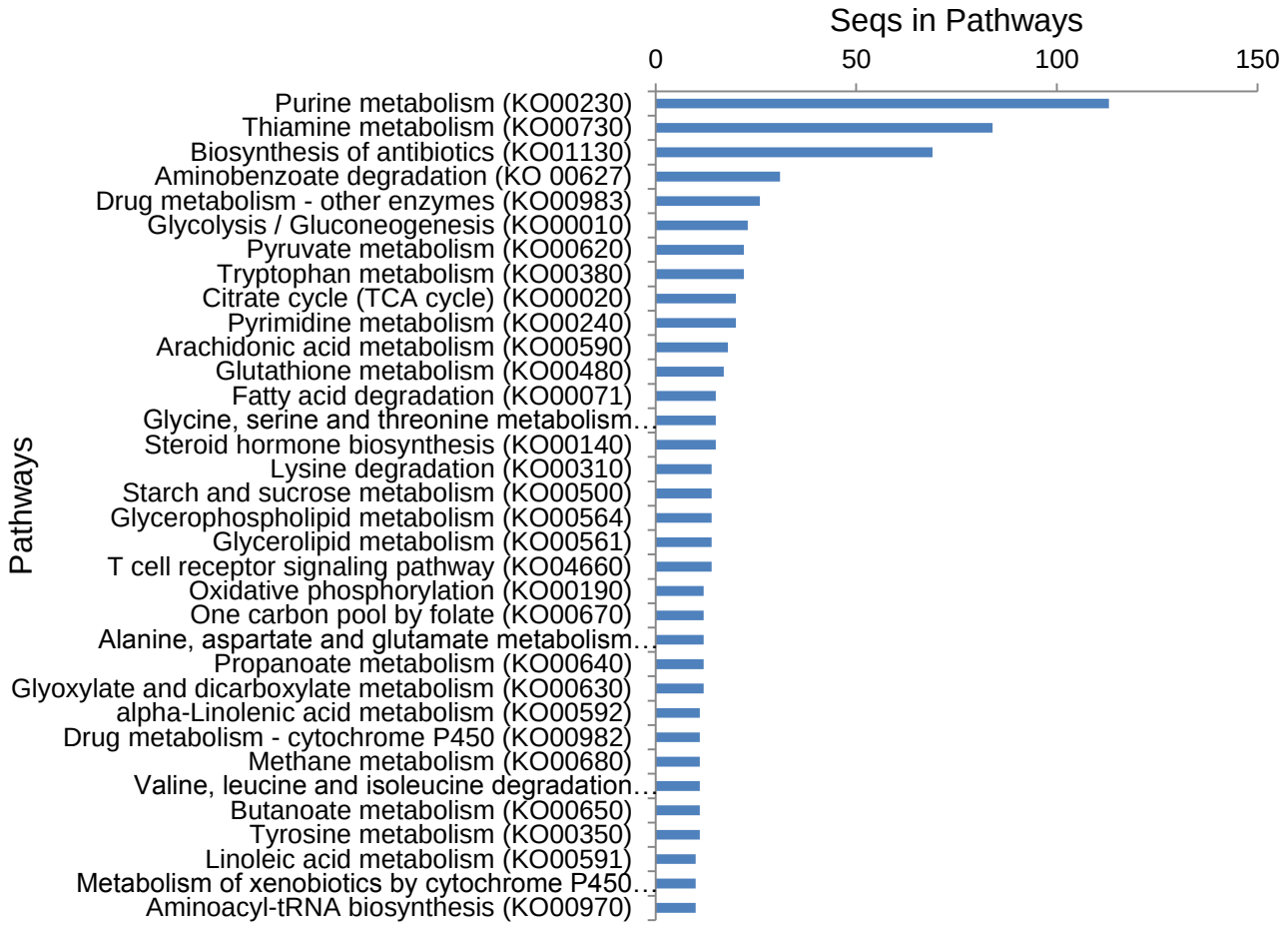

Fig 3. KEGG classification of *Piaractus mesopotamicus* liver transcriptome.

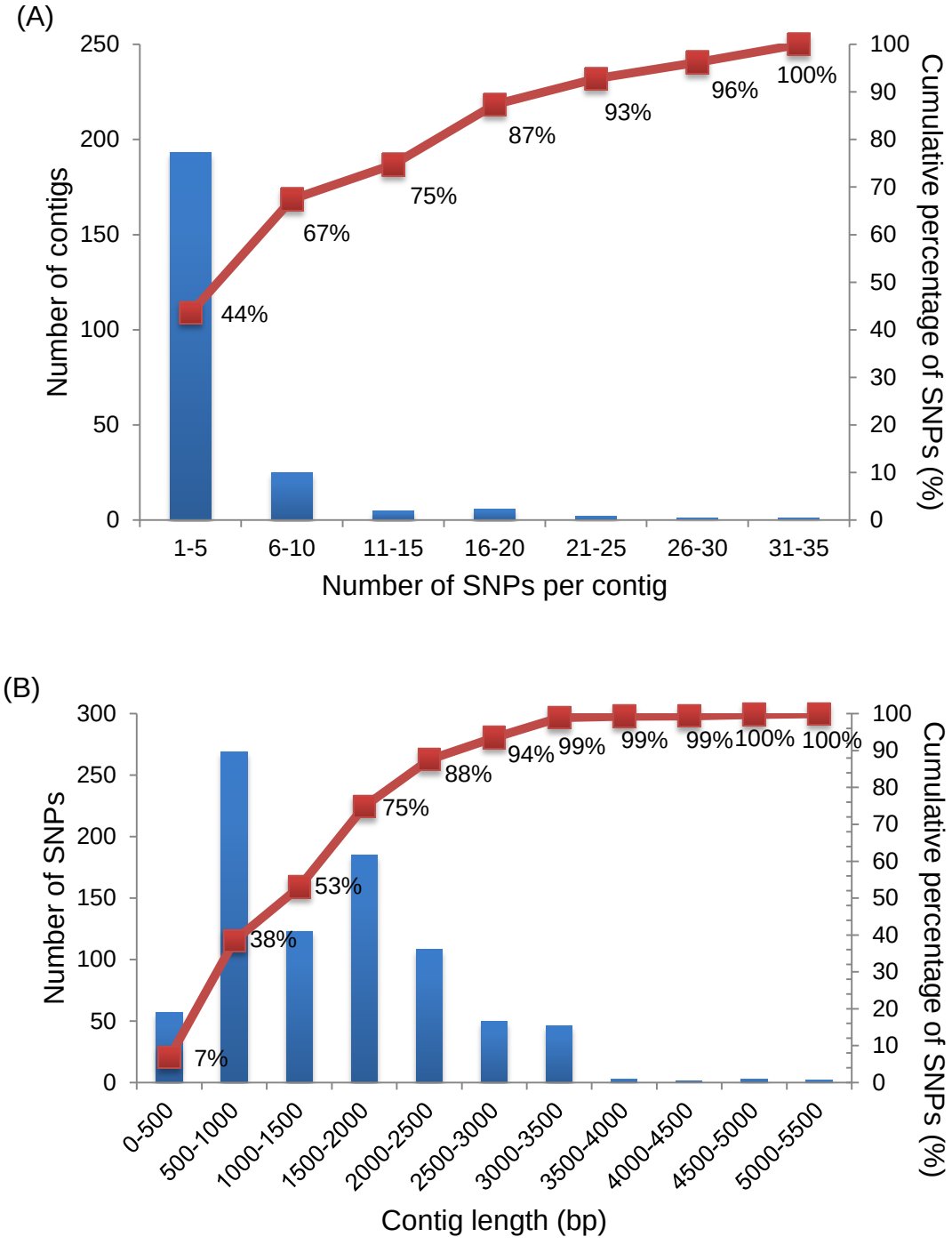

Fig 4. SNP distribution among contigs (A) and among length of contigs (B). In Fig 4A, the x-axis represents number of SNPs per contig, and in Fig 4B, the x-axis represents the contig size (base pairs). In both, the curved line denotes the cumulative percentage of SNPs assembled.

Fig 5. Frequency of SNPs according to the location in the gene. SNP classification according to the transcript region in which the mutation is situated (A) and the effect in coding region (CDS) (B), if the SNP variation kept the same functional information (synonymous SNPs) or functional changes (non-synonymous SNPs).

Capítulo 2

Genetic variability and parentage assignment assessed by SNPs in stocks of the fish pacu (*Piaractus mesopotamicus*)

Resumo

Pacu (*Piaractus mesopotamicus*) é uma espécie de peixe Neotropical amplamente distribuída na bacia do rio Prata, e comercialmente valorizada pelo seu potencial produtivo na aquicultura. Estudos relacionados à avaliação da estrutura genética de populações para esta espécie ainda são insuficientes, principalmente aplicados para programas de melhoramento genético. O objetivo do estudo foi realizar análises de variabilidade genética e parentesco em sete populações cultivadas por meio de 32 SNPs (*Single Nucleotide Polymorphisms*). Não foram observadas diferenças significativas entre as médias de heterozigosidade observada (H_{obs}), heterozigosidade esperada (H_{exp}) e frequência mínima de alelos (*MAF*) (P<0,05), quando todas as populações foram comparadas. Os valores de F_{is} foram predominantemente negativos, porém próximos de zero. Todos os loci apresentaram conformidade com o equilíbrio de Hardy-Weinberg para todas as populações, após a correção de Bonferroni (P>0,00156). Em geral, os parâmetros dos SNPs demonstraram alta variabilidade genética, que foi similar entre a população selvagem e os estoques cultivados. Além disso, análises de diferenciação genética mostraram moderado nível de estruturação genética entre populações selvagens e cultivadas de pacu (F_{st} = 0.064; K = 2), com a formação de dois grupos distintos. Os dados de AMOVA demonstraram que 93,59% da variação estavam dentro de populações. Análises de parentesco realizadas pelo coeficiente de Wang (r) mostraram que a maioria das estações de piscicultura possuíam pelo menos 40% de indivíduos aparentados, com risco de formar casais endogâmicos. Com os resultados obtidos neste estudo, conclui-se que é apropriado formar a população base do melhoramento utilizando pelos menos indivíduos de pisciculturas que compõem os dois grupos genéticos, a partir de um programa de acasalamentos direcionados e sem risco de endogamia, de forma a manter a variabilidade genética já existente dentro das populações cultivadas de pacu.

Palavras-chave: diversidade, aquicultura, endogamia, genética de populações

Abstract

Pacu (*Piaractus mesopotamicus*) is a Neotropical fish species widely distributed in the Prata basin, with high commercial value due to its productive potential in aquaculture. Studies related to the evaluation of genetic structure for this species are still insufficient, mainly applied for breeding programs. The aim of this study was to perform genetic variability and kinship analysis in seven cultivated populations of pacu through 32 SNPs (Single Nucleotide Polymorphisms). No significant differences were observed between observed heterozygosity (H_{obs}), expected heterozygosity (H_{exp}) and minimum allele frequency (*MAF*) means ($P < 0.05$), when all populations were compared. F_{is} values were predominantly negative, but close to zero. All loci showed accordance with Hardy-Weinberg equilibrium for all populations, after Bonferroni correction ($P = 0.00156$). In general, these SNPs parameters showed high genetic diversity, which was similar between the wild population and cultivated stocks. However, we detected moderate genetic structure between wild and cultivated populations (Fst = 0.064; K=2), with the presence of two distinct groups. AMOVA results demonstrated that 93.59% of the genetic variation was within of populations. Analysis performed by Wang kinship coefficient (r) showed that the majority of fish farms had at least 40% of related individuals (full and half-sib), with risk of inbreeding. With the results of the present study, we concluded that it is appropriate to begin the population basis of a breeding program using at least individuals from fish farms of the two genetic groups, through directed mating program and without inbreeding, to maintain the current genetic variability present within the farmed stocks of pacu.

Keywords: diversity, aquaculture, inbreeding, genetic populations

1. Introduction

According to the latest statistics available, the world aquaculture production reached record level in 2013 (97.2 million tons), with different intensities systems and technology levels. The fish production grew rapidly and now represents 43.1% of total aquaculture production. This growth has also increased due to expansion of incentive public policies in developing countries (FAO, 2014; 2015).

Pacu (*Piaractus mesopotamicus*) is a freshwater fish of South America with wide distribution of the Parana-Paraguay basins. Wild populations are threatened by overfishing, since this species is considered of high commercial value, with large-scale catches by the industrial and recreational fisheries (Resende, 2003). Furthermore, it is one of the most important non-model species cultivated in South America and its production has increased even in other parts of the world, such as China, Myanmar, Thailand and Vietnan (Honglang, 2007; FAO, 2010).

Genetic studies directed to the Neotropical species of aquaculture are practically absent. About pacu, they are limited to few genetic variability studies using mitochondrial DNA D-loop regions (Iervolino *et al*., 2010) and microsatellite markers (Calcagnotto *et al*., 2001; Calcagnotto and DeSalle, 2009) in wild populations. Therefore, due to the decline of natural populations and economic importance in aquaculture, a better understanding of pacu genome is necessary for the development of genetic management in wild populations, besides provide genetic resources to increase productivity in aquaculture.

Breeding programs are fundamental techniques for the sustainable development of aquaculture, allowing an increase in productive performance with reduced production costs, and better product quality through efficient use of food, water and productive areas (Gjedren and Baranski, 2009). However, less than 10% of aquaculture production is based on genetically improved stocks and limited to few species (Gjedren *et al*., 2012).

Several factors are important for breeding programs provide expressive and enduring genetic gains in fish production (Lind *et al*., 2012), including

mainly a suitable pre-breeding program. The evaluation of the genetic variability is essential for an effective control over the matings to be performed, in order to prevent problems related to narrowing of the genetic basis of certain species or stocks (Ponzoni, 2006). Changes in allele frequencies, with consequent reduction of genetic variability in selection of genetically superior organisms can result in increased mating probability of inbreeding individuals and consequent appearance of deleterious genes, decimating the potential lineage of fish production, such as reduced growth rates, low survival, morphological deformities, disease resistant problems and reduced ability to adapt to new environments (Allendorf and Phelps, 1980; Kincaid, 1983). Therefore, the analysis of how much sufficient variation can be maintained during directed mattings is important for successful management in selective breeding programs (Beaumont and Hoare, 2003).

The vast majority of studies involving genetic structure and parentage assignment researches of important species for aquaculture were initially based on the use of microsatellite markers (Herbinger *et al*., 1995; Estoup *et al*., 1998; Pardo *et al*., 2006; An *et al*., 2014; Shikano *et al*., 2015; Morzeven *et al*., 2016). However, since the advent of high-throughput sequencing techniques, genetic populational studies through SNPs (Single Nucleotide Polymorphisms) are increasingly being used in several species for aquaculture, and, in the future, they will be the DNA marker most utilized in structural population analysis (Kong *et al*., 2014; Aykanat *et al*., 2015; Jiang *et al*., 2015; Laconcha *et al*., 2015; Pocwierz-Kotus *et al*., 2015) and development of panels for parentage assignments (Abadía-Cardoso *et al*., 2013; Liu *et al*., 2016). As type I markers, SNPs originated from transcriptome analysis are associated with genes of known function and are gaining popularity to assess the functional genetic diversity of wild and cultivated stocks of non-model species (Liu and Cordes, 2004). When compared to microsatellites, SNPs are less informative due to the biallelic nature whereas microsatellites can identify multiple polymorphisms (Powell *et al*., 1996). However, the abundance of SNPs in the genome could compensate for this deficiency, including wide markers coverage in the analyses (Glaubitz *et al*., 2003). Moreover, SNPs are more suitable for automated analysis and more computer friendly due to their biallelic nature (Liu,

2011) and it has greater power to perform parental assignment when compared to microsatellite markers (Sellars *et al*., 2014).

Therefore, the main objective of the study was to perform parental assessment and genetic diversity analysis in wild and cultivated stocks of pacu using SNPs previously obtained by transcriptome sequencing (NGS) and *de novo* assembly (Mastrochirico-Filho *et al*., 2016). This research was the first study about SNP applicability in Neotropical fish species with potential for aquaculture and the results will be useful in pre-breeding programs.

2. Material and methods

2.1 Ethic Statement

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and was approved by the Animal Use Ethics Committee (CEUA), protocol nº 22.255/15. 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). Fin fragments were collected from each fish under benzocaine anesthesia and all efforts were made to minimize suffering.

2.2 Experimental population, DNA extraction and SNP analysis

Genetic variability and kinship evaluation were performed through fins sampling from 139 individuals of seven different fish farms (FF) in Sao Paulo state (Brazil) (Table 1; Figure 1). The animals were individually marked with transponders (pit-tags - passive integrated transponder tag, model full-duplex FDX-B, 134.2-kHz) and kept alive in the fish farm stations for subsequent management and genetic selection of breeders for breeding programs. The identity and exact location of the fish farm stations were preserved. All fin samples were stored in absolut ethanol at -20°C.

DNA was extracted from fin fragments using the Wizard Genomic DNA Purification Kit (Promega), according the manufacturer's protocol. The DNA concentration was quantified using the Qubit dsDNA BR Assay kit (Life Technologies) and measured in the Qubit 2.0 Fluorometer (Invitrogen) (concentration for genotyping = 10ng/µL). The genotyping analysis were performed through MassARRAY platform (Sequenom, San Diego, CA, USA), in CeGen (Genotyping National Center, Santiago de Compostela, Spain). We used 32 SNPs obtained through liver transcriptome sequencing of pacu (Mastrochirico-Filho *et al*., 2016) for genetic studies.

In the present study, for comparative analysis, we also included the results obtained from the samples collected in the River Paraná (WILD population), which were previously performed in Mastrochirico-Filho *et al*., (2016).

2.3 Data analysis

About intrapopulational analysis, observed (H_{obs}) and expected heterozygosity (H_{exp}) were calculated using Cervus 3.0.7 (Marshall *et al*., 1998). Minimum allele frequencies (*MAF*) and tests of deviation from Hardy-Weinberg equilibrium (HW) (p-value > 0.05) were performed using Genepop 4.0.11 (Rousset, 2008). Significant differences hypothesis between *MAF*, H_{obs} and H_{exp} means of populations were tested through ANOVA tests (p-value < 0.05). Conformance to Hardy-Weinberg equilibrium was checked using the complete enumeration method (Louis and Dempster, 1987) because only two alleles were identified at each locus. F_{is} parameters were estimated using Weir and Cockerham (1984) approach through FSTAT 2.9.3.2 (Goudet, 1995). Bonferroni correction was performed because multiple tests were done (Rice, 1989).

To estimate genetic differentiation between the stocks, global and pairwise F_{ST} were calculated with the Weir and Cockerham (1984) method using FSTAT version 2.9.3.2 (Goudet, 2001) and threshold index of genetic differentiation according to Wright (1951): low genetic differentiation: <0.05; moderate genetic differentiation: 0.05 – 0.25; higher genetic differentiation: >0.25. The partitioning of variation at different levels was calculated by Analysis

of Molecular Variance (AMOVA) in ARLEQUIN version 3.5.2.2, using 1,000 permutations (Excoffier *et al*., 2005). Groups to perform the AMOVA analysis were formed firstly by the isolation of the wild population in relation to the cultivated populations, and sequentially cultivated populations were isolated according to the nearest hydrographic basin (Group 1: WILD sample; Group 2: FF3; Group 3: FF1; Group 4: FF2, FF4, FF5, FF6, FF7), according to the Fig. 1. 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 program STRUCTURE version 2.3.4 (Pritchard *et al*., 2000) without prior information about population. Primarily, we determined the distribution of ΔK, an *ad hoc* statistics 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 10. The analysis was performed in 45 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 STRUCTURE HARVESTER program.

Wang estimator (r) (Wang, 2002) was calculated by the program SPAGeDI 1.3 (Hardy and Vakemans, 2002) to evaluate the pairwise relatedness between individuals of each farmed station, in order to assess the applicability of SNPs in genetic monitoring of stocks in pre-breeding programs. Threshold of r coefficient values were determined as lower coefficients ($r < 0.125$) corresponding to no parentage relationship, which the animals can be crossed without risk of inbreeding; coefficient values between $0.126 \leq r \leq 0.375$ were considered as half-sib individuals and higher coefficients ($r > 0.376$) were considered as full-sib individuals, of which are not interesting for breeding programs due to inbreeding risks.

3. Results

Population parameters which resulted from analyses of genetic variability in both natural and cultivated populations of pacu are showed in Table 2.

Monomorphic SNPs were found in some populations, being more

common in FF4 (6 monomorphic SNPs) and FF6 (4 monomorphic SNPs) stations, as opposed to WILD, FF2, FF5 and FF7 stations which had all polymorphic SNPs and presented the largest *MAF* means (0.275±0.132, 0.285±0.134, 0.273±0.127, 0.270±0.140, respectively). Finally, stations which presented the lowest *MAF* means were FF1 (0.252±0.150) and FF3 (0.253±0.159). Despite such variations, no significant difference was found between average *MAF* values (p-value < 0.05) in all populations.

Expected (H_{exp}) and observed heterozygosity (H_{obs}) means showed higher values in FF5 (H_{obs} = 0.398±0.177; H_{exp} = 0.376±0.131), FF2 (H_{obs} = 0.394±0.194; H_{exp} = 0.382±0.131) and in WILD (H_{obs} = 0.385±0.171; H_{exp} = 0.370±0.130). Inversely, FF7 and FF4 revealed lower values of H_{obs} (0.342±0.150 and 0.349±0.230, respectively) which not even exceeded the H_{exp} values (0.362±0.147 and 0.353±0.195, respectively). Despite such differences, when H_{obs} and H_{exp} were compared in each population, no significant difference was found (p < 0.05). ANOVA tests comparing H_{obs} and H_{exp} values among all populations were performed and also showed no significant differences (p-value < 0.05). The F_{is} results showed predominance of negative values, although these values were closer to zero. After performing the Bonferroni correction (p = 0.0015), all loci were in accordance with Hardy-Weinberg equilibrium. Additionally, when samples were submitted to global test for all *loci*, all populations were in accordance with Hardy-Weinberg expectations (p > 0.05).

Population differentiation was obtained through global and pairwise F_{st} parameters. The global F_{st} value was 0.064 and suggested moderate genetic differentiation among populations. Population pairwise F_{st} values were calculated for the eight pairs of populations and significant differentiation was found between all population pairs (p < 0.05) (Table 3). Overall, we did not observe high genetic differentiation between all pair of populations. Nevertheless, the populations which had higher genetic differentiation were between FF4 and FF6 (F_{st} = 0.146), followed by FF1 and FF6 (F_{st} = 0.136) and FF3 and FF4 (F_{st} = 0.104). Inversely, WILD and FF7 registered the lowest genetic differentiation (F_{st} = 0.032), followed by WILD and FF5 (F_{st} = 0.036), FF2 and FF7 (F_{st} = 0.041) and FF4 and FF5 (F_{st} = 0.042). The low genetic structure between populations was confirmed through AMOVA (values close to

zero), which showed a negative percentage of variation (-1.49%) among groups (according to the hydrographic basin) and a slightly positive variance component (7.9%) among populations within groups. Otherwise, 93.59% of the genetic variation was contained within populations.

To evaluate the level of admixture among population samples, the model-based clustering analyses were performed based on ΔK distribution, indicating that the K =2 parameter was the most likely to explain the population structure of pacu stocks (Figure 2). Therefore, results showed a clustering of populations in two groups composed by FF6 station in an isolated group; another group formed by FF1, FF4 and FF5 stations; and remaining populations (WILD, FF2, FF3, FF7) seemed to be an admixture between these two clusters (Figure 3). The analysis confirmed the pairwise F_{st} analysis, with moderate genetic composition of FF6 compared to the others fish farms and admixture of individuals with similar genetic composition when compared FF4 and FF5; FF2 and FF7 and WILD and FF7.

In order to assess the applicability of SNPs in directed matings for future breeding programs, Wang coefficient (r) was used to evaluate the kinship in individuals of the brood stock in each fish farm sampled. The results are showed in Figures 4-10. Each figure shows the Wang r coefficient values obtained among individuals of each population, where the highest ones corresponds to full-sib individuals presented in red colors ($r > 0.376$), half-sib individuals presented such yellow colors ($0.126 \leq r \leq 0.375$) and unrelated individuals are showed in green colors ($r < 0.125$). Overall, the results showed that the majority of fish farms had about 40% of related individuals (full-sib and half-sib individuals). We found the higher percentage (68%) of unrelated individuals in FF7 with a consequent lower proportion of full-sib individuals (13%), whereas FF2 showed the higher proportion of full-sib individuals (29%), but the percentage of unrelated individuals was also high (57%).

4. Discussion

The lack of genetic diversity in cultivated stocks may lead to increased susceptibility to diseases and deformities, reduction of the adaptability of

organisms to new environments and consequent reduction of productivity (Kincaid, 1983). In the present study, general parameters of genetic variability, such as heterozygosity values (H_{obs} and H_{exp}), were similar to other studies about genetic diversity of fish through SNP approach (Vera *et al*., 2013; Aykanat *et al*., 2015; Pocwierz-Kotus *et al*., 2015), which ranged from 0.04 to 0.882 and 0.04 to 0.520, respectively. The results showed no significant differences between H_{obs} and H_{exp} and *MAF* averages when populations were compared ($P<0.05$). In general, heterozygosity values indicated high genetic variability in all farmed populations studied, take into consideration that wild populations of pacu have high heterozygosity values, as described by Calcagnotto and DeSalle (2009) through microsatellite loci, and the absence of differences in genetic diversity of SNP parameters between WILD and cultivated populations. Our data suggest that it is possible to create a population basis for breeding program from cultivated individuals of fish farms that showed high genetic variability, *i.e.*, not necessarily only from wild stocks. Therefore, it is important to monitor the genetic variability of the stocks and to assess how this variation can be maintained through directed matings (Gjedren and Baranski, 2009; Portela and Huerta, 2007; Beaumont and Hoare, 2003).

The loss of genetic variability in natural populations may be due to ecological and evolutionary factors, such inbreeding, environmental changes and genetic drift (Primack and Rodrigues, 2001). In cultivated stocks, aquaculture have the tendency to reduce genetic variability due detrimental practices (Wang *et al*., 2012), such as artificial selection and acquisition of reduced number of breeders in initial population base. In the present study, we used SNPs gene-associated to assess the genetic variability, therefore, they are more susceptible to selection pressure and, consequently, to allele and heterozygosis reduction (Sun *et al*., 2014). However, all the loci of the populations were in accordance with Hardy-Weinberg equilibrium, and factors such as selection and genetic drift had no influence on stocks. In addition, all loci F_{is} values were predominantly negative, with exception in FF4 (F_{is} = 0.008) and FF7 (F_{is} = 0.056) stocks, but all values were not significantly different to zero (Table 2).

In relation to the analysis of genetic structure, when WILD was compared to FF7 (F_{st} = 0.032) and FF5 (F_{st} = 0.036), it was demonstrated a low genetic structure, as well as between FF2 and FF7 (F_{st} = 0.041) and FF4 and FF5 (F_{st} = 0.42). The low genetic differentiation between populations was confirmed through AMOVA, which did not corroborate with the hypothesis that variation could occur among groups (according to the hydrographic basin). This genetic similarity may be related originally to the lack of genetic structure present in natural populations of pacu, already registered in previous studies (Iervolino *et al*., 2010; Calcagnotto and DeSalle, 2009), because this fish has high gene flow capacity due to their migratory behavior. Additionally, genetic similarity can be explained by sharing of stocks conducted between fish farms, since producers are currently not investing in brood stock samples obtained from the natural environment because of the high level of capital required for such an investment (Hashimoto *et al*., 2012). Thus, there is a trend towards the continuous use of brood stocks composed by specimens from other fish farms, which generally provide samples without pedigree. The higher proportion of genetic variation herein observed was present within the populations (93.59%), which has also been reported in other pacu analyses (Calcagnotto and DeSalle, 2009; Iervolino *et al*., 2010).

According to our results, although there is similarity of genetic diversity using SNP parameters (minimum allele frequencies and heterozygosity), overall F_{st} estimated a genetic differentiation of 0.064, suggesting moderate structuring between populations. This information is fundamental for breeding programs, because the differences of genetic structure should be taken into consideration to compose the initial population that will be used to create the families. Moreover, according to the results of pairwise F_{st} analysis (Table 3), we observed moderate structuration of FF6 in relation to the other stocks, which was confirmed on Structure analysis (Figure 3), because this fish farm make up a distinct genetic group. This result may be explained due to breeding management practices already carried out in FF6, since this producer has been conducting directed matings to obtain genetically superior individuals.

Therefore, inbred individuals can be obtained between mating of genetically superior individuals and could affect the production of fish without

management techniques and directed mating among individuals (Gjedrem, 2005). Kinship analysis appears as an essential tool in genetic pre-breeding programs of fish, because it can minimize inbreeding rates by directing the mating of unrelated individuals (Pino-Querido *et al*., 2010). With the exception of FF7, which obtained 68% of unrelated individuals, all farm stations showed almost a minimum of 40% of kinship individuals (half + full-sib) and may result in inbreeding risk, which results in several skeletal deformities and low viability of the offspring (Su *et al*., 1996; Imsland *et al*., 2001). Therefore, the results of this research will serve as a strategy to improve the production of pacu in Brazil, and as direct subsidies for pre-breeding programs in order to minimize the risk of inbreeding and suggest a directed mating program.

This study aimed to provide initial knowledge about the genetic profile of pacu stocks in different fish farms, considering its importance to Neotropical aquaculture and the necessity to offer subsidies for the development of its production. In conclusion, this SNP set showed the applicability of these genomic markers in pre-breeding programs, particularly to delineate the formation of the best families in terms of genetic variability and genetic structure, without inbreeding risks.

5. References

Abadía-Cardozo A., Anderson E.C., Pearse D.E., Garza J.C. 2013. Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead. *Molecular Ecology* 22, 4733-4746.

Allendorf, F.W., Phelps, S.R., 1980. Loss of genetic variation in a hatchery stock of cutthroat trout. *Transactions of the American Fisheries Society* 109, 537-543.

An, H.S., Nam, M.M., Myeong, J.I., An, C.M., 2014. Genetic diversity and differentiation of the Korean starry flounder (*Platichthys stellatus*) between and within cultured stocks and wild populations inferred from microsatellite DNA analysis. *Molecular Biology Reports* 41(11), 7281-7292.

Aykanat, T., Johnston, S.E., Orell, P., Niemela, E., Erkinaro, J., 2015. Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population. *Molecular Ecology* 24(20), 5158-5174).

Beaumont, A.R., Hoare, K., 2003. Genetic considerations in the hatchery. In: Beaumont, A.R., Hoare, K. (Eds.), Biotechnology and Genetics in Fisheries and Aquaculture. Blackwell Science, Oxford, pp.73-90.

Calcagnotto, D.; DeSalle, R., 2009. Population genetic structuring in pacu (Piaractus mesopotamicus) across the Paraná-Paraguay basin: evidence from microsatellites. *Neotropical Ichthyology* 7(4), 607-616.

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, 245-247.

Estoup, A., Gharbi, K., Sancristobal, M., Chevalet, C., Haffray, P. et al. 1998. Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Sciences* 55(3), 715-725.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611-2620.

Excoffier, L., Laval, G., Schneider, S., 2005. ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47-50

FAO., 2015. *Global Aquaculture Production database updated to 2013 – Summary information*. FAO: Rome, Italy.

FAO., 2014. *The State of World Fisheries and Aquaculture 2014*. FAO: Rome, Italy, 223 pp.

FAO., 2010. *The State of World Fisheries and Aquaculture 2010*. FAO: Rome, Italy, 197 pp.

Flores Nava, A. 2007. Aquaculture seed resources in Latin America: a regional synthesis. In: Assessment of freshwater fish seed resources for sustainable aquaculture*.* FAO Fisheries Technical Paper No 501*;* Bondad-Reantaso, M.G., Ed.; FAO: Rome, Italy, 628 pp.

Gabriel, S.; Ziaugra, L., 2004. SNP genotyping using Sequenom MassARRAY 7K Platform. *Current Protocols in Human Genetics* 42:2.12, 2.12.1–2.12.16.

Gjedrem, T., 2005. Status and scope of aquaculture, in: Gjedrem, T. (Eds.), Selection and breeding programs in Aquaculture. Springer, pp.1-8.

Gjedren, T.; Robinson, N.; Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 350-353, 117-129.

Gjedren, T.; Baranski, M., 2009. *Selective Breeding in Aquaculture: An Introduction*. Springer Science & Business Media. 221 pp.

Glaubitz, J.C., Rhodes, O.E., Dewoody, J.A., 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology* 12, 1039-1047.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3), 485-6.

Hardy, O.J., Vakemans, X., 2002. SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2, 618–620.

Hashimoto, D. T., Senhorini, J. A., 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.

Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A. et al. 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245-256.

Honglang, H., 2007. Freshwater fish seed resources in China. In: *Assessment of freshwater fish seed resources for sustainable aquaculture*, FAO Fisheries Technical Paper No 501 Bondad-Reantaso, M.G., Eds.; FAO: Rome, Italy, 628 pp.

IBGE. Instituto Brasileiro de Geografia e Estatística, 2014. *Produção da Pecuária Municipal 2013, Vol 41*. Rio de Janeiro, Brasil, 1-108.

Iervolino, F.; Resende, E. K.; Hilsdorf, A.W.S., 2010. The lack of genetic differentiation of pacu (Piaractus mesopotamicus) populations in the Upper-Paraguay Basin revealed by the mitochondrial DNA D-loop region: Implications for fishery management. *Fisheries Research* 101, 27-31.

Imsland, A.K., Foss, A., Naevdal, G., Stefansson, S.O., 2001. Selection or adaptation: differences in growth performance of juvenile turbot (*Scophthalmus maximus* Rafinesque) from two close-by localities of Norway. Sarsia 86, 43-51.

Jiang, L.H., Chen, Y.J., Zhang, J.S., Zhu, J.S., Wu, C.W., 2015. Population structure of large yellow croacker (*Larimichthys crocea*) revealed by single nucleotide polymorphisms. *Biochemical Systematics and Ecology* 63, 136-142.

Kincaid, H.L., 1983. Inbreeding in fish populations used for aquaculture. *Aquaculture* 33, 215-227.

Kong, L., Bai, J., Li, Q., 2014. Comparative assessment of genomic SSR, EST-SSR and EST-SNP markers for evaluation of the genetic diversity of wild and

cultured Pacific oyster, *Crassostrea gigas* Thunberg. *Aquaculture* 420-421, 585-591.

Laconcha, U., Iriondo, M., Arrizabalaga, H., Manzano, C., Markaide, P., et al., 2015. New nuclear SNP markers unravel the genetic structure and efetive population size of albacore tuna (*Thunnus alalunga*). *Plos One* 10(6), e0128247.

Landegren, U., Nilsson, M., Kwok, P., 1998. Reading bits of genetic information: Methods for single-nucleotide polymorphism analysis. *Genome Research* 8, 769-776.

Lind, C.E.; Ponzoni, R.W.; Nguyen, N.H.; Khaw, H.L., 2012. Selective breeding in fish and conservation of genetic resources for aquaculture. *Reproduction in Domestic Animals* 47(4), 255-263.

Liu S., Palti Y, Gao G, Rexroad III C.E. 2016. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture* 452, 178-182.

Liu, Z., 2011. *Next Generation Sequencing and Whole Genome Selection in Aquaculture*; Wiley-Blackwell: Iowa, USA; 221 pp.

Liu, Z.J., Cordes, J.F., 2004. DNA marker Technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1-37.

Louis EJ, Dempster ER., 1987. An exact test for Hardy-Weinberg and multiple alleles. Biometrics 43, 805-811.

Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7, 639-655.

Mastrochirico-Filho V.A., Hata M.E., Sato L.S, Jorge P.H., Foresti F. et al. 2016. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conservation Genetic Resources*, doi 10.1007/s12686-016-0521-3.

Meistertzheim, A., Arnaud-Haond, S., Boudry, P., Thébault M., 2013. Genetic structure of wild European populations of the invasive Pacific oyster *Crassostrea gigas* due to aquaculture pratices. *Marine Biology* 160(2), 453-463.

Meng, X.H., Wang, Q.Y., Jang, I.K., Liu, P., Kong, J., 2009. Genetic differentiation in seven geographic populations of the fleshy shrimp *Penaeus* (*Fenneropenaeus*) *chinensis* based on microsatellite DNA. *Aquaculture* 287, 46-51.

Morzeven, R., Charrier, G., Boudry, P., Chauvaud, L., Breton, F., et al, 2016. Genetic structure of a commercially exploited bivalve, the great scallop *Pecten maximus*, along the European coasts. *Conservation Genetics* 17(1), 57-67.

Pan, G., Yang, J., 2010. Analysis of microsatellite DNA markers reveals no genetic differentiation between wild and hatchery populations of pacific threadfin in Hawaii. *International Journal of Biological Sciences* 6(7), 827-833.

Pardo B.G., Hermida M., Fernández C., Bouza C., Sánchez L., Martínez P. 2006. A set of highly polymorphic microsatellites useful for kinship evaluation and population analysis in turbot (*Scophthalmus maximus*). *Aquaculture Research* 37, 1578-1582.

Pierce, B.A., 2009. *Genetics: a conceptual approach*; W.H. Freeman and Co: New York and Basingstoke, 774 pp.

Pino-Querido, A., Hermida, M., Vilariño, M., Bouza, C., Martínez, P., 2010. Statistical properties and performance of pairwise relatedness estimators using turbot (*Scophthalmus maximus* L.) family data. *Aquaculture Research* 41, 528-534.

Pocwierz-Kotus, A., Bernas, R., Kent, M.P., Lien, S., Leliuna, E., 2015. Restitution and genetic differentiation of salmon populations in the southern Baltic genotyped with the Atlantic salmon 7K SNP array. *Genetics Selection Evolution* 47,39.

Ponzoni, R.W., 2006. Genetic improvement effective dissemination: Keys to prosperous and sustainable aquaculture industries. In: *Development of Aquatic Animal Genetic Improvement and Dissemination Programs: current status and action plans*; Ponzoni, R.W.; Acosta, B.O.; Ponniah, A.G., Eds.; WorldFish Center: Penang, Malásia, 114 pp.

Portela, P.M., Huerta, A.F. 2007. Genética y Genómica em Acuicultura. Serie: Publicaciones científicas y tecnológicas del Observatorio Español de Acuicultura. Ministerio de Agricultura, Pesca Y Alimentación. 889 pp.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., et al., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2, 225-238.

Primack, R.B., Rodrigues, E. 2001. Biologia da conservação. Editora Rodrigues, Londrina.

Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-949.

Resende, E.K., 2003. Migratory fishes of the Paraguay-Paraná basin excluding the Upper Paraná River. In: *Migratory fishes of South America: biology, fisheries and conservation states*; Carolsfeld, J., Harvey, B., Ross, C., Baers, A., Eds.; World Bank: Victoria, Canada, pp.99-156.

Rice WR., 1989. Analyzing tables of statistical tests. Evolution 43, 223-225.

Rousset F., 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8, 103-106.

Sellars M.J., Dierens L., McWillian S., Little B, Murphy B., et al. 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research* 45, 417-426.

Senanam, W., Pechsiri, J., Sonkaew, S., Na-Nakorn, U., Sean-In, N., Yashiro, W., 2015. Genetic relatedness and differentiation of hatchery populations of Asian seabass (*Lates calcarifer*) (Bloch, 1790) broodstock in Thailand inferred from microsatellite genetic markers. *Aquaculture Research* 46, 2897-2912.

Shikano, T., Jarvinen, A., Marjamaki, P., Kahilainen, K.K., Merila, J. 2015. Genetic variability and structuring of Arctic charr (*Salvelinus alpinus*) populations in Northern Fennoscandia. *Plos One* 10(10), e0140344.

Su, G.S., Liljedahl, L.E., Gall, G.A.E., 1996. Effects of inbreeding on growth and reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 142, 139-148.

Sun, L., Liu, S., Wang, R., Jiang, Y., Zhang, Y., et al., 2014. Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). *Plos One*, e109666.

Vera, M., Alvarez-Dios, J.A., Fernandez, C., Bouza, C., Vilas, R., Martinez, P., 2013. Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *International Journal of Molecular Sciences,* 14, 5694-5711.

Wang, L., Shi, S., Su, Y., Meng, Z., Lin, H., 2012 Loss of genetic diversity in the cultured stocks of the large yellow croaker, *Larimichthys crocea*, revealed by microsatellites. *International Journal of Molecular Sciences* 13(5), 5584-5597.

Wang J., 2002. An estimator for pairwise relatedness using molecular markers. Genetics 160, 1203-1215.

Weir BS, Cockerhsam CC., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358-1370.

Wright S. 1951. The genetical structure of populations. *Annals of Human Genetics* 15: 323-354.

Table List

Table 1. Sample fins collected at different stations for genetic variability analysis in pacu (*Piaractus mesopotamicus*). WILD corresponds to 34 samples collected in Parana river (Mastrochirico-Filho *et al*., 2016). FF1 – FF7 corresponds to fish farm sampled.

Sample populations	Number of samples	Region (Basin)
WILD	34	Parana
FF1	22	Tietê
FF2	23	Grande
FF3	19	Peixe
FF4	13	Grande
FF5	19	Grande
FF6	17	Grande
FF7	26	Grande
Total	173	4

Table 2. Genetic variability parameters of the 32 SNPs in loci of pacu populations (*Piaractus mesopotamicus*). WILD: Parana river; FF1 – FF7: farm fish stations. *MAF*: minimum allele frequency; H_o: observed heterozygosity; H_e: expected heterozygosity; *P(HW)*: Hardy Weinberg p-value; F_{is}: inbreeding coefficient. Overrall $F_{is} = F_{is}$ over all loci. No inf = No information available.

	WILD					FF1					FF2					FF3				
Locus	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}
C1013_445	0.426	0.382	0.496	0.289	0.233	0.429	0.500	0.492	1.000	0.052	0.435	0.696	0.502	0.093	-0.397	0.361	0.500	0.475	1.000	-0.055
C128_1801	0.118	0.059	0.211	0.001	0.723	0.000	0.000	0.000	0.000	0.000	0.044	0.087	0.085	1.000	-0.023	0.210	0.316	0.341	1.000	0.077
C1459_108	0.344	0.688	0.458	0.004	-0.512	0.350	0.632	0.478	0.327	-0.295	0.326	0.652	0.449	0.050	-0.467	0.267	0.533	0.405	0.507	-0.333
C147_351	0.162	0.324	0.275	0.562	-0.179	0.250	0.381	0.372	1.000	-0.068	0.326	0.652	0.449	0.050	-0.467	0.139	0.278	0.246	1.000	-0.133
C178_243	0.162	0.324	0.275	0.562	-0.179	0.432	0.619	0.508	0.662	-0.182	0.456	0.739	0.507	0.038	-0.472	0.056	0.111	0.108	1.000	-0.030
C191_480	0.353	0.471	0.464	1.000	-0.015	0.364	0.524	0.470	0.650	-0.156	0.413	0.217	0.496	0.010	0.567	0.237	0.368	0.371	1.000	0.008
C213_629	0.471	0.647	0.506	0.165	-0.285	0.455	0.524	0.508	1.000	-0.077	0.413	0.565	0.496	0.670	-0.144	0.444	0.778	0.508	0.050	-0.556
C238_1041	0.044	0.088	0.086	1.000	-0.031	0.023	0.048	0.048	No inf	No inf	0.478	0.435	0.510	0.676	0.151	0.444	0.444	0.508	0.656	0.128
C239_1594	0.338	0.500	0.454	0.706	-0.102	0.454	0.571	0.502	1.000	-0.077	0.196	0.304	0.322	1.000	0.055	0.417	0.389	0.500	0.379	0.227
C240_1549	0.206	0.294	0.332	0.599	0.115	0.318	0.571	0.455	0.361	-0.235	0.217	0.435	0.348	0.538	-0.257	0.361	0.722	0.475	0.039	-0.545
C260_818	0.471	0.706	0.506	0.036	-0.404	0.386	0.571	0.483	0.387	-0.224	0.435	0.435	0.502	0.674	0.137	0.500	0.556	0.514	1.000	-0.083
C271_399	0.029	0.059	0.058	1.000	-0.015	0.046	0.095	0.093	1.000	-0.024	0.304	0.609	0.433	0.063	-0.420	0.111	0.222	0.203	1.000	-0.097
C30_132	0.176	0.235	0.295	0.249	0.205	0.114	0.143	0.215	0.224	0.344	0.239	0.217	0.372	0.071	0.421	0.028	0.056	0.056	No inf	No inf
C348_245	0.485	0.500	0.507	1.000	0.014	0.295	0.381	0.418	1.000	0.041	0.196	0.391	0.322	0.544	-0.222	0.250	0.500	0.386	0.524	-0.308
C379_275	0.265	0.412	0.395	1.000	-0.043	0.204	0.286	0.316	1.000	0.045	0.370	0.217	0.476	0.020	0.549	0.028	0.056	0.056	No inf	No inf
C391_875	0.426	0.618	0.496	0.177	-0.249	0.273	0.571	0.418	0.143	-0.355	0.174	0.348	0.294	1.000	-0.189	0.368	0.421	0.478	0.648	0.122
C417_302	0.191	0.324	0.314	1.000	-0.031	0.204	0.333	0.345	1.000	0.045	0.130	0.261	0.232	1.000	-0.128	0.167	0.222	0.286	0.390	0.227
C41_428	0.235	0.412	0.365	0.647	-0.130	0.400	0.474	0.491	1.000	-0.016	0.370	0.565	0.476	0.652	-0.192	0.237	0.368	0.371	1.000	0.008
C437_455	0.132	0.265	0.233	1.000	-0.138	0.000	0.000	0.000	0.000	0.000	0.044	0.000	0.085	0.022	1.000	0.000	0.000	0.000	0.000	0.000
C43_831	0.309	0.382	0.433	0.686	0.119	0.068	0.143	0.136	1.000	-0.050	0.196	0.304	0.322	1.000	0.055	0.132	0.263	0.235	1.000	-0.125
C455_315	0.368	0.500	0.472	1.000	-0.060	0.386	0.571	0.483	0.387	-0.224	0.456	0.391	0.507	0.401	0.233	0.250	0.389	0.386	1.000	-0.008
C458_2209	0.044	0.088	0.086	1.000	-0.031	0.114	0.238	0.215	1.000	-0.105	0.204	0.227	0.333	0.179	0.323	0.000	0.000	0.000	0.000	0.000
C470_159	0.353	0.471	0.464	1.000	-0.015	0.225	0.421	0.341	0.528	-0.267	0.370	0.739	0.476	0.008	-0.571	0.474	0.421	0.512	0.645	0.182
C4_231	0.324	0.471	0.444	1.000	-0.060	0.250	0.190	0.372	0.081	0.413	0.391	0.435	0.487	0.676	0.109	0.222	0.444	0.356	0.529	-0.259
C564_1273	0.273	0.485	0.403	0.383	-0.208	0.409	0.524	0.494	0.682	-0.105	0.391	0.522	0.487	1.000	-0.073	0.395	0.579	0.491	0.633	-0.186
C579_153	0.235	0.471	0.365	0.152	-0.294	0.136	0.190	0.251	0.324	0.250	0.370	0.565	0.476	0.652	-0.192	0.500	0.444	0.514	0.654	0.139
C585_507	0.485	0.324	0.507	0.043	0.365	0.227	0.429	0.345	0.538	-0.273	0.429	0.286	0.502	0.074	0.436	0.132	0.263	0.235	1.000	-0.125
C5_660	0.338	0.559	0.454	0.254	-0.234	0.295	0.286	0.418	0.311	0.258	0.250	0.409	0.384	1.000	-0.068	0.263	0.421	0.398	1.000	-0.058
C627_936	0.397	0.382	0.486	0.285	0.216	0.139	0.235	0.214	1.000	-0.133	0.087	0.174	0.162	1.000	-0.073	0.472	0.167	0.513	0.005	0.681
C83_761	0.191	0.265	0.314	0.570	0.159	0.023	0.048	0.048	No inf	No inf	0.087	0.174	0.162	1.000	-0.073	0.111	0.222	0.203	1.000	-0.097
C857_201	0.279	0.324	0.409	0.233	0.211	0.341	0.476	0.455	1.000	-0.090	0.130	0.261	0.232	1.000	-0.128	0.111	0.222	0.203	1.000	-0.097
C87_1726	0.176	0.294	0.295	1.000	0.003	0.454	0.476	0.511	0.684	0.106	0.196	0.304	0.322	1.000	0.055	0.421	0.421	0.501	0.643	0.163
Overall				0.293	-0.040				0.998	-0.064				0.060	-0.034				0.983	-0.025
Mean	0.275	0.385	0.370			0.252	0.358	0.340			0.285	0.394	0.382			0.253	0.347	0.338		

Table 2. *continuation*...

	FF4					FF5					FF6					FF7				
Locus	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}	*MAF*	H_o	H_e	*P(HW)*	F_{is}
C1013_445	0.192	0.385	0.323	1.000	-0.200	0.368	0.526	0.478	1.000	-0.104	0.464	0.357	0.516	0.316	0.316	0.269	0.385	0.401	1.000	0.042
C128_1801	0.208	0.417	0.344	1.000	-0.222	0.237	0.474	0.371	0.525	-0.285	0.471	0.471	0.513	1.000	0.086	0.080	0.160	0.150	1.000	-0.067
C1459_108	0.188	0.375	0.325	1.000	-0.167	0.318	0.636	0.455	0.479	-0.428	0.294	0.588	0.428	0.241	-0.391	0.267	0.533	0.405	0.507	-0.333
C147_351	0.000	0.000	0.000	0.000	0.000	0.132	0.263	0.235	1.000	-0.125	0.441	0.882	0.508	0.003	-0.778	0.269	0.538	0.401	0.134	-0.351
C178_243	0.000	0.000	0.000	0.000	0.000	0.132	0.263	0.235	1.000	-0.125	0.324	0.529	0.451	0.608	-0.180	0.212	0.346	0.340	1.000	-0.018
C191_480	0.292	0.417	0.431	1.000	0.035	0.421	0.526	0.501	1.000	-0.053	0.029	0.059	0.059	No inf	No inf	0.300	0.360	0.429	0.631	0.163
C213_629	0.385	0.308	0.492	0.256	0.385	0.368	0.526	0.478	1.000	-0.104	0.500	0.294	0.515	0.141	0.436	0.365	0.269	0.473	0.038	0.435
C238_1041	0.000	0.000	0.000	0.000	0.000	0.105	0.211	0.193	1.000	-0.091	0.471	0.706	0.513	0.162	-0.391	0.269	0.462	0.401	0.631	-0.154
C239_1594	0.423	0.385	0.508	0.574	0.250	0.342	0.368	0.462	0.607	0.208	0.265	0.529	0.401	0.276	-0.333	0.500	0.538	0.510	1.000	-0.057
C240_1549	0.423	0.231	0.508	0.085	0.556	0.395	0.368	0.491	0.353	0.254	0.382	0.529	0.487	1.000	-0.091	0.212	0.269	0.340	0.288	0.212
C260_818	0.423	0.846	0.508	0.023	-0.714	0.263	0.421	0.398	1.000	-0.058	0.324	0.294	0.451	0.260	0.355	0.479	0.458	0.510	0.695	0.103
C271_399	0.000	0.000	0.000	0.000	0.000	0.079	0.053	0.149	0.081	0.653	0.176	0.353	0.299	1.000	-0.185	0.038	0.077	0.075	1.000	-0.020
C30_132	0.500	0.385	0.520	0.578	0.268	0.263	0.316	0.398	0.550	0.212	0.235	0.471	0.371	0.520	-0.280	0.307	0.462	0.434	1.000	-0.064
C348_245	0.385	0.615	0.492	0.566	-0.263	0.290	0.579	0.422	0.253	-0.385	0.118	0.235	0.214	1.000	-0.103	0.288	0.500	0.419	0.629	-0.199
C379_275	0.423	0.538	0.508	1.000	-0.063	0.132	0.263	0.235	1.000	-0.125	0.382	0.529	0.487	1.000	-0.091	0.231	0.385	0.362	1.000	-0.064
C391_875	0.458	0.583	0.518	1.000	-0.132	0.263	0.316	0.398	0.550	0.212	0.357	0.429	0.476	1.000	0.103	0.300	0.360	0.429	0.631	0.163
C417_302	0.269	0.385	0.409	1.000	0.062	0.237	0.263	0.371	0.235	0.297	0.118	0.235	0.214	1.000	-0.103	0.077	0.154	0.145	1.000	-0.064
C41_428	0.417	0.167	0.507	0.029	0.681	0.447	0.684	0.508	0.176	-0.360	0.324	0.294	0.451	0.260	0.355	0.260	0.440	0.393	1.000	-0.123
C437_455	0.000	0.000	0.000	0.000	0.000	0.237	0.474	0.371	0.525	-0.286	0.000	0.000	0.000	0.000	0.000	0.140	0.280	0.246	1.000	-0.143
C43_831	0.250	0.500	0.391	0.529	-0.294	0.053	0.105	0.102	1.000	-0.029	0.118	0.235	0.214	1.000	-0.103	0.100	0.200	0.184	1.000	-0.091
C455_315	0.307	0.462	0.443	1.000	-0.043	0.316	0.526	0.444	0.607	-0.192	0.412	0.588	0.499	0.624	-0.185	0.307	0.385	0.434	0.653	0.117
C458_2209	0.038	0.077	0.077	No inf	No inf	0.079	0.158	0.149	1.000	-0.058	0.000	0.000	0.000	0.000	0.000	0.019	0.038	0.038	No inf	No inf
C470_159	0.375	0.417	0.489	1.000	0.154	0.447	0.263	0.508	0.060	0.489	0.206	0.412	0.337	1.000	-0.231	0.413	0.391	0.496	0.399	0.214
C4_231	0.462	0.769	0.517	0.112	-0.519	0.421	0.632	0.501	0.356	-0.271	0.265	0.412	0.401	1.000	-0.028	0.481	0.500	0.509	1.000	0.018
C564_1273	0.208	0.417	0.344	1.000	-0.222	0.447	0.579	0.508	0.655	-0.144	0.294	0.588	0.428	0.241	-0.391	0.479	0.292	0.510	0.046	0.433
C579_153	0.307	0.308	0.443	0.508	0.314	0.184	0.263	0.309	0.489	0.151	0.500	0.647	0.515	0.358	-0.266	0.327	0.269	0.449	0.069	0.405
C585_507	0.333	0.333	0.464	0.518	0.290	0.290	0.579	0.422	0.253	-0.385	0.235	0.353	0.371	1.000	0.050	0.460	0.360	0.507	0.226	0.294
C5_660	0.333	0.667	0.464	0.216	-0.468	0.421	0.632	0.501	0.356	-0.271	0.382	0.529	0.487	1.000	-0.091	0.480	0.400	0.509	0.422	0.218
C627_936	0.167	0.333	0.303	1.000	-0.111	0.263	0.421	0.398	1.000	-0.058	0.147	0.294	0.258	1.000	-0.143	0.360	0.480	0.470	1.000	-0.021
C83_761	0.000	0.000	0.000	0.000	0.000	0.053	0.105	0.102	1.000	-0.029	0.000	0.000	0.000	0.000	0.000	0.019	0.038	0.038	No inf	No inf
C857_201	0.307	0.462	0.443	1.000	-0.044	0.447	0.579	0.508	0.655	-0.144	0.088	0.059	0.166	0.091	0.652	0.104	0.125	0.191	0.206	0.349
C87_1726	0.458	0.417	0.518	0.594	0.203	0.290	0.368	0.422	0.607	0.131	0.000	0.000	0.000	0.000	0.000	0.240	0.480	0.372	0.276	-0.297
Overall				0.892	0.008				0.988	-0.062				0.787	-0.082				0.947	0.056
Mean	0.266	0.349	0.353			0.273	0.398	0.376			0.260	0.372	0.345			0.270	0.342	0.362		

Table 3. Pairwise F_{st} and global F_{st} estimates from 32 SNPs loci in the wild and farmed populations of pacu (*Piaractus mesopotamicus*). WILD: Parana river sample; FF1-FF7: farmed stations. The significance of population pairwise F_{st} and global F_{st} tested by 1000 permutations. All significant at $P < 0.05$. Threshold values of F_{st} : low genetic differentiation (0 – 0.05); moderate genetic differentiation (0.05 – 0.25); higher genetic differentiation (>0.25) (Wright, 1951)

	WILD	FF1	FF2	FF3	FF4	FF5	FF6	FF7	Global Fst
WILD	-								
FF1	0.051	-							
FF2	0.052	0.057	-						
FF3	0.070	0.095	0.085	-					
FF4	0.053	0.067	0.096	0.104	-				
FF5	0.036	0.054	0.081	0.068	0.042	-			
FF6	0.094	0.136	0.066	0.089	0.146	0.093	-		
FF7	0.032	0.051	0.041	0.063	0.092	0.052	0.067	-	0.064

Figure List

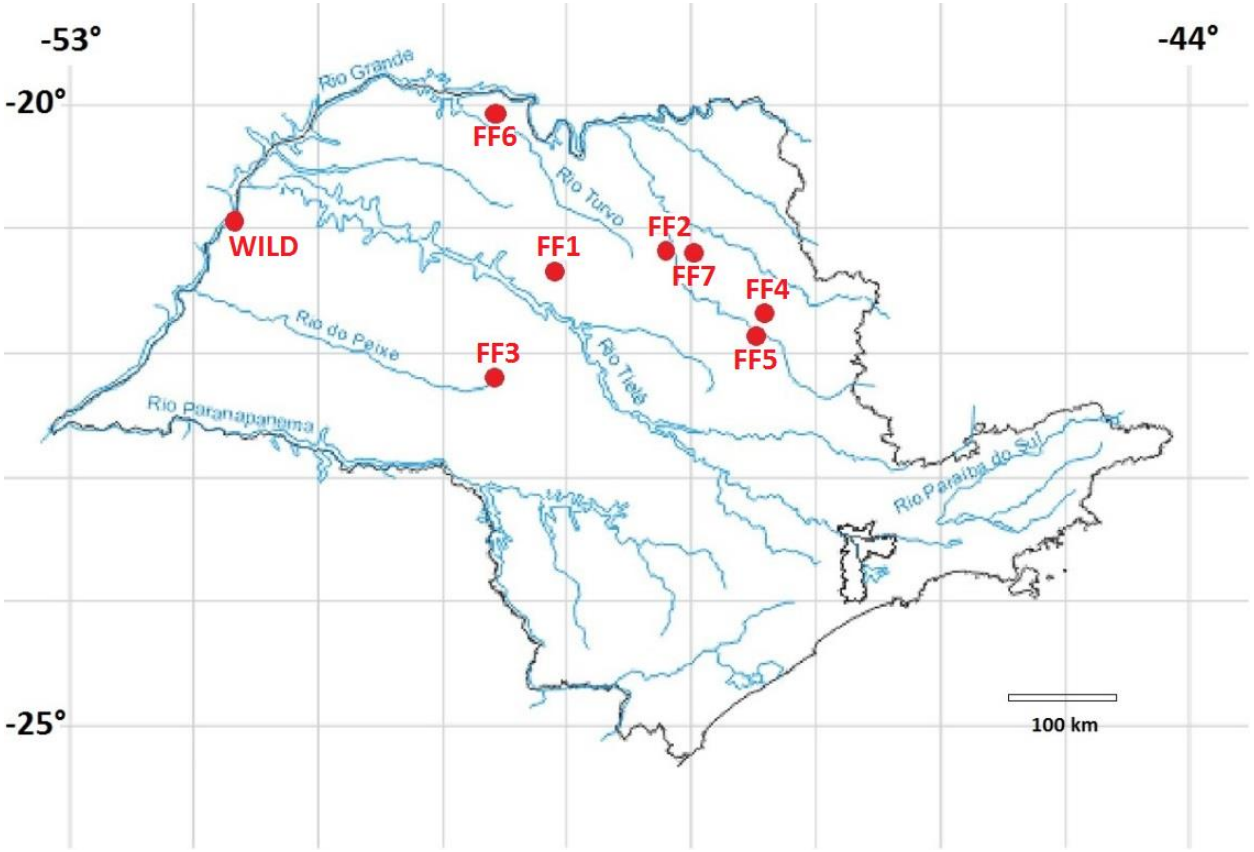

Figure 1. The geographical distribution of the seven fish farm stations (FF1 – FF7) and WILD population (Parana river) used for genetic analysis.

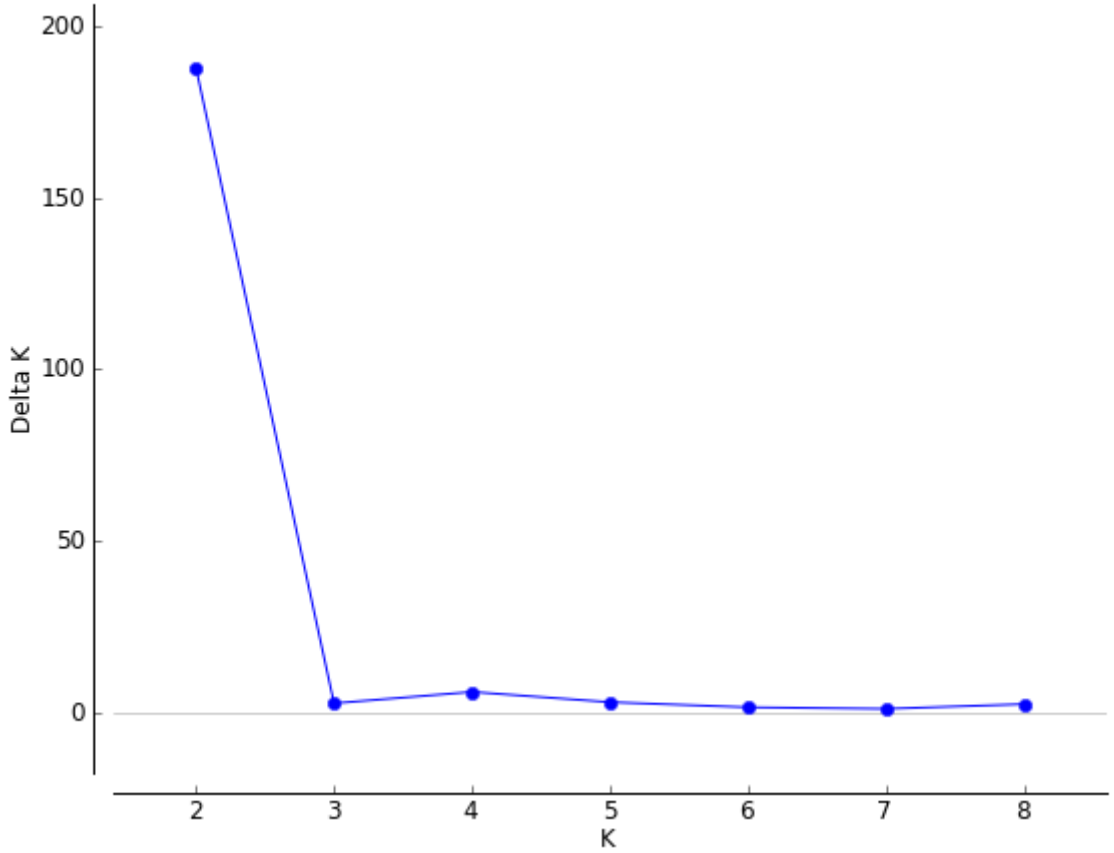

Figure 2. Magnitude of ΔK statistics in STRUCTURE analysis as a function of the number of putative genetic clusters (K) for pacu populations based on 32 SNPs loci.

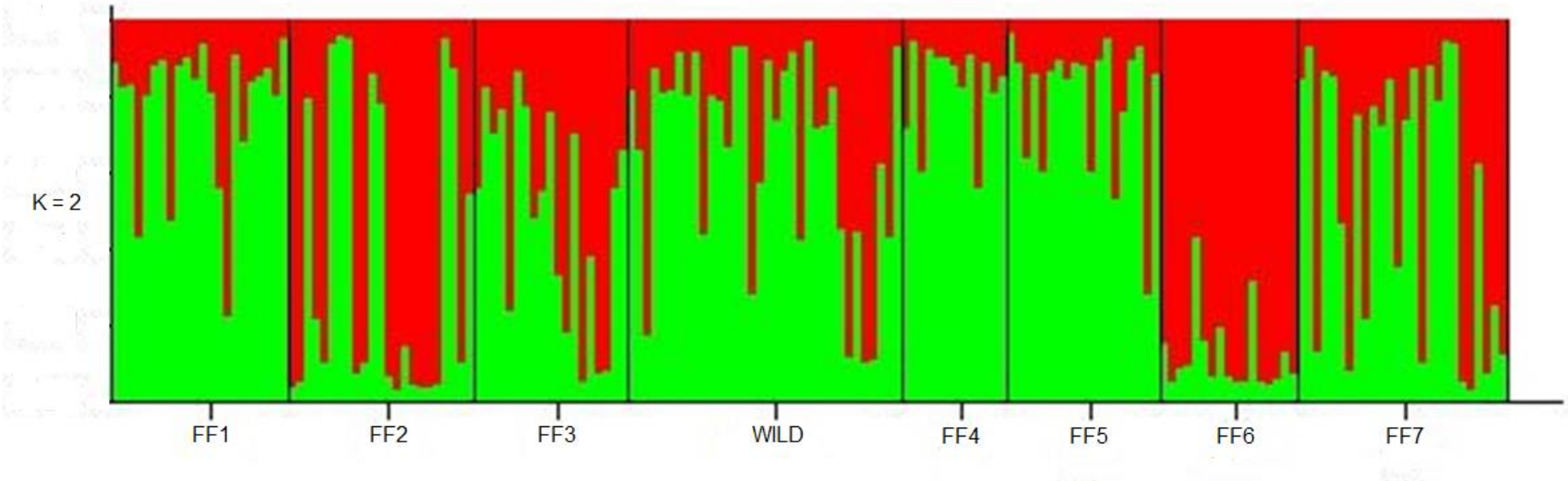

Figura 3. Coefficient of ancestry of pacu individuals collected from seven cultivated populations (FF1-FF7) and a wild population (WILD) estimated by the STRUCTURE program for K = 2 based on 32 SNPs loci. Each vertical bars represents an individual. The stations are separates by vertical black lines. The color proportions of each bar correspond to individuals estimated membership fractions of each of the clusters.

ALL LOCI	1136	1137	1138	1139	1140	1142	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1155	1156	1159	1160	1162	1164
1136																						
1137	0,5441																					
1138	0,4268	0,3660																				
1139	0,3925	0,2681	0,2989																			
1140	0,2029	0,7452	0,2874	0,1799																		
1142	0,5553	0,7436	0,2979	0,1648	0,6619																	
1144	0,0990	0,0000	0,1019	0,2255	0,0000	0,0000																
1145	0,0564	0,0019	0,3116	0,3045	0,0000	0,0060	0,0000															
1146	0,3815	0,4088	0,3133	0,0000	0,2401	0,3383	0,0000	0,0000														
1147	0,3984	0,3961	0,0455	0,3430	0,1406	0,2521	0,0000	0,0110	0,2708													
1148	0,5701	0,8700	0,2484	0,1939	0,5988	0,6882	0,0000	0,0000	0,2930	0,3271												
1149	0,2076	0,4763	0,0000	0,0000	0,2110	0,3706	0,0000	0,0000	0,0226	0,3175	0,3643											
1150	0,2987	0,3069	0,6321	0,2488	0,1376	0,3253	0,0000	0,2643	0,5948	0,1707	0,1868	0,0000										
1151	0,0000	0,0000	0,0000	0,1829	0,0000	0,0000	0,0528	0,0000	0,0000	0,0169	0,0000	0,0000	0,0000									
1152	0,2982	0,1408	0,4683	0,2588	0,1428	0,1680	0,0000	0,4561	0,3426	0,0980	0,0995	0,0000	0,5045	0,1841								
1153	0,5456	0,4015	0,0000	0,2657	0,2307	0,4237	0,0000	0,0253	0,2460	0,4344	0,2919	0,5829	0,0907	0,0000	0,1798							
1155	0,4416	0,2859	0,4865	0,0000	0,0000	0,1760	0,0000	0,0000	0,2226	0,0000	0,3448	0,0000	0,5285	0,0000	0,4246	0,0000						
1156	0,0107	0,0000	0,0000	0,3970	0,0000	0,0000	0,5382	0,0569	0,0000	0,0491	0,0000	0,0000	0,0000	0,1358	0,0995	0,0000	0,0000					
1159	0,2299	0,1168	0,3014	0,1843	0,0309	0,1066	0,0000	0,0000	0,2811	0,2780	0,1620	0,0000	0,4378	0,0000	0,3473	0,0297	0,3557	0,0054				
1160	0,7819	0,5701	0,2760	0,0961	0,3023	0,6231	0,0000	0,0000	0,4419	0,1585	0,5982	0,1675	0,1760	0,0000	0,1348	0,3924	0,0935	0,0000	0,0000			
1162	0,2039	0,2677	0,7765	0,3809	0,0846	0,2855	0,0332	0,2531	0,0319	0,0538	0,1244	0,0000	0,5571	0,0226	0,2938	0,0327	0,2107	0,0000	0,1142	0,0298		
1164	0,3811	0,2774	0,3905	0,4264	0,0253	0,1832	0,0529	0,3115	0,1846	0,4349	0,0265	0,1689	0,5160	0,1772	0,1887	0,3453	0,0143	0,0890	0,4493	0,1410	0,3428	

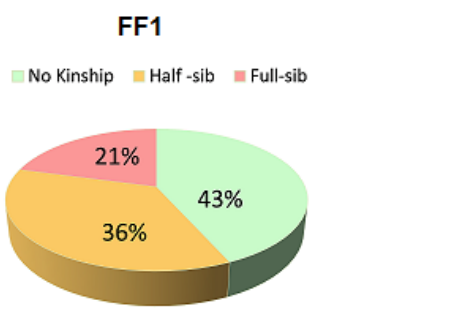

Figure 4. Kinship analysis of FF1 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	725	726	727	728	729	730	731	732	733	735	736	737	738	739	740	744	746	748	749	751	753	754	755
725																							
726	0,4751																						
727	0,0336	0,2255																					
728	0,3359	0,2234	0,0000																				
729	0,0928	0,4693	0,1040	0,0158																			
730	0,0000	0,0000	0,0000	0,0238	0,0000																		
731	0,0000	0,0000	0,0750	0,0000	0,0000	0,0000																	
732	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000																
733	0,3797	0,4256	0,0657	0,5563	0,2225	0,0000	0,0000	0,0000															
735	0,3464	0,4582	0,1801	0,0096	0,4223	0,0000	0,0000	0,0000	0,4552														
736	0,0000	0,0000	0,6800	0,1079	0,0000	0,0000	0,2211	0,0208	0,0000	0,1010													
737	0,0000	0,0000	0,5971	0,0000	0,0000	0,0000	0,2328	0,1003	0,0000	0,0162	0,5262												
738	0,2970	0,4178	0,1685	0,3875	0,6255	0,0000	0,0000	0,0000	0,3429	0,4581	0,0920	0,0000											
739	0,5575	0,7469	0,1386	0,3061	0,2995	0,0000	0,0000	0,0000	0,4242	0,5451	0,0595	0,0000	0,5036										
740	0,2634	0,4663	0,1859	0,1870	0,4168	0,0000	0,0000	0,0000	0,3073	0,5899	0,1947	0,1064	0,6233	0,6359									
744	0,6712	0,4677	0,0256	0,0865	0,1024	0,0000	0,0000	0,0000	0,2832	0,5064	0,0000	0,0274	0,2174	0,4710	0,2691								
746	0,8232	0,3217	0,0000	0,5346	0,2316	0,0000	0,0000	0,0000	0,6167	0,4048	0,0000	0,0000	0,5345	0,5839	0,4975	0,6593							
748	0,8159	0,3388	0,0000	0,5154	0,3822	0,0000	0,0000	0,0000	0,6009	0,4253	0,0000	0,0000	0,5606	0,5668	0,5219	0,6903	1,0000						
749	0,4711	0,7482	0,3927	0,2230	0,3820	0,0000	0,0000	0,0000	0,4219	0,5388	0,2320	0,0646	0,4165	0,5797	0,4685	0,6334	0,4944	0,5186					
751	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,5506	0,0054	0,0000	0,0000	0,0961	0,2774	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000				
753	0,0000	0,0000	0,4521	0,0000	0,0000	0,0000	0,2727	0,1334	0,0000	0,0000	0,6387	0,5476	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0923	0,3084			
754	0,0619	0,7516	0,1367	0,0000	0,3815	0,0000	0,0000	0,0000	0,2567	0,6273	0,0000	0,0000	0,4105	0,5831	0,5442	0,2195	0,0209	0,0256	0,5844	0,0000	0,0000		
755	0,0000	0,0000	0,3922	0,0000	0,0000	0,0000	0,0220	0,0000	0,0000	0,0471	0,5656	0,6379	0,0000	0,0000	0,1407	0,0000	0,0000	0,0000	0,0092	0,0000	0,5927	0,0000	

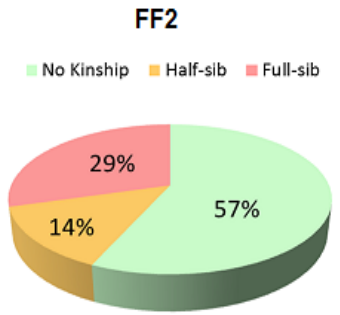

Figure 5. Kinship analysis of FF2 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	756	757	758	759	760	761	762	763	764	765	766	767	769	770	771	772	773	774	3_775
756																			
757	0,2189																		
758	0,4242	0,0000																	
759	0,4750	0,1790	0,2209																
760	0,1401	0,1173	0,0000	0,0000															
761	0,1914	0,0000	0,0892	0,0000	0,0000														
762	0,4575	0,0000	0,6335	0,1692	0,0000	0,0000													
763	0,2241	0,5841	0,0000	0,1018	0,3729	0,0000	0,0000												
764	0,5759	0,1091	0,4979	0,4477	0,2274	0,1915	0,2299	0,2852											
765	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,2385	0,2618										
766	0,3791	0,0000	0,5558	0,1107	0,0000	0,0573	0,5195	0,0096	0,2804	0,0542									
767	0,1957	0,1419	0,0000	0,0000	0,1032	0,0000	0,0000	0,4855	0,2855	0,2892	0,0000								
769	0,4129	0,0000	0,5053	0,0000	0,0000	0,0970	0,6349	0,0000	0,2742	0,0000	0,7190	0,0000							
770	0,0076	0,0000	0,1032	0,0000	0,0000	0,0000	0,0000	0,2108	0,1019	0,0174	0,3089	0,4774	0,0168						
771	0,3877	0,5845	0,0000	0,2654	0,3287	0,0651	0,0017	0,6702	0,4107	0,0656	0,0028	0,3560	0,0000	0,1987					
772	0,1837	0,0000	0,1809	0,0000	0,0171	0,0000	0,0581	0,3827	0,3707	0,1073	0,2137	0,6620	0,0058	0,6681	0,4583				
773	0,3112	0,0000	0,0645	0,0169	0,0000	0,0666	0,1869	0,1662	0,1531	0,0000	0,2607	0,1825	0,0625	0,4619	0,4154	0,3018			
774	0,5573	0,2526	0,2286	0,5052	0,0197	0,0554	0,1841	0,4263	0,6151	0,0704	0,2653	0,3262	0,1383	0,0505	0,5075	0,3107	0,2640		
3_775	0,0988	0,1771	0,0000	0,1853	0,0000	0,0000	0,0000	0,4201	0,1019	0,1001	0,0000	0,0000	0,0000	0,0000	0,5410	0,0000	0,0000	0,5382	

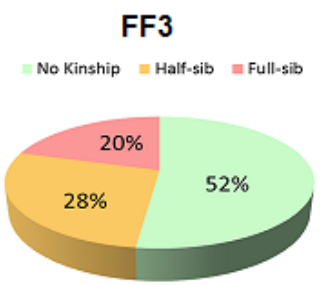

Figure 6. Kinship analysis of FF3 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	681	682	684	690	693	697	699	701	702	1233	1238	1240	5_1242
681													
682	0,0951												
684	0,3316	0,3388											
690	0,1325	0,4999	0,0382										
693	0,1828	0,0731	0,0489	0,0000									
697	0,2381	0,2945	0,0000	0,2689	0,0000								
699	0,1200	0,0000	0,0000	0,0000	0,1274	0,0000							
701	0,2855	0,1126	0,2747	0,0000	0,0000	0,1063	0,0000						
702	0,2191	0,0000	0,3020	0,0000	0,0423	0,0000	0,4708	0,0000					
1233	0,0000	0,1078	0,2254	0,1540	0,0000	0,0000	0,0201	0,0000	0,0000				
1238	0,3659	0,3355	0,1118	0,4114	0,0000	0,5303	0,0000	0,4073	0,0000	0,0000			
1240	0,1199	0,0000	0,5077	0,0000	0,0198	0,0000	0,3651	0,2854	0,5854	0,0000	0,0000		
5_1242	0,2989	0,0000	0,5562	0,0000	0,0000	0,0000	0,4686	0,0687	0,7394	0,0938	0,0000	0,9318	

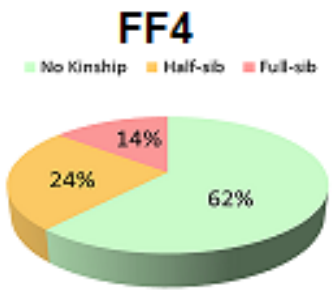

Figure 7. Kinship analysis of FF4 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	1095	1096	1097	1098	1100	1101	1104	1110	1111	1115	1120	1121	1122	1123	1124	1125	1129	1134	6_1135
1095																			
1096	0,0105																		
1097	0,0000	0,5381																	
1098	0,0000	0,0000	0,2368																
1100	0,0000	0,0261	0,0536	0,0000															
1101	0,2166	0,1233	0,0000	0,0000	0,0000														
1104	0,0000	0,6010	0,3551	0,0026	0,0000	0,2635													
1110	0,0000	0,1164	0,0000	0,0000	0,0000	0,6697	0,1705												
1111	0,0000	0,4606	0,0579	0,0000	0,0000	0,0000	0,2206	0,0000											
1115	0,0681	0,1120	0,0000	0,0000	0,0000	0,6755	0,2522	0,7605	0,0000										
1120	0,0000	0,3867	0,4521	0,2011	0,0058	0,0000	0,3920	0,0000	0,1709	0,0000									
1121	0,0000	0,4828	0,3655	0,0000	0,1069	0,2162	0,3191	0,3855	0,1039	0,3372	0,0586								
1122	0,0000	0,1383	0,0000	0,0000	0,0000	0,0000	0,0000	0,0523	0,2594	0,0000	0,0000	0,2919							
1123	0,0000	0,0899	0,0447	0,0000	0,0468	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000						
1124	0,0000	0,4996	0,3554	0,0993	0,2896	0,2844	0,6498	0,2770	0,2820	0,3193	0,4128	0,5824	0,1815	0,2485					
1125	0,2848	0,2850	0,0622	0,0000	0,2348	0,3279	0,1974	0,2427	0,0579	0,1867	0,1670	0,2807	0,0811	0,3921	0,2832				
1129	0,3709	0,1587	0,3339	0,0000	0,0000	0,3173	0,0687	0,0000	0,0000	0,1413	0,1671	0,3903	0,0000	0,0000	0,1245	0,2293			
1134	0,0000	0,2968	0,3536	0,0000	0,0056	0,0000	0,2145	0,0000	0,1751	0,0000	0,4186	0,3178	0,0899	0,1780	0,4952	0,2742	0,0000		
6_1135	0,0000	0,5230	0,1863	0,0000	0,0000	0,1852	0,6823	0,1034	0,2284	0,1747	0,3000	0,1446	0,0000	0,0116	0,4987	0,4515	0,0000	0,2251	

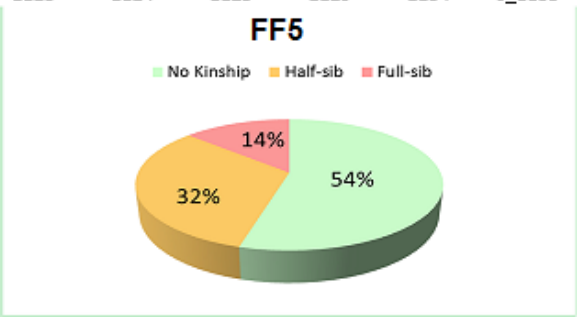

Figure 8. Kinship analysis of FF5 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	1206	1208	1209	1210	1212	1213	1214	1215	1216	1217	1219	1220	1221	1222	1223	1224	7_1225
1206																	
1208	0,1384																
1209	0,3229	0,0000															
1210	0,4246	0,0744	0,4398														
1212	0,2822	0,0000	0,4543	0,3068													
1213	0,0000	0,0294	0,0914	0,4786	0,2971												
1214	0,3748	0,4705	0,2689	0,2262	0,4085	0,0000											
1215	0,7830	0,0151	0,1053	0,1053	0,2486	0,0000	0,2456										
1216	0,0952	0,6353	0,0000	0,1542	0,0000	0,0000	0,0000	0,0000									
1217	0,3215	0,0569	0,5510	0,4905	0,2501	0,2196	0,3898	0,1039	0,0000								
1219	0,0000	0,0975	0,2466	0,3985	0,1237	0,4365	0,2079	0,0000	0,1434	0,3000							
1220	0,0202	0,1084	0,4649	0,0000	0,5076	0,2119	0,4059	0,0737	0,0000	0,2519	0,1290						
1221	0,0946	0,5357	0,0000	0,0000	0,0000	0,0000	0,0399	0,0000	0,6666	0,0000	0,0375	0,0000					
1222	0,2405	0,0000	0,2153	0,6174	0,2582	0,0000	0,3674	0,0275	0,1189	0,2743	0,3686	0,0981	0,1814				
1223	0,4952	0,1476	0,3719	0,6156	0,4160	0,2105	0,5326	0,2822	0,0948	0,5100	0,2796	0,0108	0,0873	0,5945			
1224	0,0608	0,0746	0,3473	0,4938	0,2936	0,2566	0,4087	0,0000	0,0000	0,1553	0,4136	0,2125	0,0000	0,3705	0,2950		
7_1225	0,1838	0,0000	0,1327	0,2723	0,2615	0,0582	0,3317	0,0605	0,1732	0,2460	0,3802	0,1717	0,0000	0,3469	0,2516	0,1369	

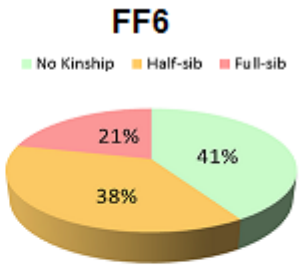

Figure 9. Kinship analysis of FF6 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship ($r < 0.125$), half-sib individuals ($0.125 \leq r \leq 0.375$), full-sib individuals ($r > 0.375$)

ALL LOCI	60	61	63	66	69	70	72	73	74	75	76	77	78	79	80	81	82	83	85	86	87	88	90	93	94	8_95
60																										
61	0,2266																									
63	0,0000	0,0000																								
66	0,5315	0,7249	0,0000																							
69	0,0488	0,5541	0,0000	0,3531																						
70	0,1316	0,0000	0,5689	0,0531	0,0000																					
72	0,0000	0,0000	0,0013	0,0270	0,0895	0,0000																				
73	0,1895	0,0247	0,0593	0,1650	0,1087	0,2428	0,0000																			
74	0,0393	0,0000	0,2156	0,0000	0,0968	0,1363	0,2550	0,1865																		
75	0,0837	0,0000	0,0000	0,2457	0,0000	0,0012	0,0000	0,0000	0,0000																	
76	0,0926	0,0000	0,3128	0,2136	0,0000	0,4036	0,1965	0,0235	0,3109	0,3723																
77	0,4258	0,6838	0,0000	0,7843	0,9421	0,0000	0,0190	0,1833	0,0968	0,0000	0,0000															
78	0,3870	0,0605	0,2225	0,0640	0,0997	0,3084	0,2621	0,4117	0,2544	0,0160	0,2379	0,1714														
79	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1513	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000													
80	0,3244	0,6283	0,0000	0,7340	0,5867	0,0000	0,0000	0,1059	0,0000	0,0170	0,0000	0,7538	0,0231	0,0000												
81	0,0000	0,0000	0,2764	0,0000	0,0000	0,1943	0,1465	0,0300	0,5325	0,0000	0,1948	0,0000	0,0000	0,3186	0,0000											
82	0,2416	0,0000	0,0000	0,4994	0,0196	0,0000	0,0000	0,0000	0,2138	0,0947	0,1392	0,6718	0,0000	0,0180	0,1247	0,0881										
83	0,0519	0,0177	0,0000	0,1669	0,0000	0,2477	0,0000	0,0728	0,0000	0,1392	0,0262	0,3142	0,1031	0,0000	0,0000	0,0000	0,0000									
85	0,3740	0,7089	0,0000	0,6875	0,5594	0,0000	0,0000	0,0654	0,0000	0,0000	0,2326	0,8009	0,0000	0,0000	0,5731	0,0000	0,2929	0,0000								
86	0,6011	0,7470	0,0000	0,8230	0,4401	0,0000	0,0000	0,0749	0,0000	0,0000	0,0408	0,8005	0,0000	0,0000	0,6231	0,0000	0,3220	0,1090	0,8708							
87	0,1369	0,0000	0,0000	0,0000	0,0000	0,1473	0,3399	0,0000	0,2430	0,0084	0,1495	0,0000	0,0931	0,1494	0,0000	0,4443	0,0000	0,0916	0,0000	0,0000						
88	0,0000	0,0000	0,2169	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0431	0,0000	0,0000	0,0000	0,0000	0,1358	0,0000	0,0000	0,0000	0,0000	0,0000					
90	0,5432	0,3159	0,0000	0,4845	0,2696	0,0000	0,1842	0,1466	0,0000	0,0000	0,2304	0,3125	0,3130	0,0000	0,2797	0,0000	0,0569	0,0000	0,5192	0,3960	0,0517	0,0000				
93	0,0000	0,0000	0,0437	0,0000	0,0000	0,0000	0,3757	0,0000	0,3708	0,0499	0,0504	0,0000	0,3803	0,0000	0,0000	0,2697	0,0000	0,0000	0,0000	0,0000	0,2855	0,0000	0,0463			
94	0,0000	0,0000	0,1624	0,0000	0,0000	0,2429	0,0000	0,0000	0,0000	0,1103	0,0589	0,0000	0,0000	0,0000	0,0000	0,1242	0,0000	0,0000	0,0000	0,0000	0,0000	0,6138	0,0000	0,0000		
8_95	0,0000	0,0000	0,3866	0,0000	0,0000	0,3018	0,0000	0,0000	0,0000	0,0719	0,1245	0,0000	0,1233	0,0000	0,0000	0,0000	0,0000	0,5036	0,0000	0,0000	0,0000	0,4737	0,0000	0,0000	0,4348	

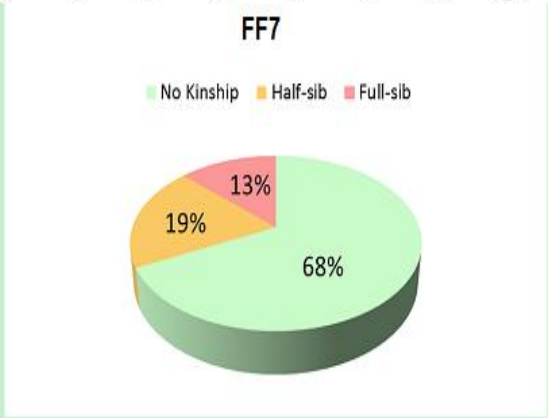

Figure 10. Kinship analysis of FF7 population. Green values represent non-related individuals, yellow values correspond to half-sib individuals and red values are full-sib individuals. Threshold values for kinship analysis: no kinship (r < 0.125), half-sib individuals (0.125 ≤ r ≤ 0.375), full-sib individuals (r > 0.375).