



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

**VALIDAÇÃO DE UM PAINEL DE SNPS DE BAIXA DENSIDADE PARA
ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA
VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus
vannamei*)**

NAYELLE MEYRE LISBOA SILVA

DISSERTAÇÃO DE MESTRADO EM CIÊNCIAS ANIMAIS

**BRASÍLIA/DF
JANEIRO DE 2020**



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

**VALIDAÇÃO DE UM PAINEL DE SNPS DE BAIXA DENSIDADE PARA
ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA
VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus
vannamei*)**

NAYELLE MEYRE LISBOA SILVA

Dr. ALEXANDRE RODRIGUES CAETANO

Dra. PATRÍCIA IANELLA

DISSERTAÇÃO DE MESTRADO EM CIÊNCIAS ANIMAIS

PUBLICAÇÃO: 216/2020

**BRASÍLIA/DF
JANEIRO DE 2020**

REFERÊNCIA BIBLIOGRÁFICA E CATALOGAÇÃO

SILVA, N. M. L. Validação de um painel de SNPs de baixa densidade para análise de paternidade e parentesco e avaliação da variabilidade genética do camarão cinza (*Litopenaeus vannamei*). Brasília: Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2020, 63 p. Dissertação de Mestrado.

Documento formal, autorizando reprodução desta dissertação de mestrado para empréstimo ou comercialização, exclusivamente para fins acadêmicos, foi passado pelo autor à Universidade de Brasília e acha-se arquivado na Secretaria do Programa. O autor e o seu orientador reservam para si os outros direitos autorais, de publicação. Nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor ou do seu orientador. Citações são estimuladas, desde que citada a fonte.

FICHA CATALOGRÁFICA

SILVA, Nayelle Meyre Lisboa. Validação de um painel de SNPs de baixa densidade para análise de paternidade e parentesco e avaliação da variabilidade genética do camarão cinza (*Litopenaeus vannamei*). Brasília: Faculdade de Agronomia e Medicina Veterinária da Universidade de Brasília, 2020. 63 p. Dissertação (Mestrado em Ciências Animais) - Faculdade de Agronomia e Medicina Veterinária da Universidade de Brasília, 2020.

1. Brasil. 2. Carcinicultura. 3. Melhoramento Genético.
4. *Pedigree*.

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

**VALIDAÇÃO DE UM PAINEL DE SNPS DE BAIXA DENSIDADE PARA
ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA
VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus vannamei*)**

NAYELLE MEYRE LISBOA SILVA

**DISSERTAÇÃO DE MESTRADO SUBMETIDA
AO PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIAS ANIMAIS, COMO PARTE DOS
REQUISITOS NECESSÁRIOS À OBTENÇÃO
DO GRAU DE MESTRE EM CIÊNCIAS
ANIMAIS.**

APROVADA POR:

ALEXANDRE RODRIGUES CAETANO, Doutor (Embrapa Recursos Genéticos e Biotecnologia) (ORIENTADOR)

CONCEPTA MARGARET MCMANUS PIMENTEL, Doutora (Universidade de Brasília) (EXAMINADOR INTERNO)

LARISSA PAOLA RODRIGUES VENANCIO, Doutora (Universidade Federal do Oeste da Bahia) (EXAMINADOR EXTERNO)

BRASÍLIA/DF, 21 DE JANEIRO DE 2020

AGRADECIMENTOS

Agradeço...

A Deus por ter me dado a vida, oportunidade de estudar, me graduar e ser mestre, pois sem o consentimento dele nada disso seria possível.

A minha família de criação que cuidou de mim até os 6 anos (entre 10-12 anos também) com amor, zelo e carinho. Influenciaram muito na formação da minha personalidade, portanto a pessoa que me tornei é principalmente devido ao que vivi nessa fase em que tudo foi tão maravilhoso.

A minha mãe biológica Rosimeire que me deu um lar e apoio em momentos específicos, pois sem isso também não seria possível chegar até aqui.

Ao meu namorado Danilo, por todos os momentos maravilhosos que me proporcionou ao seu lado durante o desenvolvimento desse trabalho, sempre amoroso, carinhoso, procurando me deixar menos tensa em momentos complicados, além de me agradar, motivar muito e sempre auxiliar da melhor maneira possível. Amo você!

Aos meus orientadores Patrícia e Alexandre que disponibilizaram auxílio e conhecimento teórico para escrita, além de aprendizados que serão utilizados para a vida.

Aos meus colegas de trabalho, Noeliton e Aline que são companhias tão agradáveis e proporcionaram bons momentos.

A Embrapa Recursos Genéticos e Biotecnologia pela disponibilidade de toda sua estrutura para realização dos trabalhos.

Ao Programa de Pós-Graduação em Ciências Animais da Universidade de Brasília, pela oportunidade de realização do Mestrado.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelo acesso ao portal de periódicos.

RESUMO

VALIDAÇÃO DE UM PAINEL DE SNPs DE BAIXA DENSIDADE PARA ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus vannamei*)

Nayelle Meyre Lisboa Silva¹, Patrícia Ianella², Alexandre Rodrigues Caetano³

1 Bióloga/ Mestranda em Ciências Animais – PPG/UnB – Brasília/DF

2 Bióloga/ Doutora – Embrapa Recursos Genéticos e Biotecnologia

3 Zootecnista/ Doutor – Embrapa Recursos Genéticos e Biotecnologia

O objetivo deste estudo foi validar um painel SNP de baixa densidade em populações de núcleo de melhoramento genético e em amostras de camarão coletadas em lojas de varejo, para atribuir paternidade e parentesco, avaliar variabilidade e estruturação genética em camarão cinza. Amostras de reprodutores de três linhagens comerciais (A, B e C) e amostras de sete lojas de varejo (SPDF1-SPDF5, PEDF e PEPB) foram avaliadas, totalizando 328 amostras. Inicialmente, a validação foi feita utilizando 144 SNPs para seleção dos 96 marcadores mais informativos que iriam compor o painel. A genotipagem dos SNPs foi realizada com plataforma Fluidigm e em seguida foram feitas análises de estruturação populacional, verificação da Heterozigosidade observada (H_o) e esperada (H_e), avaliação da qualidade, por meio do RMSD (do inglês *Root Mean Square Deviation*) e RMSE (do inglês *Root Mean Square Error*), das estimativas de parentesco e da probabilidade de não exclusão parental. Os resultados de PCA (Análise de Componentes Principais) e STRUCTURE das três linhagens do núcleo de melhoramento genético agregaram as amostras das linhagens A e B no mesmo grupo, enquanto a linhagem C apresentou uma composição diferenciada, indicando que a origem do germoplasma das linhagens A e B é o mesmo, e/ou que existe um forte fluxo gênico entre as

duas linhagens. Em relação a estruturação das 10 populações, foi observado que as amostras dos varejistas não são oriundas do produtor de larvas abordado nesse estudo e que todas as populações das lojas de varejo aparentam ter origem de um mesmo produtor. Já as estimativas de H_o e H_e indicaram alta variabilidade genética das amostras, pois a maior variação encontrada entre essas estimativas foi de 4%, sendo que a média dos valores de H_o para todas as populações foi de 0,392. Quanto a qualidade das estimativas de parentesco, foi demonstrado que o valor de RMSD calculado entre as estimativas obtidas com o painel de 96 SNPs e as adquiridas com 3.857 SNPs do chip comercial da Illumina foi de 0,08. Já o RMSE desses dois grupos variou somente 0,05, indicando que os intervalos de confiança são muito próximos. O RMSD e RMSE indicam que o painel de 96 SNPs é tão eficaz para estimar parentesco quanto a ferramenta comercial da Illumina. As probabilidades de não exclusão das 10 populações também demonstraram resultados satisfatórios para P1, P2 e P3 de $4,5 \times 10^{-7}$, $4,4 \times 10^{-4}$ e $5,2 \times 10^{-11}$, respectivamente. Portanto, o painel desenvolvido no presente estudo poderá ser utilizado em estudos amplos que demandem grandes números de atribuições de paternidade, avaliações de parentesco e análises de estruturação populacional do camarão cinza.

PALAVRAS-CHAVE: Brasil, carcinicultura, melhoramento genético, *pedigree*.

ABSTRACT

VALIDATION OF A LOW DENSITY SNPS PANEL FOR PATERNITY AND KINSHIP ANALYSIS AND EVALUATION OF GENETIC VARIABILITY IN PACIFIC WHITE SHRIMP (*Litopenaeus vannamei*)

The objective of this study was to validate a low density SNP panel in breeding nucleus populations and in shrimp samples collected in retail stores, to attribute paternity and kinship, to assess genetic variability and structure in pacific white shrimp. Broodstock samples from three commercial lines (A, B and C) and samples from seven retail stores (SPDF1-SPDF5, PEDF and PEPB) were evaluated, totaling 328 samples. Initially, the validation was done using 144 SNPs to select the 96 most informative markers that would compose the panel. Genotyping was performed with a Fluidigm EP1 platform, followed by population structure analysis, verification of the observed (H_o) and expected (H_e) heterozygosity, and calculation of Root Mean Square Deviations (RMSD) and Root Mean Square Errors (RMSE) for kinship estimates obtained with different numbers of SNP markers. PCA (Principal Component Analysis) and STRUCTURE results considering the three Broodstock lines grouped samples from lines A and B, while line C showed a different composition, indicating that the origin of the germplasm of lines A and B is the same, and / or that there is a strong gene flow between the two lines. Regarding the structuring of the 10 populations, it was observed that the retailers' samples are not from the larvae producer addressed in this study and that all the populations of the retail stores appear to have originated from the same producer. H_o and H_e estimates indicated high genetic variability of the samples, since the greatest variation found between these estimates was 4%, with the average H_o values for all populations being 0.392. Regarding the quality of kinship estimates, it was demonstrated that the RMSD value calculated between the estimates obtained with the 96 SNP panel and those acquired with 3,857 SNPs from Illumina's commercial chip was 0.08. The RMSE of these two groups varied only 0.05, indicating that the

confidence intervals are very close. The RMSD and RMSE indicate that the 96 SNP panel is as effective at estimating kinship as Illumina's commercial tool. The probabilities of non-exclusion of the 10 populations also demonstrated satisfactory results for P1, P2 and P3 of 4.5×10^{-7} , 4.4×10^{-4} and 5.2×10^{-11} , respectively. Therefore, the panel developed in the present study can be used in broad studies that require large numbers of paternity assignments, kinship assessments and analyzes of the pacific white shrimp population structure.

KEYWORDS: Brazil, genetic improvement, *pedigree*, shrimp farming.

SUMÁRIO

Capítulo 1	1
1.1 Introdução	2
1.2 Objetivos	4
1.2.1 Objetivo geral	4
1.2.2 Objetivos específicos	4
1.3 Revisão Bibliográfica	5
1.3.1 Descrição biológica do camarão cinza	5
1.3.1.1 Taxonomia	5
1.3.1.2 Morfologia	6
1.3.1.3 Distribuição natural e ciclo de vida	6
1.3.2 Produção do camarão cinza	8
1.3.3 Melhoramento genético do camarão cinza no Brasil	11
1.3.4 Tecnologias para a genotipagem de SNPs em camarão cinza.....	13
1.3.5 Uso de marcadores SNPs para análises de paternidade, determinação de parentesco e avaliação da variabilidade genética na produção de espécies aquícolas	16
Referências Bibliográficas	19
Capítulo 2: Validação de um painel de SNPs de baixa densidade para análise de paternidade e parentesco e avaliação da variabilidade genética do camarão cinza (<i>Litopenaeus vannamei</i>)	25
2.1 Resumo	26
2.2 Abstract	28
2.3 Introdução	30
2.4 Materiais e Métodos	32
2.4.1 Amostragem de camarão cinza.....	32
2.4.2 Extração de DNA	33
2.4.3 Genotipagem com plataforma Fluidigm.....	33
2.4.4 Validação do painel com amostras de reprodutores de linhagens comerciais ...	34
2.4.4.1 Estruturação populacional	35
2.2.4.2 Análise de parentesco	35
2.4.5 Análise de estruturação populacional, variabilidade genética e parentesco de amostras de camarão de varejistas.....	36
2.5 Resultados	38
2.5.1 Validação do painel com amostras de reprodutores de linhagens comerciais	38
2.5.1.1 Estruturação populacional	38

2.5.1.2 Análise de paternidade e parentesco.....	40
2.5.2 Análises de amostras de camarão de varejistas	44
2.5.2.1 Estruturação populacional e variabilidade genética	45
2.5.2.2 Análise de paternidade	49
2.6 Discussão.....	51
Conclusão	54
Referências Bibliográficas.....	55
Apêndice A – Sequências ASP1, APS2, LSP e STA dos 96 SNPs genotipados com Fluidigm.....	59
Apêndice B – N° de alelos, MAF, HW do painel com 96 SNPs e 86 SNPs das 325 amostras do produtor de larvas e das lojas de varejo.....	62

CAPÍTULO 1
INTRODUÇÃO E REVISÃO BIBLIOGRÁFICA

1.1 INTRODUÇÃO

A criação de camarões (carcinicultura) é um dos ramos mais importantes da aquicultura, com 19,53% do valor da produção aquícola nacional em 2016 (IBGE, 2019). O camarão cinza (*Litopenaeus vannamei*) destaca-se nesse ramo, devido à alta adaptabilidade, à rapidez no crescimento e à ampla faixa de tolerância à salinidade (Costa, 2004). Essa espécie é impactante para aquicultura do Brasil e do mundo, já que a produção mundial foi 4,1 milhões de toneladas em 2016, o que equivale a 53% da produção dos crustáceos (FAO, 2018).

A implementação de programas de melhoramento genético do camarão cinza contribuiu para o aumento da lucratividade e da produção devido à seleção de indivíduos que apresentam características de interesse econômico, como maiores taxas de crescimento, sobrevivência, e resistência a doenças (Castillo-Juárez et al, 2015; Gitterle et al., 2005). Os programas de melhoramento do camarão são geralmente estruturados em esquemas de seleção dentro de famílias de irmãos completos e entre elas, nas quais as dificuldades para marcação e identificação individual geram grandes desafios, restringindo os processos de formação de grupos contemporâneos de múltiplas famílias, expostas a diferentes ambientes de criação. É por meio dessa marcação e identificação individual que as informações de *pedigree* dos indivíduos das populações de melhoramento são mantidas. A manutenção desses dados de *pedigree* é essencial para a realização das avaliações genéticas e para a identificação de indivíduos geneticamente superiores, com o intuito de serem utilizados como reprodutores da próxima geração. Dá-se ênfase a ganhos genéticos em características produtivas, assim como à manutenção da variabilidade genética da população, buscando-se minimizar as taxas de cruzamentos endogâmicos (Moss e Moss, 2009).

A identificação individual do camarão cinza para rastreamento de informações de *pedigree* pode ser realizado a partir da análise de paternidade e parentesco, com marcadores

moleculares (Liu et al., 2016). Os marcadores microssatélites¹ foram utilizados para caracterizar a diversidade genética, composição de estoques de indivíduos selvagens e cultivados (Maggioni et al., 2013; Perez-Enriquez et al., 2018; Perez-Enriquez et al., 2009 Vela Avitúa, et al., 2013) e determinação de parentesco em camarão cinza (Wang, 2006). O uso de microssatélites em estudos de *L. vannamei* foi frequente nos últimos anos, no entanto, os Polimorfismos de Base Única (SNPs do inglês *Single Nucleotide Polimorphism*) passaram a ser amplamente utilizados em estudos genéticos mais recentes em outras espécies de animais (Flanagan e Jones, 2019).

Os SNPs são codominantes² e geralmente bialélicos³. O seu polimorfismo ocorre devido a mutações de base única da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). As transições⁴ são as mutações mais comuns, enquanto as transversões⁵ são menos frequentes. O uso de SNPs em análises de parentesco e caracterização genética passou a ser mais vantajoso do que a utilização de microssatélites, devido à sua grande abundância no genoma e ao seu baixo custo de genotipagem, com métodos que permitem automação e baixas taxas de erro (Caetano, 2009; Vignal et al., 2002). Flanagan e Jones (2019) observaram que um número substancial de estudos demonstrou a maior eficiência dos SNPs em relação a microssatélites em análises de parentesco.

Os painéis de SNPs de baixa densidade (com até 100 SNPs) são ferramentas úteis na realização de análises de parentesco e paternidade. Liu et al. (2016) desenvolveram um painel com 96 SNPs para atribuição eficaz de parentesco em truta arco-íris, enquanto Holman et al. (2017) concluíram que é possível determinar parentesco em salmão com 100% de acurácia utilizando um painel composto por 94 SNPs. Além disso, simulações realizadas por Anderson e Garza (2006) indicaram que 60 a 100 SNPs são suficientes para atribuição de paternidade precisa em larga escala. Portanto, o desenvolvimento e validação de painéis de SNPs de baixa densidade para camarão cinza, com o intuito de fornecer informações de *pedigree* de cada indivíduo a partir da análise de paternidade e parentesco seria de grande utilidade para o setor produtivo, uma vez que pode aprimorar os programas de melhoramento genético, melhorando as taxas de ganho genético e auxiliando na manutenção da variabilidade genética dos estoques de reprodutores.

¹ Sequências curtas (1 a 6 pb), repetidas em tandem, distribuídas aleatoriamente no genoma.

² Possibilitam diferenciar indivíduos homocigotos de heterocigotos.

³ São encontradas apenas dois variantes em uma espécie, por exemplo, um alelo corresponde a um par de bases A/T e o outro a um G/C.

⁴ Trocas de uma purina por outra purina (A " G) ou de uma pirimidina por outra pirimidina (C " T).

⁵ Troca de uma purina por uma pirimidina, ou vice-versa (C/T " A/G)

1.2. OBJETIVOS

1.2.1 Objetivo geral

Validar um painel de SNPs de baixa densidade para controle de paternidade, parentesco e avaliação de variabilidade genética em linhagens de reprodutores de núcleos de melhoramento genético e em populações de lojas de varejo do camarão cinza.

1.2.2 Objetivos específicos

- Validar um painel para análises de paternidade, parentesco e variabilidade genética de camarão cinza do setor produtivo brasileiro.
- Comparar a eficiência de estimativas de parentesco fornecidas pelo painel de baixa densidade com as obtidas com o chip comercial da Illumina Infinium® ShrimpLD-24 que tem maior número de SNPs (6K).
- Avaliar parâmetros de variabilidade e estruturação genética nas populações comerciais estudadas.

1.3 REVISÃO BIBLIOGRÁFICA

1.3.1 Descrição biológica do camarão cinza

1.3.1.1 Taxonomia

O camarão cinza, *Litopenaeus vannamei* (Boone, 1931), pertence ao filo Arthropoda, que é definido por ter associado apêndices a um exoesqueleto que é trocado periodicamente (Bailey-Brock e Moss, 1992). A classificação taxonômica de *L. vannamei* (Farfante e Kenslev, 1997; Ma et al, 2011) é a seguinte:

Domínio: Eukaryota

Reino: Metazoa

Filo: Arthropoda

Subfilo: Crustáceos

Classe: Malacostraca

Subclasse: Eumalacostraca

Ordem: Decapoda

Subordem: Dendrobranchiata

Superfamília: Penaeoidea

Família: Penaeidae

Gênero: *Litopenaeus*

Espécie: *Litopenaeus vannamei*

1.3.1.2 Morfologia

O corpo do *L. vannamei* é composto pelo cefalotórax (cabeça e tórax fundidos) e pelo abdômen. São encontrados 19 pares de segmentos (Figura 1.1), sendo que cinco pares estão localizados no céfalo, oito pares são encontrados no tórax e os últimos seis pares compõem o abdômen (Dugassa e Gaetan, 2018). Nos segmentos da cabeça estão contidos os olhos compostos e lateralmente móveis, o rostrum e os pares de antênulas e antenas. O tórax possui três pares de maxilípedes e cinco pares de pereiópodos (os três primeiros são utilizados para alimentação e os dois últimos para locomoção) (Dugassa e Gaetan, 2018). Os pleópodes, ou pernas para natação, estão localizados nos cinco primeiros segmentos do abdômen e são divididos em protopódito e exopódito. O sexto segmento do abdômen é a cauda, composta por dois pares de urópodes e pelo telson. Em conjunto eles formam um remo para natação (Dall et al., 1991).

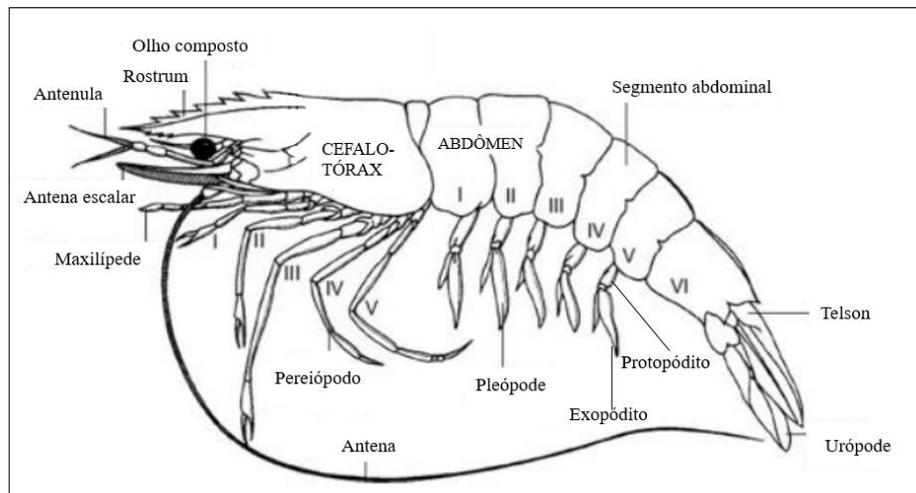


Figura 1.1: Morfologia externa do camarão cinza. Fonte: adaptado de Sucharita e Jyoti, (2013).

1.3.1.3 Distribuição Natural e Ciclo de Vida

O camarão cinza é nativo da costa leste do Oceano Pacífico desde a América do Norte (Sonora - México) até América do Sul (Tumbes - Peru), em águas cuja temperatura normalmente é superior a 20 °C ao longo do ano (FAO, 2019a) (Figura 1.2).



Figura 1.2: Distribuição natural do camarão cinza (em vermelho). Fonte: adaptado de FAO, (2019b)

O camarão cinza habita principalmente mares rasos em região tropical e subtropical. Os adultos vivem e desovam em mar aberto. O primeiro estágio larval do ciclo de vida é o náuplio, fase em que o animal se alimenta do seu próprio vitelo. Posteriormente, dá-se início o estágio larval protozoa, no qual a larva tem o corpo alongado com o cefalotórax e se alimenta de partículas menores que 10 micrômetros. No estágio larval seguinte, misis, as larvas assumem a forma de um camarão juvenil e os pleópodes começam a se desenvolver. O último estágio larval é a pós-larva, onde os indivíduos apresentam formato corporal e comportamento de um camarão adulto. Nesse momento ocorre a transição da fase marinha para estuária, pois as pós-larvas migram para a costa para passar seus estágios juvenis e subadultos em estuários costeiros, lagoas ou áreas de mangue (FAO, 2019a; Ribeiro, 2017; SENAR, 2016) (Figura 1.3).

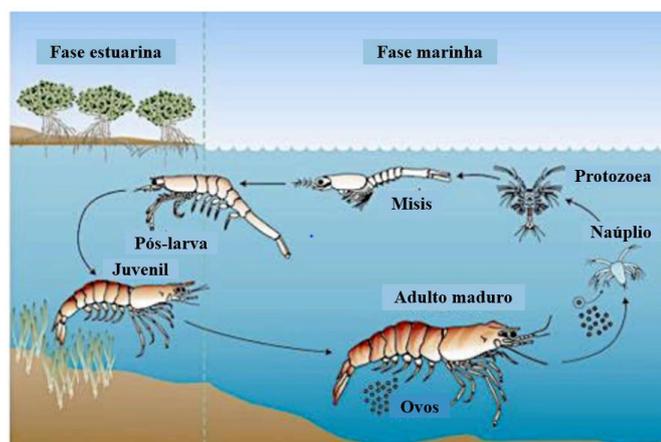


Figura 1.3: Ciclo de vida e fases larvais do camarão cinza. Fonte: adaptado de CSIRO (2011).

1.3.2 Produção do camarão cinza

O cultivo de camarão cinza está em constante desenvolvimento no mundo, representando cerca de 53% da produção dos crustáceos (FAO, 2018). Segundo a FAO (2019a), mais de 50 países estão envolvidos nessa atividade. Os principais produtores são: China, Tailândia, Indonésia, Brasil, Equador, México, Venezuela, Honduras, Guatemala, Nicarágua, Belize, Vietnã, Malásia, Taiwan, Ilhas do Pacífico, Peru, Colômbia, Costa Rica, Panamá, El Salvador, Estados Unidos da América, Índia, Filipinas, Camboja, Suriname, São Cristóvão, Jamaica, Cuba, República Dominicana, Bahamas (FAO, 2019a) (Figura 1.4).



Figura 1.4: Principais países produtores de camarão cinza. Fonte: (FAO 2019a)

Variados tipos de sistemas de cultivo de camarão cinza são usados no mundo, como viveiros escavados no solo, tanques à base de concreto, gaiolas flutuantes e cercados. Em relação as etapas de cultivo, existem basicamente quatro: larvicultura, berçário, engorda e despesca. A larvicultura é responsável pela produção de pós-larvas (Ribeiro et al., 2014), e começa com a maturação, ou seja, pelo acasalamento e desova. Machos e fêmeas são mantidos juntos em tanques apropriados até que ocorra o acasalamento. Em seguida, as fêmeas ovadas são colocadas em tanques de desova. Os náuplios, recém eclodidos, são transferidos para tanques de larvicultura e permanecem lá até atingirem o estágio de pós-larva (Figura 1.5) (Dugassa e Gaetan, 2018; SENAR, 2016).

A etapa de berçário ocorre quando os produtores utilizam um cultivo intermediário, antes do envio dos camarões para as fazendas de engorda, por um período de 10 a 30 dias. Quando chegam nas fazendas, os camarões juvenis são liberados em viveiros onde

ficam até alcançarem o tamanho e o peso ideal para comercialização. Por fim, ocorre a despesca que é a retirada da água dos tanques, coleta e abate dos camarões (Figura 1.5) (SENAR, 2016).

A realização dessas quatro etapas pode ocorrer na mesma localidade ou em lugares diferentes. Desta forma, o cultivo pode ser classificado em mono, bi ou trifásico. No cultivo monofásico as pós-larvas provenientes da larvicultura são estocadas diretamente nos viveiros de engorda e permanecem lá até a despesca final dos camarões. O cultivo bifásico é composto pela fase de berçário e de engorda, e as pós-larvas também são oriundas da larvicultura realizada por empresas especializadas. Já o cultivo trifásico é composto pelas etapas de larvicultura, berçário e engorda, sendo encontrado, principalmente, em grandes empreendimentos da carcinicultura (Figura 1.5) (Ribeiro et al., 2014).

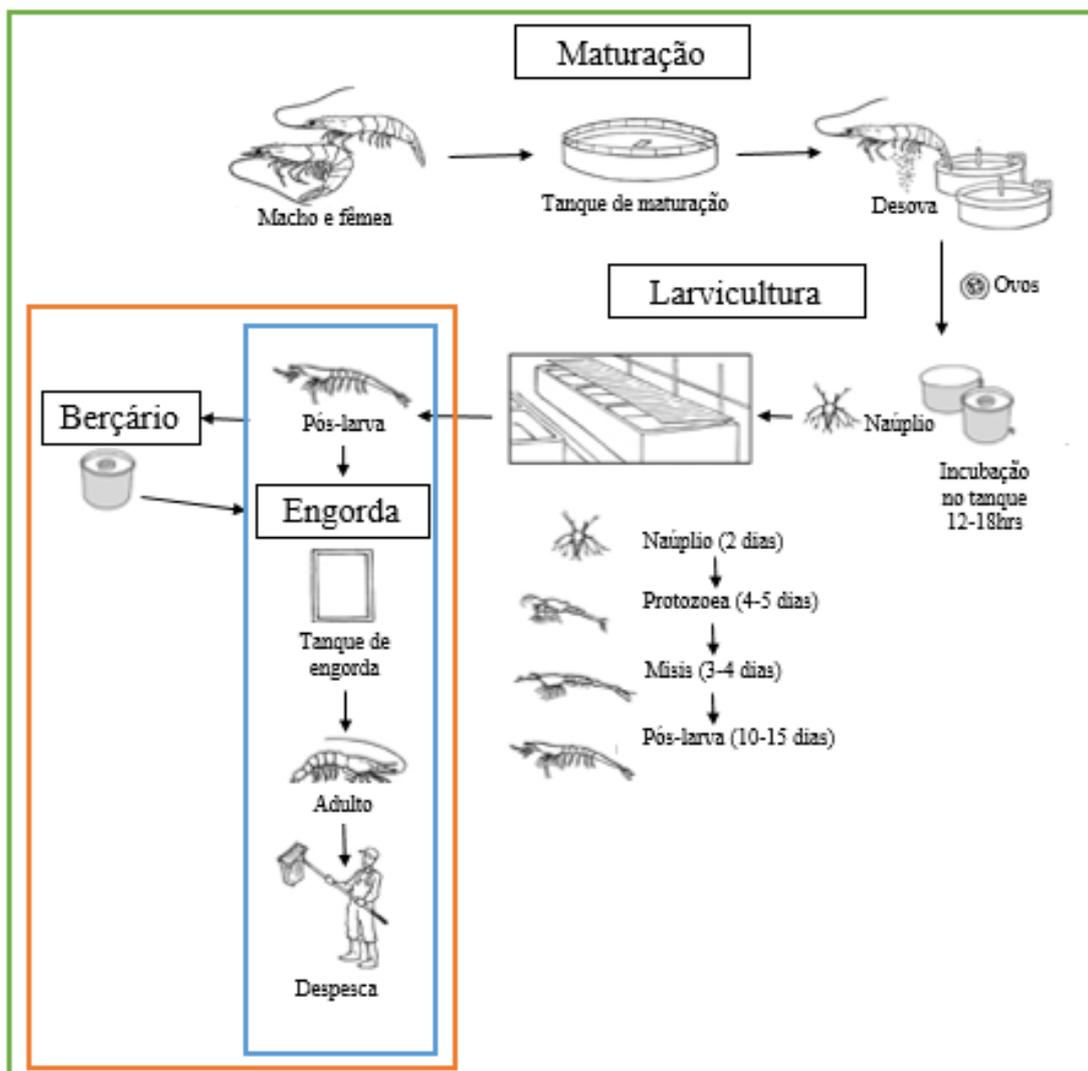


Figura 1.5: Etapas da produção do camarão cinza e fases presentes nos sistemas de cultivo mono (azul), bi (laranja) e trifásico (verde). Fonte: adaptado de (FAO 2019a).

Na fase de engorda são utilizadas técnicas que podem ser subdivididas em quatro categorias principais: extensivas, semi-intensivas, intensivas e hiperintensivas, que representam densidades de estocagem baixas, médias, altas e extremamente altas, respectivamente (FAO, 2019a). Essas técnicas se diferenciam, também, na forma de alimentação, na renovação de água, na localização das fazendas, no investimento para implantação e no rendimento obtido (Tabela 1.1) (Ribeiro et al., 2014).

Tabela 1.1: Características gerais das técnicas de cultivo na fase de engorda.

Tipo de técnica	Densidade	Tamanho lagoas/pistas	Rendimento
Extensiva	4-10PL*/m ²	5-10 ha (até 30 ha)	150-500 kg/ha/safra (1-2 safras/ano)
Semi-intensiva	10-30PL*/m ²	1-5 ha	500-2000 kg/ha/safra (2 safras/ano)
Intensiva	60-300PL*/m ²	0,1- 1 ha	7-20.000 kg/ha/safra (2-3 safras/ano)
Hiperintensiva	300-450JV ^o /m ²	282 m ²	28.000-68.000 kg/ha/safra (2-3 safras/ano)

Fonte: adaptado de (FAO 2019a)

*PL: Pós-larvas

^o Juvenis

A técnica de sistema extensivo ocorre em lagoas construídas próximas ao mar com o gerenciamento de água feito por troca de maré. A única alimentação dos camarões é natural, baseada no plâncton e nutrientes que entram no tanque por meio da circulação de água. Já na técnica semi-intensiva, os camarões são nutridos com alimentos naturais aprimorados pela fertilização em lagoas, suplementadas por dietas formuladas 2 ou 3 vezes ao dia, além de a renovação regular da água ser feita por bombeamento. Portanto, os custos de manutenção da técnica semi-intensiva são mais altos quando comparados à extensiva devido a esses fatores (FAO, 2019a; Ribeiro et al., 2014).

A técnica intensiva ocorre em fazendas localizadas em áreas sem maré. Os tanques podem ser completamente drenados, secos e preparados antes de cada lotação. É necessária aeração para circulação e oxigenação da água. A alimentação é feita por meio de dietas artificiais de 4 a 5 vezes ao dia (FAO, 2019a). Por fim, a técnica hiperintensiva é feita em sistemas fechados em estufas com pistas e sem renovação de água (ocorre somente

substituição da água evaporada), sendo assim uma técnica biossegura e ecologicamente correta, já que evita a contaminação dos corpos hídricos pelo aumento de carga orgânicas de águas oriundas dos cultivos. As técnicas intensivas e hiperintensivas apresentam maiores custos de manutenção devido à utilização de altas quantidades de rações balanceadas e à necessidade de areação artificial, gerenciamento da qualidade da água e, em alguns casos, antibióticos (FAO, 2019a; Ribeiro et al., 2014).

Dentre as quatro técnicas de cultivo abordadas, a semi-intensiva é a mais utilizada no Brasil. O clima favorável, a disponibilidade hídrica e o surgimento de novas tecnologias de cultivo de camarões favoreceram o estabelecimento do país como um dos principais produtores de camarão das Américas (Ribeiro et al., 2014). Entre 1997 e 2014, por exemplo, houve um aumento na produção de 3.600 para 85.000 toneladas (ABCC, 2015). O platô da produção ocorreu em 2003 com 90.190 toneladas, sendo a maior taxa produtiva entre todos os países produtores (Rocha, 2010). Em 2017, o Brasil ocupava o oitavo lugar na lista de produtores mundiais (Rocha, 2017). Entre as regiões brasileiras, a Nordeste é a que mais se destaca nas atividades de carcinicultura, sendo responsável por 99% do total da produção, tendo como notoriedade os estados do Rio Grande do Norte e da Bahia, além do Ceará, da Paraíba, do Piauí, de Pernambuco e de Sergipe, onde a atividade se apresenta em expansão (IBGE, 2019). Essa discrepância da produção em relação a outras regiões do país está associada, especialmente, à ampla zona costeira nordestina, às condições climáticas, à hidrologia e à topografia favoráveis para o cultivo do camarão (Castro e Pagani, 2004).

Além de ótimas condições ambientais para a carcinicultura, o Nordeste foi precursor em estudos de testes para produção e viabilidade para cultivo do gênero *Litopenaeus* no início da década de 1970, o que foi essencial para a implantação e o desenvolvimento da atividade no Brasil (Natori et al., 2011). A partir das informações obtidas por meio desses estudos foi concluído que o *L. vannamei* era a espécie mais adequada para o cultivo, devido à sua rusticidade, alta adaptabilidade, rapidez no crescimento, ampla faixa de tolerância à salinidade, entre outras razões (Costa, 2004).

1.3.3 Melhoramento genético do camarão cinza no Brasil

Atualmente, poucos programas de melhoramento genético de camarão cinza são descritos no Brasil e os existentes têm como foco a melhoria das características de crescimento e da taxa geral de sobrevivência. Um dos desafios encontrados nos programas é a manutenção

das informações de *pedigree*. A preservação desses dados pode ser feita com: i) a separação física das famílias e a utilização de elastômeros para marcação física (Figura 1.6A e B); ii) criação de grupos de múltiplas famílias e uso de marcadores moleculares (Figura 1.6C).



Figura 1.6: Estratégias de programas de melhoramento genético brasileiros para preservação das informações de *pedigree* do camarão cinza. (A) Tanques que separam as famílias fisicamente; (B) Marcação física realizada com elastômeros; (C) Fazendas de engorda com tanques onde ocorre a criação de pós-larvas de múltiplas famílias. Fonte: Teixeira et al, 2019.

Os programas que utilizam a primeira estratégia, só podem realizar a marcação individual com elastômeros quando os indivíduos apresentam em média o peso mínimo de 1,5g, atingido a partir de 5ª semana pós-eclosão. Como os dados da linhagem familiar já são mantidos logo após o nascimento, as larvas são separadas em tanques de acordo com a família (Moss e

Moss, 2009; Teixeira et al, 2019). Essas estratégias trazem algumas dificuldades, pois exigem maior investimento em infraestrutura de tanques para manter as famílias separadas até a marcação, além de haver maior probabilidade de erros nas estimativas dos parâmetros genéticos, uma vez que as variações ambientais observadas entre tanques podem ser confundidas com os efeitos genéticos das famílias.

A criação de grupos de animais de múltiplas famílias é mais desejável e pode gerar ganhos genéticos superiores, no entanto, requer a geração de informações de *pedigree* de cada indivíduo a partir da análise de paternidade e parentesco, com marcadores moleculares (Liu et al, 2016). Já existe um programa de melhoramento genético no Brasil que utiliza marcadores moleculares para a obtenção das informações de *pedigree*. Esse programa tem como principal foco a seleção de animais que toleram a pressão das cargas virais e bacterianas do ambiente e que resistam às oscilações de temperatura e salinidade. O plantel com as pós-larvas é levado para as fazendas de engorda para ser desafiado em condições idênticas àsquelas dos viveiros comerciais. Uma limitação que pode ocorrer em relação à obtenção das informações de *pedigree* nesse caso, é que o uso de marcadores moleculares pode ser proibitivo em termos de custo, uma vez que até o momento poucas tecnologias com este fim estão disponíveis para a espécie.

1.3.4 Tecnologias para a genotipagem de SNPs em camarão cinza

A plataforma Fluidigm é uma das tecnologias existentes para genotipagem de SNPs e utiliza sistema de circuitos fluídicos integrados (IFC – do inglês *Integrated Fluidics Circuits*), uma rede de canais microfluídicos, válvulas e câmaras com controle automático, que permitem diminuir drasticamente o número de passos de pipetagem e minimizar erros, o que torna o experimento mais robusto e homogêneo (Fluidigm, 2018).

A genotipagem⁶ é feita com ensaios SNP *type* utilizando os *primers*⁷ alelo específico ASP1, ASP2 (*Allele Specific Primer*), o *primer* locus⁸ específico LSP (*Locus-Specific Primer*) e as sondas⁹ marcadas com corantes FAM ou HEX. Os *primers* ASP1 e ASP2 possuem a sequência locus específica com o SNP de interesse na extremidade 3'¹⁰. As sondas

⁶ Processo pelo qual é identificado cada polimorfismo de interesse.

⁷ Fragmentos de DNA que se ligam por complementaridade ao início da sequência de DNA que se quer multiplicar.

⁸ Local determinado que o SNP ocupa num cromossomo

⁹ Fragmentos de DNA utilizados para identificar e se anelar a sequências de nucleotídeos específicos.

¹⁰ Orientação química relativa às extremidades duma cadeia de ácido nucleico.

marcadas com FAM e HEX contêm uma sequência complementar a que é encontrada na extremidade 5'¹⁰ da ASP1 e ASP2, respectivamente. Já o *primer* LSP é reverso sem marcador para o locus (Confidential: Fluidigm Internal Use Only) (Figura 1.7).

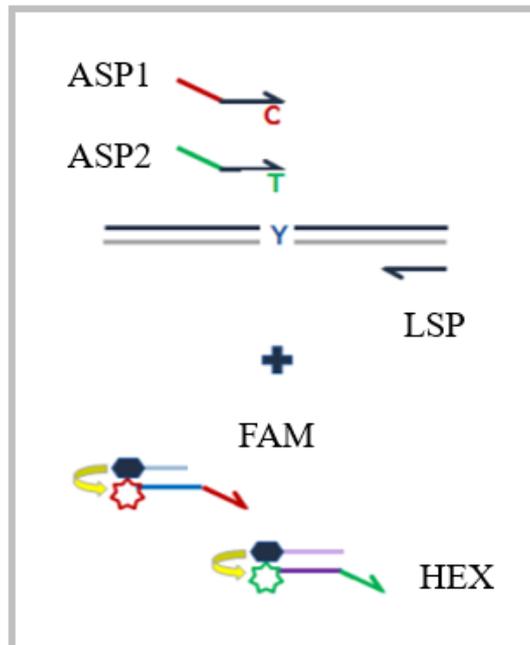


Figura 1.7: Componentes do SNP type. Fonte: Confidential: Fluidigm Internal Use Only.

Na genotipagem, os *primers* ASP1 e ASP2 em combinação com o *primer* LSP amplificam cada alelo e preparam o local para a incorporação da sonda (Figura 1.8A e B). Em ciclos subsequentes de PCR¹¹ (Reação em Cadeia da Polimerase), as sondas se anelam ao *amplicon*¹² (que tem sequência complementar à da sonda) para gerar produtos fluorescentes (Figura 1.8C). Ocorre a liberação do fluoróforo¹³ e os produtos marcados podem ser detectados pelo equipamento Fluidigm EP1 (Confidential: Fluidigm Internal Use Only) (Figura 1.8D e E).

¹¹ Técnica utilizada na biologia molecular para amplificar uma única cópia ou algumas cópias de um segmento de DNA em várias ordens de grandeza, gerando milhares a milhões de cópias de uma determinada sequência de DNA.

¹² Moléculas de DNA previamente amplificadas

¹³ Componente de uma molécula que faz com que esta seja fluorescente.

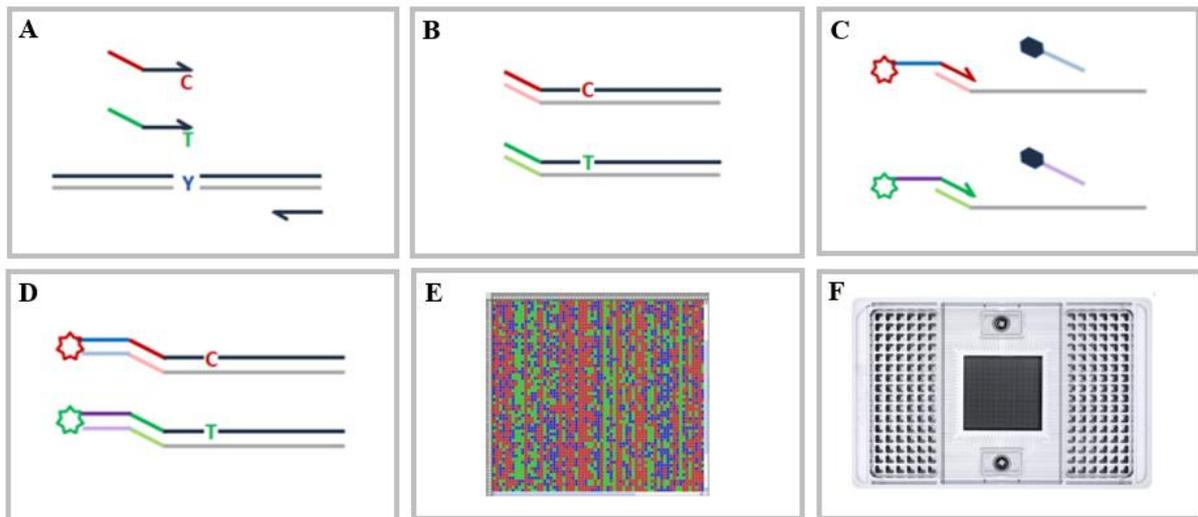


Figura 1.8: Genotipagem na plataforma Fluidigm: (A) Anelamento dos primers ASP e LSP ao alvo; (B) *Tags*¹⁴ incorporados no amplicon; (C) Sondas se anelam ao alvo; (D) Sondas incorporadas geram sinal; (E) Visualização do mapa de chamadas que exibe graficamente os resultados do genótipo em formato de chip; (F) Dynamic Array 96,96 IFC. Fonte: Confidential: Fluidigm Internal Use Only; Fluidigm (2018).

Essas genotipagens podem ser realizadas com diversos tipos de IFCs (Circuitos Fluídicos Integrados), como por exemplo, o Dynamic Array 96,96 que contém 96 poços para receber as amostras e 96 poços designados para a sonda e os *primers* pares (Figura 1.8F). A pressurização dos componentes no chip ocorre pelo uso do controlador de circuitos fluídicos integrados, sucedendo-se a combinação sistemática dos componentes em 9.216 reações paralelas (Fluidigm, 2018).

Outra tecnologia disponível para genotipagem de SNPs no camarão cinza é o chip comercial da Illumina Infinium® ShrimpLD-24 v1.0 (Figura 1.9A), que apresenta uma matriz com 6.465 SNPs distribuídos em todo o genoma (Jones et al., 2017). A superfície do chip é formada por esferas de sílica que são cuidadosamente gravadas, alojadas em poços e revestidas com cópias de sondas oligonucleotídicas¹⁵ (Figura 1.9B) (Illumina, 2012; Illumina, 2009). A genotipagem utilizando essa tecnologia, tem início quando os fragmentos de DNA do camarão cinza passam sobre o chip e a sonda de cada esfera se liga a uma sequência complementar do DNA da amostra. A ligação ocorre até alcançar uma base antes do alelo de interesse (Figura 1.9C). Posteriormente é feita uma extensão de base única em que ocorre a

¹⁴ Sequência curta de DNA complementar ao *amplicon*.

¹⁵ Fragmento curto de DNA com 20 ou menos bases de nucleotídeos.

incorporação de um dos quatro nucleotídeos (A, T, C, G) marcados com fluoróforo (Figura 1.9D). Um laser é utilizado para excitar o fluoróforo e o scanner detecta imagens de alta resolução da luz emitida (Illumina 2012; Illumina, 2009) (Figura 1.9E e F).

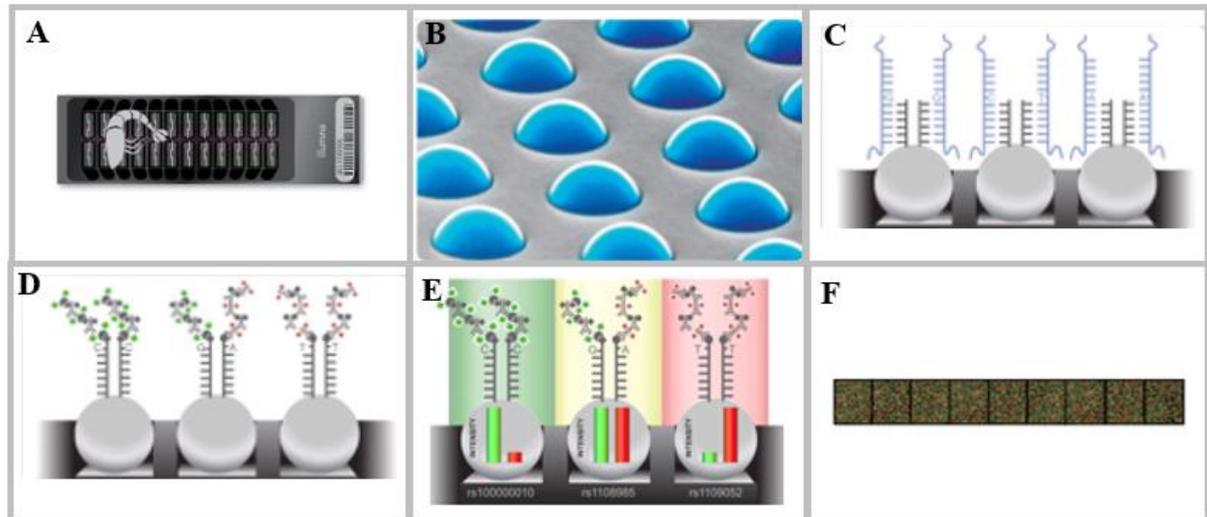


Figura 1.9: Processo de genotipagem no Infinium ShrimpLD-24 BeadChip. (A) Esferas de sílica presentes na superfície do chip; (B) Sondas presentes nas esferas se ligando a uma sequência complementar do DNA; (C) Extensão de base única para especificação do alelo; (D) Intensidade do sinal emitido para o scanner; (E) Imagem do chip emitindo fluoróforos. Fonte: Illumina (2016); Illumina (2012); Illumina, (2009).

1.3.5 Uso de marcadores SNPs para análise de paternidade, determinação de parentesco e avaliação da variabilidade genética na produção de espécies aquícolas

Nas populações comerciais de espécies aquícolas, o acasalamento entre parentes é um fenômeno frequente que ocorre, principalmente, devido à ausência de informações de *pedigree* ou pela seleção de animais geneticamente superiores para características de interesse econômico (Hashimoto et al., 2012; Kincaid, 1983). A falta de informações de *pedigree* impossibilitam o manejo genético de plantéis de reprodutores e leva ao acasalamento de indivíduos que tenham parentesco entre si (Moss e Moss, 2009). Já a seleção de animais geneticamente superiores para características economicamente importantes pode acarretar na escolha de indivíduos com alta capacidade produtiva, com genes em comum e com alto grau de parentesco, o que pode resultar no cruzamento entre indivíduos que pertencem à mesma família, ou seja, consanguinidade ou endogamia (Resende, 2007; Varela, 2015).

O acasalamento entre parentes compromete a qualidade do plantel e causa acúmulo de consanguinidade quando isso ocorre durante várias gerações (Moss e Moss, 2009). O acúmulo de endogamia aumenta a probabilidade de ocorrer a união de alelos prejudiciais em homozigose, provocando grande perda de indivíduos principalmente nos estágios iniciais do desenvolvimento, além de perdas na fase produtiva em decorrência da depressão endogâmica (Nakadate, 2003; Su et al., 1996). Esta causa uma diminuição no desempenho das taxas de crescimento, sobrevivência, resistência a doenças, dentre outras características (Moss e Moss, 2009).

A avaliação da endogamia e da diversidade genética é essencial nesse cenário porque melhora o manejo de matrizes, e auxilia os produtores a reduzir ou evitar a consanguinidade, além de aumentar a capacidade de uma população ter respostas positivas a alterações ambientais (Perez-Enriquez et al., 2009). No entanto, o controle dos acasalamentos entre parentes e a manutenção da alta variabilidade genética só são feitos quando as informações de *pedigree* dos indivíduos são conhecidas e utilizadas na orientação dos cruzamentos dos reprodutores do plantel (Moss e Moss, 2009).

A preservação das informações de *pedigree* em espécies aquícolas enfrenta obstáculos adicionais aos observados em outras espécies de interesse zootécnico, já que as formas jovens frequentemente apresentam tamanho milimétrico, e não podem ser marcadas individualmente. Em vista disso, são utilizadas estratégias que incluem a separação dos animais de um grupo familiar recém eclodidos em tanques individuais até que a marcação individual possa ocorrer (Moss e Moss, 2009). Esses processos geram custos operacionais e de capital elevados, além de aumentar a probabilidade de ocorrer erros nas estimativas dos parâmetros genéticos, já que os animais de uma mesma família são submetidos ao mesmo ambiente durante um período significativo da fase de crescimento. Métodos alternativos que utilizam ferramentas, como os marcadores moleculares, para determinação de paternidade e parentesco representam alternativas tecnicamente viáveis para circunver essas questões (Liu et. al, 2016).

Marcadores moleculares baseados em SNPs foram utilizados em estudos promissores em várias espécies aquícolas com intuito de inferir paternidade e parentesco. Thongda et al. (2018), por exemplo, empregaram SNPs de forma eficaz na avaliação da variabilidade genética e na inferência de paternidade em ostra oriental (*Crassostrea virginica*). Determinações das relações de paternidade e parentesco são feitas em diversos trabalhos, empregando diferentes métodos de análises de dados, como exclusão, atribuição fracional e categórica e reconstrução de *pedigree* (Flanagan e Jones, 2019).

Metodologias baseadas na exclusão de paternidade apresentam uma abordagem mais intuitiva, pois consideram a identificação de uma ou mais incompatibilidades mendelianas entre um par de indivíduos para concluir que um não pode ser pai do outro (Flanagan e Jones, 2019). Na literatura não é comum encontrar a utilização desse método em análises de paternidade em espécies aquícolas, no entanto, estudos utilizando essa abordagem em bovinos (Buchanan et al., 2017; McClure et al., 2015) e ovelhas (Heaton et al., 2014; Strucken et al., 2016) são frequentes.

Outra metodologia utilizada é a atribuição de paternidade. Essa abordagem é a mais comumente usada em análises de paternidade, podendo ser dividida nas categorias de alocações fracionárias e categóricas. A primeira consiste em uma técnica que aloca fracionadamente os filhos aos pais, ou seja, o possível pai apresenta maior probabilidade e recebe maior fração da prole, mas a prole também é parcialmente alocada a qualquer pai com uma probabilidade diferente de zero (Flanagan e Jones, 2019). Essa técnica foi utilizada em estudos com espécies aquícolas, como no mexilhão azul (*Mytilus galloprovincialis*), para designar famílias, estimar parâmetros genéticos e obter resposta à seleção na segunda geração de um programa de melhoramento (Nguyen et al., 2014). A alocação categórica, por outro lado, atribui cada filho inteiramente aos pais com maior probabilidade utilizando verossimilhança e baseia-se na comparação das probabilidades relativas de diferentes hipóteses sobre os relacionamentos entre parentes díades (pais e filhos) ou tríades (mãe-pai-filho) (Flanagan e Jones, 2019). Beacham et al. (2017) utilizaram alocação categórica para atribuir paternidade em salmão (*Oncorhynchus kisutch*) com 304 SNPs e 100% de acurácia, assim como Dussault e Boulding (2018) que empregaram o mesmo marcador molecular e metodologia para inferir paternidade em salmão do atlântico (*Salmo salar*).

Já a utilização do método de reconstrução de *pedigree* consiste na ideia de que o genótipo multilocus completo de um dos pais pode ser reconstruído se os genótipos do outro pai e de muitos descendentes forem conhecidos (Flanagan e Jones, 2019). Abadía-Cardoso et al. (2013), por exemplo, fizeram reconstrução de *pedigree* em truta-arco-íris (*Oncorhynchus mykiss*) e para isso utilizaram 4.895 amostras.

REFERÊNCIAS BIBLIOGRÁFICAS

Abadía-Cardoso, A. et al. (2013). Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*, 22, 4733–4746.

ABCC. (2015). Intensificação da Carcinicultura: Desafios, Caminhos e Oportunidade.

Anderson, E. C. e Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172, 2567–2582.

Bailey-Brock J. H e Moss S.M. (1992). Penaeid taxonomy, biology and Zoogeography. Marine shrimp culture: principles and practices. *Elsevier Science*, 42, 525-550.

Beacham, T. D. et al. (2017). Population and individual identification of coho salmon in british columbia through parentage-based tagging and genetic stock identification: An alternative to coded-wire tags. *Canadian Journal of Fisheries and Aquatic Sciences*, 74, 1391–1410.

Buchanan, J. W. et al. (2017). Analysis of validated and population-specific single nucleotide polymorphism parentage panels in pedigreed and commercial beef cattle populations. *Canadian Journal of Animal Science*, 97, 231–240.

Caetano, A. R. (2009). Marcadores SNP: Conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. *Revista Brasileira de Zootecnia*, 38, 64–71.

Castillo-Juárez, H. et al. (2015). Genetic improvement of Pacific white shrimp (*Penaeus (Litopenaeus)vannamei*): Perspectives for genomic selection. *Frontiers in Genetics*, 6, 1–4.

Castro, A. A. e Pagani, G. D. (2004). Secagem e composição química da cabeça de camarão (*Litopenaeus vannamei* Boone) a diferentes temperaturas. *Revista Brasileira de Produtos Agroindustriais*, 6(2), 123–129.

Costa, W. D. M. (2004). Efeito da Proteína Vegetal na Qualidade de Água dos Efluentes da Carcinicultura. Universidade Federal Rural de Pernambuco.

Dall, W. (1991). The Biology of the Penaeid. Morphology. *Advances in Marine Biology*, 27, 7–54.

CSIRO (2011). Prawn fishing industry in the Gulf of Carpentaria.

Dugassa, H. e Gaetan, D. G. (2018). Biology of White Leg Shrimp, *Penaeus vannamei*: Review. *World Journal of Fish and Marine Sciences*, 10(2), 5–17.

Dussault, F. M. e Boulding, E. G. (2018). Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: A methodology illustrated using Atlantic salmon. *Aquaculture Research*, 49, 1368–1372.

FAO. (2018). El estado mundial de la pesca y la acuicultura 2018. Cumplir los objetivos de desarrollo sostenible.

FAO. (2019a). Cultured Aquatic Species Information Programme: *Penaeus vannamei*. 1–14.

FAO. (2019b). Species Fact Sheets: *Penaeus vannamei*. 1–14.

Farfante, P.I. e Kenslev, B. (1997). Penaeoid and sergestoid shrimps and prawns of the world: Keys and dignoses for the families and genera. *The Quarterly Review of Biology*, 175, 10-55.

Flanagan, S. P. e Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, 28, 544–567.

Fluidigm. (2018). SNP Genotyping User Guide (PN 68000098 Q1).

Gitterle, T. et al. (2005). Genetic (co)variation in harvest body weight and survival in *Penaeus (Litopenaeus) vannamei* under standard commercial conditions. *Aquaculture*, 243, 83–92.

Hashimoto, D. T. et al. (2012). Cartilha de Genética na Piscicultura: Importância da variabilidade genética, marcação e coleta para análise de DNA.

Heaton, M. P. et al. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PloS One*, 9(4), 1–10.

Holman, L. E. et al. (2017). A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture*, 476(March), 59–64.

IBGE. (2019). Tabela 3940 – Produção da aquicultura, por tipo de produto. Disponível em <<https://sidra.ibge.gov.br/tabela/3940#resultado>>. Acesso em: 10 de maio, 2019.

Illumina. (2009). Infinium® HD Assay Ultra Protocol Guide. 1–225.

Illumina. (2012). Illumina Advances Genomic Research with the Infinium Assay. Disponível em: <<https://www.youtube.com/watch?v=IVG04dAAyvYet=119s>>. Acesso em: 25 de outubro, 2019.

Illumina. (2016). Infinium ShrimpLD-24 v1.0 BeadChip.

Jones, D. B. et al. (2017). A comparative integrated gene-based linkage and locus ordering by linkage disequilibrium map for the Pacific white shrimp, *Litopenaeus vannamei*. *Scientific Reports*, 7, 1–16.

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

Liu, S. et al. (2016). Development and validation of a SNP panel for parentage assignment in

rainbow trout. *Aquaculture*, 452, 178–182.

Ma, K. Y. et al. (2011). Refuting the six-genus classification of *Penaeus* (*Dendrobranchiata*, *Penaeidae*): A combined analysis of mitochondrial and nuclear genes. *Zoologica Scripta*, 40(5), 498–508.

Maggioni, R. et al. (2013). Genetic variability of marine shrimp in the Brazilian industry. *Pesquisa Agropecuária Brasileira*, 48(8), 968–974.

McClure, M. C. et al. (2015). SNP Selection for Nationwide Parentage Verification and Identification in Beef and Dairy Cattle. *International Committee for Animal Recording (ICAR)*, 175–181.

Moss, S.M. e Moss, D. R. (2009). Selective breeding of penaeid shrimp. *Shellfish Safety and Quality*.

Nakadate, M. et al. (2003). Inbreeding depression and heterosis in various quantitative traits of the guppy, *Poecilia reticulata*. *Aquaculture*, 220, 219–226.

Natori, M. M. et al. (2011). Desenvolvimento da Carcinicultura Marinha no Brasil E no Mundo : avanços tecnológicos e desafios. *Informações Econômicas*, 41(2), 61–73.

Nguyen, T. T. T. et al. (2014). Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. *Aquaculture*, 420–421, 295–301.

Perez-Enriquez, R. et al. (2009). Genetic diversity status of White shrimp *Penaeus* (*Litopenaeus*) *vannamei* broodstock in Mexico. *Aquaculture*, 297(1–4), 44–50.

Perez-Enriquez, R. et al. (2018). Genetic tracing of farmed shrimp (Decapoda: Penaeidae) in wild populations from a main aquaculture region in Mexico. *Revista de Biología Tropical*, 66(1), 381–393.

Resende, E. K. et al. (2007). Melhoramento Animal no Brasil: Uma visão crítica espécies aquáticas.

- Ribeiro, L. F. et al. (2014). Desafios da carcinicultura: aspectos legais, impactos ambientais e alternativas mitigadoras. *Journal of Integrated Coastal Zone Management*, 14(3), 365–383.
- Ribeiro, P. R. (2017). Carcinicultura no Agreste Paraibano: Qualidade da Água, um Fator Limitante. Universidade Federal da Paraíba.
- Rocha, I. P. (2010). Carcinicultura Brasileira: Processos Tecnológicos, Impactos Sócio-Econômicos, Sustentabilidade Ambiental, Entraves e Oportunidades. *February*, 1–10.
- Rocha, I. P. (2017). Carcinicultura Marinha: Realidade Mundial e Desafios Confrontados no Brasil. 77.
- SENAR. (2016). Larvicultura de camarão marinho : do náuplio a pós-larva.
- Strucken, E. M. et al. (2016). How many markers are enough? Factors influencing parentage testing in different livestock populations. *Journal of Animal Breeding and Genetics*, 133, 13–23.
- Su, G. S. et al. (1996). Effects of inbreeding on growth and reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 142, 139–148.
- Sucharita, V. e Jyoti, S. (2013). An Identification of Penaeid Prawn Species Based on Histogram Values. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(7), 807–811.
- Teixeira, A. K. G, et al. (2019). Melhoramento genético de camarão marinho livre de doenças. Disponível em: <<https://panoramadaaquicultura.com.br/melhoramento-genetico-de-camarao-marinho-livre-de-doencas/>>. Acesso em: 10 de maio, 2019.
- Thongda, W. et al. (2018). Development of SNP Panels as a New Tool to Assess the Genetic Diversity, Population Structure, and Parentage Analysis of the Eastern Oyster (*Crassostrea virginica*). *Marine Biotechnology*, 20, 385–395.
- Varela, E. S. (2015). Parentesco genético em reprodutores de tambaqui (*Colossoma*

macropomum) baseado em marcadores de DNA: perspectivas de manejo genético na ausência de pedigree. *Boletim de Pesquisa e Desenvolvimento*.

Vela Avitúa, S. et al. (2013). Decline of genetic variability in a captive population of Pacific white shrimp *Penaeus (Litopenaeus) vannamei* using microsatellite and pedigree information. *Electronic Journal of Biotechnology*, 16(4), 1–10.

Vignal, A. et al. (2002). A review on SNPs and other types of molecular markers and in their use in animal genetics. *Genetics, Selection, Evolution : GSE*, 34, 275–305.

Wang, H. X. et al. (2006). The application of microsatellite markers for parentage determination in selective breeding of Pacific white shrimp (*Litopenaeus vannamei*). *Journal Hereditas*, 28(2), 179-183.

CAPÍTULO 2

VALIDAÇÃO DE UM PAINEL DE SNPS DE BAIXA DENSIDADE PARA ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus vannamei*)

2.1 RESUMO

VALIDAÇÃO DE UM PAINEL DE SNPS DE BAIXA DENSIDADE PARA ANÁLISE DE PATERNIDADE E PARENTESCO E AVALIAÇÃO DA VARIABILIDADE GENÉTICA DO CAMARÃO CINZA (*Litopenaeus vannamei*)

Nayelle Meyre Lisboa Silva¹, Patrícia Ianella², Alexandre Rodrigues Caetano³

1 Bióloga/ Mestranda em Ciências Animais – PPG/UnB – Brasília/DF

2 Bióloga/ Doutora – Embrapa Recursos Genéticos e Biotecnologia

3 Zootecnista/ Doutor – Embrapa Recursos Genéticos e Biotecnologia

O objetivo deste estudo foi validar um painel SNP de baixa densidade em populações de núcleo de melhoramento genético e em amostras de camarão coletadas em lojas de varejo, para atribuir paternidade e parentesco, avaliar variabilidade e estruturação genética em camarão cinza. Amostras de reprodutores de três linhagens comerciais (A, B e C) e amostras de sete lojas de varejo (SPDF1-SPDF5, PEDF e PEPB) foram avaliadas, totalizando 328 amostras. Inicialmente, a validação foi feita utilizando 144 SNPs para seleção dos 96 marcadores mais informativos que iriam compor o painel. A genotipagem dos SNPs foi realizada com plataforma Fluidigm e em seguida foram feitas análises de estruturação populacional, verificação da Heterozigosidade observada (H_o) e esperada (H_e), avaliação da qualidade, por meio do RMSD (do inglês *Root Mean Square Deviation*) e RMSE (do inglês *Root Mean Square Error*), das estimativas de parentesco e da probabilidade de não exclusão parental. Os resultados de PCA (Análise de Componentes Principais) e STRUCTURE das três linhagens do núcleo de melhoramento genético agregaram as amostras das linhagens A e B no mesmo grupo, enquanto a linhagem C apresentou uma composição diferenciada, indicando que a origem do germoplasma das linhagens A e B é o mesmo, e/ou que existe um forte fluxo gênico entre as

duas linhagens. Em relação a estruturação das 10 populações, foi observado que as amostras dos varejistas não são oriundas do produtor de larvas abordado nesse estudo e que todas as populações das lojas de varejo aparentam ter origem de um mesmo produtor. Já as estimativas de H_o e H_e indicaram alta variabilidade genética das amostras, pois a maior variação encontrada entre essas estimativas foi de 4%, sendo que a média dos valores de H_o para todas as populações foi de 0,392. Quanto a qualidade das estimativas de parentesco, foi demonstrado que o valor de RMSD calculado entre as estimativas obtidas com o painel de 96 SNPs e as adquiridas com 3.857 SNPs do chip comercial da Illumina foi de 0,08. Já o RMSE desses dois grupos variou somente 0,05, indicando que os intervalos de confiança são muito próximos. O RMSD e RMSE indicam que o painel de 96 SNPs é tão eficaz para estimar parentesco quanto a ferramenta comercial da Illumina. As probabilidades de não exclusão das 10 populações também demonstraram resultados satisfatórios para P1, P2 e P3 de $4,5 \times 10^{-7}$, $4,4 \times 10^{-4}$ e $5,2 \times 10^{-11}$, respectivamente. Portanto, o painel desenvolvido no presente estudo poderá ser utilizado em estudos amplos que demandem grandes números de atribuições de paternidade, avaliações de parentesco e análises de estruturação populacional do camarão cinza.

PALAVRAS-CHAVE: Brasil, carcinicultura, melhoramento genético, *pedigree*.

2.2 ABSTRACT

VALIDATION OF A LOW DENSITY SNPS PANEL FOR PATERNITY AND KINSHIP ANALYSIS AND EVALUATION OF GENETIC VARIABILITY IN PACIFIC WHITE SHRIMP (*Litopenaeus vannamei*)

The objective of this study was to validate a low density SNP panel in breeding nucleus populations and in shrimp samples collected in retail stores, to attribute paternity and kinship, to assess genetic variability and structure in pacific white shrimp. Broodstock samples from three commercial lines (A, B and C) and samples from seven retail stores (SPDF1-SPDF5, PEDF and PEPB) were evaluated, totaling 328 samples. Initially, the validation was done using 144 SNPs to select the 96 most informative markers that would compose the panel. Genotyping was performed with a Fluidigm EP1 platform, followed by population structure analysis, verification of the observed (H_o) and expected (H_e) heterozygosity, and calculation of Root Mean Square Deviations (RMSD) and Root Mean Square Errors (RMSE) for kinship estimates obtained with different numbers of SNP markers. PCA (Principal Component Analysis) and STRUCTURE results considering the three Broodstock lines grouped samples from lines A and B, while line C showed a different composition, indicating that the origin of the germplasm of lines A and B is the same, and / or that there is a strong gene flow between the two lines. Regarding the structuring of the 10 populations, it was observed that the retailers' samples are not from the larvae producer addressed in this study and that all the populations of the retail stores appear to have originated from the same producer. H_o and H_e estimates indicated high genetic variability of the samples, since the greatest variation found between these estimates was 4%, with the average H_o values for all populations being 0.392. Regarding the quality of kinship estimates, it was demonstrated that the RMSD value calculated between the estimates

obtained with the 96 SNP panel and those acquired with 3,857 SNPs from Illumina's commercial chip was 0.08. The RMSE of these two groups varied only 0.05, indicating that the confidence intervals are very close. The RMSD and RMSE indicate that the 96 SNP panel is as effective at estimating kinship as Illumina's commercial tool. The probabilities of non-exclusion of the 10 populations also demonstrated satisfactory results for P1, P2 and P3 of 4.5×10^{-7} , 4.4×10^{-4} and 5.2×10^{-11} , respectively. Therefore, the panel developed in the present study can be used in broad studies that require large numbers of paternity assignments, kinship assessments and analyzes of the pacific white shrimp population structure.

KEYWORDS: Brazil, genetic improvement, *pedigree*, shrimp farming.

2.3 INTRODUÇÃO

O camarão cinza (*Litopenaeus vannamei*) é a espécie de penaeideo mais cultivada em todo o mundo, com produção mundial de 4,1 milhões de toneladas em 2016 (FAO, 2018). Devido à importância econômica da espécie, foram implantados programas de melhoramento genético com o objetivo de aumentar a eficiência da produção por meio da seleção de indivíduos que apresentem melhores taxas de crescimento, sobrevivência e resistência a doenças (Castillo-Juárez et al., 2015; Crock et al., 2009; Gitterle et al., 2005; Lightner et al., 2012).

Um dos desafios do melhoramento genético é manter as informações de *pedigree* dos indivíduos em avaliação, as quais são essenciais para estimar parâmetros genéticos, identificar indivíduos geneticamente superiores para características de interesse econômico, orientar acasalamentos, e assim, minimizar as taxas de consanguinidade e a consequente perda de variabilidade genética nas linhagens de reprodutores (Moss e Moss, 2009)

A preservação das informações de *pedigree* no camarão cinza é feita com a criação das larvas em tanques com grupos familiares, seguido da marcação individual de juvenis com elastômeros. Essas estratégias trazem algumas dificuldades. A primeira delas é que a marcação física só é possível quando os indivíduos apresentam em média o peso mínimo de 1,5g (Teixeira et al, 2019; Moss e Moss, 2009), já com cerca de 53 dias de vida. A segunda é que a criação de famílias individualizadas demanda maiores investimentos em infraestrutura e instalações. A terceira é o aumento no erro das estimativas dos parâmetros genéticos, uma vez que os efeitos ambientais observados entre tanques podem ser confundidos com os efeitos genéticos das famílias. Portanto, considerando os pontos citados, a criação de grupos de animais de múltiplas famílias, formados com animais ainda em estágio juvenil, é mais adequada para execução de avaliações genéticas e estimar parâmetros e valores genéticos de famílias e

indivíduos, visando a identificação e seleção de animais superiores para características de interesse econômico.

Nesses sistemas, a obtenção das informações de *pedigree* é possível com a utilização de informações de marcadores moleculares geradas a partir de amostras de animais que já atingiram a idade de interesse (Liu et al, 2016). Os marcadores microssatélites foram amplamente utilizados para atribuição de paternidade (Aungsuchawan et al., 2008; Lee, 2013), parentesco (Artiles et al, 2011) e avaliação da variabilidade genética em camarão cinza (Garcia e Alcivar-Warren, 2007; Perez-Enriquez et al., 2009). No entanto, os polimorfismos de nucleotídeo único (SNPs – do inglês *Single Nucleotide Polymorphism*) passaram a ser empregados em estudos genéticos mais recentes em *L. vannamei* (Jones et al., 2017; Perez-Enriquez e Max-Aguilar, 2016) e em outras espécies aquícolas de interesse zootécnico (Harlizius et al., 2011; Heaton et al., 2014; Heaton, et al., 2002; Rohrer et al., 2007). O amplo uso dos SNPs ocorre devido a sua abundância no genoma, ao baixo custo de genotipagem, à fácil padronização de ensaios entre laboratórios, e às menores taxas de erro de genotipagem (Vignal et al., 2002; Caetano, 2009).

Segundo Anderson e Garza (2006), o uso de 60-100 SNPs com frequência alélica menor (*Minor Allele Frequency* - MAF) alta (0,3-0,5) são suficientes para fornecer uma atribuição de paternidade eficaz. Sellars et al. (2014) compararam a precisão de painéis com 59, 63 e 122 SNP no camarão tigre (*Penaeus monodon*) e concluíram que todos os conjuntos de SNPs forneceram taxas de atribuição de paternidade adequadas, enquanto Karlsson et al. (2011) verificaram que um painel com 60 SNPs é viável para diferenciar geneticamente estoques de salmão (*Salmo salar*) selvagens e cultivados.

Assim como foi observado em outras espécies aquícolas, o desenvolvimento e validação de um painel de SNPs de baixa densidade (até 100 SNPs) para camarão cinza seria de grande utilidade para o setor produtivo. Essas ferramentas moleculares fornecem dados que podem aprimorar os programas de melhoramento genético, auxiliando na manutenção da variabilidade, facilitando a identificação individual, reduzindo a probabilidade de erros nas estimativas dos parâmetros genéticos e diminuindo os custos com a criação de famílias isoladas. Devido a todos esses fatores, o objetivo deste estudo foi validar um painel SNP de baixa densidade em populações de núcleo de melhoramento genético e em amostras de camarão coletadas em lojas de varejo, para atribuir paternidade, parentesco e avaliar a variabilidade genética em camarão cinza (*L. vannamei*).

2.4 MATERIAL E MÉTODOS

2.4.1 Amostragem de camarão cinza

Foram utilizadas 328 amostras oriundas de um produtor de larvas de camarão (parceiro da nossa equipe) que tem 12,8% de participação no mercado nacional, e de sete lojas de varejo. Um total de 137 amostras do produtor foi obtido de reprodutores de diferentes famílias (para reduzir o viés ocasionado por efeito amostral) de três linhagens denominadas A, B e C, provenientes de núcleos de melhoramento genético para crescimento rápido.

As demais 191 amostras foram coletadas em cinco supermercados (SPDF1 a SPDF5) e uma peixaria no Distrito Federal (PEDF), e em uma peixaria na Paraíba (PEPB). A coleta das amostras de PEPB foi realizada em um momento oportuno de visita ao estado. A maioria das amostragens foi realizada com camarões frescos a granel e somente os indivíduos da população SPDF3 estavam congelados e ensacados. Durante a coleta de amostras de varejistas, foi registrada a origem de produção do material, conforme relato verbal de cada comerciante (Tabela 2.1).

Tabela 2.1: Amostras e populações de *L. vannamei* utilizadas no estudo.

	Pop	Nº de indivíduos	Origem da produção relatada
Produtor de larvas	A	20	RN
	B	57	RN
	C	60	RN
	Total	137	-
Lojas de varejo	SPDF1	36	CE
	SPDF2	24	CE
	SPDF3	35	NR*
	SPDF4	24	CE
	SPDF5	24	RN
	PEDF	24	MA
	PEPB	24	RN
	Total	191	-

NR*: Não relatado

2.4.2 Extração de DNA

As amostras de camarão cinza do produtor de larvas (A, B e C) e das lojas de varejo (SPDF1, SPDF3 e PEPB) tiveram o DNA extraído de tecidos musculares com método CTAB adaptado por Boyce et al. (1989). O DNA das amostras coletadas nas demais localidades foi extraído utilizando metade de um exopódito e com método de digestão em tampão com proteinase K (50µl de 50mM KCl, 10mM Tris-HCl pH9,0, 0.1% Triton X-100, com 0,04mg de pK). A qualidade e a quantidade de DNA obtido pelo método CTAB foram avaliadas em eletroforese de gel de agarose a 1% e por espectrofotometria com NanoDrop™ 8000.

2.4.3 Genotipagem com plataforma Fluidigm

A prospecção de SNPs foi realizada, antes da execução deste trabalho, por sequenciamento de alta cobertura do genoma do camarão cinza com complexidade reduzida, utilizando tecnologia Illumina (Ianella et al., *em preparação*). Pools de 10µg de DNA dos 130 indivíduos das linhagens A, B e C foram digeridos com a enzima *Hae III*. Fragmentos de

tamanhos variáveis (400-500; 500-650; e 650-850pb) foram submetidos ao sequenciamento com plataforma Illumina, gerando mais de 1,6 milhões de *reads*. Um total de 3.895,392 SNPs foi encontrado com cobertura >50. Em seguida, os SNPs candidatos para o painel foram selecionados utilizando como critérios: i) MAF >0,45; ii) sequência mínima de 60 pb em cada lado do SNP alvo para desenvolver ensaios do tipo SNPtype™ na plataforma Fluidigm; iii) ausência de SNPs adjacentes dentro de 30 pb em cada lado do SNP alvo; iv) número de SNPs equivalente a quantidade de ensaios presentes no sistema Dynamic Array 96.96 IFCs (*Integrated Fluidic Circuits*) (Ianella et al., *em preparação*).

Posteriormente, sequências selecionadas de 96 SNPs foram submetidas à Fluidigm (Sul de San Francisco, CA, EUA) para desenho e síntese de *primers* para um painel de genotipagem SNPtype™. Quatro *primers* foram desenhados para cada ensaio, sendo dois *primers* específicos de cada alelo (ASP), um *primer* específico de locus (LSP) e um *primer* de amplificação de alvo específico (STA) (Apêndice A). A genotipagem no Fluidigm foi realizada de acordo com as instruções do fabricante (Fluidigm, 2018). Resumidamente, um *pool* contendo os *primers* STA e LSP dos 96 ensaios foi utilizado na pré-amplificação do DNA genômico, usando 1,5 µl de amostra (~ 50ng/µl) para enriquecer as sequências de SNPs de interesse. Os produtos pré-amplificados foram diluídos e utilizados na genotipagem com sistema Dynamic Array 96.96 IFCs. Este sistema reúne automaticamente reações de PCR, permitindo testes simultâneos de até 96 amostras e 96 marcadores SNP. Após a amplificação, as matrizes geradas foram lidas usando o sistema EP1 e os genótipos foram chamados automaticamente usando o software de análise de genotipagem Fluidigm SNP 4.1 com limiar mínimo de confiança de 95% (Fluidigm, 2018).

2.4.4 Validação do painel com amostras de reprodutores de linhagens comerciais

A validação inicial do painel de SNPs foi realizada com as amostras de reprodutores de linhagens comerciais de um produtor de larvas, utilizadas inicialmente para a descoberta dos marcadores, com os 96 SNPs genotipados nesse estudo e 48 SNPs que já haviam sido testados previamente (Ianella et al., *em preparação*). Desse modo, um total de 144 SNPs foi utilizado na seleção dos 96 marcadores mais informativos para compor o painel de parentesco e variabilidade genética do camarão cinza. A seleção foi baseada em análises feitas no software *SNP e Variation Suite* (SVS – *Golden Helix*). Primeiramente, foram descartadas

sete amostras e 21 SNPs que apresentaram Call Rate $<0,90$. Um total de 21 SNPs adicionais foi eliminado por apresentarem desvios de Hardy Weinberg ($p <0,01$). Dois marcadores foram excluídos por apresentarem MAF $<0,01$ e nenhum marcador foi eliminado por desequilíbrio de ligação, pois todos apresentaram $r^2 <0,10$. Após a realização dessas filtragens, restaram 130 amostras e 100 SNPs. A partir desse conjunto, foi feita a seleção final dos 96 marcadores para compor o painel, com a exclusão dos quatro SNPs com menor MAF.

2.4.4.1 Estruturação populacional

Os 96 SNPs mais informativos foram utilizados em análises de componentes principais (PCA - *Principal Component Analysis*) com o software *SNP e Variation Suite* (SVS – Golden Helix). Análises de estruturação populacional foram realizadas com o software STRUCTURE 2.3.4 (Pritchard et al., 2000), e o melhor valor de K foi obtido com o pipeline CLUMPAK (Kopelman et al., 2015). Além disso, foram feitas atribuições populacionais para as linhagens A, B e C e calculados valores de F_{st} com o programa Genalex 6.5 (Peakall e Smouse, 2012).

2.4.4.2 Análise de parentesco

O painel de 96 SNPs e as 137 amostras dos reprodutores das linhagens A, B e C foram utilizadas para obter Probabilidades de Não Exclusão Parental (probabilidade de não excluir o pai correto) e estimar parentesco com o software Coancestry (Wang, 2011). Em seguida, foram selecionadas 48 amostras com menor relação de parentesco para genotipagem com chips Illumina Infinium® ShrimpLD-24 (com 6.465 SNPs).

O controle de qualidade das amostras e dos 6.465 SNPs (Call Rate $>0,90$, marcadores bialélicos, Equilíbrio de Hardy-Weinberg (HWE) com valor de $p > 0,01$ e Frequências Alélicas (MAFs) próxima de 0,5) realizada com o software *SNP e Variation Suite* (SVS – Golden Helix) eliminou duas amostras e 2.232 marcadores, restando 46 amostras e 4.233 SNPs. Enquanto no painel com 96 SNPs, foram mantidas 46 amostras e 88 marcadores (Call Rate $>0,90$) para as análises subsequentes. Após a realização dessas filtragens, diferentes desbastes (*prunning*) dos dados foram feitos com base em estimativas de Desequilíbrio de

Ligação ($r^2 < 0,02; 0,05; 0,10; 0,20; 0,50$) para obter cinco subconjuntos de SNPs oriundos do chip Illumina.

A etapa subsequente foi realizada para estimar o parentesco entre as amostras utilizando o software Coancestry (Wang, 2011) e o estimador TriML desenvolvido por Wang (2007) para verificar a diferença da raiz do desvio quadrático médio (RMSD – do inglês *Root Mean Square Deviation*) das estimativas obtidas com os 96 SNPs e com os subconjuntos de marcadores para avaliar a eficácia do painel validado. Para isso, foi utilizada a seguinte equação:

$$RMSD = \sqrt{\frac{\sum_{n=1}^x (x_e - x_o)^2}{n}}$$

onde, x_e é a estimativa de parentesco esperada, x_o é a estimativa de parentesco observada e n é o número total dos valores das estimativas de parentesco.

O Intervalo de Confiança (IC) das estimativas de parentesco do painel com 96 SNPs e dos subconjuntos oriundos dos chips foi utilizado para calcular a raiz do erro quadrático médio (RMSE – do inglês *Root Mean Square Error*) das estimativas (Wang, 2007). A seguinte equação foi usada:

$$RMSE: \sqrt{\frac{\sum_{n=1}^x \left(\frac{CI95}{2 \times 1,96}\right)^2}{n}}$$

onde, $CI95$ é o intervalo de confiança de 95% gerado pelo estimador utilizado (Wang, 2007), e n é o número total dos valores das estimativas de parentesco.

2.4.5 Análises de estruturação populacional, variabilidade genética e parentesco de amostras de camarão de varejistas

Inicialmente, o painel com 96 SNPs e as 328 amostras das três linhagens e das sete lojas de varejo passaram por um controle de qualidade que eliminou três amostras e 10 marcadores com Call Rate $< 0,90$. Estimativas de MAF, Heterozigidade Esperada (He), Heterozigidade Observada (Ho) e Coeficiente de Endogamia (Fis) foram obtidas com o

software Genalex 6.5. Subsequentemente, foram realizadas análises de parentesco de Probabilidade de Não Exclusão Parental com o software Coancestry.

A análise de componentes principais (PCA - *Principal Component Analysis*) foi realizada no software *SNP e Variation Suite*. A estruturação populacional foi obtida com o software STRUCTURE v2.3.4 e o melhor valor de K foi verificado por meio do pipeline CLUMPAK.

2.5 RESULTADOS

2.5.1 Validação do painel com amostras de reprodutores de linhagens comerciais

2.5.1.1 Estruturação populacional

Os resultados das análises das três primeiras componentes principais foram responsáveis por 11,8% da variação total das linhagens A, B e C, em que o PC1 foi responsável por 5,47%, PC2 por 3,39% e PC3 por 2,94% das variações dos dados. A PCA agregou as amostras das linhagens A e B no mesmo grupo, enquanto a linhagem C apresentou estruturação em relação às outras linhagens (Figura 2.1A). Os resultados obtidos com o software STRUCTURE corroboraram os resultados de PCA, uma vez que a análise de agrupamentos também identificou dois grupos genéticos distintos nas linhagens do produtor de larvas, indicados pela maior probabilidade para $K=2$ (Figura 2.1B). As amostras de A e B foram atribuídas a um mesmo agrupamento, enquanto a linhagem C representou o segundo agrupamento (Figura 2.1C).

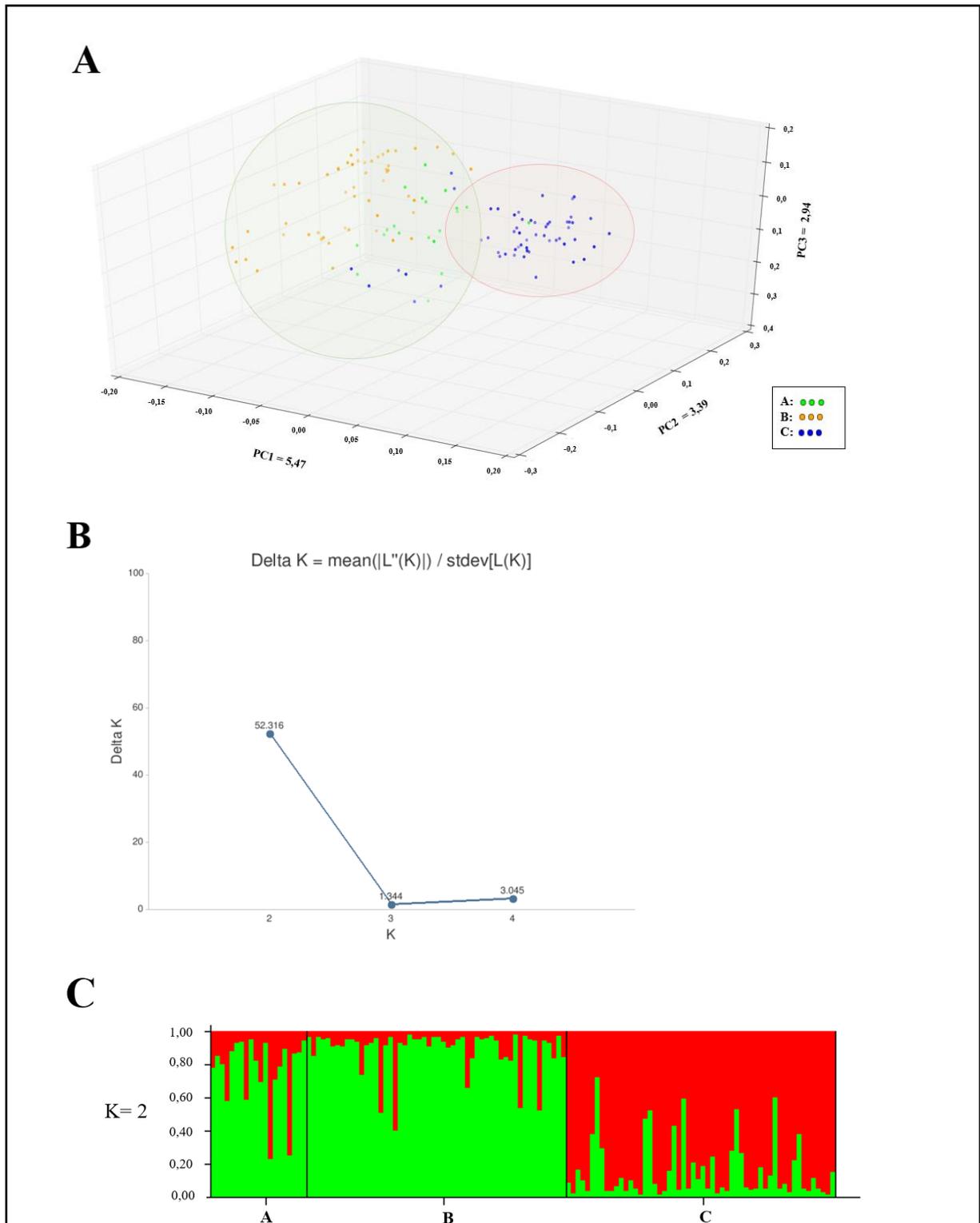


Figura 2.1: Estrutura populacional obtida com o painel de 96 SNPs e 130 amostras das três linhagens (A, B e C) do produtor de larvas. (A) Plot 3D com análises de componentes principais (PCA), em que PC1, PC2 e PC3 correspondem a 5,47, 3,39 e 2,94%, respectivamente, do total da variância observada das linhagens; (B) Representação gráfica do melhor valor de $K=2$; (C) Estruturação das três linhagens, sendo que as três populações analisadas foram separadas por uma linha vertical.

Análises subsequentes das estimativas de F_{st} variaram entre 0,030 e 0,042, sendo que a menor variação encontrada foi entre as linhagens A e B e a maior entre B e C (Tabela 2.2). Esses dados corroboraram com os dados de STRUCTURE e PCA abordados anteriormente.

Tabela 2.2: Análise de F_{st} em pares entre as três linhagens do produtor de larvas.

	A	B	C
A	0,000		
B	0,030	0,000	
C	0,039	0,042	0,000

A atribuição dos indivíduos das linhagens às suas respectivas populações utilizando o painel com 96 SNPs é mostrada na Tabela 2.3. A maioria dos indivíduos (95%) das três linhagens foi atribuída à sua população de origem, e somente 5% foi designado a outras populações. Cerca de 90% das amostras da linhagem A, 98% da B e 93% da C foram atribuídas às suas próprias linhagens, respectivamente. Dessa forma, a população B foi a que apresentou maior porcentagem de atribuição das amostras a população referência e apenas um indivíduo foi designado a outra linhagem.

Tabela 2.3: Atribuição de população com 96 SNPs para 130 indivíduos das linhagens A, B e C do produtor de larvas.

	População referência	Amostras alocadas em outras populações			
		A	B	C	Total
A	18 (90%)	1	1	2 (10%)	
B	53 (98%)		1	1 (2%)	
C	52 (93%)	4		4 (7%)	
Total	123 (95%)			7 (5%)	

2.5.1.2 Análise de paternidade e parentesco

A probabilidade de exclusão de cada SNP variou de 0,09 a 0,18 quando um dos pais é conhecido (P1), 0,03 a 0,12 quando um dos pais não é conhecido (P2) e 0,15 a 0,28 quando um par de pais putativos é excluído (P3). Os valores das probabilidades de não exclusão combinada em P1, P2 e P3 foi $9,6 \times 10^{-9}$, $8,8 \times 10^{-4}$ e $2,8 \times 10^{-13}$, respectivamente, quando os 96 marcadores são usados (Figura 2.2).

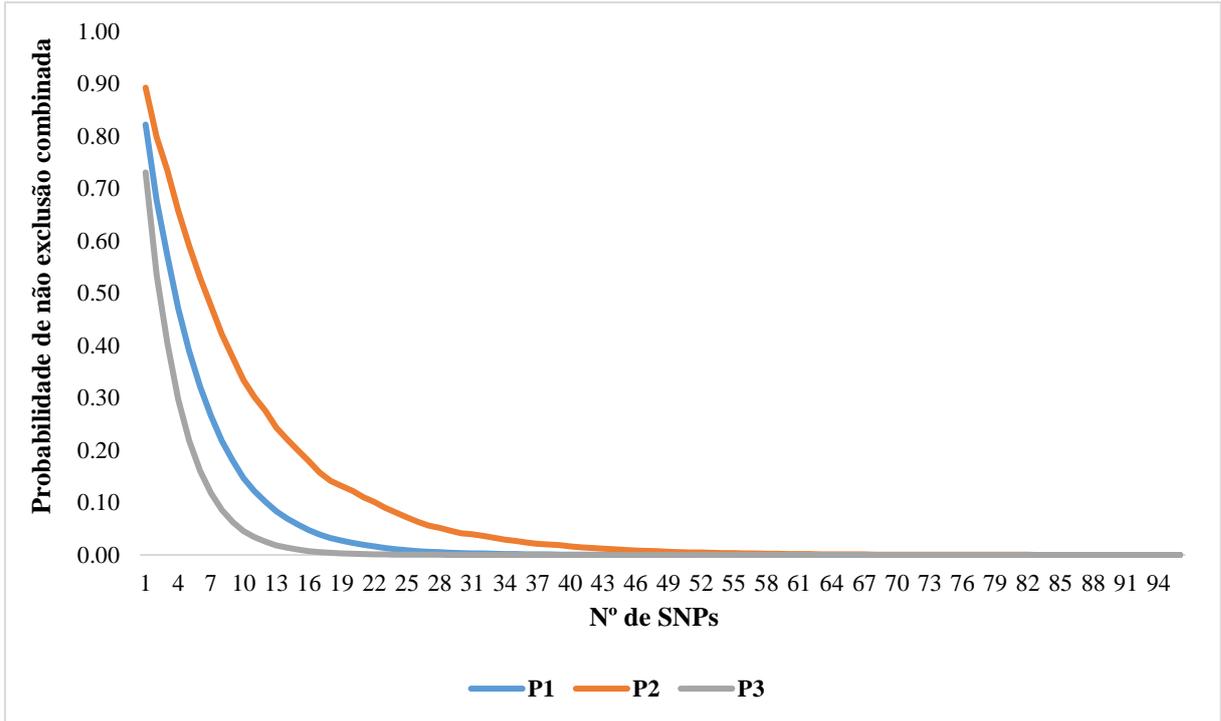


Figura 2.2: Probabilidade de não exclusão combinada de 96 SNPs analisados para um parente conhecido (P1), um parente desconhecido (P2) e exclusão de um par de pais putativos.

As análises de parentesco subsequentes foram realizadas somente com 48 amostras das três linhagens. A matriz de parentesco gerada a partir dessas amostras genotipadas com chip ShrimpLD-24 da Illumina (6.465 SNPs) pode ser observada na figura 2.3. As médias das estimativas de parentesco das linhagens A, B e C foram de 0,12, 0,08 e 0,07, respectivamente. Dentre as 48 amostras, foram escolhidas 16 de cada linhagem.

A

	A_7_10	A_7_2	A_7_4	A_7_6	A_7_7	A_7_9	A_8_1	A_8_10	A_8_2	A_8_3	A_8_4	A_8_5	A_8_6	A_8_7	A_8_8
A_7_10															
A_7_2	0														
A_7_4	0,1565	0,0104													
A_7_6	0,1121	0,1326	0,1165												
A_7_7	0,2651	0,0313	0,1069	0,062											
A_7_9	0,0637	0,2146	0,1408	0,0891	0,23										
A_8_1	0,0367	0,1452	0,1131	0,0828	0,2723	0,3551									
A_8_10	0,1821	0	0,5236	0,2492	0,1708	0,1567	0,2028								
A_8_2	0,1191	0,1445	0,124	0	0,3405	0,3072	0,2837	0							
A_8_3	0,0924	0,0787	0	0,0471	0	0,0048	0,1574	0	0						
A_8_4	0,0493	0,2548	0,1505	0	0	0	0	0,1472	0	0,2724					
A_8_5	0,165	0,088	0,1926	0,1659	0,2324	0,2708	0,2331	0,2583	0,0324	0,2013	0,0248				
A_8_6	0,1513	0	0,154	0,1544	0,2439	0,27	0,2244	0,3006	0,0132	0,1655	0,061	0,1061			
A_8_7	0	0,3607	0,0585	0,1657	0	0,0635	0,0935	0	0	0,2083	0,1735	0	0		
A_8_8	0	0,543	0,065	0,0177	0,0809	0,24	0,011	0	0,1314	0,3051	0,4088	0,0558	0,1792	0,1917	
A_8_9	0	0	0	0,1693	0	0,0159	0	0,0457	0	0,087	0,0209	0,4047	0,151	0,1752	0,0199

B

	B_02	B_04	B_05	B_06	B_09	B_15	B_16	B_21	B_31	B_35	B_45	B_48	B_50	B_51
B_02														
B_04	0,0849													
B_05	0	0,0698												
B_06	0,0207	0,1921	0,2372											
B_08	0,175	0,194	0,2193	0,1667										
B_09	0,3087	0,0459	0	0										
B_15	0,0812	0,2053	0,1681	0,1292	0,041									
B_16	0	0	0	0	0									
B_21	0,1086	0,278	0,1074	0,0887	0,1588	0,1257	0							
B_31	0,2	0,1761	0,1478	0,0375	0,1383	0	0	0,1303						
B_35	0	0	0	0,2495	0	0	0	0,3049	0,0046					
B_45	0,0868	0,0589	0,1395	0	0,0763	0,211	0	0	0,1112	0,0862				
B_48	0	0	0,112	0,0309	0	0	0,1796	0	0	0,2571	0,1776			
B_50	0,2142	0	0,3639	0,2155	0	0,1658	0,1674	0	0,0392	0	0	0,1138		
B_51	0,1632	0	0	0,0705	0,1346	0	0,0762	0,0386	0,1677	0,1518	0,0444	0	0,0446	
B_52	0	0,1716	0,1401	0,1392	0	0,0885	0	0,2041	0	0,4253	0,1679	0	0,0739	0,0357

C

	C_05	C_07	C_08	C_16	C_17	C_18	C_28	C_31	C_33	C_34	C_36	C_44	C_45	C_46	C_52
C_05															
C_06	0														
C_07	0														
C_08	0	0,1574													
C_17	0,2327	0,1333	0,0005	0,0676											
C_18	0	0,276	0,1027	0,1594	0										
C_28	0	0	0,1397	0,0876	0	0									
C_31	0	0	0	0,0019	0	0,0056	0,3511								
C_33	0,1099	0	0,4104	0,0201	0,0774	0	0	0,198							
C_34	0,0431	0,0021	0	0,1306	0,0681	0	0,3052	0,2861	0						
C_36	0,0186	0,1969	0	0,0688	0	0,3125	0	0,0411	0	0					
C_44	0,0396	0,1546	0	0	0,0221	0	0	0	0	0,0083	0,1143				
C_45	0,0593	0,1739	0,083	0,0244	0,1481	0,231	0	0	0,3138	0,0022	0	0,1904			
C_46	0,1638	0,1528	0,0642	0,115	0	0,1915	0,0379	0	0,0187	0	0,2374	0,2463	0,0725		
C_52	0,008	0	0	0,331	0,0805	0	0,0337	0,1289	0,0989	0,0414	0	0	0,0273	0,0533	
C_55	0,1108	0,0676	0,3328	0,0995	0,1413	0	0,2045	0	0,1931	0,0341	0	0	0,1707	0,0488	0,1666

Figura 2.3: Matriz de parentesco entre as três linhagens comerciais gerada com 96 SNPs. Verde – relação de parentesco $<0,1$; Amarelo – entre $0,1$ e $0,2$; Vermelho $>0,2$. (A) Linhagem A, (B) linhagem B, (C) linhagem C.

As filtragens de LD feitas com os 4.233 SNPs (marcadores que restaram após a realização do controle de qualidade) resultaram na formação de cinco subconjuntos com diferentes quantidades de marcadores, que variaram de 81 a 3.857. Esses subconjuntos receberam a denominação II por terem sido genotipados com o chip comercial da Illumina,

enquanto os 88 SNPs (marcadores resultantes do painel com 96 SNPs após a realização das filtragens) genotipados com o equipamento da Fluidigm são chamados de F-88 (Tabela 2.4).

As estimativas de parentesco obtidas com painel de 88 SNPs e com os subconjuntos foram utilizadas para calcular a raiz do desvio quadrático médio (RMSD). O valor de RMSD entre F-88 e Il-3857 foi de 0,08, já o RMSD do Il-3857 quando comparado ao Il-81 (0,08) apresenta o valor igual ao F-88. O menor valor de RMSD foi de 0,03 quando Il-1731 foi comparado com Il-3857 e Il-377. Dentre os subconjuntos de marcadores genotipados com o chip comercial da Illumina, o que apresentava 81 SNPs teve maiores valores de RMSD quando comparado aos demais Il (Tabela 2.4).

Tabela 2.4: Raiz do desvio quadrático médio (RMSD) entre as estimativas de parentesco obtidas com dados que foram genotipados com Fluidigm (FI) e com Illumina (II).

	F-88	II-81	II-138	II-377	II-1731	II-3857
F-88	-					
II-81	0,11	-				
II-138	0,09	0,06	-			
II-377	0,08	0,08	0,05	-		
II-1731	0,08	0,09	0,06	0,03	-	
II-3857	0,08	0,08	0,06	0,04	0,03	-

A raiz do erro quadrático médio (RMSE) calculada com os intervalos de confiança das estimativas de parentesco pode ser observada na figura 2.4. O valor de RMSE utilizando o painel com 88 SNPs foi 0,07 e quando usado 3.857 marcadores do chip comercial da Illumina foi 0,02, ou seja, houve uma redução de apenas 0,05 no RMSE, quando foram utilizados 88 ou 3.857 SNPs. Além disso, quando comparado o RMSE dos 3.857 marcadores e dos 81 SNPs oriundos do chip comercial, é observado à mesma redução de 0,05.

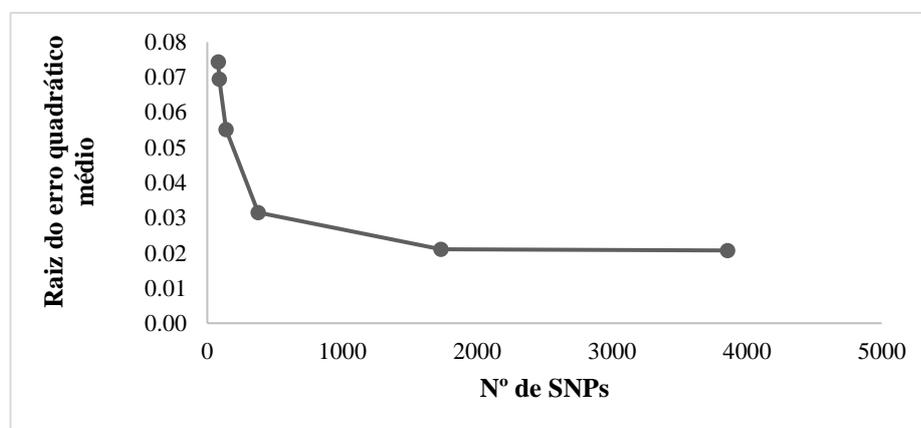


Figura 2.4: Raiz do erro quadrático médio (RMSE) dos intervalos de confiança das estimativas de parentesco obtidas com diferentes números de SNPs.

2.5.2 Análise de amostras de camarão de varejistas

Os 86 SNPs selecionados apresentaram variação de MAF entre 0,12 e 0,5 com média de 0,37, sendo que somente quatro SNPs apresentaram MAF menor que 0,2 (Apêndice B). A maioria dos marcadores do painel apresentaram MAF entre 0,3 e 0,4 (32 SNPs) e entre 0,4 e 0,5 (38 SNPs). Em relação ao Equilíbrio de *Hardy Weinberg* foram observados desvios em 22 marcadores que apresentaram o valor de $p < 0,01$ (Apêndice B), devido à um excesso de homocigotos. Na tabela 2.5 encontra-se a distribuição de MAF das 10 populações e a presença de diferenças nas médias entre as linhagens e as populações das lojas de varejo. Os valores de MAF foram significativamente diferentes ($p < 0,05$) entre: a linhagem A e seis populações das lojas de varejo; a linhagem B e cinco populações dos varejistas; e a linhagem C e duas lojas de varejo (Tabela 2.6).

Tabela 2.5: Distribuição da Menor Frequência Alélica dos 86 SNPs do painel nas linhagens A, B e C e nas populações das lojas de varejo.

Pop	Intervalos de MAF					Média	Desvio padrão
	0,00-0,10	0,10-0,20	0,20-0,30	0,30-0,40	0,40-0,50		
A	0	0	12	32	42	0,381	0,077
B	1	2	22	22	39	0,366	0,095
C	2	9	11	26	38	0,354	0,105
SPDF1	7	9	23	21	26	0,305	0,127
SPDF2	8	15	20	22	21	0,286	0,131
SPDF3	7	14	17	28	20	0,297	0,126
SPDF4	4	13	16	19	34	0,324	0,130
SPDF5	5	18	19	25	19	0,289	0,122
PEDF	9	12	19	24	22	0,297	0,132
PEPB	6	13	16	23	28	0,308	0,132

Tabela 2.6: Teste de Tukey para diferenças de médias de MAF das amostras testadas (* = $p < 0,05$).

	A	B	C	SPDF1	SPDF2	SPDF3	SPDF4	SPDF5	PEDF	PEPB
A	-									
B	0,99780	-								
C	0,88230	0,99960	-							
SPDF1	0,00121*	0,02943*	0,19160	-						
SPDF2	0,00002*	0,00046*	0,00708*	0,98640	-					
SPDF3	0,00017*	0,00612*	0,06005	1	0,99980	-				
SPDF4	0,05625	0,40530	0,84920	0,98850	0,49820	0,89140	-			
SPDF5	0,00003*	0,00109*	0,01510*	0,99720	1,00000	1	0,64680	-		
PEDF	0,00016*	0,00583*	0,05789	1	0,99980	1	0,88630	1	-	
PEPB	0,00246*	0,05029	0,27620	1	0,96330	0,99980	0,99690	0,98940	0,99980	-

2.5.2.1 Estruturação populacional e variabilidade genética

Os resultados das análises de componentes principais indicaram que as amostras das três linhagens do produtor de larvas e das lojas de varejo estão estruturadas em dois grupos. Dentre as linhagens, A foi a que apresentou menor estruturação em relação às populações das lojas de varejo. Houve uma dispersão maior entre A, B e C, enquanto a maioria das amostras das lojas de varejo ficou mais agrupada (Figura 2.5A). A estruturação populacional avaliada com STRUCTURE corroborou com os dados obtidos com a PCA, uma vez que o maior valor de ΔK foi detectado quando um modelo assumindo duas populações foi estabelecido (Figura 2.5B). Esse valor indica que as linhagens do produtor de larvas e as amostras das lojas de varejo agruparam-se em dois grupos distintos. No entanto, a linhagem A apresenta certo grau de subestruturação genética entre os dois (Figura 2.5C), assim como foi observado na PCA em que essa linhagem apresentava configuração similar. A estruturação populacional de $K=3$ demonstra novamente o agrupamento das amostras das linhagens A e B no mesmo grupo e a divergência da linhagem C em relação às outras linhagens (Figura 2.5C).

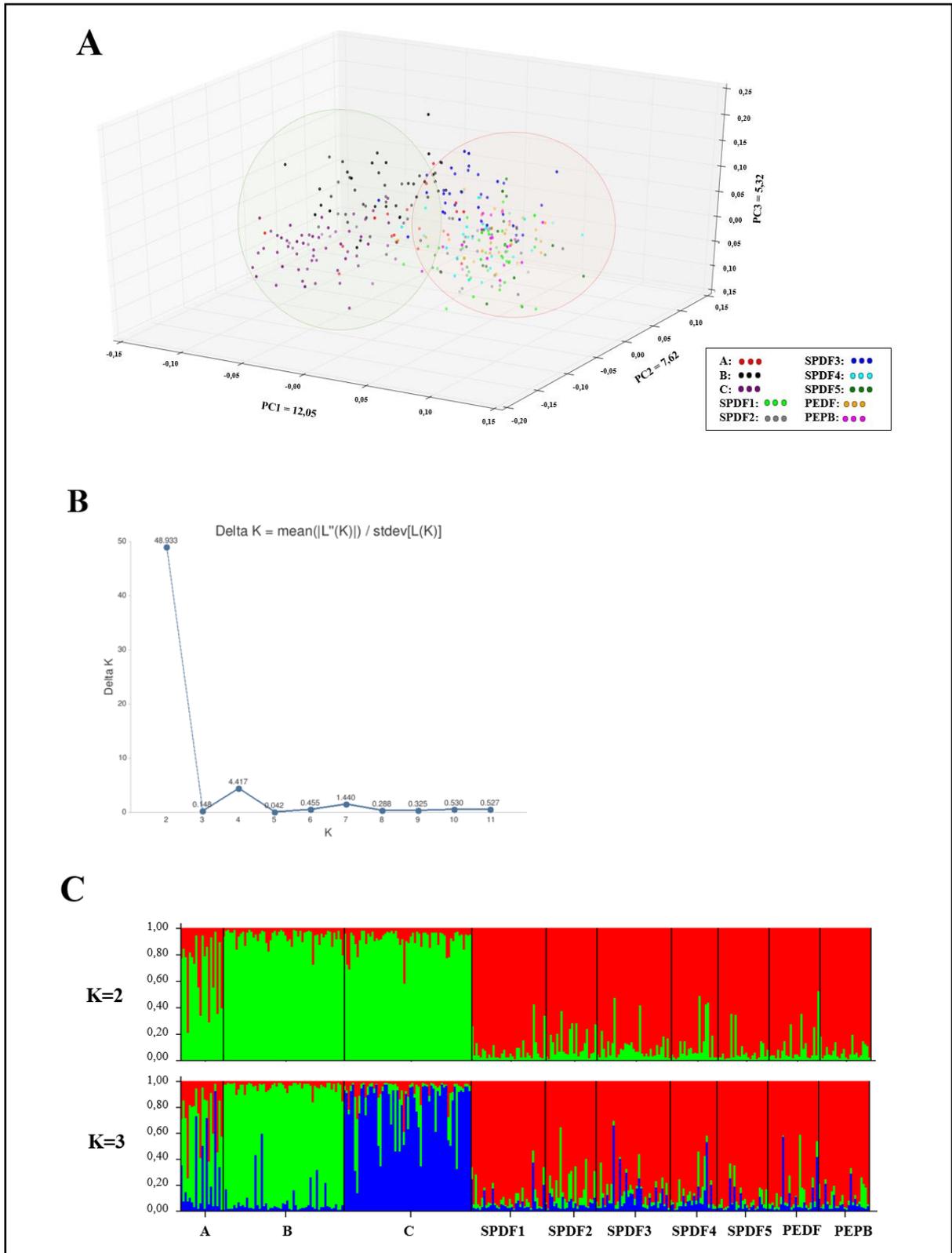


Figura 2.5: Estrutura populacional obtida com o painel de 86 SNPs e 325 amostras das linhagens (A, B e C) do produtor de larvas e das lojas de varejo (SPDF1-SPDF5, PEF e PEPB). (A) Plot com análises de componentes principais (PCA), em que PC1, PC2 e PC3 correspondem a 12,05, 7,62 e 5,32%, respectivamente, do total da variância observada das linhagens; (B) Representação gráfica do melhor valor de $K=2$; (C) Estruturação para $K=2$ e $K=3$ das 10 populações analisadas que foram separadas por uma linha vertical.

Análises subsequentes das estimativas de F_{st} das populações variaram de 0,015 a 0,071. O F_{st} mínimo de 0,015 foi encontrado entre as amostras populacionais de SPDF1 e SPDF4. O F_{st} máximo de 0,071 foi detectado entre a linhagem B e SPDF2. De forma geral, a população SPDF1 foi a que apresentou menores valores de F_{st} quando comparadas com as demais populações dos varejistas, enquanto SPDF2 apresentou as maiores estimativas de diferenciação (Tabela 2.7).

Tabela 2.7: Análise de F_{st} em pares entre as três linhagens do produtor de larvas e as populações das lojas de varejo.

	A	B	C	SPDF1	SPDF2	SPDF3	SPDF4	SPDF5	PEDF	PEPB
A	-									
B	0,029	-								
C	0,036	0,042	-							
SPDF1	0,042	0,063	0,053	-						
SPDF2	0,056	0,071	0,069	0,04	-					
SPDF3	0,046	0,068	0,054	0,034	0,056	-				
SPDF4	0,034	0,058	0,052	0,018	0,04	0,032	-			
SPDF5	0,041	0,065	0,061	0,015	0,045	0,035	0,02	-		
PEDF	0,043	0,06	0,057	0,021	0,047	0,037	0,02	0,022	-	
PEPB	0,046	0,062	0,056	0,023	0,05	0,033	0,026	0,028	0,022	-

A atribuição dos indivíduos das linhagens A, B e C e das lojas de varejo às suas respectivas populações utilizando o painel com 86 SNPs é mostrada na tabela 2.8. Cerca de 75% das amostras das 10 populações foram atribuídas à sua população de origem, e 25% foram designadas a outras populações. Ocorreu uma grande variação dessas atribuições, uma vez que na população SPDF5, somente 38% dos indivíduos foram designados à sua população, enquanto na linhagem B, 98% das amostras foram atribuídas corretamente à sua linhagem.

Tabela 2.8: Determinação de população utilizando 86 SNPs das amostras da linhagens (A, B e C) do produtor de larvas e das sete lojas de varejo.

	População			Amostras alocadas em outras populações								Total
	referência	A	B	C	SPDF1	SPDF2	SPDF3	SPDF4	SPDF5	PEDF	PEPB	
A	15 (75%)		1	2				2				5 (25%)
B	56 (98%)			1								1 (2%)
C	54 (90%)	4	2									6 (10%)
SPDF1	21 (60%)							5	6		3	14 (40%)
SPDF2	19 (79%)		1					2	1		1	5 (21%)
SPDF3	29 (83%)				2			1	1	2		6 (17%)
SPDF4	10 (45%)				3		2		1	5		12 (55%)
SPDF5	9 (38%)				5		3	3		3	1	15 (63%)
PEDF	14 (58%)	1		1	1			1	2		4	10 (42%)
PEPB	17 (71%)				2			1	1	3		7 (29%)
Total	244 (75%)											81 (25%)

A variabilidade genética das amostras analisadas foi avaliada por meio das estimativas dos índices de H_o , H_e e F_{is} . A maior variação encontrada entre a heterozigidade observada (H_o) e esperada (H_e) foi de 4%, sendo que a média dos valores de H_o para todas as populações foi de 0,392 (Tabela 2.9). O coeficiente de endogamia (F_{is}) variou de -0,07 a 0,10. Além disso, é possível observar que a média de H_o nas linhagens é 10% maior em relação às amostras das lojas de varejo. A média do F_{is} também foi diferente entre os indivíduos do produtor de larvas e dos varejistas, uma vez que variou de -0,02 a 0,06.

Tabela 2.9: Índices de variabilidade genética estimados com 86 SNPs de amostras do produtor de larvas e das lojas de varejo.

População	H_o	H_e	F_{is}
A	0,443	0,460	0,037
B	0,467	0,446	-0,046
C	0,470	0,435	-0,078
Média	0,460	0,447	-0,029
SPDF1	0,372	0,392	0,052
SPDF2	0,365	0,374	0,025
SPDF3	0,358	0,387	0,075
SPDF4	0,373	0,405	0,080
SPDF5	0,343	0,382	0,102
PEDF	0,355	0,383	0,074
PEPB	0,372	0,392	0,050
Média	0,363	0,388	0,065
Média total	0,392	0,406	0,034

2.5.2.2 Análises de paternidade

A probabilidade de exclusão de cada SNP variou de 0,08 a 0,19 quando um dos pais é conhecido (P1), 0,02 a 0,12 quando nenhum dos pais é conhecido (P2) e 0,13 a 0,28 quando um par de pais putativos é excluído (P3). Como mostrado na figura 2.6, o valor da probabilidade de exclusão combinada em P1, P2 E P3 é $4,5 \times 10^{-7}$, $4,4 \times 10^{-4}$ e $5,2 \times 10^{-11}$, respectivamente, quando os 86 marcadores são usados. Esses resultados indicam que as estimativas são confiáveis na realização de análises de paternidade e parentesco, tanto em populações de núcleo de melhoramento genético, como em amostras de camarão coletadas em lojas de varejo.

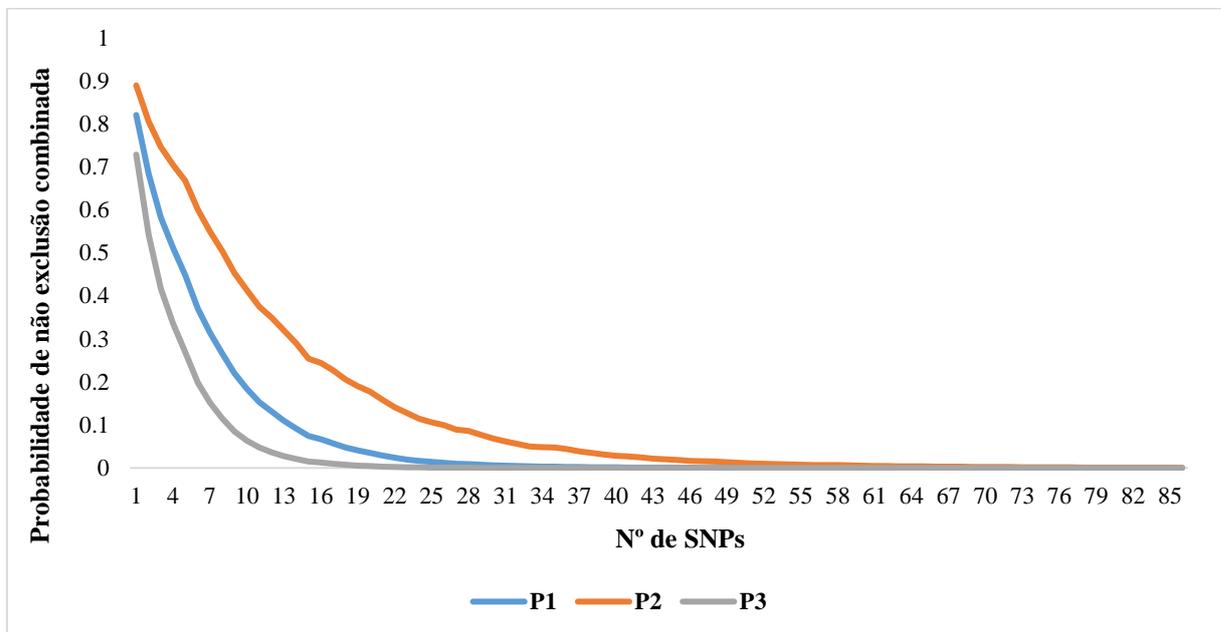


Figura 2.6: Probabilidade de não exclusão combinada de 86 SNPs e 325 amostras analisadas para um parente conhecido (P1), um parente desconhecido (P2) e exclusão de um par de pais putativos.

Além disso, é possível observar que mesmo quando as três linhagens e as outras sete populações são analisadas separadamente, o valor da probabilidade de não exclusão é similar em todas as populações (Tabela 2.10).

Tabela 2.10: Probabilidade de não exclusão dos 86 SNPs nas linhagens A, B e C (LIN) e nas populações do Ceará (CE), Rio Grande do Norte (RN), NR (Não Relatada) e Maranhão (MA).

Pop	P1	P2	P3
A	$5,4 \times 10^{-8}$	6×10^{-5}	$2,1 \times 10^{-12}$
B	$8,5 \times 10^{-8}$	$1,0 \times 10^{-4}$	$4,0 \times 10^{-12}$
C	$1,2 \times 10^{-7}$	$1,3 \times 10^{-4}$	$6,6 \times 10^{-12}$
SPDF1	$4,9 \times 10^{-7}$	$5,1 \times 10^{-4}$	$5,2 \times 10^{-11}$
SPDF2	$9,3 \times 10^{-7}$	$8,0 \times 10^{-4}$	$1,3 \times 10^{-10}$
SPDF3	$5,9 \times 10^{-7}$	$6,3 \times 10^{-4}$	$6,5 \times 10^{-11}$
SPDF4	$3,2 \times 10^{-7}$	$3,1 \times 10^{-4}$	$2,8 \times 10^{-11}$
SPDF5	$6,8 \times 10^{-7}$	$7,6 \times 10^{-4}$	$7,9 \times 10^{-11}$
PEDF	$6,9 \times 10^{-7}$	$6,4 \times 10^{-4}$	$8,5 \times 10^{-11}$
PEPB	$5,1 \times 10^{-7}$	$4,9 \times 10^{-4}$	$5,5 \times 10^{-11}$

2.6 DISCUSSÃO

O objetivo principal deste estudo foi validar um painel SNP de baixa densidade para avaliar parâmetros de variabilidade e estruturação genética, atribuir paternidade e analisar relações de parentesco em populações comerciais de *L. vannamei*. As análises de variabilidade e estruturação genética das dez populações revelaram: (i) alta variabilidade genética, (ii) quantidade limitada de consanguinidade nas populações dos varejistas, (iii) estruturação populacional entre as linhagens do produtor de larvas e as populações das lojas de varejo. O primeiro e o segundo resultado indicam que as populações são oriundas de programas de melhoramento genético que controlam o acasalamento entre parentes e mantêm alta variabilidade genética do camarão cinza (Moss e Moss, 2009). Já o terceiro resultado sugere que as amostras dos varejistas não são oriundas do produtor de larvas abordado nesse estudo e que todas as populações das lojas de varejo aparentam ter origem de um mesmo produtor.

Os resultados da análise de STRUCTURE das linhagens A, B e C com $K=2$ sugerem que a origem do germoplasma das linhagens A e B é o mesmo, e/ou que existe um forte fluxo gênico entre as duas linhagens. Já quando todas as amostras foram consideradas, com $K=3$, pode-se observar que as amostras oriundas de lojas de varejo possuem uma forte estruturação em relação as linhagens do produtor de lavas. A amostragem das linhagens foi feita em 2016, enquanto que a coleta nas lojas de varejo foi realizada em 2019. Considerando que o intervalo de geração para o camarão cinza é 12 meses, é pouco provável que a estruturação distinta observada nas amostras de varejo é decorrente de efeitos de deriva e seleção genética em três gerações. Portanto, podemos concluir que o material amostrado em lojas de varejo possui uma forte contribuição de um germoplasma de uma linhagem distinta (Perez-Enriquez et al., 2018). Adicionalmente, os resultados obtidos indicam que o painel de marcadores validado pode ser utilizado para análises de estruturação e auxiliar nas decisões de cruzamento entre linhagens oriundas de programas de melhoramento genético distintos.

As comparações entre as estimativas de parentesco obtidas com o painel de 96 SNPs e com diferentes subconjuntos de marcadores derivados do chip Illumina Infinium® ShrimpLD-24 demonstram que mesmo com um número reduzido de SNPs é possível obter estimativas confiáveis. A informatividade máxima de um conjunto de SNPs em análises de parentesco e paternidade é obtida quando os valores de MAF são próximos de 0,5, como demonstrado em simulações realizadas por Anderson e Garza (2006). Outros estudos enfatizaram a relevância do MAF, como Baruch e Weller (2008), que constataram que a probabilidade de não exclusão diminui com o aumento do MAF no mesmo número de SNPs em um painel.

A média de MAF dos marcadores observada considerando em todas as amostras variou entre 0,12 e 0,5 (Apêndice B). Já as comparações das médias das estimativas de MAF entre as dez populações revelou diferenças significativas entre as amostras coletadas em lojas de varejo e as das linhagens comerciais. No entanto, as probabilidades de não exclusão estimadas individualmente para cada população foram satisfatórias, sugerindo que o conjunto de marcadores validado pode ser utilizado na identificação de paternidade e parentesco em todas as populações analisadas. Resultados obtidos por Rengmark et al (2006) em salmão (*Salmo solar*) demonstraram probabilidade de não exclusão P_3 de $1,53 \times 10^{-2}$, mas mesmo assim 82,4% dos indivíduos foram atribuídos corretamente aos seus pais utilizando um painel com somente 26 SNPs. Já o painel desenvolvido para camarão cinza apresentou P_3 entre $1,3 \times 10^{-10}$ e $8,5 \times 10^{-11}$ nas diferentes populações testadas. Portanto, pode-se esperar uma eficácia significativamente superior a relatada por Rengmark et al (2006).

Estudos adicionais para desenvolvimento e validação de painéis de SNP para o camarão cinza foram publicados durante a execução desse trabalho. Jones et al. (2017) desenvolveu o Illumina Infinium® ShrimpLD-24 com 6.465 SNPs e Perez-Enriquez et al., (2018) validou um painel de 76 SNPs derivados de experimentos de sequenciamento nextRAD. O ShrimpLD-24 apresenta maior quantidade de marcadores, mas também um custo de genotipagem muito superior. Considerando os resultados obtidos e os demais estudos citados (Anderson e Garza, 2006; Baruch e Weller 2008) 60 a 100 SNPs são adequados para a atribuição precisa de paternidade em larga escala, e, portanto os custos adicionais para gerar genótipos para 6K SNPs não podem ser justificados. Perez-Enriquez e Max-Aguilar (2016) utilizaram o painel com 76 SNPs, com uma probabilidade de não exclusão combinada P_3 de 10×10^{-8} , para obter uma atribuição de paternidade correta em 96,9% dos indivíduos analisados. As estimativas de probabilidade de não exclusão combinada obtidas com as dez populações

testadas no presente estudo foram superiores, variando entre $4,5 \times 10^{-7}$, $4,4 \times 10^{-4}$ e $5,2 \times 10^{-11}$ para P1, P2 E P3, respectivamente indicando a maior eficácia do painel validado.

O painel de SNPs validado abre a possibilidade de estudos experimentais para obtenção das informações de *pedigree* em ambientes em que são criados grupos de animais de múltiplas famílias. A ferramenta desenvolvida poderá ser utilizada para auxiliar programas de melhoramento genético que buscam selecionar, principalmente, indivíduos que apresentem melhores taxas de sobrevivência. Nesse caso, os animais são criados em tanques abertos em que não existe estabilidade ambiental, o que torna a criação de indivíduos em tanques individuais por família inviável, já que a estimativa dos parâmetros genéticos ficaria comprometida devido à interação genótipo-ambiente (Moss e Moss, 2009).

CONCLUSÕES

Uma alternativa viável que busca simplificar a identificação individual em programas de melhoramento genético do camarão cinza é a obtenção dos dados de *pedigree* com marcadores moleculares, como o polimorfismo de nucleotídeo único (SNPs). Sendo assim, esse trabalho teve como objetivo validar um painel de 96 SNPs para análises de parentesco, avaliação de variabilidade e estruturação genética em camarão cinza.

Os resultados obtidos demonstraram a existência de alta variabilidade em todas as populações estudadas. Foi observada também estruturação entre as linhagens e as populações das lojas de varejo, o que sugere que as amostras dos varejistas não são oriundas do produtor de larvas abordado nesse estudo. O painel de marcadores validado também é eficaz em análises de paternidade, pois apresentou probabilidades de não exclusão satisfatórias e as estimativas de parentesco obtidas com os 96 SNPs foram muito próximas das adquiridas com 3.857 SNPs do chip comercial ShrimpLD-24.

Portanto, o painel validado poderá ser utilizado em estudos amplos que demandem grandes números de atribuições de paternidade, avaliações de parentesco e análises de estruturação populacional do camarão cinza. A ferramenta pode fornecer informações de *pedigree* e aprimorar os programas de melhoramento genético que buscam selecionar, principalmente, indivíduos que apresentem melhores taxas de sobrevivência, já que nesse caso os animais precisam ser criados em tanques abertos com grupos de animais de múltiplas famílias.

REFERÊNCIAS BIBLIOGRÁFICAS

Anderson, E. C. e Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172, 2567–2582.

Artiles, A. et al. (2011). Low genetic variability in the fifth introduction of *Litopenaeus vannamei* in Cuba, as estimated with microsatellite markers. *Biotechnologia Aplicada*, 28(3), 142–150.

Aungsuchawan, S. et al. (2008) Evaluation of published microsatellites for paternity analysis in the pacific white shrimp *Litopenaeus vannamei*. *ScienceAsia*, 34(1), 115–118.

Baruch, E. e Weller, J. I. (2008). Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics*, 39, 474–479.

Boyce, T. M. et al. (1989). Mitochondrial DNA in the bark weevils: Size, structure and heteroplasmy. *Genetics*, 123(4), 825–836.

Caetano, A. R. (2009). Marcadores SNP: Conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. *Revista Brasileira de Zootecnia*, 38, 64–71.

Castillo-Juárez, H. et al. (2015). Genetic improvement of Pacific white shrimp (*Penaeus (Litopenaeus) vannamei*): Perspectives for genomic selection. *Frontiers in Genetics*, 6, 1–4.

Crock, J. et al. (2009). Breeding for disease resistance of Penaeid shrimps. *Aquaculture*, 286, 1–11.

FAO. (2018). El estado mundial de la pesca y la acuicultura 2018. Cumplir los objetivos de desarrollo sostenible.

Fluidigm. (2018). SNP Genotyping User Guide (PN 68000098 Q1).

Garcia, D. K. e Alcivar-Waren, A. (2007). Characterization of 35 New Microsatellite Genetic Markers for the Pacific Whiteleg Shrimp, *Litopenaeus vannamei*: Their Usefulness for Studying Genetic Diversity of Wild and Cultured Stocks, Tracing Pedigree in Breeding Programs, and Linkage Mapping. *Journal of Shellfish Research*, 26(4), 1203–1216.

Gitterle, T. (2005). Genetic (co)variation in harvest body weight and survival in *Penaeus (Litopenaeus) vannamei* under standard commercial conditions. *Aquaculture*, 243, 83–92.

Harlizius, B. et al. (2011). A single nucleotide polymorphism set for paternal identification to reduce the costs of trait recording in commercial pig breeding. *Journal of Animal Science*, 89(6), 1661–1668.

Heaton, M. P. et al. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. *PloS one*, 9(4), 1–10.

Heaton, M. P. et al. (2002). Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome*, 13(5), 272–281.

Ianella, P. et al. (Em preparação). Single Nucleotide polymorphism identification and development of a low-density SNP panel for kinship and genetic diversity evaluation of Pacific white shrimp (*Litopenaeus vannamei*).

Jones, D. B. et al. (2017). A comparative integrated gene-based linkage and locus ordering by linkage disequilibrium map for the Pacific white shrimp, *Litopenaeus vannamei*. *Scientific Reports*, 7, 1–16.

Karlsson, S. et al. (2011). Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources*, 11, 247–253.

Kopelman, N. M. et al. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191.

Lee, Y. (2013). Paternity assignment in *Litopenaeus vannamei* using microsatellites. *University of Hawaii at Manoa*.

Lightner, D. V. (2012). Historic emergence, impact and current status of shrimp pathogens in the Americas. *Journal of Invertebrate Pathology*, 110(2), 174–183.

Liu et al. (2016). Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture*, 452, 178–182.

Moss, S.M. e Moss, D. R. (2009). Selective breeding of penaeid shrimp. *Shellfish Safety and Quality*.

Peakall, R. e Smouse, P. E. (2012). GenAlEx 6.5 : genetic analysis in Excel . Population genetic software for teaching and research — an update. *Bioinformatics*, 28(19), 2537–2539.

Perez-Enriquez, R. et al.(2018). SNP markers for the genetic characterization of Mexican shrimp broodstocks. *Genomics*, 110(6), 423–429.

Perez-Enriquez, R. et al. (2009). Genetic diversity status of White shrimp *Penaeus (Litopenaeus) vannamei* broodstock in Mexico. *Aquaculture*, 297, 44–50.

Perez-Enriquez, R. e Max-Aguilar, A. (2016). Rastreabilidad del pedigrí en camarón blanco (*Litopenaeus vannamei*) mediante marcadores genéticos: Una comparación entre microsatélites y SNP. *Ciencias Marinas*, 42(4), 227–235.

Pritchard, J. K. et al. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics Society of America Inference*, 155, 945–959.

Rengmark, A. H. et al. (2006). Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture*, 253(1-4), 229-237.

Rohrer, G. A. et al. (2007). Single nucleotide polymorphisms for pig identification and parentage exclusion. *Animal Genetics*, 38(3), 253–258.

Sellars, M. J. et al. (2014). Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research*, 45(3), 417–426.

Teixeira, A. K. G, et al. (2019). Melhoramento genético de camarão marinho livre de doenças. Disponível em: <<https://panoramadaaquicultura.com.br/melhoramento-genetico-de-camarao-marinho-livre-de-doencas>>. Acesso em: 10 maio, 2019.

Vignal, A. et al.(2002). A review on SNPs and other types of molecular markers and in their use in animal genetics. *Genetics, selection, evolution*, 34, 275–305.

Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, 89, 135–153.

Wang, J. (2011). Coancestry: a program for simulating , estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 11, 141–145.

APÊNDICE A – Sequências ASP1, ASP2, LSP e STA dos 96 SNPs genotipados com Fluidigm

SNP_NAME	ASP1	ASP2	LSP	STA
C57153922_110	CACGTAATCCTCGTCCAGGATG	CACGTAATCCTCGTCCAGGATT	ACGAGCAAGAAGGCCTCGG	TCATCTGTCTCGACGTCC
C57466369_134	ACTAATGCGCCTAATATCAGGACAAA	ACTAATGCGCCTAATATCAGGACAAAG	CCAGAGTCAGCGTCCGA	CGACTGCAGCAACGGG
C57535591_147	CCCCAATTACGGGTCTCTCATTAATA	CCCAATTACGGGTCTCTCATTAATG	TGCCCTCAACATCTACACATTCA	CCTTTCATATCCCAATTACGGG
C57577308_144	CGAGATGTCCAAACATCAGGTCA	CGAGATGTCCAAACATCAGGTCCG	CCGCTGCTAAGGAAGGCTTG	GTGACATGTTCCTGCGCA
C57578086_155	GGTGTGATTCATTGCTATGAGGTCC	AGGTGTGATTCATTGCTATGAGGTAA	GCTGCTGGATGTGTAACAATAGCC	GAGATGAGGTTGGTATTCATTGCT
C57603451_134	CGCACTGAAAAGAAATCATACTTCGAC	CGCACTGAAAAGAAATCATACTTCGAT	CGCCTTGACTTACGACGAAGAAAC	GCATAACAGTTCACAACAATCGC
C57611531_136	CCAGTTCACGCAAGGCCAAA	CCAGTTCACGCAAGGCCAAG	GCTAATGGACCCGAGTCGGC	GATCGTTCGTCTCATCCGC
C57638303_137	GTGGAATGTGGTGTTTTATATTCACCA	GTGGAATGTGGTGTTTTATATTCACCAG	AGACATCTAATATGAGAGCTTAATGCCTAATAATGAGA	AAACTATCAGTGGAAATGTGGTGT
C57673261_175	CTCGGATGACCTTGCTTCGT	CTCGGATGACCTTGCTTCGC	AGTCTGAGGAAGGCTTTCGAGA	TGCTAGATTTAACTTGGTCACTGG
C57706433_166	AATCGTCTCTCGATCGACATG	AAATCGTCTCTCGATCGACATF	CGGACTTGTAGGCAACGAAACA	TCTCACTTCAATCACCATCAAAGC
C57759483_165	TCTGCCTCCATATCAACTATCTATCTATCC	TCTGCCTCCATATCAACTATCTATCTATCT	ACACAAAGATTTGTAGATAGAGTAACAGACGG	CTGCTGTCTCATATCTCCATTT
C57819698_180	CATTATCAGTGAGGCATTTGGATTAA	CATTATCAGTGAGGCATTTGGATTAAAG	GCAATATTCAGAATATGCTCAACATGACT	CACAGCCTACAAGCCACAC
C57832144_157	TGGGTTCCAAATCCCGAGATAATAAC	TGGGTTCCAAATCCCGAGATAAAT	CATCCCAGAAAGGCAGG	CCTCTGCGACAGATGTGA
C57833982_108	ACGAATTTGCATGAGAAGGTCCG	ACGAATTTGCATGAGAAGGTCC	CGGACGCGAATGCACGA	AGACAGATAGACAGATAGATATTCTGC
C57838059_125	CACAAAGCGACATGATCATGCA	CACAAAGCGACATGATCATGCT	TCTATGCGTGAGTGTACGTGTG	CTCGTCAATGGTATTCACATCACC
C57849651_140	TTCTTGGACGAGGAGTTTGGAAT	TCTTGGACGAGGAGTTTGGAAC	GATGCACATGTCACGCCTGC	AAACATCTCATCTTGTCCAAC
C57850365_195	CTGAAACAAATCATCAACATCTCGCA	TGAAACAAATCATCAACATCTCGCG	TCTTGAAGGACAGCGCGGTA	ACGGAGACCGAGATTGGG
C57900351_108	TTTTATGGTACCCACACCATGGTATA	TTTTATGGTACCCACACCATGGTATA	TGAAATAAGGGAGATGGGCTATTATT	AGAAATTACAGTAATTGAAGGAGCGA
C57917905_171	TGCAGCAGCATGAATACCGA	GCAGCAGCATGAATACCGG	CTTGAGGGCGTCTGCTTGG	ACGAGGATGGAAAGTAGTAACT
C57923795_179	GGTAATCTCTCAATGAACCTTAGCTT	GGTAATCTCTCAATGAACCTTAGCTA	CCACCACAGACTAAGTAAACGCGA	TGTGAATATCTGAACATTTAATCCCA
C57968599_172	CAGTCTCCTTTGAGAGCGCT	AGTCTCCTTTGAGAGCGCC	TCAGCGCGATAGGGTCC	GACAGCGCAAGACAGCC
C57987359_182	GCGTGATCTTTTACATTTGTCAGGTATC	GCGTGATCTTTTACATTTGTCAGGTATG	CAACTTCATACAGGATGCCACT	ATGGTTCCTCTCTTATTATGT
C57991179_206	TCATTCACTCAATGCGGAGATCA	CATTCACTCAATGCGGAGATCG	CTTTCGAACATGATCTAAATTAACGGT	CCGGGCATTAGAGACTATGTTTA
C57994195_257	GTCCTTACTTCCAATCGACCTAAGAC	GTCCTTACTTCCAATCGACCTAAGAT	CAGCTCAAGAGGTTGTGCTCG	CACATGTGCACAGTCTTACCT
C57996339_174	TCAACAGTGGTCTTGTCTCC	TCAACAGTGGTCTTGTCTCCG	CCATCACTAAGGACGTGATTGAT	TGATGATGATCGCGTAATGATTAATT
C57997907_129	GATTACTGAACACTTGCTTTGTACGTC	GATTACTGAACACTTGCTTTGTACGTT	CCAATGAATAACTAATTCACACATTTCTATGATTGACT	TCGATGCACGGTGTCTTGG
C58003019_178	GGATGACCCAAGTTATTAAGCTTGC	GGATGACCCAAGTTATTAAGCTTGT	ACCCCGAGTTTCGCCGTA	TGCCCTTATCCATGTGCTTAAAG
C58013521_113	ACATTAATGGCAATTCAGCCTGA	CATTAATGGCAATTCAGCCTGG	GCCTCGTGGACCAGTCGA	CTGAGCCTGCATTATAGATGCTT
C58033295_194	GTTTCCAGTCTGTTTGCCTAAACC	GTTTCCAGTCTGTTTGCCTAAACT	CGCTTGTACTGACGCGGA	GCGTCTCTATTGCTGTGAAA
C58038927_235	ACTCGTCATTCTGTGATTCTGTAGAA	ACTCGTCATTCTGTGATTCTGTAGAG	ATGCGGAGAGTGTACGTC	GACCTTGTCTGCTTGGT

C58044867_165	GATCTCGGAATATTGCTCAAGCC	GATCTCGGAATATTGCTCAAGCG	TGCGAGGGCGTCTCCA	TTTCACGAGGACAATGGCAC
C58051675_249	TCATTCCTGCGTAATCATCC	TTCATTCCTGCGTAATCATCT	TCATGAACCTCATAATGACAAGTAATGAGA	ATCACTTCATTCCTGCG
C58071739_184	GCTGGCTAACCTAATTCGCA	GCTGGCTAACCTAATTCGCT	TGGCTTTGGTCCCATCCATT	CGGTGCTGAGTGGCAAA
C58078597_141	ATTTGCTACTCATGCGTATATTGTGC	CATTTGCTACTCATGCGTATATTGTGT	ATCGAGTCGAGCAAAATCCCCT	CGTAAGTGTCCAGTTATTTACC
C58082151_285	CCACGCTCCCTTCGTATTT	CCACGCTCCCTTCGTATTT	CCACGCTAACACTTCGTGCA	CCGCGTAACAGCGAAC
C58090729_165	GCTCAGACAATACACAGAAGGCTTAT	GCTCAGACAATACACAGAAGGCTTAC	AGTCGGTCTCATTGCAAAGTGAGT	TTTCTGAAAGGTGGACAAGGC
C58104727_290	GCACTTCAAGCCTGCATCA	GCACTTCAAGCCTGCATCG	CCACTCCCCTCCAGACT	CATGACGTTGGTGTCTGGAA
C58106625_158	CAAAGATAGAGTGTACGAGAGCATAGAC	CAAAGATAGAGTGTACGAGAGCATAGAT	CGGAGCCCTTACGCC	AACATGCTGATTACGAAGGTAGC
C58107721_279	CTCCGCTTATCGTCAT CCGTA	TCCGCTTATCGTCATCCGTC	GTGTTCTCCGCTCGAAGC	CGTAGCGCATCATCCCC
C58113213_318	CCACGAAAGCTAAAATATCGACGT	CCCACGAAAGCTAAAATATCGACGA	TCAATTCATAATTTATGTAGTCCAGTGAATGGGT	AATTTTGTTCATCCCACGAAAGC
C58120413_184	CACAATTAGCCTAATCTGCCGA	CACAATTAGCCTAATCTGCCGT	GTCTTGTAAATGTCTTGTCTCCCGC	CTTCCAGAAGGTAAGAAATAGGCAA
C58122813_104	AATTTAGTGTAGAAATGTTTAGAATACTGGACTTA	AATTTAGTGTAGAAATGTTTAGAATACTGGACTTT	GTCCACACTTTTGGATGTTTAAAGTAATCA	ACCTAGAACAATATTGCAGTAAGG
C58144515_334	TCTGTCGGTGTCTTGACAT	TCTGTCGGTGTCTTGACAC	GGCAGAGGAACAGATGCCGA	CGATTTCCAGACGACGCT
C58147081_320	CCATTAATCGAATGTAACAACATCAATCTAGA	CCATTAATCGAATGTAACAACATCAATCTAGC	TGATGTAGAGGTAAGATATCTGTGAGT	TTCTTCCATTAATCGAATGTAACAACA
C58149639_163	CCTGTTCTACAAATCAACTTGTAACTTTTGATC	CCTGTTCTACAAATCAACTTGTAACTTTTGATT	CCACTACCAGGTTGCACCG	CCCCTGAGTTTCTAGAAAACTCATA
C58154395_144	ATGATAAAGCATAATGCCTTTGTGAGTTT	TGATAAAGCATAATGCCTTTGTGAGTTC	CGACGGATTTGATTATATTACAAAAATGCT	CATGGTGGCGAAGAACC
C58154411_294	CGTTTGCCCATCACGACC	CGTTTGCCCATCACGACT	CTTGTGGACTCCATTCTTCG	AGAGACAGGGTATTCTACTTCGC
C58162213_358	GCACGGATTGAGTTCGCC	AGCACGGATTGAGTTCGCT	CGTGTTAATGACGGTTCGA	GCGAATCCGATTGAGTAAATGAG
C58176181_376	AGAGACGGATGAAGGCTTAACAG	AAGAGACGGATGAAGGCTTAACAT	TCCTTCACTTTATTCATAATGCGTTTAGTCC	CGTGAAAATGTGCACGTC
C58183565_107	AGCGTTCCTCTCAATTGCT	AGCGTTCCTCTCAATTGCA	TGTGTTCCCTGCGTCCA	TGCAATTGGTAATAGAGCGTTCC
C58183857_335	CTTGCATCCCTACACAGCAC	CCTTGCATCCCTACACAGCAA	AATGATCTGGCAGTGCCTGC	CGCCTTGCATTACCCGAAT
C58185845_156	CGACTGACCAGACACCTT	CGACTGACCAGACACCCG	AGGCAAGGTAGGAAGTGTGACA	ACTTGTTCGATCTTTGAACCTGC
C58187921_109	TCTTATCGCCGCATAAGCA	TCTTATCGCCGCATAAGCG	CGCCATGACCGTGCCA	TTCCCTTACGTACTGACTCG
C58188387_205	GCATTTGCGGTTACAGGTTATT	GCATTTGCGGTTACAGGTTATA	TGTACCTATAATCGTCACGGATGA	TCTCGACGCAAAGCTTTGG
C58189295_381	AGCACAATAGATCCTGATATGGTTCTT	CAGCACAATAGATCCTGATATGGTTCTTA	GCATCTTTATATATCTTCTGCCATGTTATAGCTTGT	ACAAAACCATCATATACATTGAGCA
C58204155_374	CTGACCAACAAGGTATTCATGTCT	CTGACCAACAAGGTATTCATGTCC	GCCAAGAGAACTCTCGGGTCTT	GAGGGTCCCATACTACTGTT
C58206533_414	AGACAGCTGCTTTGCCATATCTT	GACAGCTGCTTTGCCATATCTC	TGGATGACACTGCCAGTTACTTAACTATT	ATGTGAACAGAGTTATGCTTATGGTT
C58207301_355	CATCTGGGCTCGAAACTTTTCC	TCATCTGGGCTCGAAACTTTTCT	GCAAGTTGAGTTTGGCTGTGAAA	GACACCATTACCATCATCTGGG
C58207381_110	GCTCTGCCAAACGGTCATATCAC	GCTCTGCCAAACGGTCATATCTAT	AGCGACAGAACCCTTAGGCT	TGTTACACGGTGGCAAGC
C58218379_151	GTTTTCAAGAAATGCTCAGGATAGCG	GTTTTCAAGAAATGCTCAGGATAGCA	GCGTCACGAGTCTAGTACCC	GTCAGTTTTTCAAGAAATGCTCA
C58223239_283	TGCATGTAGGACCTTTGTTCAAATTT	TGCATGTAGGACCTTTGTTCAAATTC	GCATCGCAGACATCTGGGTTT	TGCTGAAAAGAGGCTTATTGCA
C58229657_296	CCTTAACCGCGCCG	ATCCTTAACCGCGCCGT	CCCAGATTAGCCGTTTGAACA	CCATCCTTAACCGCC
C58241727_209	GCGCAGATTAAGGGATTGCATG	GCGCAGATTAAGGGATTGCAT	AATGCAAAACAGAGGATAAGAATAAACT	GCTTTTATGTTCTGCATAAATTTGCG

C58243183_324	AGTTAGTCCATCTTGCCTAAATG T	GTTAGTCCATCTTGCCTAAATG C	CTGCATACTAGCAGCCAGGAG	AAGTTCATCATTCATGAAGTTAGTCCAT
C58244295_308	TGGTCCGTCGCGCTAA A	TGGTCCGTCGCGCTAA G	GCCCGTGTGCTGTTGTTGAA	CCCGTCTGCGGGTTT
C58245125_129	CTGAGGTATGTAAACTTTCTCAAATG C A	CTGAGGTATGTAAACTTTCTCAAATG C G	TATGTCATCAATGCTGATGTTGTCA	CATGTTAACCCACCACTGAGG
C58245899_308	GTCCACACTCAGTTTAGTAAGG A G	GTCCACACTCAGTTTAGTAAGG A A	GGACGGACTCGCCTTCTCT	CGCTACCACATGGTCGTG
C58247267_448	ACATGGTTAAGGAATTAAGATTGCTG T G	ACATGGTTAAGGAATTAAGATTGCTG T A	CCCAATCCGTCGGAATGGC	CTTGTGGAAACATCTCCGTGTA
C58247721_268	CATTCAAGTAGCGTATTGTTG T G	CATTCAAGTAGCGTATTGTTG T G	GAAATGTGATGGGAAATGGAACGGT	GTCCTCGATTACAGATTAATCAC
C58248675_231	CAGCTCTGCAATCGCG A G	CAGCTCTGCAATCGCG A A	CACGTGACGTATTTAGAGATTACAGATGC	TTCCAAACGAATATAGAAGTGCAG
C58250755_362	ACAATAGACCTTGCTGACAC A G	ACAATAGACCTTGCTGACAC A G	GGAATTTGTGCGGACCGCAT	ATTCTCTAACCATTCATTGATTCACAC
C58259409_462	CCCAATTATCACACCTTCTG T G	ACCCAATTATCACACCTTCTG T A	GCTTAGTATATCATATATTCATAGTATGTCAAATATTCAG	TCAGAGACCACGGCACC
C58264051_322	TCCGAGGACTTGGTCC A CT	CCGAGGACTTGGTCC A CC	GCGGAAGGTTCCACGGC	GTAATTTTGGAGCGCTTGTGTATC
C58265315_294	AGGACATAGAGAGCATACCTTCTG A T	GGACATAGAGAGCATACCTTCTG A T	CCTGCTGAACCGGAGTCTCA	AGCCTTTAACTTTGCCAATTAGTCA
scaffold16518_471	CTACTCTACTGGAGTAGAAACAATG A G	CTACTCTACTGGAGTAGAAACAATG A A	GCTGAAAGCGCTCTTCAGGA	GCATGTGTTGTTATTGCTAGGTTT
scaffold22457_284	CAAAATGCCCTCAAAGTCAC A G	CAAAATGCCCTCAAAGTCAC A G	GGGGCCATGCATGATTACTCTG	TCTTGTACAAAATGCATAACCAATTGC
scaffold27410_493	TCACTTGTCCACTTTAGCC A CA	TCACTTGTCCACTTTAGCC A CT	CAGTGTCCACTTTGGTGCCT	ACACTTGAAAGACTTCACTTGTCC
scaffold27633_199	CATCGGAGCCAATCG C G	GCATCGGAGCCAATCG C A	CTTTCGGTGTGTTGTTGCGGAG	ACAAAAGAGCATCGCGCA
scaffold27724_154	CACTCACCAACTATTC A CTCACT	ACTCACCAACTATTC A CTCAC C	AGTAGAGATCCAAGAGAGGCAA	GCGACTACCCATTC A CTC
scaffold27958_488	CATTCCACCCTTGTTTACATGTT A CTC	CATTCCACCCTTGTTTACATGTT A CTA	CGGAAATGGCTTGCCATGAGTG	TCATGTGGTAACAAGACTCTTTGC
scaffold29771_706	GTAGACAAGCCTCAC C CA	GTAGACAAGCCTCAC C CG	TGTTACAGTGAGGTTCAAGGT	CTTCTAGCATAGGCCTATCTTTACG
scaffold30207_304	CGCCGATGACATAAAATCC A CA	CGCCGATGACATAAAATCC A C	TTCTTAACATCATTGCCCTTCTG A GA	AACACGTGCCCTTATTATTC A CCA
scaffold30490_126	ATAATGCTCATTCTACAGTCATAG A CA	GATAATGCTCATTCTACAGTCATAG A CT	GATAGCGGTGTTGTTGCAGCT	GCCAACAGCGACA C GA
scaffold34871_957	CTTTATTCTATAA T CACTCCAGC C CG	CCTTTATTCTATAA T CACTCCAGC A CA	CCATTAGTAGAGGCGGAGCGA	GCATGCGGTATGCTTATG C C
scaffold35512_525	CAAGTGAATGCACATTAATATGTTG G TTT	CAAGTGAATGCACATTAATATGTTG G TTT C	CCACAGATTTCAAACCATCT C CA	TTGACCAATGATTCTAAAATATCTGCTCT
scaffold35825_1168	AGATTCACTTGGTG G TTAGATA A CA T	AGATTCACTTGGTG G TTAGATA A CA A	GAGCAAGTGTGCCCTTGTATC A G	TGATCTCTACAAGATTC A CTTGG T
scaffold36378_587	GTTTTCAGACACATCGTTT G CT T	GTTTTCAGACACATCGTTT G CT A	GTGAAAGCAGCTCTGAG C CC	GGTGTAGGTGCAG C AA C G
scaffold37592_350	GTCACAAGCCCTTCTATAT A CT T	TCACAAGCCCTTCTATAT A CT C	GTGGGGTTAAGGGTGATTAG A TC	CTGGACGTGGCCCTT T G
scaffold37974_361	ACGAGAGGAAAACAGGTAGACCTAT A T A	ACGAGAGGAAAACAGGTAGACCTAT A T T	GGCAGGACTCACACCT T T	CATCACTAACAAAACAA C GGAA A CA
scaffold38155_334	CTAAAATTGCGGGATTAAGTATCAGTAATAT A CA T	CTAAAATTGCGGGATTAAGTATCAGTAATAT A CA C	GCGTGGAGCATAAA T TCTACATTAG A CAA	CAATCTTGCC C CGAT A GC
scaffold39128_861	GAGCTAGAAGCCTGTTAATAG C A G A	AGCTAGAAGCCTGTTAATAG C A G G	CGGTGAAGGAGCCATG C TG	GCGGTACCATCCTAAGGAATAT T TT
scaffold39376_898	CTTTATTTGCGACACAAG G CT C T A	CTTTATTTGCGACACAAG G CT C T C	CAAAACACAGCAGAG G CG A	CGTGGTCTGA A CTCT G C
scaffold39558_382	GTCTGTCTGACAACTGTTG A T T G	CGTGTCTCTGACAACTGTTG A T T A	CTCAGTTAGCCTCTT C T C A	CTTCGAAAGTGAA A CG A AC G C
scaffold6417_451	ACTATGGTATTATCATTATGATACT C CG A	ACTATGGTATTATCATTATGATACT C CG G	TATCATTATCATTATGATTACTGCCATTACAATAG A TT	CTGCTATTATTGGG A AA T AATGG T CT
scaffold7867_106	TTCTTTGGCGGAAG C A G	TTCTTTGGCGGAAG C A A	GAGAGGAGCCTTGTCT T C A	CGTGT A CCCGAGT G CAA A
scaffold8390_261	CGTTAATTTAAGCTCCTCC C TT A T G	CGTTAATTTAAGCTCCTCC C TT A T T	CCACCGCATCTGG A CG A	TGATTACA C CAATCAGAC A CA C C

APÊNDICE B – N° de alelos, Menor Frequência alélica e Hardy-Weinberg do painel com 96 SNPs e 86 SNPs das 325 amostras do produtor de larvas e das lojas de varejo.

SNPs	N° de alelos	MAF	HW	Painel 96	Painel 86
C56970858_106	2	0,50	0,34	X	X
C57351482_140	2	0,30	0,00*	X	X
C57466369_134	2	0,39	0,00*	X	X
C57578086_155	2	0,49	0,00*	X	X
C57603451_134	2	0,12	0,01	X	X
C57611531_136	2	0,35	0,00*	X	X
C57638303_137	2	0,32	0,01	X	
C57647905_163	2	0,45	0,37	X	X
C57658957_160	2	0,34	0,05	X	X
C57673261_175	2	0,45	0,48	X	X
C57706433_166	2	0,38	0,96	X	X
C57761763_188	2	0,38	0,02	X	
C57819698_180	2	0,44	0,28	X	X
C57850365_195	2	0,31	0,27	X	
C57900351_108	2	0,44	0,35	X	
C57906969_178	2	0,41	0,96	X	X
C57917905_171	2	0,49	0,01	X	X
C57932233_144	2	0,22	0,42	X	X
C57987359_182	2	0,14	0,02	X	X
C57988857_210	2	0,46	0,72	X	X
C57991179_206	2	0,28	0,55	X	X
C57994195_257	2	0,31	0,04	X	X
C57996339_174	2	0,46	0,02	X	X
C57997907_129	2	0,41	0,00*	X	X
C58003019_178	2	0,39	0,07	X	
C58013521_113	2	0,47	0,66	X	X
C58020677_108	2	0,30	0,01	X	X
C58033295_194	2	0,20	0,00*	X	X
C58038927_235	2	0,33	0,82	X	X
C58044867_165	2	0,31	0,87	X	X
C58046567_190	2	0,36	0,35	X	
C58064153_284	2	0,36	0,70	X	X
C58068863_216	2	0,50	0,00*	X	X
C58071739_184	2	0,29	0,08	X	X
C58078597_141	2	0,31	0,16	X	X
C58082151_285	2	0,18	0,00*	X	X
C58090729_165	2	0,50	0,01	X	X
C58090973_205	2	0,43	0,00*	X	X
C58101525_229	2	0,42	0,80	X	X
C58104727_290	2	0,41	0,19	X	X
C58106625_158	2	0,42	0,91	X	X
C58107721_279	2	0,45	0,59	X	X
C58120413_184	2	0,33	0,00*	X	X
C58144515_334	2	0,24	0,89	X	X
C58147081_320	2	0,35	0,00*	X	X
C58149639_163	2	0,49	1,00	X	X
C58154395_144	2	0,38	0,31	X	X
C58169193_137	2	0,27	0,01	X	X

APÊNDICE B – Continuação

SNPs	Nº de alelos	MAF	HW	Painel 96	Painel 86
C58176181_376	2	0,27	0,36	X	X
C58183857_335	2	0,47	0,00*	X	X
C58185845_156	2	0,25	0,54	X	X
C58187921_109	2	0,34	0,62	X	X
C58188387_205	2	0,34	0,95	X	X
C58189295_381	2	0,33	0,10	X	X
C58198407_192	2	0,50	0,47	X	X
C58204155_374	2	0,33	0,12	X	X
C58206533_414	2	0,46	0,00*	X	
C58207301_355	2	0,36	0,08	X	X
C58218379_151	2	0,35	0,04	X	
C58220123_303	2	0,39	0,51	X	X
C58228007_131	2	0,32	0,67	X	X
C58229657_296	2	0,48	0,02	X	X
C58231007_237	2	0,29	0,00*	X	X
C58241727_209	2	0,14	0,00*	X	X
C58243183_324	2	0,45	0,01	X	X
C58245899_308	2	0,26	0,00*	X	X
C58247721_268	2	0,41	0,00*	X	X
C58248675_231	2	0,34	0,15	X	X
C58250755_362	2	0,44	0,94	X	X
C58257071_518	2	0,45	0,04	X	X
C58259409_462	2	0,38	0,00*	X	X
C58264015_701	2	0,45	0,15	X	X
C58265315_294	2	0,28	0,57	X	X
C58265895_305	2	0,44	0,79	X	X
C58267663_196	2	0,32	0,16	X	X
C58269083_813	2	0,41	0,00*	X	X
scaffold12413_216	2	0,36	0,85	X	X
scaffold22457_284	2	0,32	0,00*	X	
scaffold26055_381	2	0,46	0,04	X	X
scaffold27410_493	2	0,44	0,22	X	X
scaffold27633_199	2	0,39	0,32	X	X
scaffold27958_488	2	0,31	0,01	X	X
scaffold29436_237	2	0,49	0,13	X	X
scaffold29771_706	2	0,41	0,52	X	X
scaffold30207_304	2	0,32	0,23	X	X
scaffold30490_126	2	0,32	0,12	X	X
scaffold35825_1168	2	0,39	0,70	X	X
scaffold36378_587	2	0,45	0,18	X	X
scaffold37592_350	2	0,22	0,00*	X	X
scaffold38155_334	2	0,42	0,37	X	X
scaffold38653_880	2	0,35	0,39	X	X
scaffold39128_861	2	0,39	0,49	X	X
scaffold6417_451	2	0,39	0,00*	X	
scaffold7364_386	2	0,45	0,02	X	X
scaffold7867_106	2	0,43	0,00*	X	X
scaffold8390_261	2	0,48	0,03	X	X

* p<0,01