

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP  
CÂMPUS DE JABOTICABAL**

**ESTRUTURA GENÔMICA DE UMA POPULAÇÃO DE  
SUÍNOS BASE LANDRACE**

**Letícia Borges Joaquim**  
Zootecnista

**2016**

**UNIVERSIDADE ESTADUAL PAULISTA – UNESP  
CÂMPUS DE JABOTICABAL**

**ESTRUTURA GENÔMICA DE UMA POPULAÇÃO DE  
SUÍNOS BASE LANDRACE**

**Letícia Borges Joaquim**

**Orientador: Prof. Dr. Danísio Prado Munari**

**Coorientadores: Dra. Mônica Corrêa Ledur**

**Dr. Rodrigo Pelicioni Savegnago**

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Genética e Melhoramento Animal.

**2016**

J62e      Joaquim, Letícia Borges  
Estrutura genômica de uma população de suínos base Landrace /  
Letícia Borges Joaquim. – – Jaboticabal, 2016  
xv, 37 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista,  
Faculdade de Ciências Agrárias e Veterinárias, 2016  
Orientadora: Danísio Prado Munari  
Co-orientador: Mônica Côrrea Ledur, Rodrigo Pelicioni Savegnago  
Banca examinadora: Maurício de Alvarenga Mudadu, Nedenia  
Bonvino Stafuzza  
Bibliografia

1. Segmentos de homozigose. 2. Endogamia. 3. Estratificação da  
população. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e  
Veterinárias.

CDU 636.082:636.4

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação –  
Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO: ESTRUTURA GENÔMICA DE UMA POPULAÇÃO DE SUÍNOS  
BASE LANDRACE.**

**AUTORA: LETÍCIA BORGES JOAQUIM**

**ORIENTADOR: DANISIO PRADO MUNARI**

**COORIENTADOR: RODRIGO PELICIONI SAVEGNAGO**

**COORIENTADORA: MÔNICA CORRÊA LEDUR**

Aprovada como parte das exigências para obtenção do Título de Mestra em GENÉTICA E  
MELHORAMENTO ANIMAL, pela Comissão Examinadora:



Prof. Dr. DANISIO PRADO MUNARI

Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal



Pesquisador MAURÍCIO DE ALVARENGA MUDADU

Bioinformática / EMBRAPA - Campinas/SP



Pós-doutoranda NEDENIA BONMINO STAFUZZA

Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal

Jaboticabal, 22 de fevereiro de 2016

## **DADOS CURRICULARES DO AUTOR**

**LETÍCIA BORGES JOAQUIM** – nascida em dia 5 de maio de 1991, na cidade Tubarão, Santa Catarina, filha de Rildo Vieira Joaquim e Márcia Borges Joaquim. Iniciou o curso de Zootecnia em março de 2009 na Universidade Federal de Santa Catarina, Câmpus Florianópolis, e obteve o título de bacharel em Zootecnia em março de 2014. Em agosto de 2014, ingressou no curso de mestrado pelo Programa de Pós-graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Câmpus Jaboticabal sob orientação do Prof. Dr. Danísio Prado Munari e coorientação da Dra. Mônica Corrêa Ledur e do Dr. Rodrigo Pelicioni Savegnago. Foi bolsista CAPES pelo período de 01 de agosto de 2014 a 29 de fevereiro de 2016.

“Eu tentei 99 vezes e falhei. Mas na centésima tentativa eu consegui. Nunca desista de seus objetivos, mesmo que eles pareçam impossíveis. A próxima tentativa pode ser a vitoriosa”.

Albert Einstein

Aos meus pais, Rildo e Márcia, por sempre me apoiarem e nunca medirem esforços para que eu chegasse até aqui. A minha irmã, Patrícia, pelo companheirismo e amor imensurável.

**DEDICO**

## AGRADECIMENTOS

A toda minha família, que soube apoiar as minhas escolhas e compreenderam a minha ausência. Aos meus pais Rildo e Márcia por serem modelos de honestidade, altruísmo e força e vontade. À minha irmã, Patrícia, por estar sempre presente mesmo distante na maioria das vezes (obrigada pela inúmeras chamadas pelo “FaceTime”). Aos meus avos, Ary e Ana, por toda inspiração, apoio e mimos, Pedro e Ivonete (*in memoriam*), por toda dedicação em minha criação. Aos meus tios e primos pelos momentos alegres compartilhados.

Ao meu orientador, Prof. Dr. Danísio Prado Munari, por me aceitar e receber tão bem durante esse período. Pela confiança em mim depositada na execução desse trabalho e pelos ensinamentos que muito contribuíram para conclusão de mais uma etapa da minha formação.

Aos meus coorientadores, Dra. Mônica Ledur e Dr. Rodrigo Savegnago, pelo auxílio no desenvolvimento deste trabalho, atenção e dedicação. À Mônica e toda sua equipe Embrapa Suínos e Aves, pelo fornecimento dos dados e por me acolherem carinhosamente. Obrigada especialmente aos Dr. Maurício Cantão e Dr. Ricardo Zanella e à Dra. Mônica, sem vocês o desenvolvimento desse trabalho seria muito mais difícil, por tanto obrigada pelos conhecimentos ensinados.

Ao querido Prof. Dr. Renato Irgang, por todos os ensinamentos que muito contribuíram para minha formação acadêmica e pessoal. Obrigada por todos os incentivos e pelas oportunidades de trabalho.

À minha família de Jaboticabal, Tati e Jaque, por me receberem quando o mestrado era apenas uma possibilidade e me abrigarem quando tornou-se realidade. À Tati pelo companheirismo, puxões de orelha, preocupação, e por dividir os cuidados dos bestinhas, Bentinho e Zag, que sempre nos alegram, especialmente nos momentos mais difíceis. À Jaque pela amizade, momentos de distração, e por dividir sua família comigo.

A todos os amigos de departamento de Ciências Exatas e da Pós-Graduação de Genética e Melhoramento Animal. Em especial: Tatiane, Jaqueline, Jorge, Ana Paula, Priscila, Thiago Bruno, Marcus, Guilherme Nascimento, Rodrigo, Salvador, Daniela Barreto, Bianca e Laisa, vocês tornaram essa caminha mais leve e divertida.

Além de agradecer pela diversão dividida, gostaria de agradecê-los por todo aprendizado que me proporcionaram.

Aos antigos amigos, Aline, Mariane, Marina, Max, Elisa, Gabi, Henrique, Dani, Kamila, Karol, pela amizade verdadeira que mesmo com a distância permanece. Obrigada pelas palavras de incentivo.

Aos queridos, Mario e Eliane, por todos os momentos que compartilhamos e pelo carinho e cuidado que sempre tiveram comigo.

Aos membros da banca do Exame Geral de Qualificação, Dr. Ricardo Ventura e Dr. Marcos Eli Buzanskas, pelas sugestões que enriqueceram esse trabalho.

Aos membros da banca de defesa da dissertação, Dr. Maurício Mudadu e Dra. Nedenia Stafuzza, pelas valiosas sugestões.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida no curso de mestrado.

A todos que contribuíram em algum momento, mas minha memória não permitiu citar.

## SUMÁRIO

|                                         |    |
|-----------------------------------------|----|
| RESUMO.....                             | ii |
| ABSTRACT .....                          | iv |
| LISTA DE TABELAS .....                  | v  |
| LISTA DE FIGURAS .....                  | vi |
| 1 INTRODUÇÃO .....                      | 1  |
| 2 REVISÃO BIBLIOGRÁFICA.....            | 2  |
| 2.1 Segmentos de homozigose.....        | 2  |
| 2.2 Desequilíbrio de ligação .....      | 6  |
| 2.3 Estratificação da população .....   | 8  |
| 3 MATERIAL E MÉTODOS .....              | 11 |
| 3.1 Animais e Extração do DNA.....      | 11 |
| 3.2 Genotipagem.....                    | 11 |
| 3.3 Desequilíbrio de ligação (LD) ..... | 12 |
| 3.4 Coeficiente de endogamia.....       | 12 |
| 3.5 Segmentos de Homozigose .....       | 12 |
| 3.6 Estratificação da População.....    | 13 |
| 4 RESULTADOS E DISCUSSÃO.....           | 14 |
| 4.1 Desequilíbrio de Ligação.....       | 16 |
| 4.2 Coeficiente de endogamia.....       | 17 |
| 4.3 Estratificação da População.....    | 23 |
| 5 CONCLUSÃO.....                        | 27 |
| REFERÊNCIAS.....                        | 28 |
| 7 APÊNDICE .....                        | 36 |

## ESTRUTURA GENÔMICA DE UMA POPULAÇÃO DE SUÍNOS BASE LANDRACE.

**RESUMO** - Os painéis de marcadores de alta densidade têm demonstrado a sua funcionalidade nos estudos de estrutura da população e conservação genética. Esses painéis permitem avaliar similaridades no padrão do desequilíbrio de ligação em toda população, assim como informações sobre parentesco da população. Segmentos de homozigose são utilizados como indicativo da estrutura da população e fornecem informações sobre o histórico demográfico e eventos de endogamia da mesma. O objetivo deste trabalho foi estudar a estrutura genômica de uma linhagem sintética base Landrace por meio de (i) análises de desequilíbrio de ligação; (ii) estimação do coeficiente de endogamia utilizando dados genômicos e de pedigree; (iii) análise do número e tamanho de segmentos de homozigose e (iv) determinação da estratificação da população. Foram utilizados registros de 300 fêmeas e 25 machos de uma linhagem fêmea sintética base Landrace genotipados com o painel Illumina PorcineSNP60 v2 BeadChip. A edição dos dados foi realizada no programa PLINK v.1.9 para remoção de marcadores SNPs (“Single Nucleotide Polymorphism”) que falharam em mais de 10% das amostras (“callrate”) e amostras que falharam em mais de 10% dos marcadores. A extensão do desequilíbrio de ligação foi avaliada entre todos os pares SNPs adjacentes presentes nos cromossomos autossômicos por meio da medida  $r^2$ . O coeficiente de endogamia foi calculado usando registros de pedigree e de dados genômicos. Os segmentos de homozigose foram detectados para os cromossomos autossômicos com o programa computacional PLINK v.1.9, considerando pelo menos 50 SNPs homozigotos dentro de tamanho mínimo de 1.000 Kb por animal, permitindo um SNP heterozigoto e um SNP faltante/perdido dentro de uma janela de 50 SNPs. Os segmentos de homozigose detectados foram utilizados para cálculo do coeficiente de endogamia genômica e como indicativo do histórico da população. A estratificação da população foi avaliada utilizando análises de componentes principais e pelo modelo de ancestralidade por metodologia bayesiana aplicado no programa STRUCTURE. Na análise de ancestralidade foram testados valores de K (número de “clusters”) variando de um a oito usando período de *burn-in* de 1.000 iterações e Cadeia de Markov e Monte Carlo de 10.000 iterações. As análises foram repetidas dez vezes para cada K e a determinação do melhor número para K foi estimada utilizando a estatística Delta K. O valor médio de  $r^2$  encontrado para todos os SNPs adjacentes que estão a uma distância menor que 100 Kb foi de  $0,291 \pm 0,312$ . Os coeficientes de endogamia médios obtidos a partir dos segmentos de homozigose e de registros de pedigree foram de 0,119 e 0,00011, respectivamente. A baixa correlação ( $r < 0,04$ ) encontrada entre os coeficientes de endogamia pode ser explicada pelo efeito de recombinação gênica sobre aqueles estimados a partir dos segmentos de homozigose que não é considerada nas estimativas obtidas à partir de registros de pedigree e devido aos erros de identificação dos animais no pedigree. A identificação de um grande número de longos segmentos de homozigose pode ser indicativa de endogamia recente na população estudada. O estudo da estratificação da amostra indicou que a mesma estaria dividida em duas populações ( $k=2$ ), sendo que a separação pode ser explicada pelo cruzamento entre as raças ocidentais e orientais utilizadas na formação da linhagem. Os dados de genotipagem permitiram

concluir que a população estudada possui endogamia recente, sugerindo-se que seja priorizado acasalamento entre indivíduos menos aparentados para assegurar a manutenção da diversidade genética.

**Palavras-chave:** desequilíbrio de ligação, endogamia, estratificação da população, pedigree, segmentos de homozigose, *Sus scrofa domesticus*

## GENOMIC STRUCTURE OF ADMIXED LANDRACE PIG POPULATION.

**ABSTRACT** – High-density single nucleotide polymorphism (SNP) panels have been used in genomic studies, such as population structure and conservation genetic studies. These panels allow to assess similarities in the patterns of linkage disequilibrium across populations and to estimate relatedness between populations. Runs of homozygosity are used as indicative of population structure and it provides information about demographic history and recent inbreeding. The aim of this study was to describe the genomic structure of a synthetic Landrace line by (i) linkage disequilibrium (LD) analyses; (ii) inbreeding estimates through pedigree and genomic data; (iii) analysing the number and length of runs of homozygosity (ROH); and (iv) determination of population structure. A total of 300 females and 25 males from synthetic Landrace line were genotyped using Illumina PorcineSNP60 v2 BeadChip. Data editing was performed using PLINK v.1.9. The SNPs and samples with a call rate lower than 0.90 were excluded from the data set. Only the autosomal chromosomes were considered for LD and ROH analyses. The LD between all pairs of SNPs were measured by the means of the genotype correlation coefficient ( $r^2$ ) and it determined the decay of LD with physical distance. The coefficient of inbreeding was calculated using genomic and pedigree data. ROH were detected using PLINK v.1.9 considering the follow parameters: a minimum ROH of 50 SNPs with a minimum length of 1000 (Kb), one heterozygous SNP and one missing SNP were allowed within the sliding window of 50 SNPs. The individual genomic inbreeding and population history were identified from estimated ROH. The population structure was evaluated using principal component analysis and Bayesian admixture model. The number of clusters (K) was tested from two to eight considering a burning period of 1,000 followed by 10,000 Markov chain Monte Carlo repetitions and replicated ten times for each K. The best K was estimated using the Delta K statistic. For all SNPs adjacent less than 100 kilobase (kb) apart, the average  $r^2$  was  $0.291 \pm 0.312$ . The average inbreeding coefficient, calculated by ROH and pedigree analyses, were 0.119 and 0.00011, respectively. The low correlation between the inbreeding coefficients can be justified by the genetic recombination effect over those estimated from runs of homozygosity that were no considered on the estimates from the traditional pedigree. The high number of long ROH is an evidence of recent inbreeding in this population. The population structure analysis revealed K=2 was the best number of clusters and separation. This result can be explained by the crossbreeding between the Eastern and Western breeds used in the formation of the line. The genotyping data helped confirm that the inbreeding in the studied population is recent, which suggests that mating between individuals less related occurred to ensure the maintenance of genetic diversity.

**Keywords:** inbreeding, linkage disequilibrium, pedigree analysis, population structure, runs of homozygosity, *Sus scrofa domestica*

## LISTA DE TABELAS

|                                                                                                                                              | Página |
|----------------------------------------------------------------------------------------------------------------------------------------------|--------|
| <b>Tabela 1.</b> Análise descritiva dos marcadores ao longo do genoma, por cromossomo.<br>.....                                              | 15     |
| <b>Tabela 2.</b> Número de animais e tamanho dos segmentos de homozigose<br>compartilhados por mais de 50% da população por cromossomo. .... | 22     |
| <b>Tabela 3.</b> Número total de genes por segmentos de homozigose e suas<br>classificações. ....                                            | 22     |

## LISTA DE FIGURAS

### Página

|                                                                                                                                                                                                                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1.</b> Média de desequilíbrio de ligação ( $r^2$ ) ao longo de diferentes distâncias (Kb) entre marcadores em uma linhagem suína de base Landrace. ....                                                                                                                                         | 17 |
| <b>Figura 2.</b> Histograma do coeficiente de endogamia de uma linhagem suína de base Landrace utilizando dados genômicos (A) e registros de pedigree (B). ...                                                                                                                                            | 18 |
| <b>Figura 3.</b> Relação entre o número e tamanho das segmentos de homozigose em uma linhagem suína base Landrace.....                                                                                                                                                                                    | 19 |
| <b>Figura 4.</b> Número de regiões de homozigose compartilhadas por cromossomos em uma linhagem suína base Landrace.....                                                                                                                                                                                  | 21 |
| <b>Figura 5.</b> Cariograma dos segmentos de homozigose (ROH) compartilhados entre os indivíduos considerando três classes: 1: ROH compartilhados por 10 a 20% dos indivíduos; 2: ROH compartilhados por 21 a 50% indivíduos; e 3: ROH compartilhados por mais de 50% dos indivíduos.....                 | 21 |
| <b>Figura 6.</b> Estatística <i>Delta K</i> baseada no método de clusterização de Evanno et al. (2005). ....                                                                                                                                                                                              | 24 |
| <b>Figura 7.</b> Constituição genética de animais de uma linha fêmea sintética base Landrace, em análise para $K = 2$ . As duas cores são referentes aos grupos definidos (“clusters”) com os resultados do programa STRUCTURE (Pritchard et al., 2000): vermelho = “Cluster” 1; verde = “Cluster” 2..... | 24 |
| <b>Figura 8.</b> Estrutura da população de uma linhagem suína base Landrace revelada pela Análise de Componentes Principais com os SNP com $r^2 \leq 0,2$ . ....                                                                                                                                          | 26 |

## 1 INTRODUÇÃO

A suinocultura tem sido destaque no cenário mundial agropecuário durante muitos anos. Em 2014, 109,5 milhões de toneladas de carne suína foram consumidos (USDA, 2015), representando uma das mais importantes fontes de proteína animal para o consumo humano (MIELE, WAQUIL, 2007). Além disso, os suínos também são utilizados como interessante modelo animal para estudo de doenças humanas (GIELING et al., 2011). Portanto, trabalhos têm sido desenvolvidos com a espécie para analisar a base molecular das características complexas, tais como crescimento, metabolismo energético, reprodução e comportamento (AI, HUANG, REN, 2013).

Em 2009, o painel com 61.565 marcadores do tipo SNP ("Single Nucleotide Polymorphism") foi disponibilizado para os suínos e tornou-se uma poderosa ferramenta para estudos de associação genômica ampla e seleção genômica de características economicamente importantes (EENENNAAM et al., 2013). Além da sua aplicabilidade em estudos com fins econômicos, os quais visam encontrar associações genéticas com características de relevância na produção, esses painéis de alta densidade de SNPs também têm demonstrado sua utilidade na avaliação da estrutura de populações e conservação genética de espécies. Muitas pesquisas utilizando painéis de SNPs estão sendo desenvolvidas para estudar a diversidade e a conservação genética de raças suínas comerciais e locais, de origem europeia e asiática (AI, HUANG, REN, 2013; HERERRO-MEDRANO et al., 2013).

O surgimento dos painéis de SNP possibilitaram a realização de estudo para avaliar os níveis de recombinação gênica, que além de serem necessários em estudos de seleção genômica (O'BRIEN et al., 2014), permitem estimar o tamanho efetivo da população a partir de dados genômicos (HAYES et al., 2003). A estratificação populacional, observada em raças compostas, foi reconhecida como fator de confundimento em estudos de associação genômica ampla, logo estimativas de ancestralidade, oriundas de dados genotípicos, podem ser usadas como covariáveis para corrigir a estratificação populacional (ALEXANDER, NOVEMBRE, LANGE, 2009; WANG et al., 2009). Antes da disponibilidade de grande quantidade

de dados de marcadores, essas estimativas, assim como as estimativas do coeficiente de endogamia, eram obtidas somente a partir do pedigree. No entanto, mesmo quando os dados de pedigree estão disponíveis e são acurados, existem desvantagens relacionadas à falha na detecção do efeito dos parentes distantes e do processo meiótico (McQUILLAN et al., 2008). Esses painéis de alta densidade também permitem avaliar as similaridades no padrão do desequilíbrio de ligação (LD, “Linkage Disequilibrium”) em toda população, fornecendo informações sobre o parentesco da população (WANG et al., 2013). Assim como o LD, os segmentos de homozigose (ROH, “Runs of Homozygosity”) também apresentam indicativo sobre a estrutura da população, fornecendo informações sobre o histórico demográfico e eventos de endogamia da mesma (BOSSE et al., 2012; PURFIELD et al., 2012). Em humanos, mapeamentos de homozigosidade estão sendo realizados para identificar os ROH e associá-los com doenças, tendo sido identificados segmentos de homozigose associados com a doença de Alzheimer (HOWRIGAN, SIMONSON, KELLER, 2011; YANG et al., 2012).

Dado que o estudo da estrutura da população é importante em análises genéticas, uma vez que fornece informações sobre a ancestralidade genética de indivíduos e história evolutiva de populações, o objetivo deste trabalho foi estudar a estrutura genômica de uma linhagem sintética de base Landrace por meio de (i) análises de desequilíbrio de ligação; (ii) estimação do coeficiente de endogamia utilizando dados genômicos e de pedigree; (iii) análise do número e tamanho de segmentos ROH e (iv) determinação da estratificação da população.

## **2 REVISÃO BIBLIOGRÁFICA**

### **2.1 Segmentos de homozigose**

A disponibilidade de painéis de SNP de alta densidade surgiu como uma ferramenta essencial para os estudos de associação genômica ampla e seleção de características economicamente importantes. Além do uso para fins da produção

agropecuária, estes painéis têm demonstrado a sua funcionalidade no estudo da estrutura e conservação das populações (HERERRO-MEDRANO et al., 2013).

A análise dos dados genéticos individuais possibilitou a identificação de segmentos de homozigose e revelou a ocorrência comum desses segmentos nas populações mundiais. Portanto, há interesse em compreender esses segmentos de homozigose e são realizadas análises para investigar tanto questões relacionadas com a teoria básica de genética de populações quanto em estudos de susceptibilidade à doenças. Em estudos de genética de populações, trabalhos foram realizados para analisar a distribuição, localização e prevalência dos ROH com objetivo de inferir sobre a estrutura e histórico da população e sobre a seleção natural (HOWRIGAN, SIMONSON, KELLER, 2011).

Os segmentos de homozigose surgem quando os pais de um animal possuem um ancestral comum relativamente recente, isto porque quando pais são aparentados esses compartilham grande parte dos seus genomas que são idênticos por descendência. Caso ambos os pais transmitam um segmento idêntico para sua progênie, a progênie será homozigota para o mesmo, dando origem aos ROH (KIRIN et al., 2010; PURFIELD et al., 2012; KIM et al., 2013). Entretanto, os pais podem passar um segmento cromossômico idêntico para sua progênie mesmo quando o parentesco entre esses for muito distante. Portanto, o tamanho de um segmento de homozigose contínuo depende do grau de parentesco dos progenitores, em que longos ROH são originários de endogamia recente, enquanto que, ROH devido ao acasalamento de parentes distante são geralmente muito menores (KIRIN et al., 2010).

Quando dados genômicos são utilizados, os segmentos de homozigose são definidos como longos, contendo SNPs homozigóticos (FERENČAKOVIĆ et al., 2012), sendo permitido que essa região rica em homozigose seja interrompida por um pequeno número de genótipos heterozigotos decorrentes de erros de genotipagem, genótipos ausentes ou mutações. Em geral, esses segmentos de homozigose são resultado de uma série de mecanismos que incluem alterações cromossômicas, homozigose como consequência da endogamia e seleção natural (YANG et al., 2012).

Uma limitação dos estudos sobre ROH utilizando dados genômicos é a falta de critério para a definição de um segmento de homozigose (HOWRIGAN, SIMONSON, KELLER, 2011). Essa situação tem sido observada na maioria dos trabalhos realizados com segmentos de homozigose, a exemplo dos estudos da estrutura da população de suínos, nos quais Ai, Huang e Ren (2013) avaliaram ROH com tamanho mínimo de 500 Kb, permitindo um SNP heterozigoto e cinco SNPs faltantes por janela. Para Hererro-Medrano et al. (2013), ROH foram definidos por um tamanho mínimo de 10 Kb e 20 SNPs homozigotos, permitindo apenas um SNP heterozigoto. Este mesmo critério foi utilizado por Bosse et al. (2012), variando apenas na remoção dos SNPs que apresentaram MAF ("Minor Allele Frequency") menor que 5%. Estas diferenças entre as definições de ROH dificultam comparações entre os resultados dos estudos, e a falta de um critério consenso eleva a probabilidade de encontrar resultados falso-positivos devido ao potencial em escolher a definição mais significativa para ROH entre todos os critérios investigados (HOWRIGAN, SIMONSON, KELLER, 2011).

O primeiro estudo de segmentos de homozigose foi realizado em humanos por Gibson, Morton e Collins (2006) e em bovinos por Sölkner et al. (2010). Em populações humanas, o uso de ROH é um método estabelecido para analisar a endogamia da população e avaliar a susceptibilidade a doenças por mapeamentos homozigóticos (PURFIELD et al., 2012). O mapeamento de homozigose visa identificar ROH associados com doenças, tendo sido originalmente desenvolvido para mapear genes responsáveis por doenças recessivas de linhagens puras em humanos. Estudos demonstraram que este mapeamento é um método estatisticamente poderoso ao identificar genes de susceptibilidade associados a doenças complexas, cânceres e características fenotípicas (YANG et al., 2012).

Em humanos, análises de maior escala utilizando dados genômicos foram desenvolvidas em fenótipos complexos, tais como no estudo da esquizofrenia, transtorno bipolar, Parkinson, Alzheimer, câncer colorretal, de mama e próstata. Outras características, como exemplo de altura e peso, também foram avaliadas usando análises de segmentos de homozigose (HOWRIGAN, SIMONSON, KELLER, 2011).

Os estudos com animais se concentram na avaliação de ROH como indicativo da estrutura e histórico da população a partir da análise de parâmetros, tais como coeficiente de endogamia. Devido à recente endogamia, ROH tendem a ser longos, porque a possibilidade de recombinação para quebrar os segmentos que são idênticos por descendência é pequena. Por outro lado, ROH de origem antiga são geralmente menores, já que os segmentos cromossômicos têm sido quebrados por sucessivas meioses. No entanto, alguns segmentos longos podem persistir por ocorrerem em regiões do genoma que apresentam baixas taxas de recombinação, como por exemplo, regiões próximas ao centrômero (KIRIN et al., 2010).

O acasalamento de indivíduos aparentados, endogamia, é uma importante questão na produção animal, uma vez que leva ao aumento na frequência de alelos em homozigose, elevando o risco de indivíduos serem homozigotos para alelos deletérios e reduzindo a diversidade genética (SCRAGGS et al., 2013). A taxa de endogamia tem aumentado com a aplicação intensiva de seleção nas populações e, é por isso que as informações do coeficiente de endogamia devem ser consideradas nos programas de melhoramento genético animal para controlar o aumento do nível de endogamia. O coeficiente de endogamia é definido pela probabilidade de que dois alelos de um indivíduo sejam idênticos por descendência (IBD – “Identical By Descent”) em relação a uma população base, na qual todos os alelos são considerados independentes (ZHANG et al., 2015). A abordagem clássica utilizada para estimar o coeficiente de endogamia ( $F_x$ ) em populações animais é obtida a partir de registros de pedigree; contudo é considerado difícil controlar a genealogia não apenas em populações selvagens, mas também naquelas que são criadas extensivamente (SILIÓ et al., 2013).

De acordo com Howrigan, Simonson e Keller (2011), o coeficiente de endogamia estimado a partir dos segmentos de homozigose é melhor na detecção de mutações recessivas e raras em comparação a outras alternativas, como o coeficiente de endogamia definido pelo método “SNP-by-SNP” e daquelas definidas com o uso do pedigree. A estimativa do coeficiente de endogamia utilizando a metodologia “SNP-by-SNP” é obtida com base na proporção dos genótipos homozigotos contudo a mesma apresenta a desvantagem de não diferenciar os genótipos entre idênticos por descendência e idênticos por estado (IBS – “Identical

By State”), sendo que os segmentos de homozigose são uma alternativa para quantificar homozigose individual que melhor reflete IBD (SAURA et al., 2015). Além disso, as vantagens da endogamia obtida com uso de ROH ( $F_{ROH}$ ) em comparação a endogamia com base no pedigree ( $F_{PED}$ ) são:  $F_{ROH}$  pode prever a porcentagem real do genoma que é homozigota com maior precisão, em comparação com  $F_{PED}$ ;  $F_{ROH}$  pode capturar homozigose decorrente de ancestrais comuns muito distantes (por exemplo, mais de 50 gerações anteriores);  $F_{ROH}$  pode ser estimado em qualquer indivíduo genotipado, apesar de registros de pedigree não estarem disponíveis e  $F_{ROH}$  oferece a possibilidade de examinar a distribuição de homozigose em todo o genoma e identificar locais específicos no genoma que apresentam maiores níveis de homozigose (KAMIRI, 2013).

## 2.2 Desequilíbrio de ligação

A disponibilidade de um grande número de SNPs genotipados também permite o estudo do desequilíbrio de ligação que, assim como os segmentos de homozigose, fornece informações sobre o passado e presente da população. O desequilíbrio de ligação mede o grau de associação não aleatória de alelos em diferentes loci em uma população, descrito inicialmente como desequilíbrio gamético por Lewontin e Kojima (1960). Muitos estudos em humanos e animais têm sido direcionados para a caracterização do padrão do LD, uma vez que a extensão do LD pode revelar informações sobre o histórico populacional e desenvolvimento de raças, estando relacionado ainda com o tamanho efetivo da população (WANG et al., 2013).

Segundo O’Brien et al. (2014), o LD é descrito por medidas estatísticas, nas quais se destacam  $D'$  (LEWONTIN, 1964) e  $r^2$  (HILL; ROBERTSON, 1968) como sendo as mais importantes e amplamente utilizadas. A medida de  $r^2$  foi desenvolvida para ser mais robusta, permitindo o uso com marcadores bialélicos e exigindo números menores de amostras para a obtenção de resultados mais acurados em relação ao  $D'$  (O'BRIEN et al., 2014). Além disso, estudos demonstram que a medida de  $D'$  é fortemente dependente dos níveis dos alelos menos frequentes (MAF), enquanto que o  $r^2$  é menos dependente (AMARAL et al., 2008).

O nível do LD entre dois loci em uma população animal é afetado por um conjunto de eventos da história da população, tais como os processos naturais de deriva genética, cruzamento de populações previamente isoladas (migração) e redução no tamanho efetivo da população. Além disso, a seleção natural ou artificial e os eventos de recombinação também alteram a extensão e variação do LD (O'BRIEN et al., 2014; DU, CLUTTER, LOHUIS, 2007).

Os processos de seleção provocam aumento no nível do desequilíbrio de ligação nas regiões genômicas afetadas, já que favorecem o aumento na frequência de determinado alelo no “pool” de genes da população, e assim aumenta a correlação com outros alelos de alta frequência, levando alelos ligados a apresentarem alta prevalência. Por outro lado, há a redução no nível de LD com a rápida expansão da população e com eventos recorrentes de recombinação gênica (O'BRIEN et al., 2014). O cruzamento entre populações cria um novo LD entre loci que não estão em desequilíbrio de ligação em toda a população parental e altera a extensão do LD dos loci que já estavam em LD. O desequilíbrio de ligação entre loci não ligados, criado por cruzamento, pode ser rapidamente passado para as próximas gerações e transmitido lentamente no caso de loci vizinhos (DU, CLUTTER, LOHUIS, 2007).

O conhecimento sobre o desequilíbrio de ligação em populações comerciais apresenta relevância no mapeamento de características economicamente importantes, visto que a seleção genômica depende da existência de LD entre as variantes causais e os marcadores genéticos. O mapeamento de doenças complexas por meio de estudos de associação genômica ampla (GWAS, “Genome-Wide Association Study”) depende do padrão do desequilíbrio de ligação em todo o genoma (BADKE et al., 2012; WANG et al., 2013). Além do mais, o LD entre o marcador e um QTL (“Quantitative Trait Loci”) é apontado como um dos fatores que afetam a acurácia do valor genético genômico, sendo que a extensão do desequilíbrio de ligação de uma população ajuda a determinar o número de marcadores que são necessários para o sucesso da seleção genômica e para GWAS (VERONEZE et al., 2013).

O estudo da extensão do desequilíbrio de ligação também pode auxiliar nas estimativas do tamanho efetivo da população atual e anterior, principalmente quando

os registros de pedigree não estão disponíveis. Contudo, apesar do grande interesse em estudos de conservação, a estimação deste parâmetro é complexa, uma vez que a estimação do tamanho efetivo da população assume uma população ideal que está isolada, na ausência de migração, em crescimento linear e constante e sob acasalamento ao acaso. Apesar do conhecimento de que essas pressuposições não são geralmente atendidas nas populações naturais, as estimativas do tamanho efetivo da população são amplamente aplicadas em estudos populacionais (HERERRO-MEDRANO et al., 2013).

## **2.3 Estratificação da população**

O estudo da estrutura genética de uma população tem atraído interesse crescente nas diferentes áreas do conhecimento, tais como biologia de populações, ecologia molecular e genética médica (JOMBART, DEVILLARD, BALLOUX, 2010). Tecnologias como painéis de SNP de alta densidade e sequenciamento de nova geração têm facilitado a produção de um grande número de dados que permitiram a investigação das relações genéticas em humanos e em outros organismos. Esses dados tem o potencial de gerar um novo nível de conhecimento sobre a estrutura da população, fornecendo informações sobre padrões de dispersão e eventos históricos recentes e antigos da população de interesse (LAWSON et al., 2012a).

Um objetivo importante da genética de populações é resumir, a partir de dados moleculares, as relações entre indivíduos. As duas abordagens mais comuns para analisar a estrutura da população a partir de dados genéticos são exemplificadas pelas Análises de Componentes Principais (PCA, “Principal Component Analysis”), considerada uma abordagem não-paramétrica, e a análise intra-específica de estratificação da população, que explicitamente modela a estrutura da população em agrupamentos (“Clusters”). Essas duas análises são usualmente utilizadas em estudos em que a estrutura da população é avaliada (LAWSON et al., 2012a; LAWSON et al., 2012b).

De acordo com Novembre e Stephens (2008), Cavalli-Sforza, Menozzi e Piazza (1993) foram pioneiros na aplicação da Análise de Componentes Principais em estudos de genética das populações, utilizando PCA para produzir mapas que

resumissem a variação genética humana por regiões continentais. A abordagem de PCA baseia-se na análise de uma matriz de covariância/correlação entre marcadores, que pode ser definida de diferentes maneiras, cujas entradas quantificam a semelhança genética entre pares de indivíduos. Os componentes principais (PC, "Principal Component") dessa matriz representam as direções do espaço máximo da amostra que explicam o padrão observado de similaridade genética. A visualização dos padrões principais da estrutura nos dados pode ser alcançada por representação gráfica de sucessivos PC, em que grupos de indivíduos podem ser interpretados como populações genéticas, enquanto que a mistura de duas populações resulta em grupos de indivíduos alinhados ao longo de uma linha, apesar de que outros eventos históricos também podem produzir sinais de PC idênticos, além de outras questões que podem dificultar a interpretação dos componentes principais (LAWSON et al., 2012a).

O método "STRUCTURE", introduzido por Pritchard, Stephens e Donnelly. (2000), tem sido amplamente aplicado para descrever populações cruzadas e detectar a estrutura genética de amostras com fins médicos, estudos de mistura de populações, análise de hibridação e em estudos de migração e análise da proporção alélica (NOVEMBRE, RAMACHANDRAN, 2011). Além disso, esse método de também tem sido aplicado na detecção da estrutura de populações crípticas naturais (EVANNO, REGNAUT, GOUDET, 2005). Segundo Galleti Jr. et al. (2008), espécies críptico genéticas são morfologicamente idênticas, mas constituem unidades evolutivas independentes. O método utiliza a abordagem Bayesiana para assumir um modelo no qual há K populações, em que K pode ser desconhecido, sendo que cada uma das K populações é caracterizada por um conjunto de diferentes frequências alélicas em cada loci. Esse método atribui indivíduos para populações com base em seus genótipos, enquanto simultaneamente estima a frequência alélica da população. O mesmo pode ser aplicado em vários tipos de marcadores, como microsatélites e SNPs, mas assume que o loci do marcador está em equilíbrio de ligação com outro na população (PRITCHARD, STEPHENS, DONNELLY, 2000). Esse modelo baseia-se no pressuposto de subdivisão populacional, que é muitas vezes difícil de verificar e pode restringir sua aplicabilidade. Além disso, a estimativa de um grande número de parâmetros pode exigir tempo considerável na análise

computacional de grandes conjuntos de dados (JOMBART, DEVILLARD, BALLOUX, 2010).

A partir da estimativa de divergência genética entre subpopulações e a população total ( $F_{st}$ ), descrita por Wright (1965), Alexander et al. (2013) sugeriram as quantidades necessárias de marcadores SNPs para a correção do efeito de estratificação populacional nos estudos de associação genômica. Assim, de acordo com os autores, para populações com origem em diferentes continentes ( $F_{st} > 0,05$ ), 10.000 marcadores do tipo SNP seriam suficientes para realizar a correção para estratificação populacional, enquanto que mais de 100.000 marcadores seriam necessários para a mesma correção se as populações tivessem origem dentro do mesmo continente ( $F_{st} < 0,01$ ). O coeficiente  $F_{st}$  representa tanto a proporção de diversidade genética atribuída às diferenças de frequências alélicas entre as populações quanto as correlações entre frequências alélicas nas populações e na população total, em que estimativas de  $F_{st}$  próximas de 1 podem indicar completo isolamento reprodutivo das populações, enquanto que aquelas próximo de zero indicam cruzamentos totalmente ao acaso (HOLSINGER, WEIR, 2009).

Em suínos, estudos sobre estrutura da população foram aplicados em programas de conservação em animais da península Ibérica, avaliando a estratificação da população entre suínos domesticados e selvagens de populações nativas da região, a partir da análise de componentes principais (HERERRO-MEDRANO et al., 2013). Esse tipo de análise também pode ser aplicada em estudos de diversidade genética, como no trabalho realizado por Ai, Huang e Ren (2013), que avaliaram a diversidade genética entre suínos de origem ocidental e oriental. Além disso, estudos de estratificação da população ainda foram aplicados utilizando o método “STRUCTURE” em trabalhos com raças suínas brasileiras, a exemplo de Silva (2014), que buscou avaliar a distribuição geográfica e a variabilidade genética de suínos Monteiros no pantanal da Nhecolândia e do Mato Grosso do Sul.

### 3 MATERIAL E MÉTODOS

#### 3.1 Animais e Extração do DNA

As amostras de tecido da orelha utilizadas neste estudo foram colhidas de 300 fêmeas e 25 machos de uma linhagem fêmea sintética de base Landrace, denominada comercialmente de Redone. A linhagem foi desenvolvida inicialmente a partir do cruzamento envolvendo a raça Landrace e outras raças orientais e ocidentais, na proporção de 50,00% Landrace, 12,50% Meishan, 12,50% Xia Jing, 8,33% Hampshire, 8,33% Pietrain e 8,33% Large White (LOPES, 2010; RAMACCIOTTE, 2015).

Os animais foram escolhidos para serem representativos do rebanho pertencente ao Programa de Melhoramento Genético de Suínos de uma empresa do estado de Santa Catarina, Brasil. O DNA genômico foi extraído de 200mg do tecido colhido, com o kit PureLink® Genomic DNA Mini Kit (Invitrogen), conforme protocolo utilizado no Laboratório de Genética Animal da Embrapa Suínos e Aves, Concórdia, SC. A quantificação, qualidade e integridade do DNA foram avaliadas por meio do espectrofotômetro NanoDrop 1000 (Thermo Scientific) e as amostras foram diluídas para uma concentração final de 500 ng/ $\mu$ L.

#### 3.2 Genotipagem

A genotipagem foi realizada em laboratório comercial (Deoxi Biotecnologia Ltda, Araçatuba, São Paulo, Brasil) com o painel Illumina PorcineSNP60 v2 BeadChip, de acordo com o protocolo do fabricante. Esse painel compreende 61.565 SNPs distribuídos pelo genoma suíno. O controle de qualidade foi realizado utilizando o programa PLINK v.1.9 (PURCELL et al., 2007) para remoção de marcadores que falharam em mais de 10% das amostras e amostras que falharam em mais de 10% dos marcadores. Nenhum filtro adicional foi considerado no estudo, uma vez que a remoção de alelos raros em análises de estrutura de população, pode levar a obtenção de resultados superestimados ou subestimados quando se trabalha com amostras das populações (BOSSE et al., 2012).

---

\*RAMACCIOTTE, R. (Choice Genetics). Comunicação pessoal, 2015.

### 3.3 Desequilíbrio de ligação (LD)

A extensão do LD existente na população foi avaliada utilizando-se a medida  $r^2$  (HILL; ROBERTSON, 1968), a qual é padronizada para a frequência alélica. Para todos os pares de SNPs autossômicos, as medidas de  $r^2$  foram calculadas utilizando o comando `-ld-window 60.000 -ld-window-r2 0` no programa computacional PLINK v.1.9 (PURCELL et al., 2007). Tais comandos garantem que serão geradas medidas de  $r^2$  para todos os SNPs em janelas de 60.000 Kb. O gráfico da extensão do desequilíbrio de ligação, mensurado por  $r^2$  médio em função da distância em quilo bases (Kb) entre os 18 cromossomos autossômicos, foi gerado utilizando o programa estatístico R versão 3.1.1 (R Core Team, 2014). Para calcular a média do desequilíbrio de ligação ( $r^2$ ) foram considerados apenas os SNPs adjacentes que estão a uma distância menor que 100 Kb, utilizando o programa PLINK.

### 3.4 Coeficiente de endogamia

O coeficiente de endogamia utilizando dados de pedigree ( $F_{ped}$ ) foi estimado de acordo com a metodologia proposta por Wright (1931) com o pacote pedigree do programa R versão 3.1.1 (COSTER, 2012). Na análise, três gerações foram consideradas e o gráfico do histograma do coeficiente de endogamia foi obtido utilizando-se o mesmo programa computacional.

### 3.5 Segmentos de Homozigose

Os segmentos de homozigose foram detectados utilizando o programa computacional PLINK v.1.9 (PURCELL et al., 2007), considerando o controle de qualidade anteriormente descrito adicionado de uma nova filtragem para remoção dos SNP sem posição definida e de SNP posicionados em cromossomos não-autossômicos. Os registros dos SNPs dos cromossomos sexuais foram excluídos porque a recombinação nestes cromossomos tende a ocorrer de forma diferente e também porque o cromossomo X em suínos apresenta o mapa de resolução genética menor em relação aos cromossomos autossômicos (BOSSE et al., 2012).

Os segmentos que apresentaram pelo menos 50 SNPs homozigotos em distâncias mínimas de 1.000 Kb por animal foram considerados como ROH, permitindo observar um SNP heterozigoto e um SNP faltante/perdido dentro de uma janela de 50 SNPs (SCRAGGS et al., 2013).

A partir da identificação dos segmentos de homozigose, o coeficiente de endogamia genômico para cada indivíduo ( $F_{ROH}$ ) foi estimado. Este coeficiente foi calculado usando a seguinte equação:

$$F_{ROH} = \frac{\sum_k \text{Tamanho} (ROH_k)}{L}$$

em que, “k” é número de ROH identificados para cada indivíduo multiplicado pelo tamanho médio de seus ROHs e “L” é o tamanho total do genoma suíno (2.808.525 Kb, Scrofa10.2, 2014). Logo,  $F_{ROH}$  é definido como o tamanho do genoma oriundo da análise de ROH dividido pelo tamanho total do genoma. O histograma do coeficiente de endogamia e os gráficos do número de segmentos de homozigose em função do tamanho dos segmentos, número de segmentos compartilhados entre os indivíduos e cariograma para três classes de porcentagem dos segmentos de homozigose compartilhados entre os indivíduos foram plotados utilizando o programa R versão 3.1.1 (R Core Team, 2014).

Os segmentos de homozigose compartilhados por mais que 50% das amostras foram analisados para identificação de regiões conservadas, buscando-se estudar genes que possuíssem possível associação com características adaptativas tais como sobrevivência e resistência a doenças que estão presentes nas mesmas. Os genes encontrados nessas regiões foram investigados por meio de análises nas plataformas de dados Ensembl Genome Browser (<http://www.ensembl.org/index.html>), National Center for Biotechnology Information Gene (<http://www.ncbi.nlm.nih.gov/gene>) e PANTHER (<http://www.pantherdb.org/>) para estudos das suas funções biológicas e vias metabólicas.

### 3.6 Estratificação da População

Para avaliar a estratificação da população foi aplicado o modelo de ancestralidade usando o programa STRUCTURE v.2.3.4 (HUBISZ et al., 2009). Os valores de K (número de populações) de um a oito Foram testados, usando período

de *burn-in* de 1.000 iterações e Cadeia de Markov e Monte Carlo (MCMC, “Markov Chain Monte Carlo”) de 10.000 iterações. As análises foram repetidas dez vezes para cada K e a determinação do número correto de K foi estimada utilizando a estatística Delta K, a qual se baseia na taxa de alteração do logaritmo da probabilidade dos dados entre sucessivos valores de K (EVANNO et al., 2005). O Delta K foi calculado utilizando STRUCTURE HARVESTER v.0.6.94 (EARL, VONHOLD, 2011), sendo que o maior valor de Delta K determinou o número correto de K.

A detecção das proporções de genes da população estudada também foi realizada utilizando-se o procedimento de verossimilhança, em que foram estimados os parâmetros da matriz de coeficientes de ancestralidade (Q) e da matriz de frequências alélicas populacionais (F). As análises foram realizadas utilizando-se o programa computacional ADMIXTURE (ALEXANDRE et al., 2009). A medida de  $F_{st}$ , descrita por Holsinger e Weir (2009) como a correlação entre os gametas escolhidos aleatoriamente a partir da mesma subpopulação em relação a toda população, foi obtida a partir das análises realizadas no programa ADMIXTURE.

Para descrever a estrutura da população foram realizadas análises de componentes principais utilizando o programa PLINK v.1.9 (PURCELL et al., 2007). As análises de Componentes Principais foram realizadas com todos os SNPs, e também excluindo os SNPs que apresentaram  $r^2$  maior que 0,20 para eliminar redundância entre marcadores em alto LD. A partir da análise de Componentes Principais, o gráfico bidimensional foi construído usando o programa R versão 3.1.1 (R Core Team, 2014).

#### **4 RESULTADOS E DISCUSSÃO**

Após controle de qualidade, um animal foi excluído do estudo, uma vez que mais de 10% dos seus SNPs falharam. Assim foram obtidos 57.770 SNPs informativos para a realização das análises subsequentes, incluindo 1.498 SNPs no cromossomo X, 15 SNPs no cromossomo Y e 4.076 SNPs sem posição

cromossômica definida. O número total de SNPs monomórficos foi de 5.497, representando 9,51% do total de SNPs utilizados nas análises (Tabela 1).

**Tabela 1.** Análise descritiva dos marcadores ao longo do genoma, por cromossomo.

| CHR            | Tamanho<br>CHR <sup>2</sup> | nº<br>SNP/CHR <sup>3</sup> | %<br>SNP/CHR <sup>4</sup> | nº SNP<br>mono. <sup>5</sup> | % SNP<br>mono. <sup>6</sup> | MAF<br>(média) <sup>7</sup> |
|----------------|-----------------------------|----------------------------|---------------------------|------------------------------|-----------------------------|-----------------------------|
| 0 <sup>1</sup> | -                           | 4076                       | 7,06                      | 365                          | 8,95                        | 0,2527                      |
| 1              | 315,32                      | 6624                       | 11,47                     | 748                          | 11,29                       | 0,2313                      |
| 2              | 162,57                      | 3321                       | 5,75                      | 259                          | 7,80                        | 0,251                       |
| 3              | 144,79                      | 2841                       | 4,92                      | 243                          | 8,55                        | 0,2454                      |
| 4              | 143,47                      | 3541                       | 6,13                      | 456                          | 12,88                       | 0,2385                      |
| 5              | 111,51                      | 2360                       | 4,09                      | 266                          | 11,27                       | 0,2386                      |
| 6              | 157,77                      | 3215                       | 5,57                      | 256                          | 7,96                        | 0,2594                      |
| 7              | 134,76                      | 3297                       | 5,71                      | 314                          | 9,52                        | 0,2366                      |
| 8              | 148,49                      | 2795                       | 4,84                      | 253                          | 9,05                        | 0,2513                      |
| 9              | 153,67                      | 3228                       | 5,59                      | 227                          | 7,03                        | 0,2529                      |
| 10             | 79,1                        | 1752                       | 3,03                      | 128                          | 7,31                        | 0,2661                      |
| 11             | 87,69                       | 1862                       | 3,22                      | 151                          | 8,11                        | 0,2476                      |
| 12             | 63,59                       | 1560                       | 2,70                      | 104                          | 6,67                        | 0,2545                      |
| 13             | 218,64                      | 4072                       | 7,05                      | 367                          | 9,01                        | 0,2604                      |
| 14             | 153,85                      | 3860                       | 6,68                      | 329                          | 8,52                        | 0,2701                      |
| 15             | 157,68                      | 2950                       | 5,11                      | 334                          | 11,32                       | 0,2328                      |
| 16             | 86,9                        | 1869                       | 3,24                      | 159                          | 8,51                        | 0,2566                      |
| 17             | 69,7                        | 1693                       | 2,93                      | 138                          | 8,15                        | 0,2242                      |
| 18             | 61,22                       | 1341                       | 2,32                      | 103                          | 7,68                        | 0,2455                      |
| X              | 144,29                      | 1498                       | 2,59                      | 297                          | 19,83                       | 0,1898                      |
| Y              | 1,64                        | 15                         | 0,03                      | 0                            | 0                           | 0                           |

<sup>1</sup>Cromossomo 0 representa indefinição da posição dos marcadores; <sup>2</sup>Tamanho dos cromossomos em Megabase; <sup>3</sup>Número de SNPs por cromossomo; <sup>4</sup>Porcentagem de SNPs por cromossomo; <sup>5</sup>Número de SNPs monomórficos; <sup>6</sup>Porcentagem de SNPs monomórficos; <sup>7</sup>Frequência alélica mínima por cromossomo.

Ai, Huang e Ren (2013), utilizando Porcine60K Beadchip (Illumina) para genotipar 304 suínos de origem ocidental e oriental, encontraram 5.859 SNPs sem posição definida. Uma possível explicação para ausência de posição física de SNPs pode estar relacionada com a construção do chip, uma vez que 30% dos polimorfismos utilizados foram selecionados de uma lista de SNPs que não puderam ser mapeados na construção do genoma suíno versão 7.0, utilizado em sua montagem (RAMOS et al., 2009).

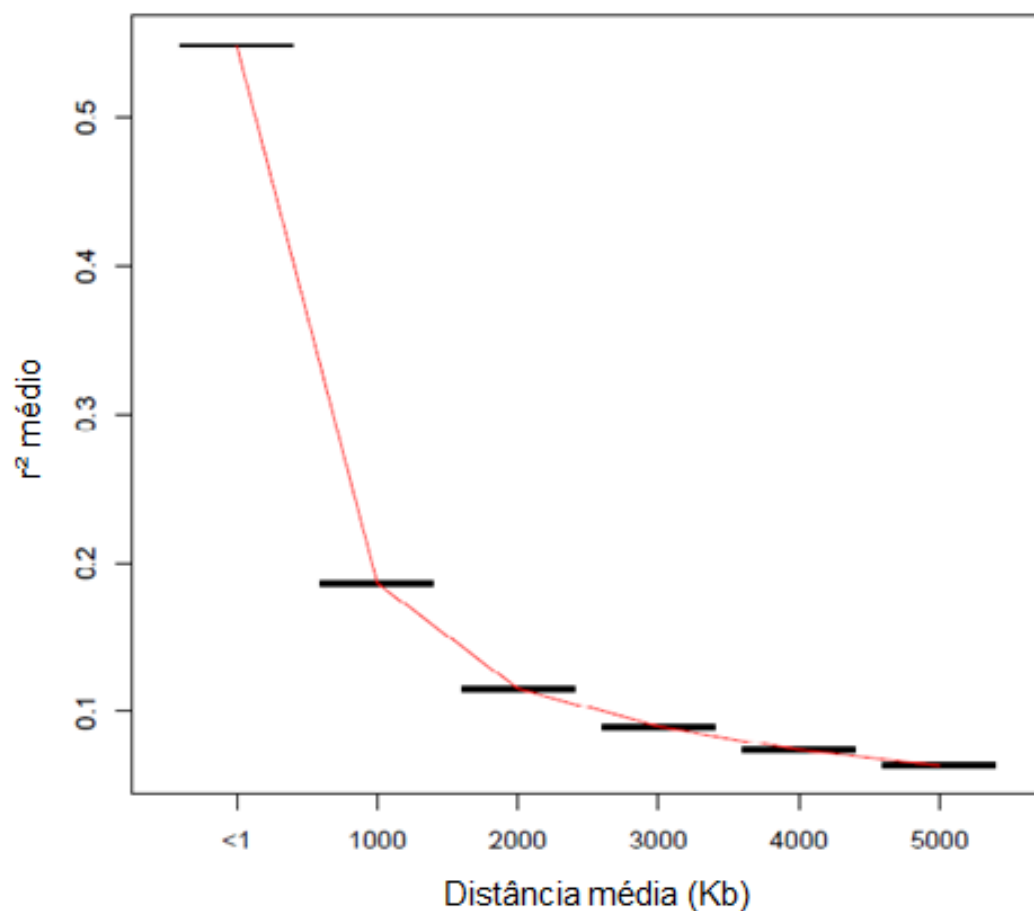
Marcadores monomórficos não foram excluídos das análises uma vez que poderiam ser informativos para estudos da estrutura da população. Em trabalho

realizado com diversas raças suínas foram observadas diferenças no número total de marcadores monomórficos entre os animais originários de diferentes continentes, relatando em média 15.101 e 5.925 SNPs monomórficos para raças de origem oriental e ocidental, respectivamente (AI, HUANG, REN, 2013). Visto que o número de SNPs monomórficos encontrados na linhagem estudada foi de 5.497, pode-se observar que esse resultado encontrado corrobora com o valor médio estimado de SNPs monomórficos para as raças ocidentais.

#### **4.1 Desequilíbrio de Ligação**

Um total de 52.196 SNPs foi utilizado para calcular a média do desequilíbrio de ligação ( $r^2$ ) entre todos os SNPs adjacentes, que estão a uma distância menor que 100 Kb. Para a população estudada, a média e respectivo desvio padrão de  $r^2$  estimado foi de  $0,29 \pm 0,312$ , indicando alta variabilidade em torno do valor médio estimado para o desequilíbrio. Estudos anteriores utilizando o painel Porcine60K Beadchip da Illumina, para descrever o padrão do desequilíbrio de ligação, encontraram  $r^2$  médio de 0,32 entre marcadores que estavam a uma distância de 0,5 Mb e excluindo SNPs para critérios de MAF e equilíbrio de Hardy-Weinberg (WANG et al., 2013). Badke et al. (2012), utilizando o programa BEAGLE e excluindo SNPs para os mesmos critérios citados, descreveram  $r^2$  médio de 0,19 para marcadores que estavam espaçados a 0,5 Mb em populações de suínos Landrace.

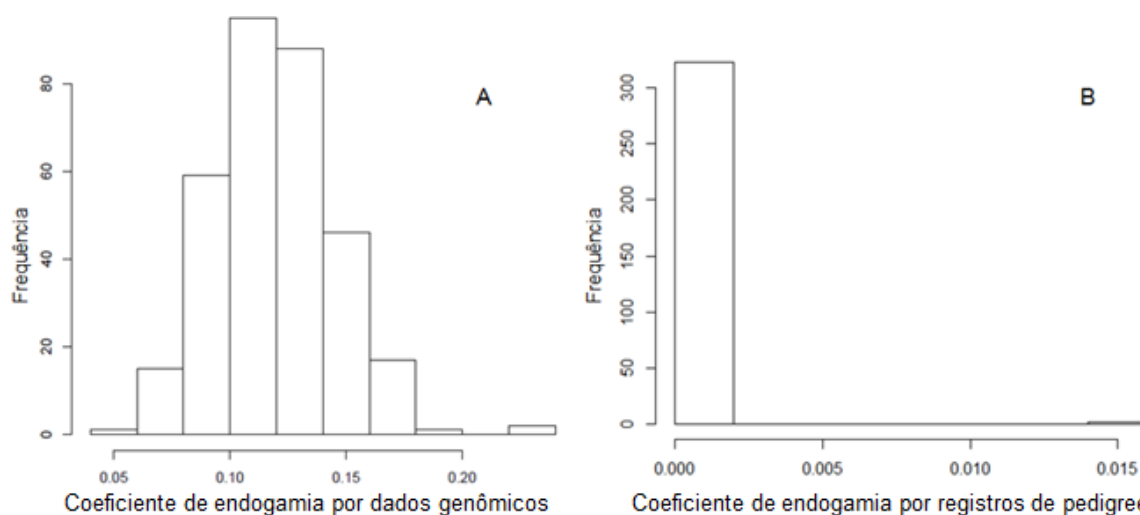
Observa-se que, à medida que a distância entre os marcadores aumenta os valores de desequilíbrio de ligação diminuem, sendo que a partir de distâncias superiores a 1 Mb o  $r^2$  médio foi inferior a 0,20 (Figura 1; Apêndice 1). O resultado obtido neste trabalho foi similar em comparação aos encontrados por Ai, Huang e Ren (2013) e Badke et al. (2012), que também observaram médias de  $r^2$  maiores para distâncias menores que 1Mb. Esse padrão de queda no desequilíbrio de ligação ao longo de maiores distâncias segue o resultado esperado, uma vez que com o aumento na distância entre marcadores no mesmo cromossomo, maior será possibilidade de ocorrerem eventos de recombinação e, portanto, a quebra da ligação entre os marcadores (DU, CLUTTER, LOHUIS, 2007).



**Figura 1.** Média de desequilíbrio de ligação ( $r^2$ ) ao longo de diferentes distâncias (Kb) entre marcadores em uma linhagem suína de base Landrace.

#### 4.2 Coeficiente de endogamia

Em média, os segmentos de homozigose cobriram 11,89% do genoma da população estudada, sendo encontrada média de 54 ROHs/genoma, com valores mínimo e máximo de 28 ROHs/genoma e 74 ROHs/genoma, respectivamente e tamanho variando entre 152.110 Kb a 652.500 Kb, com média de 333.900 Kb. O coeficiente de endogamia ( $F_{ROH}$ ) médio obtido a partir dos segmentos de homozigose foi de 0,119, apresentando valor mínimo de 0,054 e máximo de 0,232, enquanto que para análise com base no pedigree foi de 0,00011, variando entre 0 e 0,015 (Figura 2), o que indica baixa correlação entre o  $F_{ROH}$  e  $F_{PED}$  ( $r < 0,04$ ).



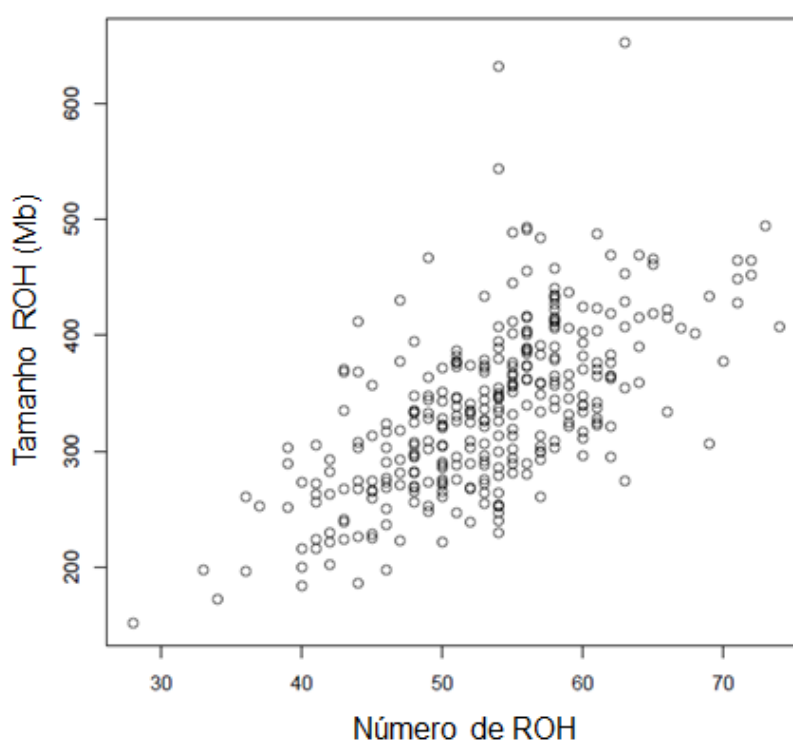
**Figura 2.** Histograma do coeficiente de endogamia de uma linhagem suína de base Landrace utilizando dados genômicos (A) e registros de pedigree (B).

Alguns estudos têm relatado correlações de baixa magnitude entre as estimativas do coeficiente de endogamia obtidas a partir do pedigree e com base em dados de genotipagem de diferentes espécies (LOPES et al., 2013; SCRAGGS et al., 2013). Isto pode ser explicado pelo efeito da variação da recombinação gênica que não é considerada quando a endogamia é calculada a partir de dados de pedigree. Além disso, estimativas com base nos registros do pedigree negligenciam os efeitos da endogamia acumulada ao longo das gerações e não consideram os erros de pedigree acumulados entre as gerações passadas.

A correlação entre as estimativas do coeficiente de endogamia com base nos segmentos de homozigose e nos registros de pedigree encontrada no trabalho foi relativamente menor em comparação a encontrada por Lopes et al. (2013), a qual foi de 0,27 entre estimativas de endogamia calculadas a partir de dados de genotipagem (utilizando frequências alélicas) e de pedigree para suínos da raça Large White. A baixa magnitude encontrada para a correlação entre as estimativas do coeficiente de endogamia no presente trabalho pode ser explicada pelo uso do pedigree em que somente foram consideradas três gerações, e também pelos possíveis erros de identificação dos animais no pedigree da população estudada, sugerindo que houve subestimação do  $F_{PED}$ .

O alto número de ROH longos (>5 Mb) (Figura 3) pode ser indicativo de que a população não sofreu cruzamento recente, caso contrário o cruzamento entre raças

distintas teria provocado a quebra desses segmentos. Herrero-Medrano et al. (2013), estudando raças Ibéricas, encontraram ROH com tamanho médio de 50 a 100 Mb, considerando longos segmentos de homozigose aqueles que tivessem tamanho superior a 100 Mb, enquanto que, para Bosse et al. (2012) longos segmentos de homozigose apresentam tamanho superior a 5 Mb. Esses autores, trabalhando com suínos selvagens e comerciais dos continentes asiático e europeu, verificaram maior proporção de ROH médios (0,1 a 5 Mb) para raças comerciais europeias.

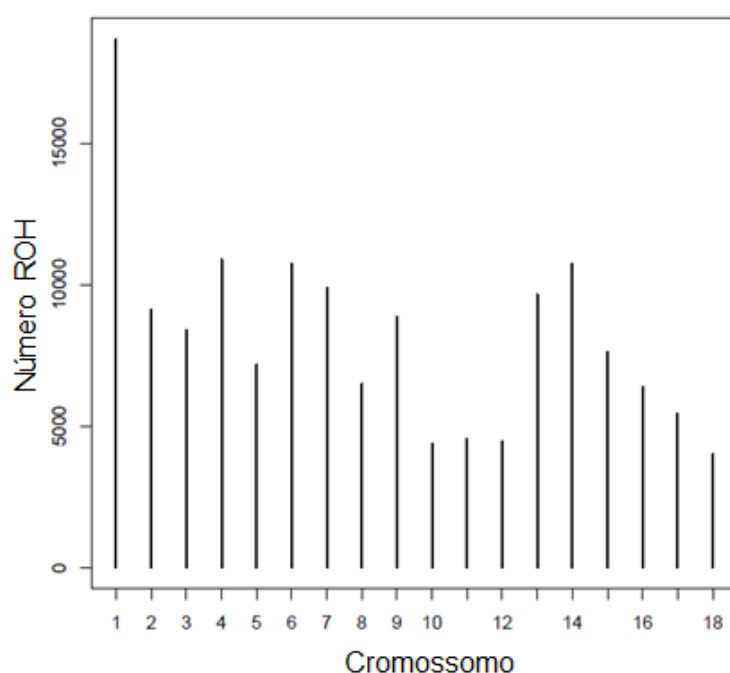


**Figura 3.** Relação entre o número e tamanho médio das segmentos de homozigose em uma linhagem suína base Landrace.

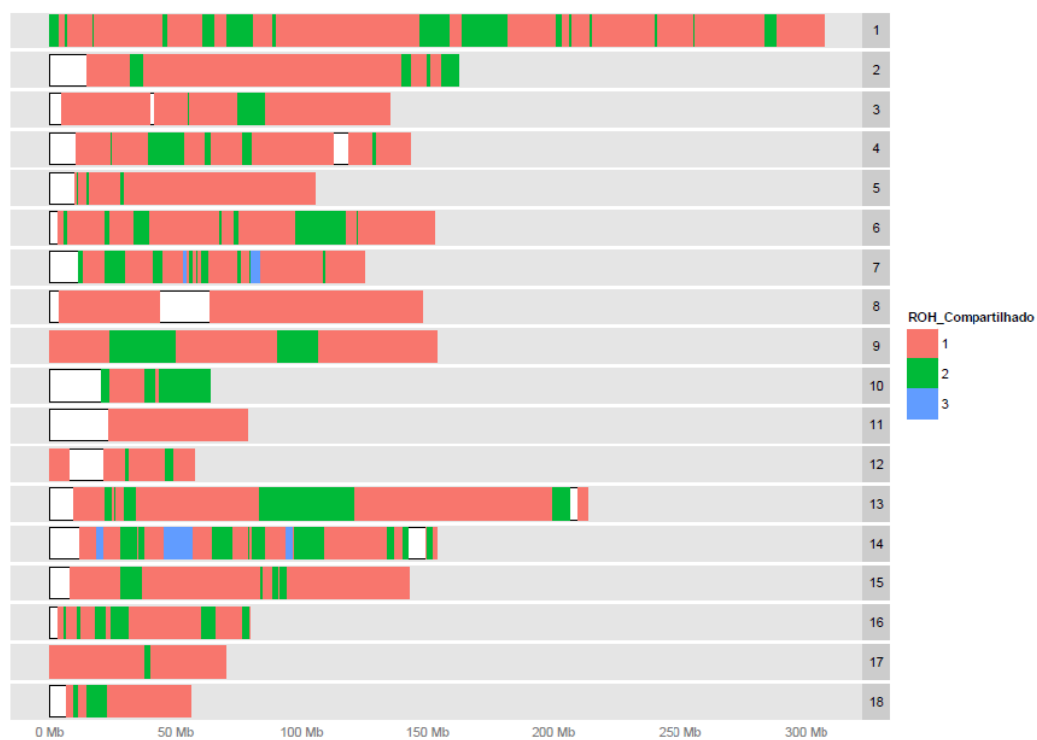
A alta prevalência de longos segmentos de homozigose pode indicar também endogamia recente na população estudada, corroborando com o resultado encontrado para  $F_{ROH}$ . Fisher (1954) observou que o tamanho esperado de um segmento de DNA, que é idêntico por descendência, segue distribuição exponencial com média igual a  $\frac{1}{2}g$  Morgans, em que “g” é o número de gerações passadas desde o ancestral comum. Logo, longos ROH (>5 Mb) refletem endogamia a partir de um ancestral comum que está a menos que 10 gerações passadas.

A endogamia é um importante parâmetro em genética de populações, uma vez que leva ao aumento da frequência de homozigose, acarreta o aumento do risco de genes deletérios recessivos serem co-expressos e ocasiona depressão endogâmica (ZHANG et al., 2014). A depressão endogâmica pode ser definida como o aumento da homozigose e da frequência de alelos prejudiciais da progênie de pais aparentados, que normalmente leva a redução do desempenho fenotípico do indivíduo e da variabilidade populacional (CHARLESWORTH, WILLIS, 2009; SILIÓ et al., 2013). Os efeitos deletérios causados pela depressão endogâmica tem sido avaliados em várias espécies animais. Em suínos, Saura et al. (2015) avaliaram a depressão endogâmica em características reprodutivas com uso da análise de regressão entre o fenótipo da característica e o coeficiente de endogamia, observando depressão endogâmica em regiões específicas que contém genes associados com tamanho de leitegada em suínos Ibéricos.

Entre todos os segmentos de homozigose observados, não foram encontrados segmentos idênticos que fossem compartilhados por todos os animais da amostra estudada. Os cromossomos 1, 4, 6 e 14 foram os que apresentaram o maior número de segmentos de homozigose compartilhados na população (Figura 4), destacando-se o cromossomo 1, que apresentou mais de 15.000 ROH compartilhados. O alto número de segmentos de homozigose compartilhados no cromossomo 1 pode ser justificado pelo fato de que este é o maior cromossomo do genoma suíno, apresentando maior número de marcadores em relação aos cromossomos mais curtos (Tabela 1). Além disso, toda a extensão do cromossomo 1 foi coberta por ROH que foram compartilhados por mais de 10% da população, enquanto em que na maioria dos cromossomos houveram trechos que não apresentaram segmentos compartilhados em pelo menos 10% da amostra (Figura 5). A falta de segmentos de homozigose nesses trechos pode estar relacionada com eventos de recombinação, uma vez que, segundo Tortereau et al. (2012), as taxas de recombinação variam entre cromossomos e ao longo dos cromossomos, em que as regiões com altas taxas de recombinação tendem a se agrupar próximo das extremidades cromossômicas, independente da posição do centrômero, como foi observado no presente trabalho (Figura 5).



**Figura 4.** Número de regiões de homozigose compartilhadas entre indivíduos por cromossomos em uma linhagem suína base Landrace.



**Figura 5.** Cariograma dos segmentos de homozigose (ROH) compartilhados entre os indivíduos considerando três classes: **1:** ROH compartilhados por 10 a 20% dos indivíduos; **2:** ROH compartilhados por 21 a 50% indivíduos; e **3:** ROH compartilhados por mais de 50% dos indivíduos.

Nos cromossomos 7 e 14 estão localizadas as regiões que apresentaram ROH compartilhados por mais de 50% da amostra da população (Figura 5, Tabela 2). O segmento de homozigose compartilhado de maior tamanho e a região presente em maior porcentagem de animais estudados estão presentes no cromossomo 14. A análise dessas regiões indicou que as mesmas contém um total de 1.904 genes, dos quais aproximadamente 76% são genes descritos como codificadores de proteínas (Tabela 3).

**Tabela 2.** Número de animais e tamanho dos segmentos de homozigose compartilhados por mais de 50% da população por cromossomo.

| Segmento | CHR <sup>1</sup> | Nº animais <sup>2</sup> | % <sup>3</sup> | Tamanho <sup>4</sup> | Início <sup>5</sup> | Fim <sup>6</sup> |
|----------|------------------|-------------------------|----------------|----------------------|---------------------|------------------|
| 1        | SSC14            | 216                     | 66,67          | 10.584.744           | 45.682.666          | 56.267.410       |
| 2        | SSC14            | 192                     | 59,26          | 20.107.004           | 94.087.498          | 114.194.502      |
| 3        | SSC14            | 172                     | 53,08          | 77.679.150           | 19.064.824          | 96.743.979       |
| 4        | SSC7             | 170                     | 52,47          | 3.062.532            | 80.285.802          | 83.348.334       |
| 5        | SSC7             | 168                     | 51,85          | 18.515.307           | 53.190.798          | 71.706.105       |

<sup>1</sup>Cromossomo; <sup>2</sup>Número de animais com o mesmo segmento; <sup>3</sup>Porcentagem de animais com o mesmo segmento; <sup>4</sup>Tamanho do segmento compartilhado (bp); <sup>5</sup>Início do segmento compartilhado (bp); <sup>6</sup>Final do segmento compartilhado.

**Tabela 3.** Número total de genes por segmentos de homozigose e suas classificações.

| Segmento | Codificadores de proteínas | Proteínas não caracterizadas | miRNA <sup>1</sup> | Total genes |
|----------|----------------------------|------------------------------|--------------------|-------------|
| 1        | 192                        | 53                           | 3                  | 248         |
| 2        | 167                        | 51                           | 1                  | 219         |
| 3        | 752                        | 244                          | 4                  | 1000        |
| 4        | 137                        | 15                           | 0                  | 152         |
| 5        | 216                        | 68                           | 1                  | 285         |

<sup>1</sup>MicroRNA

O estudo dos genes presentes nesses segmentos de homozigose possibilitou identificar aqueles genes que possam estar relacionados com características de sobrevivência. O gene MAPK1 (“Mitogen-activated Protein kinase 1”), também conhecido por ERK2 (“Extracellular Signal-regulated Kinase”), presente no segmento 1, é componente crítico nas vias metabólicas de transdução de sinal celular modulando uma variedade de junções celulares (WEI, LIU, 2009). O mesmo foi descrito como importante regulador de infecções, como casos de infecção pelo Circovírus Suíno do tipo 2 (PVC2 – “Porcine circovirus type 2”), em que sua ativação

é necessária para replicação viral em cultura (WEI, LIU, 2009).

Os genes CXCL12 (“Chemokine (C-X-C motif) ligand 12”) e TFAM (Transcription factor A mitochondrial”) foram localizados no segmento 2. O gene CXCL12 atua em conjunto com genes do sistema imune em suínos, o qual tem sido associado com resistência à doenças em Landrace (WANG et al., 2012). O gene TFAM tem papel importante na gametogênese e na pré-implantação e desenvolvimento embrionário (ANTELMAN et al., 2008).

O gene SFTPD (“Surfactant Protein D”), localizado na região do segmento 3, atua na defesa do organismo contra microrganismos e toxinas. A molécula SP-D (pertence ao grupo de lectina colagenosas de mamíferos) apresenta forte atividade neutralizante para o vírus da influenza em suínos (HILLAIRE et al., 2012; HERÍAS et al., 2007).

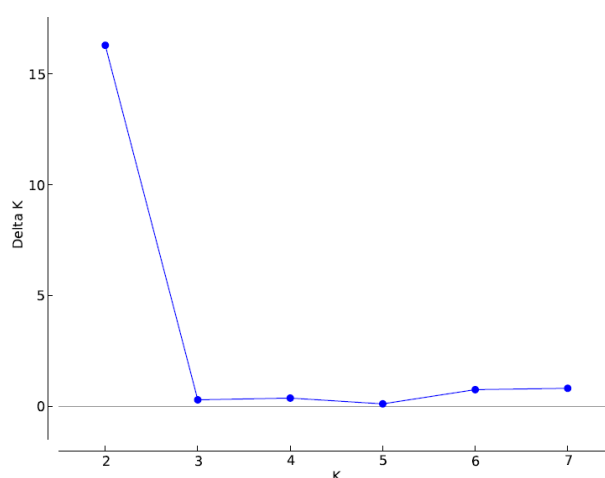
A região do segmento 5 abriga o gene NFKBIA (“Nuclear factor of Kappa light polypeptide gene enhancer in B-cells inhibitor, Alpha”) que apresenta relação com a susceptibilidade ao vírus da Peste Suína Africana (ASFV – “African swine fever virus”). A variação no seu componente RELA (“V-rel reticuloendotheliosis viral oncogene homolog A”) tem o potencial para ser a base da diferença entre tolerância e morte à infecção por ASFV (PALGRAVE et al., 2011).

De acordo com Nothnagel et al. (2010), regiões de homozigose encontradas em mais de 50% dos indivíduos de uma população podem indicar a ocorrência de forte seleção, a qual pode ter favorecido a conservação da espécie. A presença de genes ligados às funções reprodutivas, a exemplo do gene TFAM, em uma região de homozigose compartilhada por mais de 50% da população amostrada pode indicar que esta seja uma região conservada do genoma suíno, também favorecida pela seleção.

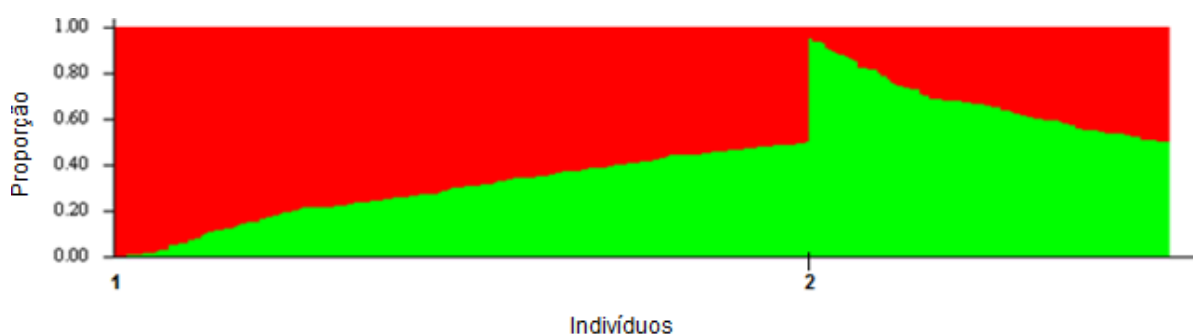
#### **4.3 Estratificação da População**

A partir da análise Bayesiana, implementada no software STRUCTURE, se demonstrou que o número provável de populações, representado pelo valor do Delta K (EVANNO, REGNAUT, GOUDET, 2005), foi igual a dois (Figura 6). A análise de estratificação da população diferenciou claramente a amostra em duas populações,

sendo os alelos da população agrupados, em maior parte (59%), no “cluster” 1 (Figura 6). Os valores máximo e mínimo e a média da proporção alélica para os animais amostrados para o “cluster” 1 foram 0,996 e 0,05 e  $0,59 \pm 39,302$  respectivamente, enquanto que para o “cluster” 2 esses valores foram estimados em 0,95 e 0,004 e  $0,41 \pm 54,296$ . Os resultados obtidos a partir do procedimento de verossimilhança, no programa ADMIXTURE, foram muito próximos aos obtidos na bayesiana com STRUCTURE. Por isso, foram apresentados somente os resultados do programa STRUCTURE (Apêndices 3 a 5).



**Figura 6.** Estatística *Delta K* baseada no método de clusterização de Evanno et al. (2005).



**Figura 7.** Constituição genética de animais de uma linha fêmea sintética base Landrace, em análise para  $K = 2$ . As duas cores são referentes aos grupos definidos (“clusters”) com os resultados do programa STRUCTURE (Pritchard et al., 2000): vermelho = “Cluster” 1; verde = “Cluster” 2.

O valor médio do  $F_{st}$  genômico indicou que o grau de divergência genética entre a população estudada foi de 0,10. Esse valor foi menor do que 0,22

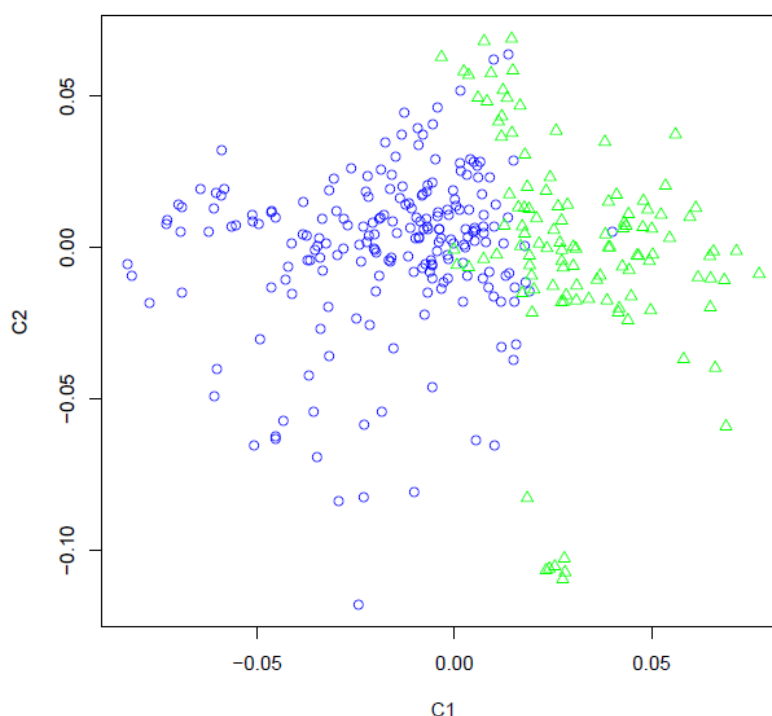
encontrado por HERERRO-MEDRANO et al. (2013) para porcos domésticos e selvagens da Península Ibérica, e inferior ao valor de  $F_{st}$  estimado em 0,32 para suínos com origem ocidental e chinesa (AI, HUANG, REN, 2013). O baixo grau de divergência genética pode ser explicado pela alta prevalência de longos ROH que resulta em maior porcentagem de endogamia e, portanto, menor diversidade genética. O resultado encontrado indica que aproximadamente 10.000 SNPs seriam suficientes para corrigir o efeito de estratificação populacional em estudos de GWAS aplicados na população estudada.

As análises de componentes principais com todos os SNPs e apenas com os SNPs que apresentaram desequilíbrio de ligação menor que 0,20 não apresentaram resultado com diferença expressiva. A partir da análise do gráfico bidimensional (Figura 8) observou-se que a população foi dividida em dois grupos, com base no primeiro componente principal, em que os animais (representados pelos pontos) localizados nas extremidades são aqueles que apresentam maior proporção de alelos do “cluster” 1 ou do “cluster” 2 (Figura 7). Dado que o componente principal pode ser considerado um índice, cada indivíduo possui um valor do índice dentro de cada componente principal, denominado escore individual do componente principal, representado em um gráfico bidimensional (Figura 8). Assim, escores individuais no primeiro componente principal indicam que os animais com escores individuais maiores ou iguais a 0,02 são classificados no grupo verde e animais com escores individuais menores que 0,02 são classificados no grupo azul, sendo 0,02 um valor arbitrário obtido a partir da análise visual do gráfico. Animais do grupo azul e verde possuem, respectivamente, maior proporção de raças orientais e ocidentais.

Quando estudadas as causas para o resultado de K, foi possível inferir que a separação entre duas populações pode ser explicada pelo cruzamento entre as raças ocidentais e orientais utilizadas na formação da linhagem estudada. Na primeira população (População 1; Figura 6) foram agrupados os animais que apresentam maior proporção de alelos das linhagens ocidentais, enquanto que na segunda população (População 2; Figura 6) foram representados os animais que apresentam maior proporção de alelos de linhagens orientais.

O cruzamento entre raças é uma estratégia amplamente utilizada no melhoramento de animais de interesse zootécnico. Em suínos, na criação de linhas

fêmea, é comum a realização de cruzamentos entre raças ocidentais e orientais, sendo utilizados em maior frequência os genótipos derivados das raças brancas Landrace e Large White e em menor escala os genótipos de raças de origem Chinesa, como a Meishan. As linhas fêmeas são desenvolvidas para produzir porcas com capacidade de desmamar um grande número de leitões por parto e por ano (SILVA et al., 2003; FÁVERO, FIGUEIREIDO, 2009), justificando a utilização das raças ocidentais Landrace, Large White, Hampshire e Pietran e as raças orientais Meishan e Xia Jing na formação da linhagem Redone, uma vez que as mesmas se destacam por apresentar excelente prolificidade e boa habilidade materna (IRGANG, 2014). A partir deste cruzamento, espera-se menor proporção de alelos de linhagens orientais, contudo foi observada variação na proporção esperada de alelos, sendo que este desvio pode ser atribuído a eventos de recombinação gênica nos cromossomos parentais que podem ocorrer durante a meiose (SÖLKNER et al., 2010). Além disso, a variação na constituição genética dos animais quanto a distribuição alélica das raças orientais e ocidentais se deve a constituição da amostra, em que se procurou representar toda a variabilidade genética da linhagem.



**Figura 8.** Estrutura da população de uma linhagem suína base Landrace revelada pela Análise de Componentes Principais com os SNP com  $r^2 \leq 0,2$ .

## 5 CONCLUSÃO

Os resultados evidenciaram a importância da genotipagem de SNPs em alta densidade para fornecer conhecimentos sobre a estrutura da população e revelaram que a amostra estudada está dividida em duas populações. A estimativa do coeficiente de endogamia genômico indicou que a diversidade genética da população estudada deve ser constantemente investigada. Os acasalamentos entre indivíduos menos aparentados devem ser priorizados para garantir a eficiência do programa de melhoramento genético aplicado na população e evitar a depressão endogâmica sobre características de importância econômica.

## REFERÊNCIAS

AI, H.; HUANG, L.; REN, J. Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. **Plos One**, v. 8, e56001, 2013.

ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Research**, v. 19, n. 9, p. 1655-1664, 2009.

ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. **Admixture 1.23 Software Manual**. 11p., 2013. Disponível em: <https://www.genetics.ucla.edu/software/admixture/admixture-manual.pdf>. Acesso em: 08 de janeiro de 2016.

AMARAL, A. J.; MEGENS, H-J.; CROOIJMANS, R. P. M. A.; HEUVEN, H. C. M.; GROENEN, M. A. M. Linkage Disequilibrium Decay and Haplotype Block Structure in the Pig. **Genetics**, v. 179, p. 569-579, 2008.

ANTELMAN, J.; MANANDHAR, G.; YI, Y. J.; LI, R.; WHITWORTH, K. M.; SUTOVSKY, M.; AGCA, C.; PRATHER, R. S.; SUTOVSKY, P. Expression of mitochondrial transcription factor A (TFAM) during porcine gametogenesis and preimplantation embryo development. **Journal of Cellular Physiology**, v. 217, p. 529-543, 2008.

BADKE, Y. M.; BATES, R. O.; ERNST, C. W.; SCHWAB, C.; STEIBEL, J. P. Estimation of linkage disequilibrium in four US pig breeds. **BMC Genomics**, v. 13, 2012.

BOSSE, M.; MEGENS, H-J.; MADSEN, O.; PAUDEL, Y.; FRANTZ, L.A.F.; SCHOOK, L. B.; CROOIJMANS, R. P. M. A.; GROENEN, M. A. M.; Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. **Plos Genetics**, v. 8, e1003100, 2012.

CAVALLI-SFORZA, L.L.; MENOZZI, P.; PIAZZA, A. Demic expansions and human evolution. **Science**, v. 259, p. 639–646, 1993.

CHARLESWORTH, D.; WILLIS, J. H. The genetics of inbreeding depression. **Nature Reviews Genetics**, v. 10, n. 11, p.783-796, 2009.

COSTER, A. 2012. pedigree: Pedigree functions. R package version 1.4.

DU, F-X.; CLUTTER, A. C.; LOHUIS, M. M. Characterizing Linkage Disequilibrium in Pig Populations. **International Journal of Biological Sciences**, v. 7, p. 166-178, 2007.

EARL, D. A.; VONHOLDT, B. M. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. **Conservation Genetics Resources**, v. 4, p. 359–361, 2011.

EENENNAM, A. L. V.; WEIGEL, K. A.; YOUNG, A. E; CLEVELAND, M. A.; DEKKERS, J. C.M. Applied Animal Genomics: Results from the Field. **The Annual Review of Animal Biosciences**, v. 2, p. 105-139, 2013.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology Resources**, v. 14, p. 2611–2620, 2005.

FÁVERO, J. A.; FIGUEIREDO, E. A. P. de. Evolução do melhoramento genético de suínos no Brasil. **Ceres**, v. 56, p. 420-427, 2009.

FERENČAKOVIĆ, M.; HAMZIC', E.; GREDLER, B.; SOLBERG, T.R.; KLEMETSDAL, G.; CURIK, I.; SÖLKNER, J. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. **Journal of Animal Breeding and Genetics**, v. 130, p. 286-293, 2012.

FISHER, R. A. A fuller theory of “junctions” in inbreeding. **Heredity**, v. 8, p. 187–197, 1954.

GALETTI JR, P. M.; RODRIGUES, F. P.; SOLÉ-CAVA, A.; MIYAKI, C. Y.; CARVALHO, D.; EIZIRIK, E.; VEASEY, E. A.; SANTOS, F. R.; FARIAS, I. P.; VIANNA, J. A.; OLIVEIRA, L. R.; WEBER, L. I.; ALMEIDA-TOLEDO, L. F.; FRANCISCO, M. R.; REDONDO, R. A. F.; SICILIANO, S.; DEL LAMA, S. N.; FREITAS, T. R.O.; HRBEK, T.; MOLINA, W. F. Genética da conservação brasileira. In: Frankham, R.; Ballou, J.D.; Briscoe, D.A. **Fundamentos de Genética da Conservação**. Ribeirão Preto: Editora SBG. cap. 10, p. 244-274, 2008

GIBSON, J.; MORTON, N. E.; COLLINS, A. Extended tracts of homozygosity in outbred human populations. **Human Molecular Genetics**, v. 15, p. 789–795, 2006.

GIELING, E. T.; SCHUURMAN, T.; NORDQUIST, R. E.; VAN DER STAAY, F. J. The pig as a model animal for studying cognition and neurobehavioral disorders. **Current Topics in Behavioral Neurosciences**, v. 5, p. 359–383, 2011.

HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; GODDARD, M. E. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. **Genome Research**, v. 13 p. 635–643, 2003.

HERRERO-MEDRANO, J.; MEGENS, H-J.; GROENEN, M. A. M.; RAMIS, G.; BOSSE, M.; PÉREZ-ENCISO, M.; CROOIJMANS, R.P.M.A. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. **BMC Genetics**, v. 14, p. 106-119, 2013.

HERÍAS, M. V.; HOGENKAMP, A.; VAN ASTEN, A. J.; TERSTEEG, M. H.; VAN EIJK, M.; HAAGSMAN, H. P. Expression sites of the collectin SP-D suggest its importance in first line host defence: Power of combining in situ hybridisation, RT-PCR and immunohistochemistry. **Molecular Immunology**, v. 44, n. 13, p.3324-3332, 2007.

HILL, W. G. G.; ROBERTSON, A. Linkage Disequilibrium in Finite Populations. **Theoretical and Applied Genetics**, v. 38, p. 226–231, 1968.

HILLAIRE, M. L.; VAN, E. M.; NIEUWKOOP, N.J.; VOGELZANG-VAN, T. S. E.; FOUCHIER, R. A.; OSTERHAUS, A. D.; HAAGSMAN, H. P.; RIMMELZWAAN, G. F. The number and position of N-linked glycosylation sites in the hemagglutinin determine differential recognition of seasonal and 2009 pandemic H1N1 influenza virus by porcine surfactant protein D. **Virus Research**, v. 169, n. 1, p.301-305, 2012.

HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. **Nature Reviews Genetics**, v. 10, n. 9, p.639-650, 2009.

HOWRIGAN, D. P.; SIMONSON, M.; KELLER, M. C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC Genomics**, v. 12, p. 460-475, 2011.

HUBISZ, M.J.; FALUSH, D.; STEPHENS, M.; PRITCHARD, J.K. Inferring weak population structure with the assistance of sample group information. **Molecular Ecology Resources**, v. 9, p. 1322–1332, 2009.

IRGANG, R. Melhoramento Genético Aplicado à Produção de Suínos: Raças e linhagens na produção de suínos. In: FERREIRA, A. H.; CARRARO, B.;

DALLANORA, D.; MACHADO, G.; MACHADO, I. P.; PINHEIRO, R.; ROHR, S. **Produção de suínos: teoria e prática**. Brasília: Qualitá, 2014. Cap. 2. p. 51-60.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **Bmc Genet**, v. 11, p.94-109, 2010

KAMIRI, Z. **Runs of Homozygosity patterns in Taurine and Indicine cattle breeds**. 2013. 53 f. Thesis (Major Animal Breeding and Genetics) - European Master in Animal Breeding and Genetics, Vienna, 2013.

KIM, E. S.; COLE, J.B.; HUSON, H.; WIGGANS, G. R.; VAN TASSELL, C. P.; CROOKER, B. A.; LIU, G.; DA, Y.; SONSTEGARD, T. S. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. **Plos One**, v. 8, e80813, 2013.

KIRIN, M.; MCQUILLAN, R.; FRANKLIN, C. S.; CAMPBELL, H.; MCKEIGUE, P. M.; WILSON J. F. Genomic Runs of Homozygosity Record Population History and Consanguinity. **Plos One**, v. 5, e13996, 2010.

LAWSON, D.J.; HELLENTHAL, G.; MYERS, S.; FALUSH, D. Inference of Population Structure using Dense Haplotype Data. **Plos Genetics**, v. 8, p. e1002453, 2012a.

LAWSON, D. J.; FALUSH, D. Population Identification Using Genetic Data. **Annual Review Of Genomics And Human Genetics**, v. 13, p.337-361, 2012b.

LEWONTIN, R. C.; KOJIMA, K. The evolutionary dynamics of complex polymorphisms. **Evolution**, v. 14, n. 4, p.458-472, 1960

LEWONTIN, R.C. The interaction of selection and linkage. I. General considerations: heterotic models. **Genetics**, v. 49, p. 49–67, 1964

LOPES, P. S. Melhoramento genético de suínos. In: VIII SIMPÓSIO BRASILEIRO DE MELHORAMENTO ANIMAL, Não use números Romanos ou letras, use somente números Arábicos., 2010, Maringá, Pr. **Anais - Palestras**. Maringá, 2010. Disponível em: <<http://sbmaonline.org.br/anais/viii/palestras/pdfs/3.pdf>>. Acesso em: 4 nov. 2015.

LOPES, M. S.; SILVA, F. F.; HARLIZIUS, B.; DUIJVESTIEN, N.; LOPES, P. S. GUIMARÃES, S. E. F.; KNOL, E. F. Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. **BMC Genetics**, v. 14, 2013.

MCQUILLAN, R.; LEUTENEGGER, A. L.; ABDEL-RAHMAN, R.; FRANKLIN, C. S.; PERICIC, M.; BARAC-LAUC, L.; SMOLEJ-NARANCIC, N.; JANICIJEVIC, B.; POLASEK, O.; TENESA, A.; MACLEOD, A. K.; FARRINGTON, S. M.; RUDAN, P.; HAYWARD, C.; VITART, V.; RUDAN, I.; WILD, S. H.; DUNLOP, M. G.; WRIGHT, A. F.; CAMPBELL, H.; WILSON, J. F. Runs of Homozygosity in European Populations. **The American Journal of Human Genetics**, v. 83, p. 359-372, 2008.

MIELE, M.; WAQUIL, P. Cadeia produtiva da carne suína no Brasil. **Política Agrícola**, Brasília, p.75-87, 2007.

NOTHNAGEL, M.; LU, T. T.; KAYSER, M.; KRAWCZAK, M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. **Human Molecular Genetics**, v. 19, n. 15, p.2927-2935, 2010.

NOVEMBRE, J.; STEPHENS, M. Interpreting principal component analyses of spatial population genetic variation. **Nature Genetics**, v. 40, p.646-649, 2008.

NOVEMBRE, J.; RAMACHANDRAN, S. Perspectives on Human Population Structure at the Cusp of the Sequencing Era. **Annual Review Of Genomics And Human Genetics**, v. 12, p.245-274, 2011.

O'BRIEN, A. M. P.; MÉSZÁROS, G.; UTSUNOMIYA, Y. T.; SONSTERGARD, T. S.; GARCIA, F. J.; TESSELL, C. P. V.; CAVALHEIRO, R.; SILVA, M. V. B.; SOLKNER, J. Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. **Livestock Science**, v. 166, p. 121-132, 2014.

PALGRAVE, C. J.; GILMOUR, L.; LOWDEN, C. S.; LILLICO, S. G.; MELLENCAMP, M. A.; WHITELAW, C. B. A. Species-Specific Variation in RELA Underlies Differences in NF- $\kappa$ B Activity: a Potential Role in African Swine Fever Pathogenesis. **Journal of Virology**, v. 85, n. 12, p.6008-6014, 2011.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of Population Structure Using Multilocus Genotype Data. **Genetics**, v. 155, p.945–959, 2000.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: a tool set for whole genome association and population-based linkage analyses. **The American Journal of Human Genetics**, v.81, p. 559-575, 2007

PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G.; Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, 2012.

RAMOS, A. M.; CROOIJMANS, R. P. M. A.; AFFARA, N. A.; AMARAL, A. J.; ARCHIBALD, A. L.; BEEVER, J. E.; BENDIXEN, C.; CHURCHER, C.; CLARK, R.; DEHAIS, P.; HANSEN, M. S.; HEDEGAARD, J.; HU, Z-L.; KERSTENS, H. H.; LAW, A. S.; MEGENS, H-J.; MILAN, D.; NONNEMAN, D. J.; ROHRER, G. A.; ROTHSCILD, M. F.; SMITH, T. P. L.; SCHNABEL, R. D.; VAN TASSELL, C. P.; TAYLOR, J. F.; WIEDMANN, R. T.; SCHOOK, L. B.; GROENEN, M. A. M. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. **Plos One**, v.8, p. e6524, 2009.

SAURA, M. FERNÁNDEZ, A.; VARONA, L.; FERNÁNDEZ, A. I.; CARA, M. A. R. de; BARRAGÁN, C.; VILLANUEVA, B. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. **Genetics Selection Evolution**, v. 47, 2015.

SCRAGGS, E.; ZANELLA, R.; WOJTOWICZ, A.; TAYLOR, J.F.; GASKINS, C.T.; REEVES, J.J.; de AVILA, J.M.; NEIBERGS, H.L. Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. **Journal of Animal Breeding and Genetics**, v. 131, n. 1, p.3-10, 2013.

SILVA, M. V. da; LOPES, P. S.; GUIMARÃES, S. E.; TORRES, R. de A. Utilização de marcadores genéticos em suínos. I. Características reprodutivas e de resistência a doenças. **Archivos Latinoamericanos de Producción Animal**, v. 11, 2003.

SILVA, E. C. da. **GENÉTICA DA CONSERVAÇÃO DE SUÍNOS LOCALMENTE ADAPTADOS NO BRASIL: USO DE FERRAMENTAS GENÔMICAS E GEOGRÁFICAS**. 2014. 154 f. Tese (Doutorado em Ciências Animais) - Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, Brasília, 2014.

SILIÓ, L.; RODRÍGUEZ, M. C.; FERNÁNDEZ, A.; BARRAGÁN, C.; BENÍTEZ, R.; ÓVILO, C.; FERNÁNDEZ, A.L. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. **Journal Of Animal Breeding And Genetics**, v. 130, p.349-360, 2013.

SÖLKNER, J.; FRKONJA, A.; RAADSMA, H.W.; JONAS, E.; THALLER, G.; GOOTWINE, E.; SEROUSSI, E.; FUERST, C.; EGGER-DANNER, C.; GREDLER, B. Estimation of Individual Levels of Admixture in Crossbred Populations from SNP Chip Data: Examples with Sheep and Cattle Populations. **Interbull Bulletin**, n. 42, p. 62-66, 2010.

TORTEREAU, F.; SERVIN, B.; FRANTZ, L.; MEGENS, H-J.; MILAN, D.; ROHRER, G.; WIEDMANN, R.; BEEVER, J.; ARCHIBALD, A. L.; SCHOOK, L. B.; GROENEN M. A. M. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. **BMC Genomics**, v. 13, n. 1, p.586-598, 2012.

UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. Livestock and Poultry: Word Markets and trade. Miami, 2015. Disponível em: <[http://apps.fas.usda.gov/psdonline/circulars/livestock\\_poultry.pdf](http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf)>. Acesso em: 24 Abril. 2015.

VERONEZE, R.; LOPES, P. S.; S. GUIMARÃES, E. F.; SILVA, F. F.; LOPES, M. S.; HARLIZIUS, B.; KNOL. E.F. Linkage disequilibrium and haplotype block structure in six commercial pig lines. **Journal of Animal Science**, v. 91, p.3493-3501, 2013.

WANG, D.; WANG, D.; SUN, Y.; STANG, P.; BERLIN, J. A.; WILCOX, M. A.; LI, Q. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. **BMC Proceedings**, v. 3, S109, 2009.

WANG, S. J.; LIU, W. J.; SARGENT, C. A.; ZHAO, S. H.; LIU, H. B.; LIU, X. D.; WANG, C.; HUA, G. H.; YANG, L. G.; AFFARA, N. A.; ZHANG, S. J. Effects of the polymorphisms of Mx1, BAT2 and CXCL12 genes on immunological traits in pigs. **Molecular Biology Reports**, v. 39, p. 2417-2427, 2012.

WANG, L.; SØRENSEN, P.; JANSS, L.; OSTERSEN, T.; EDWARDS, D. Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. **BMC Genetics**, v. 14, p. 115-126, 2013.

WEI, L.; LIU, J. Porcine circovirus type 2 replication is impaired by inhibition of the extracellular signal-regulated kinase (ERK) signaling pathway. **Virology**, v. 396, p. 203-209, 2009.

WRIGHT, S. Evolution in Medelian genetics. **Genetics**, v. 16, p. 97-159, 1931.

WRIGHT, S. The interpretation of population structure by F-statistics with special regard to systems of mating. **Evolution**, v.19, n. 3. p. 395-420, 1965.

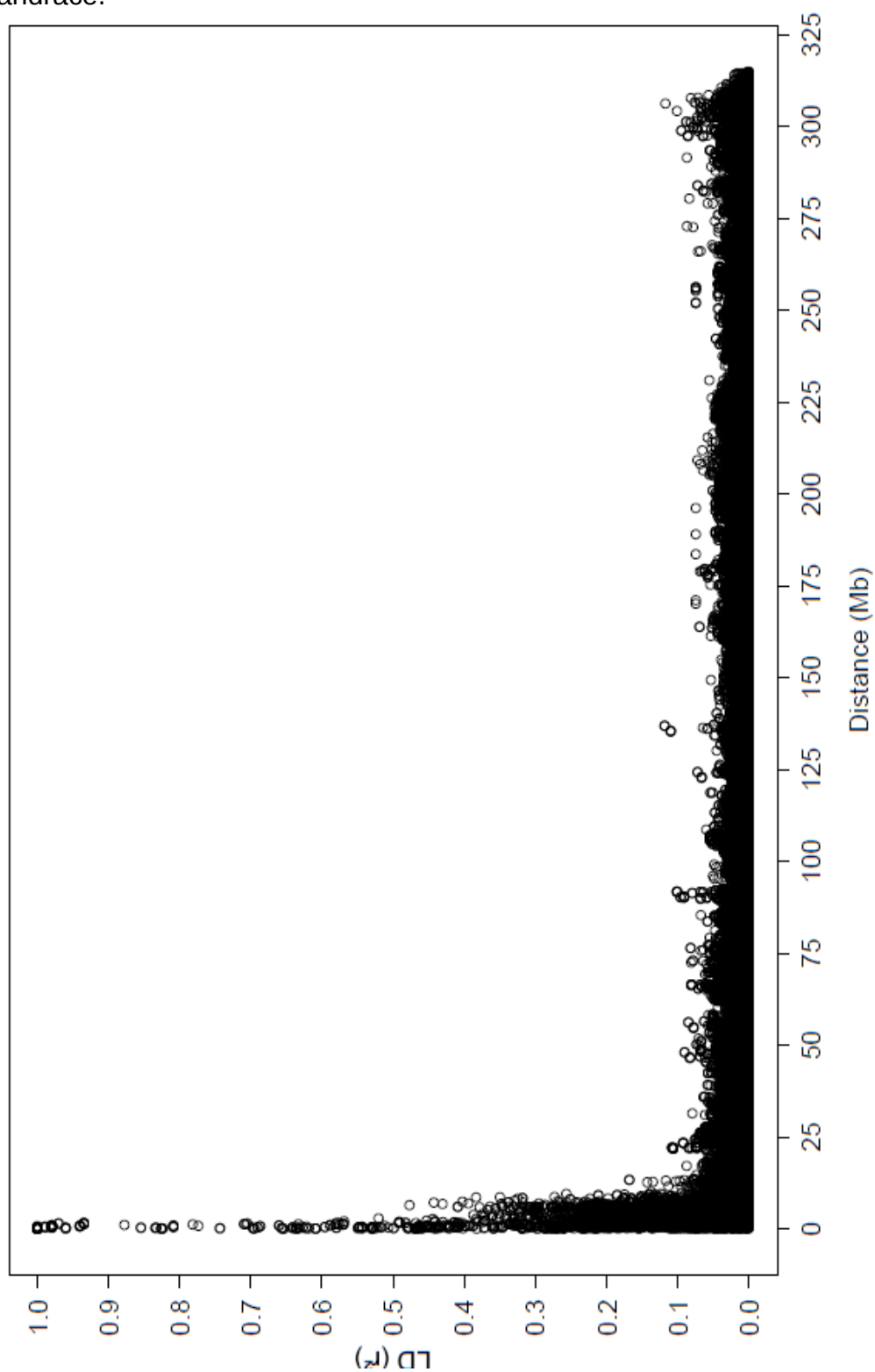
YANG, H. C.; CHANG, L. C.; LIANG, Y. J.; LIN, C. H.; WANG, P. L. A Genome-Wide Homozygosity Association Study Identifies Runs of Homozygosity Associated with Rheumatoid Arthritis in the Human Major Histocompatibility Complex. **Plos One**, v. 7, e34840, 2012.

ZHANG, Y.; J. YOUNG, M.; WANG, C.; SUN, X.; WOLC, A.; DEKKERS, J. C. M. Inbreeding by Pedigree and Genomic Markers in Selection Lines of Pigs. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., 2014, Vancouver. **Proceedings**. 2014.

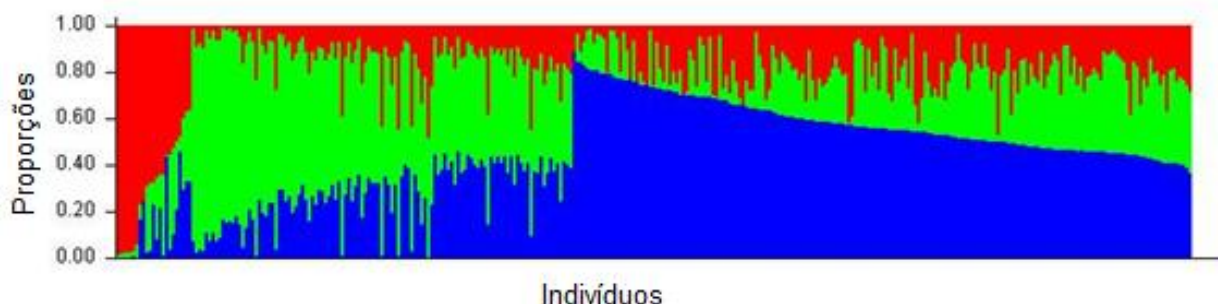
ZHANG, Q.; CALUS, M. P. L.; GULDBRANDTSEN, B.; LUND, M.; SAHANA, G.. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. **BMC Genetics**, v. 16, n. 1, 2015.

## 7 APÊNDICE

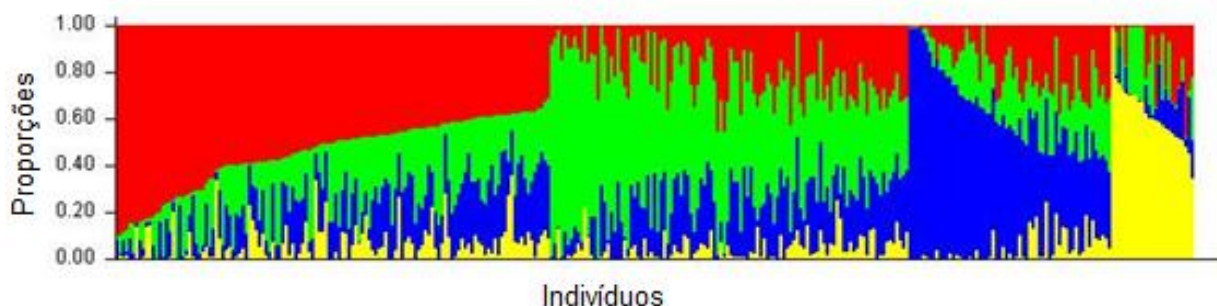
**Apêndice 1.** Gráfico de Desequilíbrio de Ligação entre marcadores de suínos da raça Landrace.



**Apêndice 3.** Constituição genética de animais de uma linha fêmea sintética base Landrace, em análise para  $K = 3$ . As três cores são referentes aos grupos definidos (“clusters”) com os resultados do programa STRUCTURE (Pritchard et al., 2000): vermelho = “Cluster” 1; verde = “Cluster” 2; azul = “Cluster” 3.



**Apêndice 4.** Constituição genética de animais de uma linha fêmea sintética base Landrace, em análise para  $K = 4$ . As quatro cores são referentes aos grupos definidos (“clusters”) com os resultados do programa STRUCTURE (Pritchard et al., 2000): vermelho = “Cluster” 1; verde = “Cluster” 2; azul = “Cluster” 3; amarela = “Cluster” 4.



**Apêndice 5.** Constituição genética de animais de uma linha fêmea sintética base Landrace, em análise para  $K = 5$ . As cinco cores são referentes aos grupos definidos (“clusters”) com os resultados do programa STRUCTURE (Pritchard et al., 2000): vermelho = “Cluster” 1; verde = “Cluster” 2; azul = “Cluster” 3; amarela = “Cluster” 4; rosa = “Cluster” 5.

