

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

**GENÉTICA DE PAISAGEM DE SUÍNOS NO BRASIL: IDENTIFICAÇÃO DE  
ASSINATURAS DE SELEÇÃO PARA ESTUDOS DE CONSERVAÇÃO E  
CARACTERIZAÇÃO DE REBANHOS**

ROBSON JOSÉ CESCNETO

Médico Veterinário/UFPeL

Mestre em Ciências/UFPeL

Tese apresentada ao Programa de Pós-Graduação em Zootecnia da UFRGS como requisito para obtenção do título de Doutor em Ciências, área de concentração Produção Animal.

Porto Alegre (RS), Brasil

Julho de 2016

### CIP - Catalogação na Publicação

Cesconeto, Robson José

GENÉTICA DE PAISAGEM DE SUÍNOS NO BRASIL:  
Identificação de assinaturas de seleção para estudos  
de conservação e caracterização de rebanhos / Robson  
José Cesconeto. -- 2016.  
131 f.

Orientador: José Braccini Neto.

Coorientadores: C. Mc Manus, S. Joost.

Tese (Doutorado) -- Universidade Federal do Rio  
Grande do Sul, Faculdade de Agronomia, Programa de  
Pós-Graduação em Zootecnia, Porto Alegre, BR-RS, 2016.

1. Suínos brasileiros. 2. Genética de Populações.  
3. Estrutura Genética de Populações Naturais. 4.  
Recursos Genéticos. 5. Sus scrofa. I. Braccini Neto,  
José, orient. II. Mc Manus, C., coorient. III.  
Joost, S., coorient. IV. Título.

ROBSON JOSÉ CESCNETO  
Medico Veterinário e Mestre em Produção Animal

## TESE

Submetida como parte dos requisitos  
para obtenção do Grau de

### DOUTOR EM ZOOTECNIA

Programa de Pós-Graduação em Zootecnia  
Faculdade de Agronomia  
Universidade Federal do Rio Grande do Sul  
Porto Alegre (RS), Brasil

Aprovado em: 05.07.2016  
Pela Banca Examinadora

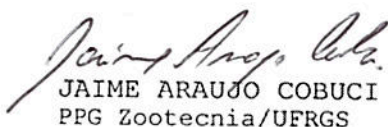
Homologado em: 14.12.2016  
Por



JOSÉ BRACCINI NETO  
PPG Zootecnia/UFRGS  
ORIENTADOR



PAULO CÉSAR DE FACCIÓ CARVALHO  
Coordenador do Programa de  
Pós-Graduação em Zootecnia



JAIME ARAUJO COBUCI  
PPG Zootecnia/UFRGS



SAMUEL REZENDE PAIVA  
EMBRAPA



PATRÍCIA IANELLA  
EMBRAPA



PEDRO ALBERTO SELBACH  
Diretor da Faculdade de Agronomia

## AGRADECIMENTOS

A Deus, pelo dom da vida, pela inspiração e, por não me permitir desanimar diante das dificuldades encontradas durante a caminhada;

Aos meus pais pelo que sou hoje;

Àquela que me apoiou e “suportou” desde a minha graduação, mas principalmente desde a decisão de iniciar o doutoramento até o fim deste, e espero, por muito tempo ainda;

Aos “Mestres”, pois sem eles, jamais teria chego até aqui;

Aos colegas do PGZ-UFRGS, pela convivência e apoio mútuo nas horas de dificuldade;

Aos amigos franceses, belgas, suíços, italianos, cubanos, chineses, que fiz durante minha estadia na Europa;

A todos aqueles, que colaboraram direta ou indiretamente na realização deste trabalho;

Ao Programa de Pós-Graduação em Zootecnia da Universidade Federal do Rio Grande do Sul e a École Polytechnique Fédérale de Lausanne, pela estrutura disponibilizada para a realização deste trabalho.

À CAPES, pelas bolsas de estudos no Brasil e no exterior

Agradeço, em especial, aos meus orientadores José Braccini e Stephane Joost por todo apoio recebido.

Muito Obrigado!

*Ao semeador, cabe apenas semear.*

## **Genética de paisagem de Suínos no Brasil: identificação de assinaturas de seleção para estudos de conservação e caracterização de rebanhos<sup>1</sup>**

Autor: Robson José Cesconeto

Orientador: José Braccini Neto

Estudos em genética de paisagem, das espécies zootécnicas, podem impulsionar o entendimento dos processos adaptativos, bem como, a maneira que o ambiente afeta o sucesso destas populações. O objetivo principal do projeto foi identificar assinaturas de seleção no genoma populações de suínos naturalizados brasileiros. Procurando obter a maior representatividade da variabilidade genética e ambiental dos suínos dentro do território brasileiros, amostras de DNA de pelo menos um animal e de pelo menos um grupo genético foram obtidas dentro do Banco de Germoplasma da EMBRAPA, totalizando 191 animais de 18 grupos genéticos suínos brasileiras que foram genotipadas, e classificadas de acordo com sua origem dentro de raça ou grupo genético, estado bioma, bacia hidrográfica, tipo de solo, ecoregião e tipo de vegetação. Após um controle de qualidade os genótipos resultantes foram utilizados no cálculo das estatísticas  $F$  de Wright, equilíbrio de Hardy-Weinberg. Análise de componentes principais, coeficiente de endogamia, análise da variância molecular, testes de Mantel, e a determinação do número ideal de populações. Os dados ambientais foram convertidos em layers através do Qgis e utilizados na detecção de assinaturas de seleção através do BayeScan v2.1 (Foll & Gaggiotti 2008) e do Samβada (Stucki *et al.* 2014). Os resultados obtidos mostram que as populações estudadas têm uma estrutura variada entre e dentro si. A maior parte da variabilidade genética esta presente nos indivíduos dentro de grupos genéticos, e indivíduos dentro de estados. Existe diferenciação genética dos suínos dentro das variáveis ambientais classificatórias. As raças Monteiro e Marajó foram as que mostraram maiores níveis de estruturação. Os componentes principais mostram proximidade entre as raças comerciais, assim como o elevado grau de variação na composição genéticas das demais raças, com marcante separação dos animais Monteiro e Marajó. Níveis elevados de endogamia foram encontrados. Foram encontrados pelo menos 8 assinaturas de seleção no genoma das raças suínas localmente adaptadas. A frequência dos genótipos destes marcadores divide o território brasileiro em duas regiões latitudinais distintas. Os níveis de estruturação das populações demonstram uma grande variabilidade genética entre e dentro as raças. As marcas de seleção encontradas demonstram a influência do meio ambiente no sucesso adaptativo destes animais ao território brasileiro.

**Palavras-chave:** Suínos brasileiros, genética de populações, estrutura genética de populações naturais, recursos genéticos, *Sus scrofa*, genética de conservação.

## **Landscape Genetic of Brazilian Autochthonous Swine Breeds: an Approach to Detect Signatures of Natural Selection to Studies of Conservation and Breed Characterization<sup>1</sup>**

Author: Robson José Cesconeto

Advisor: José Braccini Neto

Study of landscape genetics in livestock species can help understand the adaptive processes and the way that the environment affects the success of these populations. The main objective was the identification of genetic signatures of selection in the Brazilian autochthone swine breeds. We aim the higher representation of genetic and environmental variability from Brazilian territory sampling at least one animal from one genetic group per sampling point. The samples were obtained on CENARGEN-EMBRAPA Germplasm Bank in a total of 191 DNA samples from 19 Brazilian locally adapted swine genetic groups, classified into state, region, biome, hydrological basins, soil type, ecoregion and vegetation type. These samples were genotyped and after a quality control, we calculated Wright' F-statistics, individual inbreeding coefficient, Hardy-Weinberg equilibrium, Principal components, IBD, Mantel test, an ideal number of subpopulations and an analysis of molecular variance. Environmental data was converted in layers through Qgis and used to detect signatures of selection by BayeScan v2.1 (Foll & Gaggiotti 2008) and Samβada ((Stucki *et al.* 2014).

Population studied had genetic structure among different subpopulations, in a same category. The biggest part of genetic variability is the individual within breed and in individual within state. We found different intensity of population structure in the categories studied, been the he Monteiro and Marajó breeds that ones with highest values from population genetic structure. Principal component analyses showed proximity between commercial breeds, as well as the high degree of genetic variation among other breeds, with great detachment of Monteiro and Marajó animals from others. High inbreeding levels were found. Signatures of selection were found on the genome of Brazilian locally adapted swine pig. The genotypic frequency of these signatures of selection divides the Brazilian territory into two distinct latitudinal regions. The population structure levels demonstrate a wide genetic variability inside and among races. The signatures of natural selection found demonstrate the influence of the environment to successful adaptation of swine in Brazil.

**Keywords:** Brazilian pig, population genetics, natural environment genetic structure, animal genetic resources, *Sus scrofa*, population structure, conservation genetics

## SUMÁRIO

		pp
	<b>Agradecimentos.</b> .....	3
	<b>RESUMO.</b> .....	5
	<b>ABSTRACT.</b> .....	6
	<b>LISTA DE TABELAS.</b> .....	9
	<b>LISTA DE FIGURAS.</b> .....	10
	<b>LISTA DE ABREVIATURAS.</b> .....	12
	<b>CAPÍTULO I.</b> .....	13
1.1	INTRODUÇÃO GERAL. ....	14
1.2	REVISÃO BIBLIOGRÁFICA. ....	16
1.2.1	Genética de Paisagens. ....	16
1.2.1.1	Amostragem. ....	19
1.2.1.2	Analises em Geoestatística. ....	23
1.2.1.2.1	Geoestatística. ....	23
1.2.1.2.2	Desagrupamento. ....	25
1.2.1.2.3	Função Variograma/Semi-variograma. ....	26
1.2.1.2.4	Krigagem. ....	27
1.2.2	Genética de Populações. ....	27
1.2.2.1	Teorema de Hardy-Weimberg. ....	29
1.2.2.2	Alterações nas frequências gênicas. ....	33
1.2.2.2.1	Introduzindo alelos. ....	33
1.2.2.2.1.1	Mutação. ....	33
1.2.2.2.1.2	Migração/Emigração. ....	34
1.2.2.2.2	Eliminação de alelos. ....	35
1.2.2.2.2.1	Deriva Aleatória. ....	35
1.2.2.2.2.2	Efeito de Wahlund. ....	37
1.2.2.2.2.3	Seleção natural. ....	37
1.2.2.2.2.4	Endocruzamento. ....	38
1.2.2.2.2.5	Subdivisão. ....	39
1.2.2.2.2.6	Seleção Contra o Híbrido. ....	40
1.2.2.3	Estatísticas de Wright .....	40
1.3	HIPÓTESES E OBJETIVOS. ....	42
2	<b>CAPÍTULO II.</b> .....	43
	Population structure of Brazilian locally adapted swine genetic groups using PorcineSNP60 BeadChip. ....	44
3	<b>CAPÍTULO III.</b> .....	74
	Landscape Genomic Approach to Detect Selection Signatures in Locally Adapted Brazilian Swine Genetic Groups. ....	75
4	<b>CAPITULO IV.</b> .....	105



4.1	CONSIDERAÇÕES FINAIS. ....	106
4.2	REFERÊNCIAS BIBLIOGRÁFICAS. ....	109
5	<b>APÊNDICE.</b> ....	119
	Porquê Preservar a Raças Suínas Brasileiras. ....	120
6	<b>VITA.</b> ....	130

## LISTA DE TABELAS

<b>CAPÍTULO II</b> .....	43
Table 1: Sampling frequency of genetic group and local. ....	66
Table 2: Description of categorical variables for locally adapted swine populations. ....	67
Table 3: Pairwise $F_{ST}$ for Brazilian locally adapted Swine genetic groups. ....	68
Table 4: Pairwise $F_{ST}$ between samples from different Brazilian States. ....	69
Table 5: $F_{IS}$ values for Brazilian locally adapted swine genetic groups and for States in Brazil after 10000 permutations. ....	69
Table S1: Pairwise $F_{ST}$ for Brazilian locally adapted swine genetic groups by kind of vegetation. ....	70
Table S2: Pairwise $F_{ST}$ for Brazilian locally adapted swine genetic groups by kinds of soil. ....	71
Table S3: Pairwise $F_{ST}$ for Brazilian locally adapted swine genetic groups by Ecoregions ....	72
Table S4: Pairwise $F_{ST}$ for Brazilian locally adapted swine genetic groups by hydrological basin. ....	73
<b>CAPÍTULO III</b> .....	74
Table 1: Samβada output to Environmental association to 5 markers detected as signatures of selection. ....	88
Table 2: Biological function of markers identified as selection signatures in Brazilian locally adapted swine breeds. ....	89
Table 3: Regional frequencies for genotypes from markers detected as signature of selection. ....	90
Table S1: Number of individuals per locally adapted genetic group sampled per Brazilian State. ....	91
Table S2: Environmental variables in the present study (Geographic Data Base). ....	92
Table S3: AMOVA values among Brazilian regions. ....	92
Table S4: AMOVA values among commercial and locally adapted genetic groups. ....	92

## LISTA DE FIGURAS

<b>CAPÍTULO I</b> .....	13
Figura 1: Integração das informações em genética de paisagem.....	19
Figura 2: A - Representação bi-dimensional de três pontos amostrais e seus respectivos vetores B- Representação tri-dimensional d dos mesmos três pontos amostrais e seus respectivos vetores. ....	20
Figura 3: Comparação de três Diferentes resoluções espaciais onde os quadros representam 01 Arc Seg* (Vermelho), 30 Arc Seg (Preto), 1 Arc Min (Amarelo). ....	21
Figura 4: Representação visual das possíveis abordagens do processo de amostragem em genética de paisagens.....	22
Figura 5: Esquemática do processo de progressão cartesiana.....	23
Figura 6: Relação entre a Função de Variograma e a Função de Covariância.	26
Figura 7: Representação esquemática do fenômeno de Migração.....	35
Figura 8: Representação esquemática de deriva.....	37
Figura 9: Relação entre as estatísticas F de Wright.....	41
<b>CAPÍTULO II</b> .....	43
Figure 1: Map of swine breed sampling: A- Political regions and Biomes; B- Brazilian Ecoreigons; C- Brazilian Hydrological Basins. ....	61
Figure 2: 3-D Representation of the first three Principal Components from allelic frequencies of 19,422 SNPs of 181 animals sampled, according to the genetic group of origin.. ....	62
Figure 3: Cluster Analysis of Brazilian Autochthones Swines Breeds based on genetic distances between breeds using pairwise $F_{ST}$ . ....	63
Figure 4: Identity by Descent coefficients Heat Map from 181 samples grouped by breed. ....	64
Figure 5: Population structure assessed by Structure. Bar plot. Each individual is represented by a vertical bar, often partitioned into colored segments with the length of each segment representing the proportion of the individual's genome from K = 2, 7, or 18 populations. ....	65
<b>CAPÍTULO III</b> .....	74

Figure 1: Frequencies of number of model for each marker detected as selection signature in Samβada. ....	94
Figure 2: Correlogram of Maximum (Max), Minimum (Min) and Average values of Moran's I from all markers. ....	95
Figure 3: Correlogram from Genotypes of the markers identified as selection signatures in Brazilian locally adapted swine breeds by BayeScan and Samβada. ....	96
Figure 4: Genotype distribution maps from markers identified as selection signatures by Samβada/BayeScan. (a) Marker Asga0029202 in Bio18 (precipitation of Warmest Quarter). (b) Marker Alga0054315 in Minimum Temperature in May and (c) Marker Alga0054315 in Maximum Temperature in May. ....	97
Figure S1: Sampling location for Brazilian locally adapted swine breeds with biomes and states on the Brazilian territory.....	98
Figure S2: Density of points (samples) around Mantel correlation line. ....	99

**LISTA DE ABREVIATURAS**

AMOVA	Análise hierárquica da variância molecular
Call Rate	Taxa de Chamada
Cm	centiMorgam
DNA	Ácido Desoxirribonucleico
GIS	Geographic Information System
H <sub>0</sub>	Hipótese alternativa
H <sub>1</sub>	Hipótese nula
H <sub>e</sub>	Heterozigose Esperada
H <sub>o</sub>	Heterozigose Observada
HWE	Hardy-Weinberg Equilibrium
IBD	Isolation by Distance
Kb	kilo base
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
PCA	Principal Component Analysis
QTL	Quantitative Traci Loci
SNP	Single-nucleotide polymorphism

## CAPÍTULO I

“...Por longe que o homem vá  
Jamais fugirá de si...”

*(Jaime Caetano Braum)*

## 1.1 INTRODUÇÃO GERAL

Genética de paisagem é o termo utilizado para descrever o comportamento genético das populações naturais, dentro dos ecossistemas onde ocorrem, com o uso de modernas ferramentas estatístico-espaciais de análise de bancos de dados genéticos georreferenciados, aprimorando assim o estudo da genética de populações (MANEL *et al.*, 2003). Em zootecnia conhecer o comportamento genético das populações naturalizadas que sofrem pouca ou nenhuma seleção artificial, auxilia no entendimento dos processos adaptativos das espécies. Da mesma maneira, auxilia na seleção de animais melhor adaptados às condições climático\ambientais, dentro das regiões produtoras de interesse, pois algumas características estão intimamente ligadas a determinadas regiões geográficas (CAÑÓN *et al.*, 2001). Segundo (HANOTTE; DESSIE; KEMP, 2010) a genética de paisagens promove uma ligação entre as informações genéticas e geo-ambientais (fig. 1) para identificar material genético precioso.

A chegada do gênero *SUIS* ao continente americano se deu no século XVI, trazido pelos colonizadores europeus (CAVALCANTI, 1985; CROSBY, 2003). A importância dos suínos como fonte de proteína para a alimentação humana e rápida adaptabilidade destes animais em nosso continente, possibilitou a criação e dispersão de animais com características produtivas favoráveis ao nosso ambiente. Muitas das raças trazidas pelos imigrantes permanecem até hoje, compondo diversos grupos genéticos de suínos localmente adaptados, sendo criadas praticamente sem seleção além daquela realizada pela natureza (SOLLERO *et al.*, 2009). Porém, numa escala evolutiva de tempo, a presença dos suínos em território brasileiro ocorre há muito pouco tempo. A rápida adaptação aos desafios ambientais, como o clima, parasitas, escassez de alimentos e manejo, além das alterações fenotípicas corporais observadas nos diferentes grupos genéticos, deve-se provavelmente às alterações não sinônimas em sequências fixas e a uma rápida mudança nas frequências alélicas, e não em regiões regulatórias ou codificantes, como observado em uma escala evolucionária maior.

Após as mudanças no mercado na década de 70, praticamente todo o material genético suíno registrado utilizado nos criatórios comerciais brasileiros, apesar de multiplicado e selecionado no Brasil, é derivado das principais raças comerciais criadas no mundo (ABCS, 2014; BLACKBURN, 2012). Atualmente os programas de melhoramento genético de suínos dependem praticamente em sua totalidade, da importação de animais puros da Europa ou da América do Norte.

Recentemente, algumas empresas de melhoramento genético suíno introduziram em seus programas, animais de origem asiática (FAVERO; FIGUEIREDO, 2009). Isso se deu pelo minucioso estudo de característica de interesse, como o desenvolvimento fetal (FORD, 1997), e identificação de áreas do genoma de interesse (SILVA *et al.*, 2003), o que também poderia ser realizado a partir das raças crioulas nacionais.

Uma caracterização racial minuciosa deve ser feita em três níveis, genético, fenotípico e produtivo. Para a caracterização genotípica, marcadores moleculares são as ferramentas mais acessíveis. Dos vários marcadores utilizados na pesquisa genética animal, os marcadores de QTL reportados são imprecisos na localização de genes de interesse uma vez que possuem baixa resolução (1-50 cM), e a correspondência com regiões de interesse é esparsa (FAHRENKRUG *et al.*, 2002). Marcadores de microssatélites apesar de abundantes, e servirem para a identificação individual/paternidade (PAIVA *et al.*, 2004; SILVEIRA *et al.*, 2008), são instáveis e possuem pouca cobertura do genoma dos indivíduos. A descoberta de polimorfismos de base única possibilitou uma maior cobertura e resolução do genoma individual, pois são abundantes e estáveis. Marcadores SNP possibilitam a obtenção de mapas genéticos densos, diretamente pela genotipagem individual com painéis de alta densidade (800kb) ou indiretamente pelo processo de imputação a partir de genotipagens com painéis de baixa densidade (60kb) (MARCHINI; HOWIE, 2010).

Apesar da existência de alguns rebanhos de conservação dos grupos genéticos de suínos localmente adaptados brasileiros, tanto a localização como o nível de pureza racial dos rebanhos da maioria dos grupos



genéticos, ainda é pouco conhecido. Identificar os rebanhos e atestar a pureza racial dos mesmos é fundamental para ter-se confiança na caracterização fenotípica e produtiva dos grupos genéticos de suínos localmente adaptados brasileiros e são passos posteriores à identificação genética desses grupos.

Desta forma, conhecer a estrutura genética da população de suínos localmente adaptados brasileiros e entender a influência do meio ambiente na estruturação e fixação de alelos nos grupos genéticos de suínos localmente adaptados brasileiros irá contribuir tanto com os programas de melhoramento genético, como nos programas de conservação, auxiliando no constante desenvolvimento da suinocultura nacional.

## **1.2 REVISÃO BIBLIOGRAFICA**

### **1.2.1 Genética de Paisagem**

O sucesso na colonização de um ambiente por uma espécie, seja ela animal ou vegetal, se deve principalmente a uma interação favorável entre a composição genética da população e as condições ambientais à que os indivíduos são expostos ao longo do tempo.

O uso da genética de populações oportuniza um acesso a variabilidade genética das populações pelo estudo das frequências gênicas, intercâmbio de genes entre as populações, entre outros parâmetros importantes para o conhecimento da história genética das populações.

Porém, os fatores que levam à estruturação genética, assim como, à fixação de determinados alelos em detrimento de outros, não são completamente explicados somente pela deriva aleatória ou pelo efeito carona, uma vez que, mecanismos de seleção natural podem ser determinantes neste processo. Estes fenômenos (seleção natural, deriva e migração) estão constantemente atuando na composição das frequências gênicas dentro das populações. Dessa forma, estudar a heterogeneidade da paisagem identificando as barreiras para o fluxo gênico, assim como a maneira que o meio ambiente interfere no processo de adaptação das populações naturais é o objetivo de um estudo de genética de paisagem.

Nesse contexto a paisagem é um complexo conjunto de variáveis ambientais que podem ser analisadas em conjunto ou isoladamente a fim de identificar continuidades espaciais que expliquem a distribuição dos genótipos. A ausência de variação através do espaço amostral, assim como, a presença de correlação ou interdependência entre estas variáveis ambientais representam risco as conclusões em um estudo em escala de paisagem, por aumentar a chance de falsas descobertas. Da mesma forma, as diferentes precisões na coleta das diferentes variáveis podem somar certo nível de viés as conclusões. Existem basicamente duas formas de abordar estes problemas: a exclusão de variáveis altamente correlacionadas através de critérios como o fator de inflação da variância; e criar novas variáveis a partir das variáveis originais com o uso da análise de componentes principais.

Uma vez determinadas as variáveis ambientais importantes para atender os objetivos do estudo, cabe ao pesquisador determinar as ferramentas estatísticas capazes de detectar correlações entre as frequências genéticas e as variações ambientais. Análise multivariada da variância pode ser utilizada em casos em que existem mais de uma variável dependente (BARKER; BARKER 1984). Regressões múltiplas podem ser usadas quando se espera uma resposta linear das variáveis respostas, enquanto a regressão logística permite estimar e testar a influência de covariáveis na variável dependente (JOOST *et al.*, 2010). Análise de agrupamentos, correlações canônicas e análise espacial de componentes principais podem ser utilizadas para identificar semelhanças entre as variáveis. Determinar se a covariação entre as variáveis é fruto do puro acaso ou está condicionada a uma relação de causa e efeito é o último passo nesse tipo de abordagem, e deve ser capaz de explicar biologicamente as frequências alélicas observadas. A descontinuidade genética em relação às distâncias geográficas, dentro da paisagem, pode ser medida por meio do teste de Mantel, simples ou parcial, determinando a permeabilidade da paisagem em relação ao fluxo genético. Porém, por não levar em conta a não interdependência entre as variáveis dependentes e independentes, os métodos comparação (e.g. Akaike information criterion) entre os possíveis modelos não é viável (MANEL; HOLDEREGGER, 2013). A

incorporação da matriz de covariância entre as frequências alélicas, pelo uso de modelos mistos, poderia ser utilizada para resolver o problema da não interdependência (GOLDBERG; WAITS 2009). Estratégias para estimar o fluxo genético entre as populações, como o uso de regressão múltipla, análise de redundância canônica poderiam ser utilizadas, de forma a corrigir as deficiências do uso do teste de Mantel (RAUFASTE; ROUSSET, 2001). Utilizando o índice de correlação espacial global (e local) podemos acessar a continuidade espacial das frequências alélicas e estabelecer padrões de dispersão globais ou locais. Com a sobreposição dos índices de autocorrelação com os mapas ambientais é possível identificar padrões de dispersão e barreiras ao fluxo genético. Autocorrelação espacial ocorre quando amostras em um espaço geográfico tendem a serem mais similares ou dissimilares entre si do que o esperado simplesmente pelo acaso, para uma determinada variável (LEGENDRE; LEGENDRE 1998). Em estudos de genética de paisagem, este fenômeno pode ser causado pela endogamia, como uma resposta dos indivíduos ao ambiente ou ambos (DINIZ-FILHO *et al.*, 2009).

Para que se possa entender a magnitude dos efeitos ambientais dentro dos processos de seleção natural, faz-se mister o entendimento da distribuição geográfica das populações e dos diferentes alelos, dentro e entre, as populações. Esse entendimento se faz por meio do estudo das condições ambientais ao longo dos anos e da identificação dos fatores ambientais determinantes para o sucesso adaptativo da espécie em estudo.

A união das ferramentas de genética de populações, ecologia de paisagem e geografia espacial possibilita um melhor entendimento destes processos. Podemos dizer que Genética de Paisagens é o estudo da influência da paisagem (meio ambiente) na composição genética das populações, bem como, sua interferência nos processos de fixação e dispersão genética (fig. 1).



Figura 1: Integração das informações em genética de paisagem

Em suma, Genética de Paisagem é a almagamação entre genética de populações e ecologia de paisagens (MANEL *et al.*, 2003); ou o ramo da ciência que explica a quantidade dos efeitos da paisagem, configuração e matriz de qualidade do fluxo genético e variação genética espacial (STORFER *et al.*, 2007).

#### 1.2.1.1 Amostragem

Um dos principais pontos ao iniciar um estudo genético em nível de paisagem, é o correto delineamento da amostragem (HALL; BEISSINGER, 2014). Para (STEIN; ETTEMA, 2003) uma estratégia cuidadosamente planejada para a coleta de dados deve ser realizada no estudo dos fenômenos biológicos utilizando análises espaciais que envolvem meio ambiente, questões ecológicas e genética da paisagem. De fato, os dados espaciais possuem dois atributos igualmente importantes: 1- frequência de um determinado marcador molecular e 2- a sua localização, posição no espaço: longitude e latitude (SCHRÖDER, 2006). Uma eficiente coleta de dados deve ser realizada de forma a obter a maior representatividade espacial da região em estudo, assim como, a melhor representatividade da diversidade genética da espécie em estudo, minimizando os efeitos das análises em várias escalas temporo-espaciais do delineamento (MANEL *et al.*, 2003; MANEL *et al.*, 2010).

Nos estudos de genética de populações cada amostra indica um ponto específico no espaço bi ou tridimensional (latitude, longitude e altitude), podendo ser representado por um vetor de localização (fig. 2).

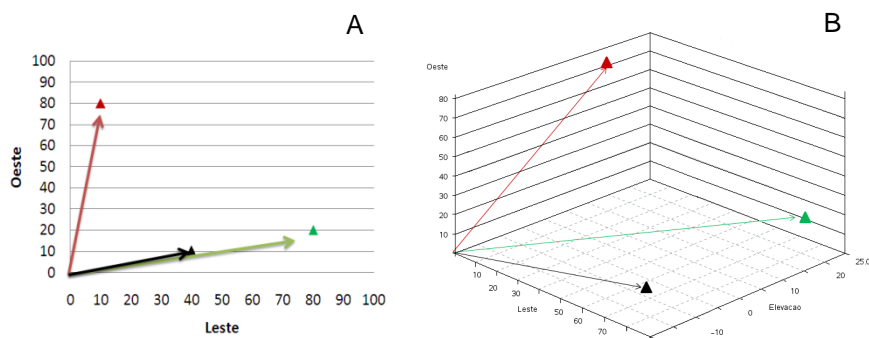


Figura 2: A - Representação bi-dimensional de três pontos amostrais e seus respectivos vetores B- Representação tri-dimensional dos mesmos três pontos amostrais e seus respectivos vetores

A qualidade da amostragem é um dos fatores determinantes para garantir o poder estatístico da amostra selecionada. Uma correta amostragem deve, além de captar a variação genética, promover uma eficiente ligação (linkagem), com a variação ambiental que influencia os eventos genéticos que afetam as populações estudadas, permitindo a produção de modelos espaciais realísticos (JOOST *et al.*, 2010). Tal qual a resolução de uma fotografia, a resolução da amostragem e dos dados ambientais (fig. 3), deve variar conforme os objetivos e espécies estudadas, gerando modelos espaciais que equilibrem a dimensionalidade dos dados genéticos (ocorrência, território, propagação, etc) e dos fenômenos ambientais envolvidos (LEEMPOEL *et al.*, 2015). Nos sistemas de informação geográfica (GIS), a informação geográfica é representada nos modelos por meio de pixels, em modelos de imagens contínuas e/ou rasters, ou meio de pontos, segmentos e polígonos quando operando em modo vetorial discreto (JOOST *et al.*, 2010). Resoluções espaciais finas (menores que 4 metros) permitem um grande incremento na acurácia e precisão da detecção, classificação e delimitação das características da paisagem como habitat, tanto em duas como em três dimensões (GREENBERG; DOBROWSKI; USTIN, 2005; WULDER *et al.*, 2004). Porém,

quando trabalhamos com espécies domésticas devemos levar em conta os sistemas de produção, tamanho dos rebanhos, comportamento da espécie entre outros, na determinação da melhor resolução espacial da amostragem (JOOST *et al.*, 2010). (STORFER *et al.*, 2007) ressaltam a importância de uma perfeita relação entre a escala dos dados ambientais utilizados e as questões formuladas para as hipóteses em estudo.

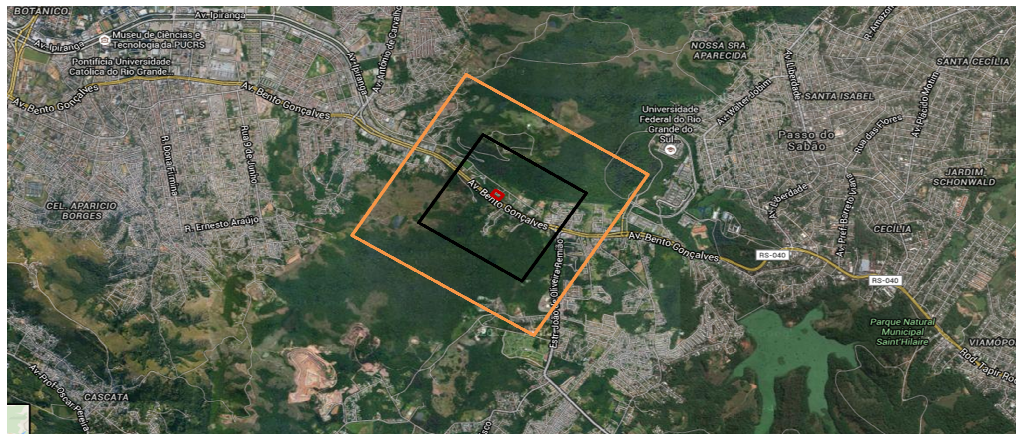


Figura 3: Comparação de três Diferentes resoluções espaciais onde os quadros representam 01 Arc Seg\* (Vermelho), 30 Arc Seg (Preto), 1 Arc Min (Amarelo). \*Arc Seg= Um Segundo de Arco da curvatura terrestre. Corresponde à 30,87 metros (cada grau da curvatura terrestre e igual a 111,12 Km)

Existem vários esquemas de amostragem possíveis em genética de paisagem (fig. 4). Uma amostragem sistemática, pode se mostrar mais eficiente para cobrir uma área de amostragem grande do que uma amostragem aleatória, mas pode se tornar viesada em áreas onde as variações ambientais assumem um gradiente semelhante ao intervalo de coleta das amostras. Da mesma forma uma amostragem completamente aleatória se mostra benéfica no estudo de espécies continuamente distribuídas, ou em parcelas do ambiente estudado, gerando uma teia de observações capaz de explicar as inter-relações entre os pontos amostrados. Porém, este tipo de amostragem pode gerar uma sobre-amostragem/sub-amostragem em algumas regiões deixando de capturar a totalidade da variação ambiental/genética da área em estudo.

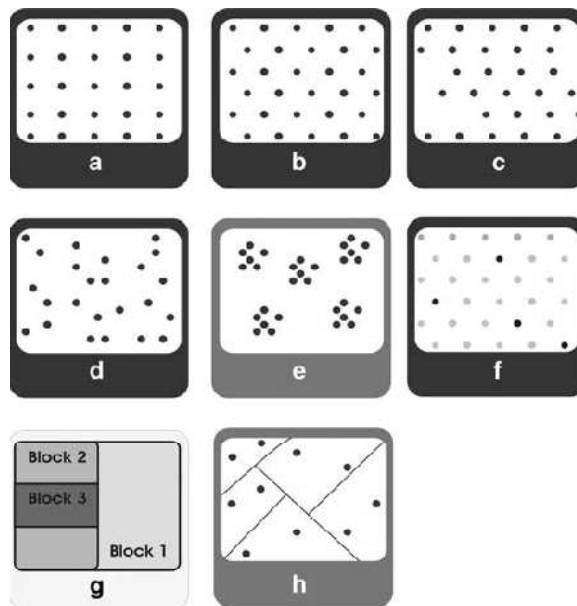


Figura 4: Representação visual das possíveis abordagens do processo de amostragem em genética de paisagens. a- Amostragem sistemática. b- Amostragem sistemática desalinhada. c- Amostragem sistemática com início aleatório. d- Aleatória simples. e- Aleatória dentro de parcelas. f- Duas escalas de variáveis. g- Ambiente dividido em blocos. h- Amostras coletadas em diferentes extratos de um ambiente particionado com base em várias variáveis. Fonte: (STORFER *et al.*, 2007)

Ainda dentro da amostragem, coletar vários pontos próximos como ao longo de rodovias, que por vezes se faz necessária, pode levar a estimações viesadas (YAMAMOTO 2013), tanto pela coleta de animais aparentados como pela ausência de variação ambiental. O uso de *grids* fixos ou móveis, conforme sugerido por (JOOST *et al.*, 2010; MANEL *et al.*, 2007), se mostra eficaz na identificação de barreiras ao fluxo genético.

Para que se possa fazer a correta análise de distâncias, precisamos projetar os pontos de localização obtidos de um sistema de referência geodésico em uma superfície plana (fig. 5), criando um mapa de distribuição espacial do globo terrestre. Para se obter um mapa preciso, necessitamos adotar um modelo matemático simplificado que melhor represente o globo terrestre, projetar os elementos da superfície sobre o modelo selecionado e relacionar os pontos do modelo matemático ao plano de projeção (IBGE, 2016).

Porém, uma vez que a projeção de uma superfície curva sobre um plano gera distorções, devemos considerar a melhor projeção de acordo com os objetivos do mapa produzido, assim como, lidar com as distorções produzidas por cada tipo de projeção.

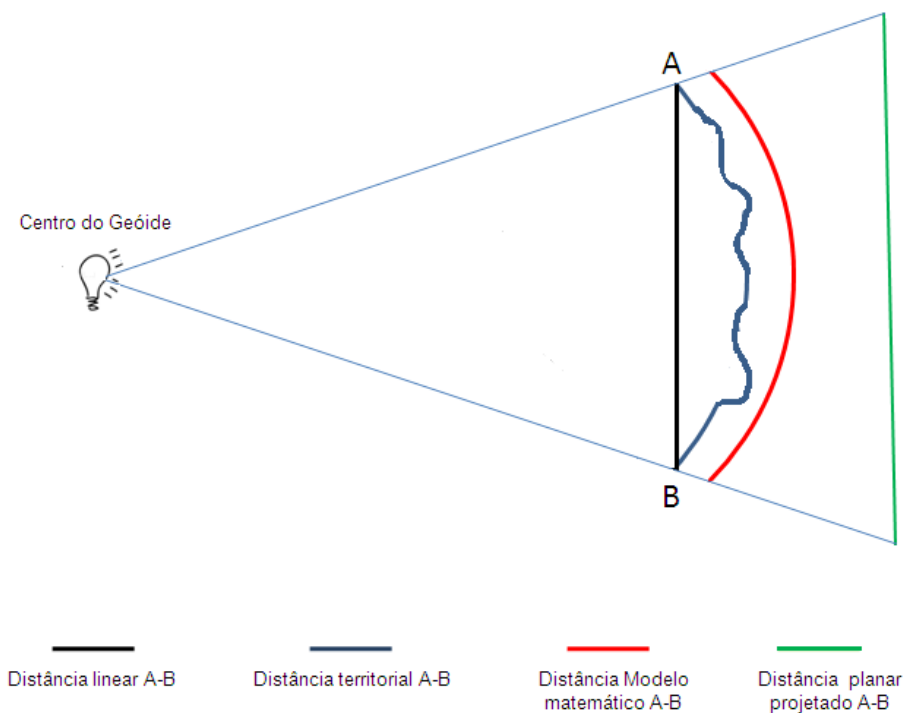


Figura 5: Esquemática do processo de projeção cartésiana

Dentro da amostragem espacial os conceitos de amostragem baseada no modelo e no design são abordados por (DE GRUIJTER; TER BRAAK, 1990).

### 1.2.1.2 Análises em Geoestatística

#### 1.2.1.2.1 Geoestatística

Quando realizamos uma amostragem de fenômenos estocásticos em duas ou mais dimensões, o interesse não se limita aos valores encontrados nos pontos de amostragem, mas sim, em toda a superfície de ocorrência das variáveis. Desta forma, os pontos amostrados são usados para determinar os prováveis valores existentes entre os pontos de amostragem por meio de um processo de interpolação.



Geoestatística pode ser definida como a aplicação formalista de modelos aleatórios no reconhecimento e estimação de fenômenos naturais (MATHERON, 1963). Para (HOHN, 1991) a geoestatística permite a descrição das continuidades espaciais, adaptando as técnicas clássicas de regressão. Permite ainda, por meio da análise da dependência espacial, estimar os valores das variáveis em estudo para os pontos não amostrados dentro do espaço, por meio de estimadores não viesados.

Tanto em biologia quanto na estatística geográfica trabalhamos com variáveis aleatórias. Variáveis aleatórias são aquelas que não podem ser determinadas pelos eventos anteriores a ela, tal qual no lançamento de dados, onde o lançamento anterior não afeta as probabilidades de ocorrência nos lançamentos posteriores. Desta forma ficam implícitas duas condições para que uma variável seja considerada uma variável aleatória: 1- repetições infinitas da amostra, e 2- que cada uma das amostras seja independente das outras.

Quando as amostras são coletadas no espaço, os valores realizados para cada uma das unidades observacionais representam na verdade, um ponto na distribuição de probabilidade de ocorrência de valores para a variável naquele local de amostragem, determinada por uma propriedade  $Z(x)$ , com média  $m$  e variância  $s^2$ . O conjunto de valores da variável aleatória  $x$  nos diferentes pontos amostrados é determinado por uma função aleatória:

Algebricamente:

$$X:\Omega\rightarrow\mathbb{R}.$$

Em que:

$X$  é a probabilidade de ocorrência da variável aleatória  $x$ ;

$\Omega$  é o espaço amostral;

$\mathbb{R}$  representa o intervalo de valores possíveis para a variável.

A variável aleatória será contínua, se existir uma função  $f(x) \geq 0$ , tal que:

$$F(x) = \int_{-\infty}^x f(t) dt, \forall x \in \mathbb{R};$$

com média:

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx,$$

e variância:

$$\text{Var}(x) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x) dx,$$

Assim, ao amostrarmos um mesmo ponto infinitas vezes no espaço podemos obter valores diferentes em cada uma das amostragens, caso contrário, será considerada uma variável aleatória discreta, definida por:

$$\{x_1, x_2, x_3, \dots\} \subseteq \mathbb{R} \mid X(w) \in \{x_1, x_2, x_3, \dots\}, \forall w \in \Omega$$

com uma função de probabilidade:

$$p(x_j) = P(x)({x_j}), \quad j=1,2,j \text{ e } p(x)=0 \text{ para } x \notin \{x_1, x_2, x_3, \dots\};$$

com média  $E(x) = \sum x_i p_i$ , e variância  $\text{Var}(x) = \sum (x_i - E(x))^2 p_i$ .

#### 1.2.1.2.2 Desagrupamento

O processo de amostragem, por diversos motivos, pode se distanciar em demasia do plano inicial, ou ainda, quando há presença de valores extremos nas observações, o resultado é uma amostragem semi-regular com agrupamento de pontos. Este tipo de situação, apesar de indesejada, pode ser muito comum na pesquisa geológica (PYRCZ; DEUTSCH, 2003).

Estes agrupamentos acabam por enviesar as frequências amostrais das regiões onde estes agrupamentos ocorrem, afetando as estimativas dos pontos não amostrados. As técnicas de desagrupamento procuram dar pesos aos valores amostrados, conforme sua distribuição espacial e, com isso, diminuir o viés causado pelo excesso/escassez de amostras. Dos métodos citados por (YAMAMOTO, 2013), o método de desagrupamento poligonal calcula polígonos de interesse ao redor dos pontos de dados, com os pesos de cada ponto correspondendo ao produto da área do polígono de dados e a área de interesse (obtida por exemplo, por meio de Diagramas de Voronoi). Este processo corrige a média e variância amostral, eliminando a influência dos agrupamentos.

### 1.2.1.2.3 Função Variograma/Semi-variograma

Devido à natureza das variáveis aleatórias espaciais, somente uma realização pode ser obtida para cada ponto de amostragem, tornando impossível o cálculo de estatísticas descritivas. Aceitando diferentes graus de estacionaridade da função aleatória do conjunto de dados amostrados, podemos assumir, por exemplo, que sendo a média constante, ela não depende do ponto amostral e, sim, da média aritmética de todas as realizações. Na hipótese de estacionaridade de segunda ordem (momento de segunda ordem estacionário), admitimos que a média das variáveis aleatórias exista e não depende do suporte dos pontos amostrais "x". Admitidos também, que a correlação entre duas variáveis aleatórias depende somente da distância espacial, "h", que as separa. É importante salientar, que a covariância neste caso, se refere à relação dos valores de uma mesma variável aleatória obtidos em pontos diferentes em uma direção definida. Em outras palavras, estima o comportamento da variável em função da distância dos pontos amostrados. De forma geral, é esperado que a variância seja pequena em pontos próximos, aumentando com o aumento de h; enquanto que a covariância seja alta para pontos próximos diminuindo com o aumento da separação espacial (h) entre os pontos amostrais (fig. 6).

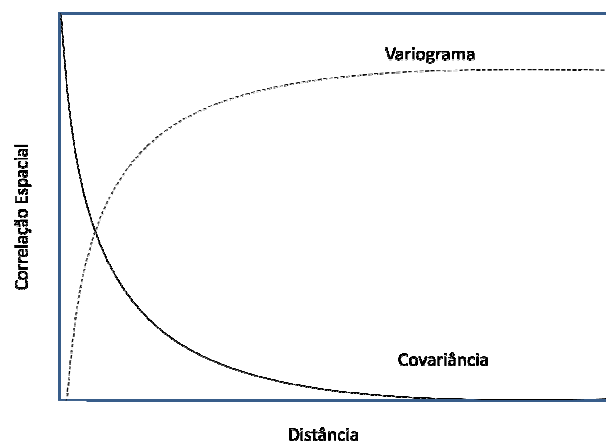


Figura 6 Relação entre a Função de Variograma e a Função de Covariância

Uma vez que a média aritmética da função aleatória  $Z(x)$  pode ser definida por:

$$m = E[Z(x)] = \frac{1}{2} \sum_{i=1}^n Z(x_i)$$

A variância associada à esta média será calculada por:

$$Var[Z(x)] = E\{[Z(x) - m]^2\}$$

com covariância:

$$C(h) = E\{[Z(x+h) - m][Z(x) - m]\}$$

e a função do semi-variograma será definida por:

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - m] - Z(x)]^2\}$$

Note que diferentemente da variância, que está associada à média da variável aleatória, a função do semi-variograma mede a variância entre dois pontos a uma distância “h” um do outro, em uma direção específica no espaço.

#### 1.2.1.2.4 Krigagem

Consiste em obter estimativas não viesadas e de mínima variância, por meio de amostras interdependentes, associadas aos valores estimados, para os pontos adjacentes (não amostrados) no tempo/espaço amostral. Assim, são gerados mapas de distribuição com superfície contínua de dados suavizados, com contrastes entre os polígonos minimizados. O principal objetivo deste processo é a modelagem do fenômeno espacial.

Estas estimativas podem ser obtidas por modelagem linear, usando os dados originais, ou com o uso de modelagem não linear sobre os dados transformados, de acordo como tipo de transformação.

#### 1.2.2 Genética de Populações

Como os genes se comportam quando na população?

É o ramo da genética que estuda o comportamento dos genes (tanto alelos, como genótipos), quando estudados no âmbito das populações. A genética de populações não se limita apenas ao estudo das características qualitativas, abrangendo o estudo de todo o conjunto gênico dos indivíduos de uma população, dos eventos que afetam as frequências genéticas, e de como esses eventos a afetam. Segundo (BEIGUELMAN, 2008), é o ramo da

Genética que visa investigar a dinâmica dos genes nas populações naturais, buscando a elucidação dos mecanismos que alteram a composição gênica destas populações.

População vem da palavra grega **Populus**, que significa povo. Segundo o Dicionário Larousse, população é um substantivo feminino, definido como os habitantes de um país, região cidade, etc. Todos os indivíduos que constituem um determinado espaço ou categoria espacial. Para (ODUM, 1959) a palavra população serve para descrever um grupo de pessoas que vivem num mesmo lugar/área, seja como um todo (qualitativamente), ou com a soma dos indivíduos num grupo (quantitativamente). Para (VECHT, 1963), população é toda estrutura biológica e, certo nível de complexidade, enquanto para (DOBZHANSKY, 1950), uma população é um conjunto de indivíduos que se reproduzem sexuadamente, compartilhando um conjunto de informação genética, e mantendo um patrimônio genético em comum.

Podemos assim, definir uma população como um conjunto de indivíduos que se reproduzem sexuadamente, compartilhando um conjunto de informações genéticas, e mantendo um patrimônio genético em comum, que pode ser representado por um grupo de indivíduos com pelo menos uma característica em comum.

População mendeliana é um conceito fundamental dentro de genética de populações. Podemos defini-la como: um grupo de indivíduos da mesma espécie que se reproduzem entre si e, por isso, apresentam propriedades comuns numa dimensão de espaço e tempo. Uma população mendeliana “ideal” deve ter as seguintes características:

1- Tamanho. Deve possuir um número infinitamente grande de indivíduos a fim de manter as proporções estatísticas.

2- Acasalamentos ao acaso (panmíticos). O cruzamento entre dois indivíduos deve acontecer ao acaso, e todas as possibilidades de cruzamentos devem ter a mesma chance de ocorrer.

3- Ausência de seleção natural, mutações, fluxo migratório e deriva genética.

Como a população mendeliana se refere, principalmente, à composição genética dos indivíduos, podemos considerar uma população de animais como uma população de genes, e podemos descrever um grupo de indivíduos por meio de suas frequências alélicas.

A genética de populações estuda a distribuição e as mudanças nas frequências alélicas/gênicas das populações, primeiramente com modelos simples, baseados na população ideal ou população mendeliana, sendo as forças naturais, que interferem nestas frequências, introduzidas de forma gradual nos modelos.

As frequências alélicas se mantêm estáveis ao longo das gerações, independente se os alelos estudados são raros, pouco ou muito frequentes dentro de uma população mendeliana. Este comportamento foi demonstrado inicialmente pelo matemático inglês Godfrey Harold Hardy na Inglaterra, contando argumentando, no vol XXVII, 49-50 da revista *Science*, críticas publicadas sobre as frequências esperadas de braquidactilia publicadas no vol 1, pag 165 do *Proceedings Royal Society of Medicine*; e pelo físico e obstetra alemão Wilhelm Weinberg em uma palestra para o encontro da *Society for the Natural History of the Fatherland in Wurttemberg* (STERN, 1943).

### **1.2.2.1 Teorema de Hardy-Weimberg**

O teorema de Hardy-Weimberg demonstra que o equilíbrio entre as frequências alélicas na população, desconsiderando a ação das forças de mutação, deriva, seleção e migração, se mantêm constante ao longo das gerações.

Até 1960 predominavam duas teorias sobre a evolução genética, uma considerava que a maior parte da variabilidade era neutra sendo fixada na população por deriva aleatória, e a segunda teoria afirmava que a seleção natural eliminava a variação, só restando aqueles alelos que apresentassem algum valor adaptativo. Se considerássemos a seleção natural isoladamente, ela só seria capaz de provocar evolução se houvesse uma grande (infinita) quantidade de variabilidade genética inicial, que poderia ser perdida muito rapidamente, em algumas gerações.

Em 1964 Kimura e Crow, formularam o modelo de alelos infinitos, onde a taxa de evolução não teria correlação com o tamanho da população quando a taxa de mutação estivesse em uma faixa intermediária. Para as mutações neutras, a taxa de mutação seria independente do tamanho da população.

Mais tarde, Kimura demonstra matematicamente, pela chamada Teoria Neutra, que pode haver evolução somente através de um equilíbrio entre as taxa de mutação e a deriva genética, sendo que a maioria das mutações teria valor adaptativo neutro e a fixação de um ou outro alelo se devia somente à deriva aleatória. Por esta teoria, todos os alelos compartilhados por uma população vêm de um ancestral comum, em um passado mais ou menos distante, dentro de um modelo retrospectivo.

$$\text{ProbCoancestralidade}(t) = (1 - 1/2Ne)^{t-1} * (1/2Ne)$$

Em que:

Ne é o tamanho da população efetiva.

Naqueles modelos em que as quatro principais forças evolutivas não estão presentes, a teoria da coalescência tem comportamento similar à curva de Poisson, onde a probabilidade de um evento ocorrer, em um determinado período de tempo, e, se eles são independentes, é determinada pela formula:

$$F(k:\lambda) = (e^{-\lambda} \lambda^k) / k!$$

Pelo principio de Hardy-Weinberg, em uma população mendeliana, sem sobreposição de gerações, sendo as frequências alélicas de um par qualquer de autossomos  $A_1$  e  $A_2$ , consideremos  $p$  a frequência para o alelo  $A_1$  e  $q$  para o alelo  $A_2$ :

$$p = \frac{2A_1A_1 + A_1A_2}{2n}$$

$$q = \frac{2A_2A_2 + A_1A_2}{2n}$$

Se:

$$p + q = 100\%$$

então:

$$q = 1 - p$$

A partir de todas as combinações possíveis de alelos nos gametas

Tipo do Acasalamento	Frequência do Acasalamento-	Genótipo da Progenie		
		A <sub>1</sub> A <sub>1</sub>	A <sub>1</sub> A <sub>2</sub>	A <sub>2</sub> A <sub>2</sub>
A <sub>1</sub> A <sub>1</sub> x A <sub>1</sub> A <sub>1</sub>	p <sup>4</sup>	p <sup>4</sup>		
A <sub>1</sub> A <sub>1</sub> x A <sub>1</sub> A <sub>2</sub>	4p <sup>3</sup> q	2p <sup>3</sup> q	2p <sup>3</sup> q	
A <sub>1</sub> A <sub>1</sub> x A <sub>2</sub> A <sub>2</sub>	2p <sup>2</sup> q <sup>2</sup>		2p <sup>2</sup> q <sup>2</sup>	
A <sub>1</sub> A <sub>2</sub> x A <sub>1</sub> A <sub>2</sub>	4p <sup>2</sup> q <sup>2</sup>	p <sup>2</sup> q <sup>2</sup>	2p <sup>2</sup> q <sup>2</sup>	p <sup>2</sup> q <sup>2</sup>
A <sub>1</sub> A <sub>2</sub> x A <sub>2</sub> A <sub>2</sub>	4pq <sup>3</sup>		2pq <sup>3</sup>	2pq <sup>3</sup>
A <sub>2</sub> A <sub>2</sub> x A <sub>2</sub> A <sub>2</sub>	q <sup>4</sup>			q <sup>4</sup>

Derivando temos:

$$f(A_1A_1)=q^2$$

$$f(A_2A_2)=p^2$$

$$f(A_1A_2)=2pq$$

Com distribuição:

$$q^2+p^2+2pq=1 \text{ ou } (p+q)^2= q^2+p^2+2pq$$

Por meio destas equações é possível demonstrar que, na população considerada, as frequências gênicas permanecem constantes ao longo das gerações, e que uma única geração em panmixia é suficiente para que as frequências gênicas da população entrem em equilíbrio. Estas frequências entrarão em equilíbrio, a partir da combinação aleatória dos gametas, na primeira geração de cruzamentos ao acaso.

Para determinar a significância dos valores obtidos para o equilíbrio de Hardy Weimberg (EHW), verificando se os valores dos desvios no HWE obtidos são derivados do acaso ou não ( $H_0$ = os marcadores não seguem o EHW;  $H_1$ = os marcadores seguem o EHW), realizamos um teste de qui-quadrado ( $\chi^2$ ) de Pearson. Considerando que, o número de graus de liberdade do teste de  $\chi^2$  é igual ao número de classes esperadas, menos o número necessário de informações da amostras para o cálculo das proporções nessas classes, e que, para realizar o teste  $\chi^2$  no EHW são utilizadas duas informações: tamanho da amostra (n) e frequência de um dos alelos, temos então 3 (genótipos/classes) -2 informações (n e frequência)= 1 grau de



liberdade. Entretanto, quando trabalhamos com grande número de marcadores ou indivíduos, a probabilidade de erro tipo I fica aumentada (WIGGINTON; CUTLER; ABECASIS, 2005). Nessas amostras, existe uma maior probabilidade de baixa eficiência de genotipagem para alguns marcadores, assim como de presença de alelos raros. Estes alelos assumem uma distribuição discreta, desrespeitando a pressuposição de uma distribuição de qui-quadrado (ROHLFS; WEIR, 2008), devido à insuficiência da amostra para todos os marcadores ou a sua baixa frequência na população. Nesse sentido, vários testes para HWE são descritos na literatura (ver (EMIGH, 1980), variando em relação ao seu poder e exatidão nas suas aproximações, dentre eles o teste exato de Fischer. (ZAPATA; ALVAREZ, 1997), citam diversos autores que consideraram o teste exato de Fischer extremamente conservador e com baixo poder, como teste para tabelas independentes 2x2, ao mesmo tempo em que (TOCHER, 1950) o considera mais uniformemente poderoso quando valores marginais são amostrados.

Se as populações naturais se comportassem tal como a população mendeliana ideal, elas estariam sujeitas a uma inércia evolutiva, uma vez que as frequências gênicas não sofreriam mudanças. Felizmente, as populações estão sujeitas a fatores evolutivos capazes de alterar as frequências genéticas, seja pela introdução de novos alelos (mutação, migração), eliminação de alelos (deriva e seleção natural), ou seja aumentando a frequência de homozigotos (endocruzamento/subdivisão).

De posse das frequências alélicas, podemos estimar a probabilidade condicional de um alelo na população e, assim, estimar a frequência de portadores de um determinado gene recessivo entre os indivíduos de fenótipo dominante, por meio de:

$$P(A_1A_2 | A_1_) = \frac{2pq}{P^2 + 2pq}$$

e, com isso, estimar a probabilidade de um acasalamento produzir descendentes homozigoto recessivos para determinado gene de interesse.

## 1.2.2.2 Alterações nas frequências gênicas

### 1.2.2.2.1 Introduzindo alelos

#### 1.2.2.2.1.1 Mutação

Mutação é toda alteração aleatória e repentina na constituição genética de um indivíduo. Pode ocorrer tanto como uma alteração cromossômica ou como uma alteração gênica. Quando ocorrem nas células somáticas, em indivíduos de reprodução exclusivamente sexuada, não são transmitidas aos descendentes. Porém, quando ocorrem nas células germinativas serão transmitidas a descendência.

As mutações gênicas ocorrem durante a fase S do ciclo celular, por alteração da sequência de uma ou mais bases, não sendo corrigidas adequadamente pelos mecanismos de reparo da replicação.

Essas alterações na sequência gênica podem ter efeito na síntese protéica, alterando a sequência de aminoácidos na proteína produzida, modificando a estrutura tridimensional da proteína, alterando ou não, a sua função no organismo. Devido degeneração do código genético, muitas das mutações são silenciosas, isto é, uma alteração da sequência de nucleotídeos não altera a sequência de aminoácidos presentes no produto da transcrição, não ocasionando alteração no fenótipo dos indivíduos.

O fenótipo expressado pelo indivíduo, portador de uma mutação, é fator determinante na sua sobrevivência. Caso o fenótipo do indivíduo mutante seja uma doença ou uma má-formação que inviabilize a vida, provavelmente, ela será perdida na mesma geração pela morte do indivíduo mutante ou simplesmente pelo seu baixo *fitness*. Por outro lado, se houver alguma vantagem no fenótipo do indivíduo mutante dentro do ambiente onde ele está inserido, a mutação será rapidamente fixada pela população. Entretanto, segundo a teoria neutra formulada por Kimura em 1969, a maioria das mutações não tem efeito sobre a adaptação e, somente uma pequena parte delas seria vantajosa, sendo fixada na população e outra pequena parte seria deletéria e sendo eliminada da população também por força da seleção natural.

De acordo com a teoria neutra, a probabilidade de uma mutação ser fixada na população depende, principalmente, do tamanho da população, sofrendo influência direta da deriva aleatória e da frequência de novas mutações na população.

#### **1.2.2.1.2 Migração/Emigração**

É um processo que atua como facilitador do processo evolutivo, pois aumenta o conjunto gênico sujeito a seleção, e, diminui ou retarda os efeitos da deriva aleatória, principalmente em populações pequenas e/ou isoladas (KIRK; FREELAND, 2011). Caracteriza-se pela movimentação (introdução ou saída) de indivíduos entre as populações de uma mesma espécie (fig. 7), e conseqüentemente genótipos/alelos, causando uma alteração no tamanho da população efetiva ( $N_e$ ). Quando a população alvo recebe novos indivíduos, recebe junto com eles novos alelos, e com isso, as frequências gênicas são alteradas. Da mesma forma, quando ocorre emigração há perda de alelos e as frequências gênicas são alteradas. Porém estas alterações nas frequências alélicas, em decorrência da migração, somente irão ocorrer se as frequências gênicas nas sub-populações forem diferentes, sendo mais suscetíveis de ocorrer se as sub-populações estiverem isoladas umas das outras ao longo do tempo. Os eventos de migração podem atuar retardando o processo de especiação, por diminuir as diferenças nas frequências alélicas entre as sub-populações (NORTH *et al.*, 2011).

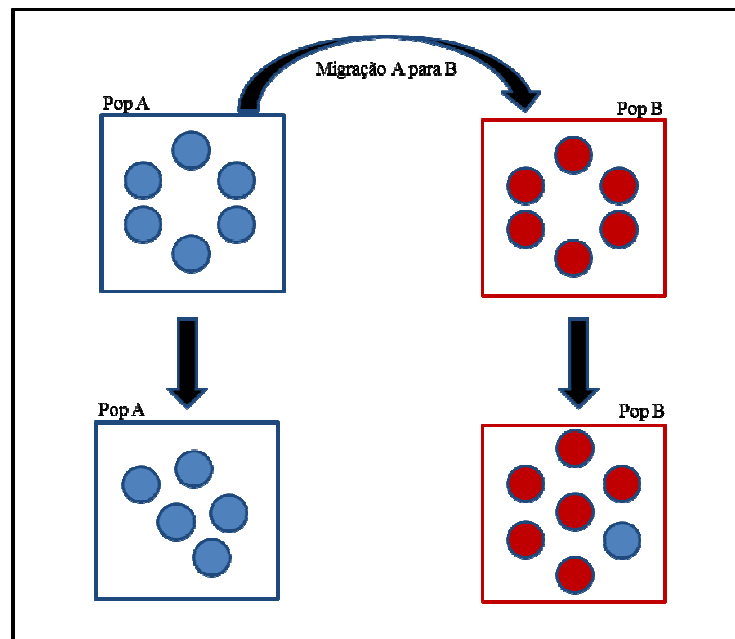


Figura 7: Representação esquemática do fenômeno de Migração

#### 1.2.2.2.2 Eliminação de alelos

##### 1.2.2.2.2.1 Deriva Aleatória

Pode ser definida como o processo que leva a mudanças nas frequências gênicas de uma população, derivado do acaso, a cada geração. Estas mudanças levam tanto a fixação como a eliminação de alelos por puro acaso. Em outras palavras, é o desvio na probabilidade de um alelo estar presente na próxima geração. Quanto maior for a amostra, menores serão os desvios, e vice-versa.

O processo de deriva poderia ser comparado com o processo de sorteio aleatório com reposição onde, cada sorteio corresponderia aos acasalamentos aleatórios e, as chances de um número ser sorteado, às chances de um alelo estar presente no gameta que irá originar um novo indivíduo. Como o número de alelos que compõe a amostra é limitado, alguns alelos não estarão presentes no primeiro gameta, mas no próximo sorteio todos os alelos voltam a ter a mesma chance, e assim por diante. Acontece que, o número de sorteios é limitado (descendentes por geração) e alguns alelos

serão sorteados nenhuma vez, enquanto outros, serão sorteados todas às vezes, por puro acaso.

A deriva genética aleatória (fig. 8) é influenciada pelo tamanho efetivo da população fundadora e pelos eventos de gargalo de garrafa.

A população fundadora representa todo o conjunto genético de uma população, quanto menor ela for, menor o pool gênico inicial, maiores os desvios nas probabilidades dos alelos estarem presentes nas gerações seguintes. Entretanto, se esta população for relativamente grande estes desvios serão menores e os efeitos da deriva serão diluídos ao longo das gerações.

Os gargalos de garrafa (*bottleneck*) consistem em uma drástica diminuição no tamanho da população, causando uma diminuição no número de genótipos (NEI; MARUYAMA; CHAKRABORTY, 1975), com perda de alelos, principalmente aqueles alelos raros (NEI, 2005). Por produzir um rápido declínio no pool gênico da população afetada, os gargalos de garrafa implicam em perda de variabilidade genética total na população afetada. A taxa de heterozigotos diminui drasticamente até um mínimo, nas gerações imediatamente posteriores ao gargalos de garrafa, quando então começa a aumentar vagarosamente, acompanhando o aumento da população, segundo a taxa de mutação por geração esperada para a espécie (NEI; MARUYAMA; CHAKRABORTY, 1975).

Eventos do tipo *bottleneck* podem ser provocados por desastres ambientais, como um furacão, terremoto, erupção vulcânica ou provocados pelo homem, como a caça de uma espécie até quase sua extinção, ou pela destruição do habitat, resultando na morte de grande número de indivíduos. Uma vez que o tamanho efetivo da população e o número de alelos diminuem, a população afetada pelo *bottleneck* sofre um nível mais elevado de deriva (LYNCH; CONERY; BURGER, 1995; WRIGHT, 1931, 1978).

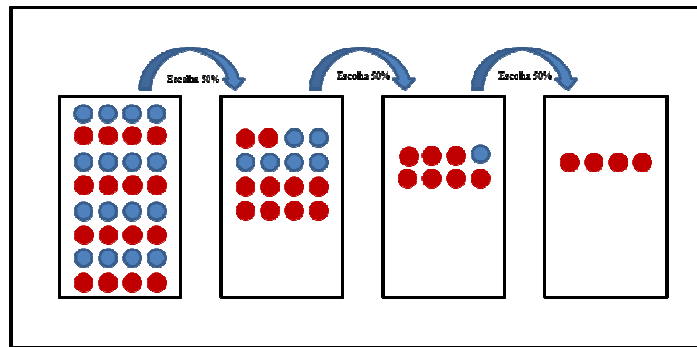


Figura 8: Representação esquemática de deriva.

#### 1.2.2.2.2 Efeito de Wahlund

Quando analisamos duas subpopulações isoladas com frequências alélicas diferentes, como se fossem uma única população, esperamos que frequências médias dos genótipos não correspondam às frequências médias dos alelos, ou seja, a frequência média do homocigoto não é igual ao quadrado da frequência média dos alelos (FREIRE-MAIA, 1974). Assim o efeito Wahlund é um aumento no número de heterocigotos observados em relação ao número esperado pelo EHW, quando analisados isolados populacionais em conjunto.

#### 1.2.2.2.3 Seleção natural

A seleção natural promove alterações direcionais nas frequências alélicas/genotípicas das populações, favorecendo os fenótipos mais adaptados às condições ecológicas do habitat populacional. Ao atuar nos fenótipos mais adaptados, a seleção natural está na verdade, selecionando os alelos que produziram estes fenótipos e eliminando os alelos desfavoráveis (desconsiderando-se os fenótipos não hereditários).

A expressão mais bem adaptada refere-se à maior probabilidade de, em um determinado ambiente, um determinado indivíduo deixar descendentes viáveis. Os indivíduos mais bem adaptados (maior *fitness*) a um ambiente têm chance maior de sobreviver e de deixar descendentes.

O coeficiente de adaptação ou o *fitness* individual pode ser obtido por:

$$w=1-S$$

Em que:

$w=fitness$

S= coeficiente de seleção natural

Seleção é maior que a deriva quando:

$$S > 1/Ne$$

Então:

$$Fixação = 2S$$

$$Tempo\ de\ fixação = \log(sNe)/S$$

O processo de seleção natural depende, principalmente de três variáveis: *fitness* ou valor adaptativo, sobrevivência e sucesso reprodutivo.

Por valor adaptativo, entendemos a presença de uma característica transmitida à prole que aumente as chances de sobrevivência e reprodução de um organismo em relação à condição alternativa dessa mesma característica. Portanto, teoricamente considerado como uma medida em longo prazo do desempenho relativo (CHARLESWORTH, 1994).

Por selecionar os genótipos de maior valor adaptativo, diminuindo o número de genótipos dentro das populações, a seleção natural tende a diminuir a variabilidade genética. Desse modo, quanto mais intensa for a seleção natural sobre uma determinada população, menor será a sua variabilidade, pois apenas alguns genótipos serão selecionados. Mesmo em ambientes bem estabilizados a seleção natural continua a atuar sobre as frequências alélicas.

#### 1.2.2.2.4 Endocruzamento

Quando indivíduos mais aparentados que a média da população se acasalam, a frequência dos alelos é alterada, aumentando a frequência de genótipos homozigotos com o passar das gerações, sem que as frequências alélicas sejam alteradas. Alelos deletérios recessivos tornam-se mais frequentes reduzindo o *fitness* da população.

Endocruzamentos podem acontecer quando a população fundadora é muito pequena e todos os indivíduos compartilham os mesmos ancestrais, como em rebanhos iniciados por poucos animais (algumas fêmeas e um único macho), ou em decorrência de eventos de gargalo de garrafa. É usado pela agroindústria para a formação de linhagens puras, que posteriormente serão usadas nos cruzamentos industriais. Por meio de endocruzamentos, características de interesse são fixadas na população. Outro importante uso da consanguinidade são os testes de progênie, quando são realizados retrocruzamentos que visam desde a determinação do valor genéticos das matrizes, estimação dos parâmetros genéticos até a conservação de populações vegetais.

#### **1.2.2.2.5 Subdivisão**

Quando duas populações de uma mesma espécie estão isoladas por barreiras naturais ou não, sendo portanto, impedidas de reproduzir entre si. Em função desse isolamento geográfico não há fluxo genético entre as subpopulações, os efeitos da deriva genética são mais intensos e estas populações acabam por fixar alelos distintos, levando a uma homogeneidade genética dos indivíduos dentro de cada uma das subpopulações.

A fragmentação do ambiente, principal causa de subdivisão populacional, é o processo pelo qual um ambiente é subdividido em partes ou manchas, mais ou menos conectadas, podendo ter ocorrência natural ou por ação antropomórfica. Os fragmentos podem apresentar diferentes graus de desafio desde a obtenção de abrigo/alimento, ou a presença de predadores, que influenciam diretamente o *fitness* individual (CERQUEIRA *et al.*, 2003), levando a estruturação da população e com o passar das gerações, dependendo da pressão exercida pela seleção natural, pode levar a especiação.

Por vezes, por falta de conhecimento prévio das barreiras naturais, analisamos duas populações isoladas (mesmo que estas estejam em HWE) em conjunto, observando a ocorrência do efeito Wahlund. Nessa situação a Heterozigose Observada será menor que Heterozigose Esperada, embora a



população se encontre em HWE. Nessa situação o pesquisador é levado a acreditar na existência de conectividade entre as duas subpopulações quando na verdade elas estão isoladas.

#### **1.2.2.2.6 Seleção Contra o Híbrido.**

Considerando duas populações (A e B), ao migrar de A para B os indivíduos levam consigo genes favoráveis (no ambiente A) para o ambiente B. Se aqueles genótipos favoráveis ao meio A forem deletérios em B haverá seleção negativa contra os híbridos, e os alelos de A se perderam com o tempo, assim como, os aqueles alelos em desequilíbrio de ligação (haplótipos).

#### **1.2.2.3 Estatísticas de Wright**

A variabilidade genética em uma população é influenciada pelo modo de reprodução, sistema de acasalamento, tamanho da população, distribuição geográfica e fluxo gênico (HAMRICK, 1982), além de ser estruturada no tempo e no espaço.

A estruturação de uma população pode ser medida pelo uso da heterozigose esperada no HWE. Na década de 1920, Sewall Wright sugeriu o uso do coeficiente de endogamia para estimar os diferentes níveis de estruturação de uma população

Os níveis de hierarquia populacional avaliados pelas estatísticas F exploram as relações entre  $H_E$  e  $H_O$  dos indivíduos, das subpopulações e da população total.

A partir da relação entre a Heterozigose Observada na população e a Esperada pelo HWE:

$$F=1-H_O/nH_E$$

estimamos a probabilidade de que dois alelos, de um mesmo locus qualquer, de um indivíduo sejam idênticos por descendência (IBD).

Ao utilizarmos a média de Heterozigose observada no indivíduo e a heterozigose da subpopulação, obtemos uma estimativa da heterozigose individual em relação aos outros indivíduos que compõe a subpopulação ( $F_{IS}$ ), ou seja, o quão homocigoto o indivíduo é em relação a população de origem.

Da mesma forma, podemos utilizar a média de Heterozigose observada do indivíduo e a heterozigose total da população ( $F_{IT}$ ), obtendo assim, a estimativa de quão homozigoto o indivíduo é em relação à população total.

E por último, quando usamos a média de Heterozigose observada na subpopulação em relação à Heterozigose Esperada na população total ( $F_{ST}$ ), obtemos uma medida da homozigose do rebanho em relação à população total.

A relação entre os níveis de estruturação (fig. 9) é dada pela equação:

$$(1-F_{IS})(1-F_{ST})=(1-F_{IT}) \quad (3)$$

O índice de fixação pode ser expresso como o produto entre a variância genética subpopulacional e a variância genética total (4)

$$F_{ST}=\sigma^2_s/\sigma^2_t \quad (4)$$

Que pode ser lido como a fração da variância genética que pode ser explicada pela estruturação da população. Em outras palavras, por meio das estatísticas F se mede a probabilidade de dois alelos escolhidos ao acaso serem diferentes por descendência.

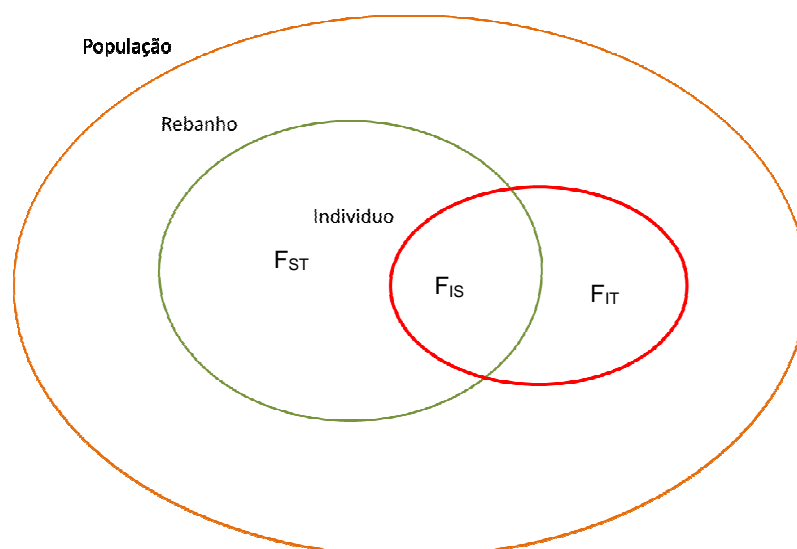


Figura 9: Relação entre as estatísticas F de Wright

### 1.3 HIPÓTESES E OBJETIVOS

#### **Hipóteses**

a- As variáveis geográficas influenciam na frequência dos alelos dos marcadores genéticos, levando à estruturações na população de suínos brasileiros localmente adaptados.

b- As variáveis ambientais influenciam na seleção e fixação dos alelos dos marcadores genéticos, permitindo identificar assinaturas de seleção com auxílio do Porcine SNP60 BeadChip.

#### **Objetivos**

O objetivo da presente é identificar a distribuição da variabilidade genética das principais raças e grupos genéticos (comerciais e localmente adaptados) de suínos do Brasil e correlacionar a fatores ambientais, identificando marcas de seleção no genoma destes animais. Identificar padrões de distribuição geográfica das marcas de seleção de forma a torná-las úteis nos programas de melhoramento genético suíno.

Objetivos específicos:

(1) Avaliar a estruturação da população de suínos brasileiros nas principais variáveis ambientais brasileiras;

(2) Identificar assinaturas de seleção no genoma dos suínos localmente adaptados ao ambiente brasileiro.

## **CAPÍTULO II<sup>1</sup>**

“I turn my back to the wind  
To catch my breath,  
Before I start off again...”

*(Neil Peart)*

**Population structure of Brazilian locally adapted swine genetic groups using PorcineSNP60 BeadChip<sup>1</sup>**

**Cesconeto, R. J.<sup>1\*</sup> ; Mc Manus, C.<sup>2</sup> ; Paiva, S.R.<sup>3</sup> ; Joost, S.<sup>4</sup> ; Silva, I.C.M.<sup>5</sup> ; Braccini, J.<sup>6</sup>**

\* CP 19, Zip: 88870-000 Orleans SC Brazil. email: cesconeto@hotmail.com

1-Animal Science PhD Student at UFRGS, 2- Professor at UNB, 3 International Relationship Secretary of EMBRAPA, 4 Professor at EPFL, 5- Post Doc at Animal Science Department of UFRGS, 6- Professor at-UFRGS

**Key Words:** Brazilian pig, Molecular Markers, Population Genetics, Environment structure

**Abstract**

In order to understand the genetic population structure of Brazilian locally adapted swine genetic groups within different Brazilian environments, we genotyped DNA samples from 191 animals of 12 Brazilian locally adapted swine genetic groups (Baé, Canastra, Canastrão, Carucho, Casco de Mula, Crioulo, Mammelado, Marajó, Mestiço, Monteiro, Nilo, Rabo de Peixe) 2 Brazilian locally adapted swine breeds and 4 commercial breeds (Duroc, Landrace, Large White, Pietran) and a group of crossbred animals. After applying quality control filters, analyses of molecular variance, Wright's F-statistics within genetic group, state, ecoregion, hydrological basins, biomes, soil types and vegetation types were calculated. Allelic frequencies were used to estimate the number of genetic populations. The main source of variation was the individual in all analyzed levels of structure. Conservation strategies need to maintain the highest possible individual variability for the formation of conservation and management units. Population genetic structure was detected within genetic groups, ecoregion, hydrological basins, biomes, soil types and vegetation types. The swine genetic groups Monteiro, and Marajó revealed very high levels of

inbreeding. Brazilian locally adapted swine breeds have high genetic variability and evident genetic structure in the analyzed environmental levels.

### **Introduction**

Many locally adapted swine genetic groups arose from animals introduced by the colonizers (Mariante & Cavalcante 2006), and they were widely used up to the middle of the last century. As occurred in Asiatic countries (Kim *et al.* 2005), local breeds that were not economically viable for industry, as well as changes in the Brazilian market during the 70's, led to the importation of European and North-American swine breeds, focusing on pork production. In Europe, in the 1980s, the consumer interest for meat derivatives produced from local breeds led to an increase in number and size of these herds (Gama *et al.* 2013). The increase of Brazilian consumer purchasing power in the last decades awoke interest for different products, thus stimulating various initiatives based on national swine genetic groups (Niederle & Gelain 2013). However, the number of registers of these animals is close to zero (ABCS 2014) and most national breeds are found in subsistence herds which supply lard and pork that are differentiated by due to their flavor and tenderness.

The biologic success of swine in Brazilian territory was shown by the dramatic phenotypic changes that occurred in the swine population since its introduction. Their rapid adaptation to the Brazilian environment was probably not caused by evolutionary processes, such as speciation or changes in protein coding regions, but by the complex process of drift, natural selection or limited genetic flow among geographic regions or environments. Evidence of separation of the swine breeds inside the different geographic regions helps

understand adaptive processes (McManus *et al.* 2013). Understanding the populational dynamics throughout the Brazilian territory by means a populational genetic study could help to preserve this important livestock and contribute to studies on breed differentiation and help conservation programs. Although there are many papers studying the evolution of swine breeds (Groenen *et al.* 2012), or the population structure of swine herds worldwide (Vicente *et al.* 2008; Burgos-Paz *et al.* 2012; Druml *et al.* 2012; Gama *et al.* 2013) as well as studies concerning Brazilian adapted swine genetic groups (Silva-Filha *et al.* 2005; Sollero *et al.* 2009), there are few works concerning dense panels of markers with cover over the Brazilian territory.

Therefore, understanding if the different Brazilian environments could influence locally adapted Brazilian swine genetic groups causing subdivision, limiting genetic flow among herds is important to develop conservation strategies for Brazilian swine genetic resources, and quantify the environmental importance on genetic composition of swine herds. So, the hypothesis of this paper is that there is a genetic structure in the locally adapted swine population in Brazil due to different environments over the Brazilian territory.

### **Material and methods**

DNA samples from 191 animals (fig 1), were selected from Embrapa's locally adapted and commercial swine breed DNA and tissue database (Table 1) of EMBRAPA Recursos Genéticos e Biotecnologia in Brasília, DF, Brazil. The sampling was based on maximum genetic and environmental variability as possible with at least one animal per genetic group and sampling point. The genetic groups sampled included 4 globally

commercial swine breeds (Landrace, Large White, Duroc and Pietran), 2 recognized locally adapted swine breeds (Piau and Moura) and 11 genetic groups not recognized as locally adapted swine breeds (Baé, Canastra, Canastrão, Caruncho, Casco de Burro, Crioulo, Mammelado, Marajó, Monteiro, Nilo, and Rabo de Peixe), as well as 1 group of crossbred animals (Mestiço).

The samples were genotyped by means of Illumina PorcineSNP60 BeadChip markers (Ramos *et al.* 2009) with 61565 SPNs. The raw database was submitted to quality control in *SNP & Variation Suite* (Golden Helix, Inc., Bozeman, MT, USA 2015), where non-mapped SNPs and SNPs located on X or Y chromosomes were removed. Additional filters with a Call Rate equal to 85% for samples and 95% for SNPs were applied and also filters for Minor Allele Frequency (MAF) at 0.05, and for Missing at 0.05. After quality control, using Plink v1.9 (Purcell, 2015), the 181 animals (95%) and 39960 markers (64.91%), that remained in the database were submitted to a Hardy-Weinberg Equilibrium (HWE) pruning at 0.001 significant threshold and Linkage Disequilibrium (LD) pruning using a window size=50, window increment=5 and  $r^2$  threshold=0.30, which selected 19,422 SNP markers for the 181 samples.

Wright's F-statistics were calculated to determine the frequencies of heterozygotes and identify the levels of swine population structuring into the categories of breed, state, hydrological basin, type of soil, biome, vegetation and ecoregions (Table 2) using ARLEQUIN 3.5.2 (Excoffier *et al.* 2005). Identity by descent (IBD) was estimated using *SNP & Variation Suite* (Golden Helix, Inc., Bozeman, MT, USA 2015).



To reduce the genetic variation of the allele frequencies from the markers for each genetic group, showing the genetic similarity among animal from different groups, a Principal Components Analysis (PCA) was performed, concentrating allele frequency variation in orthogonal axes by means of the statistic package for R using Adegnet v 1.3 (Jombart 2008) and plotted using Graphics 3D procedure (ProcG3D) of SAS 9.3 (SAS Institute Inc. 2011). The analysis of hierarchical variance was performed to subdivide the allele variation in its intra-individual, intra-class (breed, state) and intra-group (commercial or autochthones) components using the module AMOVA in ARLEQUIN 3.5.2 (Excoffier *et al.* 2005).

To determine the minimum number of clusters, thus establishing the least and most divergent groups inside the naturalized swine population, the genetic structure matrix was submitted to Cluster Procedure (ProcCLUSTER) of SAS 9.3 (SAS Institute Inc. 2011) with the minimum Ward variance as the grouping methodology and comparing the Pseudo-F and Pseudo-F square values. The composition and number of populations, determinate by individual genetic frequencies of the loci, were estimated through Admixture software (Alexander *et al.* 2009)

### **Results**

From a total of 61,565 markers in the raw database, 39960 markers (65%) remained after the filtering for MAF and Call rate. These were distributed in autosomal chromosomes with 2,217 SNPs per chromosome and an average distance of 1 SNP to each 61.631 kb. After filtering for HWE and LD we had 19,422 markers in the database.

The pairwise Wright's F-statistics analysis of the genetic groups (Table 3) revealed that the  $F_{ST}$  varied from -0.008 between Baé and Canastra to 0.248 between Monteiro and Duroc. When we calculated the same statistics among the State populations, ignoring the genetic group to which the samples belonged (Table 4), we observed values of  $F_{ST}$  varying from 0.01 between the states of Santa Catarina and Paraíba, to 0.18 between the states of Mato Grosso and Paraíba.

Among the vegetation types (Table S1), the Zones of Ecological Tension were the only ones that presented some kind of structure related to other types of vegetation ( $F_{ST} = 0.051$  to  $0.159$ ), even though they represent only 1.26% of samples. Zones of Ecological Tension are areas located at the limit of two different types of ecosystem that are subject to human influence or not. Neossolo quartzarenico and espodosolo ferrocárbico soils, with 3.31% and 5.52% of individuals respectively sampled, were among the types of soil (Table S2), that showed some kind of population structure ( $F_{ST} = 0.01$  to  $0.166$ ). Among the different ecoregions (Table S3) we found a minor structure in every comparison. The samples from Pantanal specifically (11.05% of total samples) revealed a moderate structuring in pairwise comparison ( $F_{ST} = 0.057$  to  $0.185$ ) with the other ecoregions. The Paraguai hydrographical basin (11.05% of samples), "Atlântico Leste" hydrographical basin (1.61 % of samples), "Atlântico Sudeste" hydrographical basin (1.61% of samples) and "Amazônica" hydrographical basin (15.47% of samples) revealed a moderate degree of structuring among themselves (Table S4).

$F_{IS}$  values and pseudo  $F$  values (Table 5) ranged from -0.08 to 0.215 (inside the genetic groups) and from -0.137 to 0.155 (inside the States) which could indicate that inbreeding levels of locally adapted genetic groups vary considerably inside subpopulations from an excess of heterozygotes in some groups to inbreeding in others. From the locally adapted genetic groups, Bae, Caruncho, Nilo, and Piau revealed the highest  $F_{IS}$  values, which means low levels of genetic differentiation from individuals within each genetic group. Commercial breeds presented low levels of inbreeding, which differs from what was expected, since these breeds undergo intense selection and standardization. Inside the States, the highest inbreeding levels were observed in Santa Catarina,, Pernambuco, Bahia, Goiais and Minas Gerais (Table 5), while the animals from Mato Grosso do Sul, Pará and Paraíba revealed negative values of  $F_{IS}$ . Pseudo  $F$  values from permutations in  $F_{IS}$  analyses indicated that some  $F_{IS}$  values were not the real values for these populations.

It was not possible to completely explain the genetic variability among the genetic groups by means of the principal components analysis (Fig. 2). The animals from Monteiro genetic group to the right of the graph (positive segment of the first axis) and from Marajóara genetic groups on the top to the center (positive segment of the second axis) showed highest divergence from the other genetic groups, while the remaining animals formed a third group over the first orthogonal axis (Fig. 2). Every swine genetic group presented some degree of relation to the third axis (Fig. 2), with the exception of the Monteiro, but this is not enough to group them according to their respective genetic groups.

The analysis of groupings in the  $F_{ST}$  matrix - genetic structuring (Fig. 3) revealed the formation of 10 groups. The genetic structures of the commercial breeds (Pietran, Large White and Landrace) are more similar among themselves than in comparison to the other genetic groups. On the other hand, Bae, Nilo, Canastra, Casco de Burro and Piau have a similar genetic structure in relation to the other genetic groups. The analysis of animal groupings using the inbreeding coefficient demonstrated little similarity among the individuals, since the Pseudo-F values indicated 150 groups established from 181 samples.

The analysis of hierarchical variance (AMOVA) of the sampling locations inside the geopolitical regions evidenced that the allele variation among the regions was only 0.77%. Among the states, inside the regions, the variation was 5.17%, and among individuals the allele variation was 94.07% of the detected variance. A similar data structure was found in the AMOVA among individuals inside the genetic groups. In this analysis, the individual variance reached 86.64% of allele variance, while the variances of individuals inside the genetic groups and among the genetic groups was 6.53% and 6.11% respectively. The Identity by Descent (IBD) analyses among populations was low, but inside some groups, such as Monteiro and Marajóara, the IBD coefficient was high (Fig 4).

It was possible to determine the theoretical number of ideal populations using the genetic frequencies of the loci to infer the influence of the genetic groups in the composition and number of populations. The individuals were adequately allocated inside their original population by means of a  $k = 18$ .

The lower cross-validation error was found with a  $k=8$ . Using this approach, all commercial animals were allocated inside a single population. However, the same was not observed for the other breeds that had their representatives allocated in various clusters (fig 5).

### **Discussion**

High positive  $F_{ST}$  values indicated a subdivision of the locally adapted swine population among breeds (Table 3) and among States (Table 4).  $F_{ST}$  pairwise values among breeds showed a similar pattern to those found by (Burgos-Paz *et al.* 2012) in Colombian swine populations and to that found between the French and the Spanish swine populations according to Boitard *et al.* (2010). This type of geographic and breed structuring is similar to that found in Asia according to (Kim *et al.* 2005) who observed a high level of homozygotes among Korean animals, but great divergence from those from southern China. These lightly negative  $F_{ST}$  found among genetic groups Baé and Canastra could indicate a deficit of heterozygotes, as well as lower fixation of different alleles.

$F_{ST}$  values among populations from Santa Catarina and Paraíba (Table 4) was low and was not expected because the state of Paraíba is geographically distant from Santa Catarina. These values may occur due to importation of animals and other facts, such as temperature range, which merits investigation.

Animals from the Monteiro, a genetic group found principally in the states of Mato Grosso do Sul and Mato Grosso (Fig 1), indicated the highest

$F_{ST}$  values in relation to the other breeds. High  $F_{ST}$  values occurs when a group or a breed are geographically isolated when compared to the other breeds. Burgos-Paz *et al.* (2012) found high values for Piau in their work, different from us. However, they sampled Piau animals only at a simple site, what could be interpreted as an isolated herd, when there was a connectivity among different Piau herds or genetic flow with another genetic groups.

The pairwise  $F_{ST}$  values among commercial breeds and locally adapted genetic groups suggests a minor number of shared alleles among the commercial and adapted breeds. Lower  $F_{ST}$  values were observed among the genetic groups with a broader geographical distribution, such as Piau, when compared to those within similar a regional distribution, such as Rabo de Peixe. This behavior is expected since small groups with regional distribution have a smaller probability to interchange genes or crossbreed with other groups than that with a border distribution.

The detected genetic structuring between the genetic groups and the States means that there is a genetic variability within these levels which, after studies for specific genetic and phenotypic characterization, could be used to set exclusive genetic patterns for each population and thus enable the genetic certification of swine derivatives with regional and breed certification by farmer associations or other organizations. Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI) are goods with high added value that could stimulate Brazilian locally adapted swine breeds farmers, improving the profitability of these establishments, raising the number of animals, and helping conservation programs. The structuring found in the different types of soil

(Table S2), ecoregions (Table S3) (fig 1B) and hydrographical basins (Table S4) (fig 1C) may occur due to divergences in mineral, humidity and rainfall availability necessary elements for the animals and vegetation that is used as feed.

The high values of  $F_{IS}$  (Table 5) found in the sampled population diverge from those found by Gama *et al.* (2013), working with autochthone Portuguese and Spanish swine. These authors state that the effort of farmers to mate unrelated animals is the cause of low  $F_{IS}$  values. The high indexes of inbreeding that were found in this study may arise due to the founder effect, which according to Pham *et al.* (2014) assign a small number of heterozygotes found in Mong Cai animals in relation to other Vietnamese autochthone breeds to the fact of the first group came from a mating station. Favero & Figueiredo (2009) mention the introduction of new animals into herds as the only way of improvement adopted by farmers in the 70s when a decrease in the number of Brazilian locally adapted swine herds was observed.

The grouping analysis indicates that the genetic structuring level of the commercial breeds is very similar, since they were allocated into contiguous clusters, in spite of the separation interval among clusters and of different  $F_{ST}$  values (Fig. 3). This separation is in accordance with the phenotypic differences of the commercial breeds as found by (McManus *et al.* 2010). The number of clusters suggested by Pseudo F values ( $n=10$ ) indicates that the genetic variation of the breeds is specific to each one, with exception for Baé, Nilo, Canastra, Casco de Burro and Piau, which were grouped in a single cluster to the left side of the Figure 3, away from all commercial breeds and which means

that the factors that affect the genetic structure are shared among them. The principal component analysis agreed with the other analyses when it indicated that Monteiro and Marajó breeds are genetically distant from the others that presented high genetic variability among and within breeds.

The number of groups suggested by ADMIXTURE (Fig. 5) reveals that the swine genetic groups share alleles, possibly caused by interracial mating, common in subsistence breeding as found in locally adapted swine genetic groups in Brazil, leading to allele introgression from external populations. Similar results were found by (Sollero *et al.* 2009) working with microsatellite markers and Brazilian locally adapted swine breeds. Separating the animals into two groups ( $k=2$ ), it was only possible to separate animals of the Monteiro breed from other populations. This may be caused by interracial mating or by the introduction of alleles from commercial breeds into locally adapted populations, which may not have happened with the Monteiro genetic group.

The high inbreeding level in individuals revealed with the Identity by Descent analyses from the Monteiro and Marajó breeds reveals that these populations (Fig. 4) have their genetic variability endangered, despite being numerous and having many sampled herds (Table 1). High levels of inbreeding are caused by small founder populations, mating with close relatives and/or bottleneck events. In the case of the Monteiro, a breed found only in the Pantanal (Mato Grosso and Mato Grosso do Sul), a small founder population may be explained by the history of its formation, which comes from animals abandoned by settlers during the Guerra da Tríplice Aliança in 1864 (Cavalcanti



1985; Mauro & Pereira 2004) and breeding captivity (Sollero *et al.* 2009). A bottleneck may be due to the hunting of these animals by farmers (Desbiez *et al.* 2009). (Mauro 2002) emphasizes the large oscillation of herds of pigs in virtue of the survival of piglets, because of the pressure through hunting depending on sub-region and period of the year. Large felines from this region may also hunt these herds (Cavalcanti & Gese 2010). The highly disseminated practice of capture of young males, their castration and release for late hunting (Mauro *et al.* 2004) reduces the effective population of males. The degree of inbreeding associated with high positive  $F_{ST}$  values for the Monteiro population, at an interracial and regional level, may increase inbreeding depression even more in a few generations, which causes serious consequences for the breed. In the case of Marajó animals, which are raised exclusively on the island of Marajóara, the geographic isolation and small founder population may be the origin of the high inbreeding levels found.

The reduction of genetic variability in Brazilian adapted swine breeds is a matter of concern not only for researchers, but also commercial companies, as 500 years of adaptation to environmental conditions in Brazil may help raise productivity of commercial herds. According to (Martínez *et al.* 2015), who worked with European autochthonous caprine breeds, the genetic conservation programs for Brazilian adapted swine should acknowledge the individual variability of the animals within and among breeds, thus avoiding the origin of inbred pig herds and maintaining global genetic and allelic diversity. Divergent natural selection, dispersal and gene flow, as well as other processes affecting adaptive differentiation of local demes according to Kawecki & Ebert (2004)

may be the cause of regional and environmental structuring observed in Brazilian locally adapted swine breeds. This genetic structure must be considered in an effort to avoid genetic erosion of the breeds and also in the formation of conservation units and the management of these breeds. These units must preserve neutral and adaptive diversity to conserve their potential to react to the future challenges (Medugorac *et al.* 2011) that are important for the development of more efficient and sustainable animal production systems (Yaro *et al.* 2016).

Based on the 60k panel of SNPs, locally adapted swine genetic groups in Brazil present varied genetic structure in the studied environmental levels. These environments influence allele fixation among different swine sub populations and this variability to preserve their potential to react the ambient challenges and should be explored as a source of variability to increase the productivity of herds. The largest fraction of genetic variability of locally adapted swine breeds is observed between animals and this must be considered when selecting individuals for conservation programs. Genetic variability allows separating animals into genetic groups. As they present high indexes of inbreeding, the breeds Marajóara and Monteiro composed isolated populations, must be focus of future strategies of conservation and management to avoid negative impact caused by endogamic depression.

#### **CONFLICT OF INTEREST**

The authors declare no conflict of interest.

#### **ACKNOWLEDGEMENTS**

We respectfully thank CENARGEN-EMBRAPA (Centro Nacional de Recursos Genéticos) for the provision of DNA samples from Brazilian Pigs Germoplasm Banks well as Elizabete C. Silva and Gleison R. De Biazio for their support in the preparation and submission of material for genotyping. Thanks also to CAPES, CNPq and INCT-Pecuária for support of this research.

### Bibliography

- ABCS (2014) RELATÓRIO DO SERVIÇO DE REGISTRO GENEALÓGICO DE SUÍNOS 2014. , 46.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, **19**, 1655–1664.
- Boitard S, Chevalet C, Mercat M-J *et al.* (2010) Genetic variability, structure and assignment of Spanish and French pig populations based on a large sampling: Variability of Spanish and French pig populations. *Animal Genetics*, **41**, 608–618.
- Burgos-Paz W, Souza C, Megens H *et al.* (2012) Porcine colonization of the Americas: a 60k SNP story.
- Cavalcanti SS (1985) *Produção de suínos*. Instituto Campineiro de Ensino Agrícola.
- Cavalcanti SMC, Gese EM (2010) Kill rates and predation patterns of jaguars ( *Panthera onca* ) in the southern Pantanal, Brazil. *Journal of Mammalogy*, **91**, 722–736.
- Desbiez ALJ, Keuroghlian A, Piovezan U, Bodme RE (2009) Ecologia de Populações de Porco Monteiro no Pantanal do Brasil Population Ecology of Feral Pigs in the Brazilian Pantanal. , 40.
- Druml T, Salajpal K, Dikic M *et al.* (2012) Genetic diversity, population structure and subdivision of local Balkan pig breeds in Austria, Croatia, Serbia and Bosnia-Herzegovina and its practical value in conservation programs. *Genetics Selection Evolution*, **44**, 1–9.
- Excoffier L, Laval G, Stefan S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **2005:1**, 47–50.

- Favero JA, Figueiredo EAP de (2009) Evolução do melhoramento genético de suínos no Brasil. *Ceres*, **56**.
- Gama LT, Martínez AM, Carolino I *et al.* (2013) Genetic structure, relationships and admixture with wild relatives in native pig breeds from Iberia and its islands. *Genet Sel Evol*, **45**, e18.
- Golden Helix, Inc., Bozeman, MT, USA (2015) *SNP & Variation Suite*. Golden Helix, Inc, Bozeman, MT.
- Groenen MAM, Archibald AL, Uenishi H *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, **491**, 393–398.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
- Kim TH, Kim KS, Choi BH *et al.* (2005) Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. *Journal of animal science*, **83**, 2255–2263.
- Mariante A da S, Cavalcante N (2006) *Animais do Descobrimento: raças domésticas da história do Brasil*. Embrapa Sede / Embrapa Recursos Genéticos e Biotecnologia, Brasília.
- Martínez AM, Gama LT, Delgado JV *et al.* (2015) The Southwestern fringe of Europe as an important reservoir of caprine biodiversity. *Genetics Selection Evolution*, **47**, 86.
- Mauro R (2002) FAUNISTIC STUDIES IN THE EMBRAPA PANTANAL. *Arch. Zootec*, **51**, 175–185.
- Mauro R, Pereira M (2004) Perspectivas da Criação de Porcos Monteiro no Brasil.
- Mauro R, Pereira M, Paes R (2004) Perspectivas da Criação de Porcos Monteiro no Brasil. In: *Anais dos Simpósios da 41a. Reunião Anual da Sociedade Brasileira de Zootecnia*, pp. 511–516. Campo Grande - MS.
- McManus C, Hermuche P, Paiva SR *et al.* (2013) Geographical distribution of sheep breeds in Brazil and their relationship with climatic and environmental factors as risk classification for conservation. *Journal of Science and Technology*, **1**, 15.

- McManus C, Paiva SR, Silva AVR *et al.* (2010) Phenotypic characterization of naturalized swine breeds in Brazil, Uruguay and Colombia. *Brazilian Archives of Biology and Technology*, **53**, 583–591.
- Medugorac I, Veit-Kensch CE, Ramljak J *et al.* (2011) Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and Evolution*, **1**, 408–420.
- Niederle PA, Gelain J (2013) Geographical indications in Brazilian food markets: Quality conventions, institutionalization, and path dependence. *Journal of Rural Social Sciences*, **28**, 26.
- Pham LD, Do DN, Nam LQ *et al.* (2014) Molecular genetic diversity and genetic structure of Vietnamese indigenous pig populations. *Journal of Animal Breeding and Genetics*, **131**, 379–386.
- Purcell S (2014) PLINK: Whole genome data analysis toolset.
- Ramos AM, Crooijmans RPMA, Affara NA *et al.* (2009) Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology (L Orban, Ed.). *PLoS ONE*, **4**, e6524.
- SAS Institute Inc. (2011) *SAS/STAT® 9.3 User's Guide*. SAS Inst., Cary, NC.
- Silva-Filha OL, Alves DNM, Souza JF *et al.* (2005) Caracterização da criação de suínos locais em sistema de utilização tradicional no Estado da Paraíba, Brasil. *Arch Zootec*, **54**, 523–8.
- Sollero BP, Paiva SR, Faria DA *et al.* (2009) Genetic diversity of Brazilian pig breeds evidenced by microsatellite markers. *Livestock Science*, **123**, 8–15.
- Vicente AA, Carolino MI, Sousa MCO *et al.* (2008) Genetic diversity in native and commercial breeds of pigs in Portugal assessed by microsatellites. *Journal of Animal Science*, **86**, 2496–2507.
- Yaro M, Munyard KA, Stear MJ, Groth DM (2016) Molecular identification of livestock breeds: a tool for modern conservation biology. *Biological Reviews*, n/a-n/a.

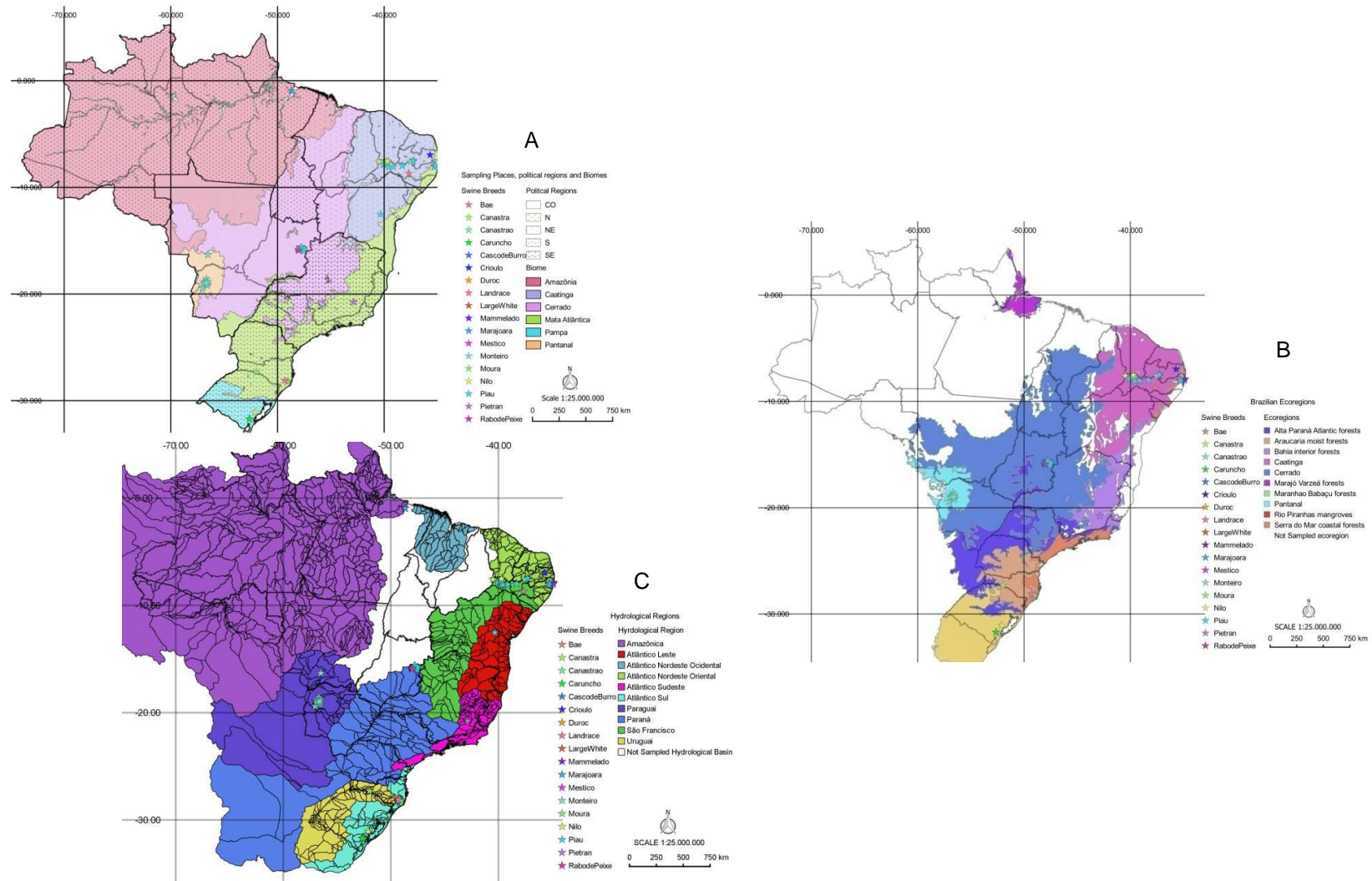


Figure 1: Map of swine breed sampling : A- Political regions and Biomes; B- Brazilian Ecoregions; C- Brazilian Hydrological Basins.

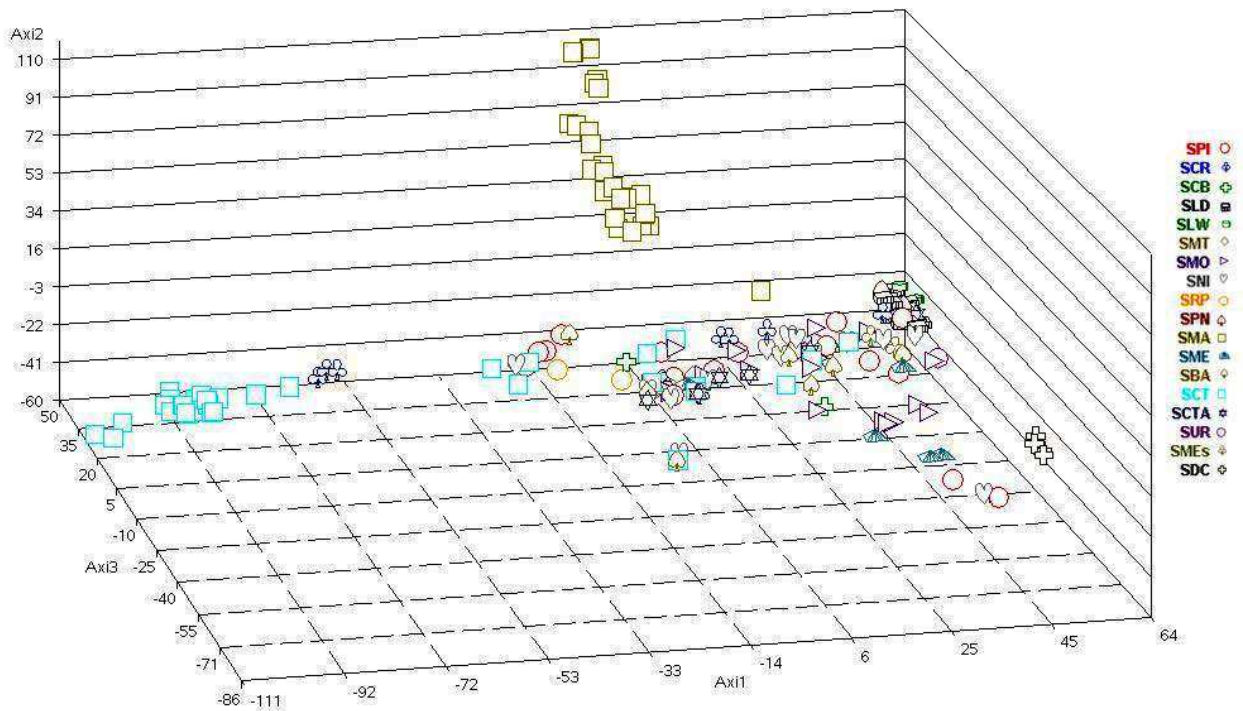


Figure 2: 3-D Representation of the first three Principal Components from allelic frequencies of 19,422 SNPs of 181 animals sampled, according to the genetic group of origin.

SBA= Baé; SCB= Casco de Burro; SCR= Caruncho; SMEc= Crioulo; SCT= Canastra; SCTA= Canastra; SDC= Duroc; SLD= Landrace; SLW= Large Withe; SMA= Marajó; SME= Mestiço; SMO= Moura; SMT= Monteiro; SNI= Nilo; SPI= Piau; SPN= Pietran; SRP= Rabo de Peixe

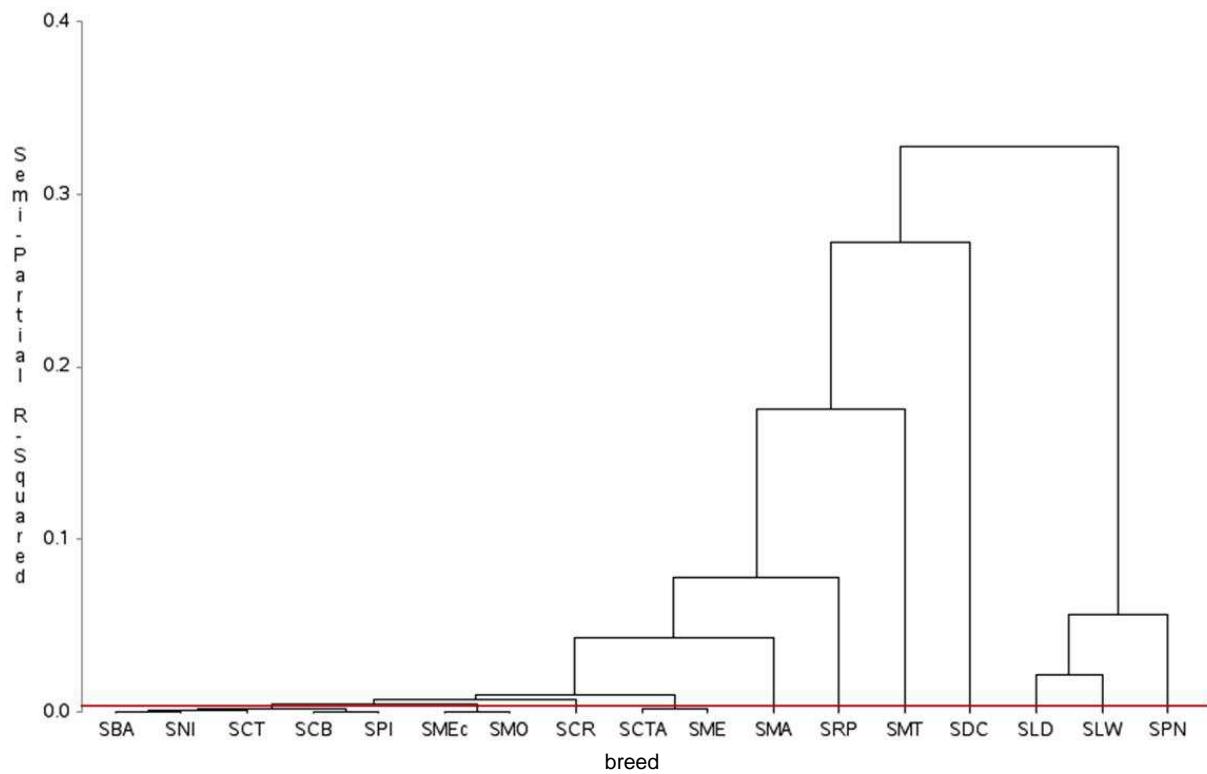


Figure 3: Cluster Analysis of Brazilian locally adapted swine genetic groups based on genetic distances between breeds using pairwise  $F_{ST}$

SBA= Baé; SCB= Casco de Burro; SCR= Caruncho; SMEc= Crioulo; SCT= Canastra; SCTA= Canastrao; SDC= Duroc; SLD= Landrace; SLW= Large Withe; SMA= Marajo; SME= Mestiço; SMO= Moura; SMT= Monteiro; SNI= Nilo; SPI= Piau; SPN= Pietran; SRP= Rabo de Peixe



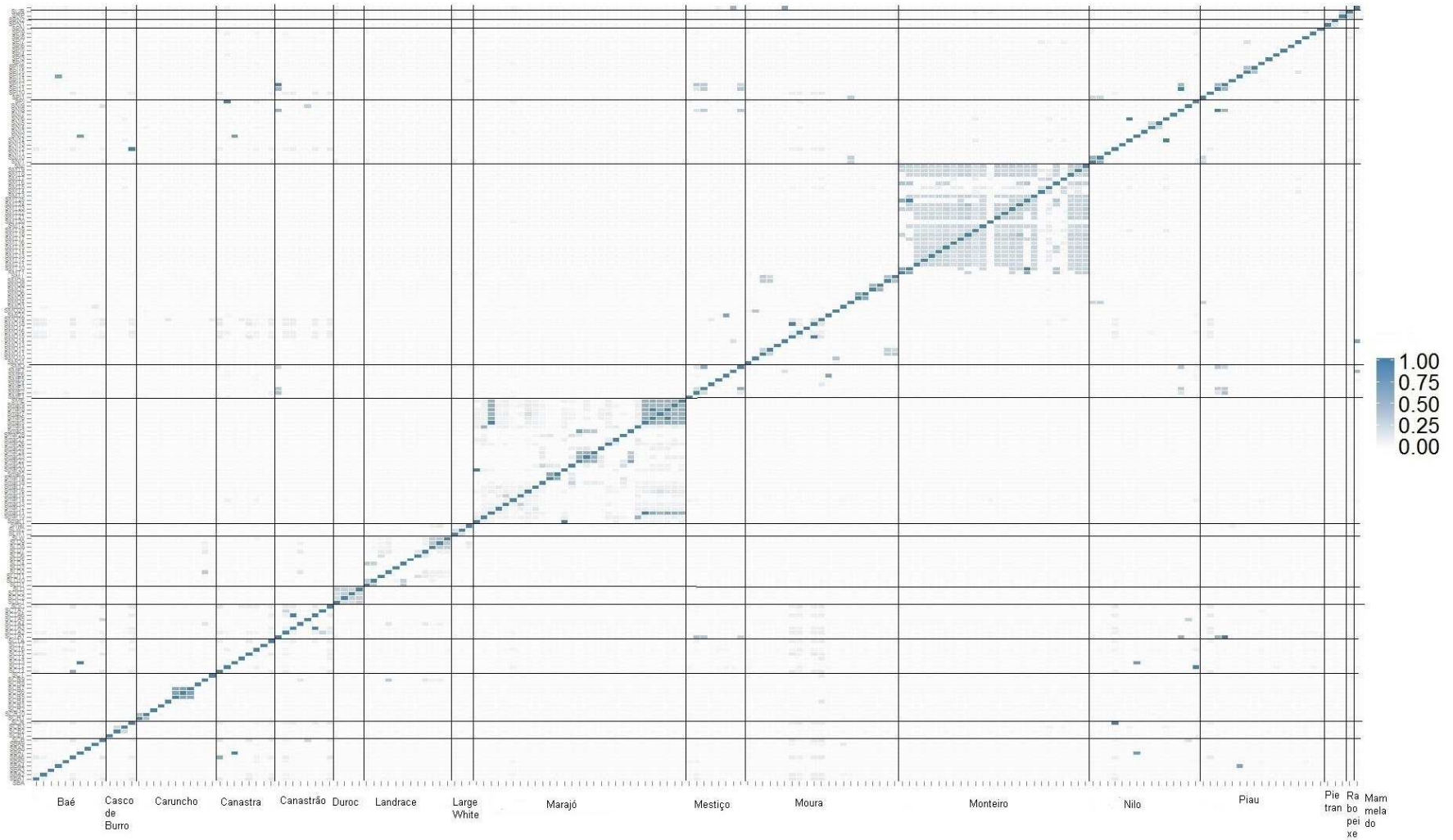


Figure 4: Identity by Descent coefficients Heat Map from 181 samples grouped by breed.

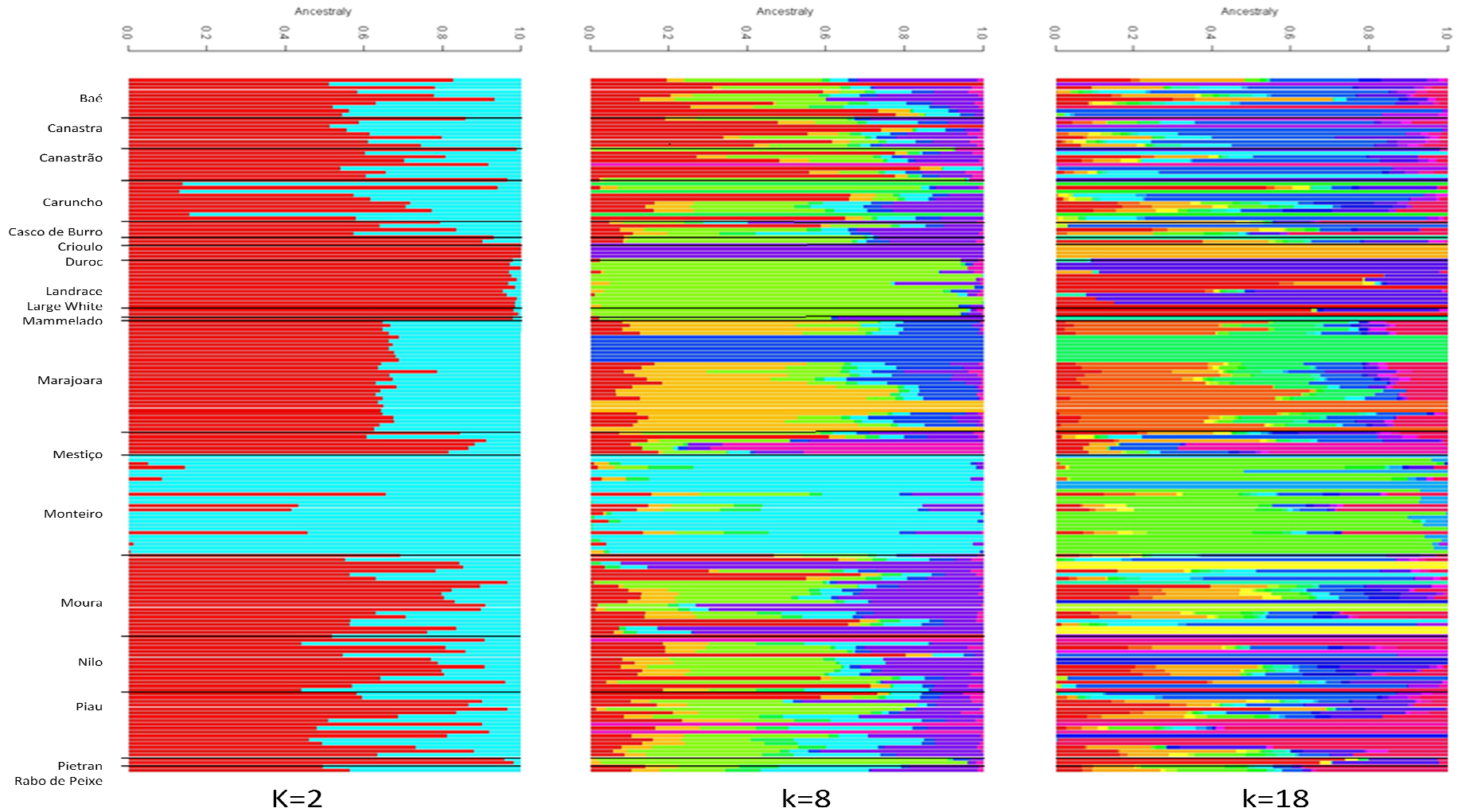


Figure 5: Population structure assessed by Structure. Bar plot. Each individual is represented by a vertical bar, often partitioned into colored segments with the length of each segment representing the proportion of the individual's genome from  $K = 2, 8,$  or  $18$  populations

SBA= Baé; SCB= Casco de Burro; SCR= Caruncho; SMEc= Crioulo; SCT= Canastra; SCTA= Canastrao; SDC= Duroc; SLD= Landrace; SLW= Large Withe; SMA= Marajo; SME= Mestiço; SMO= Moura; SMT= Monteiro; SNI= Nilo; SPI= Piau; SPN= Pietran; SRP= Rabo de Peixe

Table 1: Sampling frequency of swine genetic group and local.

Region	State	Samples		Genetic Group	Samples	
		n <sup>o</sup>	%		n <sup>o</sup>	%
N	PA	29	15.2	Marajo	29	15.2
	BA	3	1.57	Piau	3	1.57
NE	PE	66	34.6	Baé	10	5.24
				Caruncho	4	2.09
				Canastra	10	5.24
				Canastrao	8	4.19
				Landrace	3	1.57
				Mestiço	5	2.62
	PB	2	1.05	Moura	8	4.19
				Nilo	9	4.71
				Piau	8	4.19
				Mammelado	1	0.52
				Crioulo	2	1.05
				Casco de Burro	4	2.09
CO	GO	30	15.7	Caruncho	3	1.57
				Landrace	3	1.57
				Large White	1	0.52
				Moura	1	0.52
				Monteiro	5	2.62
	MS	18	9.42	Nilo	3	1.57
				Piau	8	4.19
				Rabo de Peixe	2	1.05
				Monteiro	18	9.42
				Monteiro	3	1.57
SD	MT	3	1.57	Piau	1	0.52
	MG	3	1.57	Pietran	2	1.05
S	RS	14	7.33	Caruncho	3	1.57
				Mestiço	1	0.52
				Moura	4	2.09
				Nilo	5	2.62
	SC	23	12	Piau	1	0.52
				Casco de Burro	1	0.52
				Duroc	4	2.09
				Landrace	7	3.66
Large Withe	2	1.05				
Moura	9	4.71				

N=North; NE=Northeast; CO=Midwest; SD=Southwest; S=South

PA=Pará; BA=Bahia; PE=Pernambuco; PB=Paraíba; GO=Goiás; MS=Mato Grosso do Sul; MT=Mato Grosso; MG=Minas Gerais; RS=Rio Grande do Sul; SC=Santa Catarina.

Table 2: Description of categorical variables for locally adapted swine populations.

Variable	Description	Subdivision Sampled
Sate	Internal division of Brazilian territory based on historical and political-social characteristics.	PA=Pará; BA=Bahia; PE=Pernambuco; PB=Paraíba; GO=Goiás; MS=Mato Grosso do Sul; MT=Mato Grosso; MG= Minas Gerais; RS=Rio Grande do Sul; SC=Santa Catarina
Hydrological Basin	Water drainage of water of precipitation water for one watercourse and smaller rivers that flow into the sea	Hydrological region Amazônica; Hydrological region Atlântico Leste; Hydrological region Atlântico Nordeste Oriental; Hydrological region Atlântico Sudeste; Hydrological region Atlântico Sul; Hydrological region of Paraguai; Hydrological region of Paraná; Hydrological region of São Francisco; Hydrological region of Uruguai
Type of Soil	Soil classification according to their physical and chemical properties	Red-yellow argilossols; Haplic Cambisol; Espodosol ferrocubic; Yellow Oxisol; Red Oxisol; Red/Yellow Oxisol; Luvisol Chromic ; Litholic Neosol; Quartzipsamment; Planosol Haplic; Eutric Planosol
Biome	Describes major ecological systems defined mainly by climate	Amazônia; Caatinga; Cerrado; Mata Atlântica; Pampa; Pantanal
Vegetation	Phytogeographic classification of Brazilian vegetation	Formações Pioneiras area; Ecological tension area; Estepe; Estacional Decidual forest; Estacional Semidecidual forest; Ombrofila Aberta forest; Ombrofila Densa forest; Ombrofila Mista forest; Savana
Ecoregion	Basic unit for planning of biodiversity conservation priorities	Alta Parana Atlantic forests; Araucaria moist forests; Bahia interior forests; Caatinga; Cerrado; Marajó Varzea forests; Pantanal; Pernambuco interior forests; Rio Piranhas mangroves; Serra do Mar coastal forests; Uruguayan savanna

Table 3: Pairwise  $F_{ST}$  for Brazilian locally adapted Swine genetic groups.

	SBA	SCT	SCTA	SCR	SCB	SMEc	SMT	SMO	SNI	SPI	SMA	SME	SRP	SDC	SLD	SLW	SPN
SBA	0																
SCT	-0.008	0.000															
SCTA	0.015	0.013	0.000														
SCR	0.028	0.031	0.040	0.000													
SCB	0.023	0.026	0.035	0.039	0.000												
SMEc	0.026	0.034	0.037	0.044	0.031	0.000											
SMT	0.096	0.102	0.117	0.094	0.113	0.155	0.000										
SMO	0.017	0.018	0.026	0.031	0.019	0.009	0.100	0.000									
SNI	0.009	0.007	0.019	0.028	0.015	0.019	0.095	0.015	0.000								
SPI	0.012	0.018	0.017	0.028	0.014	0.018	0.089	0.016	0.007	0.000							
SMA	0.051	0.052	0.062	0.061	0.062	0.068	0.115	0.054	0.049	0.048	0.000						
SME	0.025	0.027	0.020	0.045	0.035	0.025	0.130	0.025	0.020	0.012	0.067	0.000					
SRP	0.081	0.093	0.098	0.088	0.084	0.109	0.147	0.078	0.076	0.068	0.114	0.097	0.000				
SDC	0.128	0.137	0.145	0.156	0.111	0.129	0.248	0.092	0.118	0.113	0.167	0.118	0.216	0.000			
SLD	0.067	0.073	0.065	0.075	0.066	0.026	0.166	0.061	0.057	0.051	0.098	0.060	0.128	0.160	0.000		
SLW	0.097	0.108	0.105	0.112	0.101	0.055	0.214	0.088	0.084	0.078	0.128	0.096	0.174	0.216	0.056	0.000	
SPN	0.112	0.121	0.115	0.119	0.105	0.080	0.230	0.101	0.099	0.089	0.140	0.107	0.200	0.236	0.069	0.098	0

SBA= Baé; SCB= Casco de Burro; SCR= Caruncho; SMEc= Crioulo; SCT= Canastra; SCTA= Canastra; SDC= Duroc; SLD= Landrace; SLW= LargeWhite; SMA= Marajo; SME= Mestiço; SMO= Moura; SMT= Monteiro; SNI= Nilo; SPI= Piau; SPN= Pietran; SRP= Rabo de Peixe

Table 4: Pairwise  $F_{ST}$  between samples from different Brazilian States.

	RS	SC	PE	PB	BA	GO	MS	MT	MG	PA
RS	0.000									
SC	0.034	0.000								
PE	0.034	0.033	0.000							
PB	0.024	0.010	0.015	0.000						
BA	0.085	0.089	0.060	0.103	0.000					
GO	0.036	0.036	0.019	0.029	0.056	0.000				
MS	0.135	0.143	0.102	0.180	0.153	0.072	0.000			
MT	0.071	0.073	0.054	0.083	0.101	0.037	0.108	0.000		
MG	0.058	0.043	0.050	0.035	0.089	0.045	0.175	0.089	0.000	
PA	0.064	0.072	0.048	0.068	0.096	0.052	0.128	0.086	0.086	0.000

PA=Pará; BA=Bahia; PE=Pernambuco; PB=Paraíba; GO=Goiás; MS=Mato Grosso do Sul; MT=Mato Grosso; MG=Minas Gerais; RS=Rio Grande do Sul; SC=Santa Catarina.

Table 5:  $F_{IS}$  values for Brazilian locally adapted swine genetic groups and for States in Brazil after 10000 permutations.

Genetic Group	$F_{IS}$	$P(\text{Rand } F_{IS} \geq \text{Obs } F_{IS})$
Baé	0.156	0.003
Canastra	0.075	0.175
Canastrão	0.054	0.257
Caruncho	0.215	0.002
Casco de Burro	0.028	0.423
Crioulo	-0.012	0.662
Mammelado	0.000	1.000
Monteiro	0.049	0.098
Nilo	0.143	0.005
Marajoara	-0.010	0.587
Mestiço	0.023	0.435
Rabo de Peixe	-0.080	0.664
Duroc	-0.038	0.782
Landrace	-0.004	0.548
Large White	-0.012	0.669
Pietran	-0.005	0.673
Piau	0.150	0.00001
Moura	0.082	0.032

State	$F_{IS}$	$P(\text{Rand } F_{IS} \geq \text{Obs } F_{IS})$
RS	0.029	0.308
SC	0.109	0.005
PE	0.121	0.000
PB	-0.012	0.670
BA	0.107	0.203
GO	0.155	0.000
MS	0.005	0.462
MT	-0.137	0.662
MG	0.105	0.397
PA	-0.010	0.592

PA=Pará; BA=Bahia; PE=Pernambuco; PB=Paraíba; GO=Goiás; MS=Mato Grosso do Sul; MT=Mato Grosso; MG=Minas Gerais; RS=Rio Grande do Sul; SC=Santa Catarina.

Table S1: Pairwise  $F_{ST}$  for Brazilian locally adapted swine genetic groups by kind of vegetation.

Vegetation (n/freq)	Estepe	Estacional Semidecidual Forest	Savana	Ombrofila Aberta forest	Estacional Decidual forest	Areas of Formações Pioneiras	Ecological tension area	Ombrofila Dense forest	Ombrofila Mixed forest
Estepe (56/30.94%)	0.000								
Estacional Semidecidual Forest (15/8.29%)	0.020	0.000							
Savana (77/42.54%)	0.021	0.030	0.000						
Ombrofila Aberta forest (3/1.66%)	0.029	0.018	0.043	0.000					
Estacional Decidual forest (12/6.63%)	0.053	0.042	0.062	0.036	0.000				
Areas of Formações Pioneiras (4/2.21%)	0.031	0.020	0.046	-0.030	0.034	0.000			
Ecological tension area (3/1.66%)	0.090	0.110	0.051	0.157	0.152	0.159	0.000		
Ombrofila Dense forest (6/3.31%)	0.054	0.050	0.056	0.056	0.051	0.057	0.151	0.000	
Ombrofila Mixed forest (5/2.76%)	0.074	0.052	0.081	0.036	0.072	0.039	0.187	0.096	0.000

n=sample size; freq=sampling frequency

Table S2: Pairwise  $F_{ST}$  for Brazilian locally adapted swine genetic groups by kinds of soil.

Kind of Soil (n/freq)	Neossolo Litalico	Latossolo Vermelho- Amarelo	Neossolo Quartzaranico	Argilossolo Vermelho- Amarelo	Luvissolo Cramico	Latossolo Vermelho	Cambissolo Haplico	Espodossolo Ferrocarbico	Planossolo Haplico	Latossolo Amarelo	Planossolo Natrigo
Neossolo Litalico (27/14.92%)	0.000										
Latossolo Vermelho- Amarelo (13/7.18%)	0.009	0.000									
Neossolo Quartzaranico (6/3.31%)	0.074	0.073	0.000								
Argilossolo Vermelho- Amarelo (46/25.41%)	0.013	0.008	0.083	0.000							
Luvissolo Cramico (13/7.18%)	0.005	0.007	0.072	0.010	0.000						
Latossolo Vermelho (23/12.71%)	0.018	0.015	0.049	0.020	0.015	0.000					
Cambissolo Haplico (12/6.63%)	0.051	0.044	0.138	0.032	0.051	0.055	0.000				
Espodossolo Ferrocarbico (10/5.52%)	0.101	0.101	0.010	0.106	0.100	0.066	0.166	0.000			
Planossolo Haplico (1/0.55%)	0.069	0.063	0.010	0.075	0.066	0.038	0.128	0.010	0.000		
Latossolo Amarelo (28/15.47%)	0.049	0.050	0.102	0.049	0.047	0.053	0.091	0.124	0.100	0.000	
Planossolo Natrigo (2/1.10%)	0.013	-0.014	0.081	0.010	0.010	0.000	0.054	0.121	0.079	0.051	0.000

n=sample size; freq=sampling frequency



Table S3: Pairwise  $F_{ST}$  for Brazilian locally adapted swine genetic groups by Ecoregions

Ecoregion (n/freq)	Caatinga	Pernambuco interior forests	Cerrado	Pernambuco coastal forests	Alta Parana Atlantic forests	Rio Piranhas mangroves	Pantanal	Bahia interior forests	Marajó várzea forests	Uruguayan savanna	Serra do Mar coastal forests	Araucaria moist forests
Caatinga (56/30.94%)	0.000											
Pernambuco interior forests (9/4.97%)	0.028	0.000										
Cerrado (23/12.71%)	0.018	0.045	0.000									
Pernambuco coastal forests (9/4.97%)	0.022	0.038	0.032	0.000								
Alta Parana Atlantic forests (12/6.63%)	0.053	0.059	0.055	0.027	0.000							
Rio Piranhas mangroves (5/2.76%)	0.035	0.045	0.040	-0.029	0.040	0.000						
Pantanal (20/11.05%)	0.085	0.134	0.057	0.154	0.153	0.154	0.000					
Bahia interior forests (3/1.66%)	0.052	0.067	0.047	0.035	0.070	0.041	0.151	0.000				
Marajó várzea forests (28/15.47%)	0.048	0.076	0.053	0.068	0.091	0.077	0.113	0.086	0.000			
Uruguayan savanna (14/7.73%)	0.036	0.052	0.037	0.024	0.047	0.039	0.118	0.058	0.064	0.000		
Serra do Mar coastal forests (6/3.31%)	0.054	0.071	0.052	0.050	0.051	0.061	0.145	0.078	0.087	0.057	0.000	
Araucaria moist forests (5/2.76%)	0.074	0.077	0.074	0.041	0.072	0.037	0.185	0.051	0.109	0.070	0.096	0.000

n=sample size; freq=sampling frequency

Table S4: Pairwise  $F_{ST}$  for Brazilian locally adapted swine genetic groups by hydrological basin

Hydrological Basin (n/freq)	São Francisco	Atlântico Nordeste Oriental	Paraná	Uruguai	Paraguai	Atlântico Leste	Atlântico Sudeste	Amazônica	Atlântico Sul
São Francisco (59/32.6%)	0.000								
Atlântico Nordeste Oriental (14/7.73%)	0.015	0.000							
Paraná (17/9.39%)	0.026	0.050	0.000						
Uruguai (12/6.63%)	0.047	0.039	0.073	0.000					
Paraguai (20/11.05%)	0.085	0.121	0.048	0.153	0.000				
Atlântico Leste (3/1.66%)	0.058	0.080	0.063	0.112	0.130	0.000			
Atlântico Sudeste (3/1.66%)	0.046	0.043	0.071	0.070	0.151	0.089	0.000		
Amazônica (28/15.47%)	0.047	0.063	0.062	0.091	0.113	0.096	0.086	0.000	
Atlântico Sul (25/13.81%)	0.023	0.021	0.044	0.033	0.108	0.074	0.038	0.057	0.000

n=sample size; freq=sampling frequency

### **CAPÍTULO III<sup>1</sup>**

“..Mais c'est long le chemin

Ma vie

Qu'il est long le chemin !”

*(Alain Barrière)*

## **Landscape Genomic Approach to Detect Selection Signatures in Locally Adapted Brazilian Swine Genetic Groups**

Cesconeto, R.J.<sup>1</sup>, Joost, S.<sup>2</sup>, McManus, C.<sup>3</sup>, Paiva, S.R.<sup>4</sup>, Braccini, J.<sup>5</sup>

1- Doctoral student at UFRGS; 2- Professor at EPFL; 3- Professor at UNB; 4- Research at EMBRAPA-CENARGEN; 5- Professor at UFRGS.

1- Corresponding author: Robson José Cesconeto; CP19, Orleans, SC, BR  
Zip:88870-000; +55 (48) 3466-2964 csconeto@hotmail.com

**Key words:** Animal Genetic Resources, *Sus scrofa*, molecular markers, population structure, conservation genetics

### **Abstract**

Samples of 191 animals from 18 different Brazilian locally adapted swine genetic groups were genotyped using Porcine SNP60 BeadChip in order to identify selection signatures related to the Brazilian environmental variables. Using the BayeScan approach, 71 SNP markers were identified as  $F_{ST}$  outliers and 60 genotypes (58 markers) were found by Samβada method in 371 logistic models correlated with 112 environmental variables. Five markers were identified in both methods, with a Kappa value of 0.073 (95% CI: 0.011 to 0.134). The frequency of these markers shows a clear north-south country division that reflects Brazilian environmental differences in temperature, solar radiation and precipitation. Global spatial territory correlation for environmental variables corroborates this finding (average Moran's  $I = 0.89$ , range from 0.55 to 0.97). The distribution of alleles over the territory was not strongly correlated with the breed/genetic groups. These results are congruent with previous mtDNA studies and should be used to direct germplasm collection for the National gene bank.

## Introduction

After the introduction to Brazil of Portuguese, Spanish and Asian swine breeds in early 1500's, these animals spread throughout the territory, and were used as an important source of meat and fat by farmers and the general population (Mariante *et al.* 2003). These animals have given origin to several locally adapted swine breeds which were used intensely until the 70's, when changes in the market led to the introduction of North American and European breeds specialized in meat production. Nowadays, the locally adapted breeds are present only on small farms, with low input levels.

These local adapted breeds became adapted to the Brazilian territory through equilibrium between evolutionary (mainly genetic drift) and local environment forces. These breeds are a reservoir of genetic variability (Giovambattista *et al.* 2001) and, principally, source of traits selected and fixed mainly by influence of the environment (Hall & Ruane 1993). Mirkena *et al.*(2010) discuss genetic influence on disease tolerance/resistance in small ruminants and, in addition to other factors, cite advantages in increased fitness of locally adapted breeds. The areas of the genome that have been, or still are, under selection (Luikart *et al.* 2003; Storz 2005; Vitalis *et al.* 2014) can be estimated using Single Nucleotide Polymorphisms (SNPs) spread throughout the genome by theoretical populational  $F_{ST}$  outliers, that are assumed to be signatures of natural selection (Lewontin & Krakauer 1973; Luikart *et al.* 2003; Joost *et al.* 2007; Lotterhos & Whitlock 2014). These signatures of selection can help us understand the complex relation between adapted swine genetic groups and the environment, as well as the process of adaptation of the swine in the

Brazilian territory and to overcome the challenges in swine management in a country with continental dimensions and different climatic conditions such as Brazil. In a constantly changing world, the identification of those signatures may be a key to promote more sustainable animal production, improving gains in productivity and welfare, as well as decreasing sanitary expenses with medication and management (Mirkena *et al.* 2010; Shabtay 2015). In addition, these results might be an auxiliary tool to help the enrichment of National gene banks and conservation programs.

The hypothesis of the present study is that environment influenced successful adaptation of swine in the Brazilian territory and left detectable signatures of natural selection. To test this hypothesis a medium SNP chip array was used in animals distributed over the main Brazilian regions.

## **Material and methods**

### **Sampling**

The sample selection (Table S1 and Fig. S1) was structured to represent the Brazilian territory, capturing high spatial representation of the environment and genetic territorial dispersion of the swine breeds, with at least one sample from each Political Region (5 regions). A total of 191 samples of non-related animals from 18 different swine genetic groups (13 Locally adapted Brazilian Swine genetic groups, 4 commercial or global breeds and one group formed by crossbred animals) was randomly selected. All samples used in this experiment are deposited in Embrapa's Gene Bank (<http://aleloanimal.cenargen.embrapa.br>) located at Embrapa Genetic Resources and Biotechnology Center, Brasilia, DF.

### **Genotyping and Quality Control**

The DNA samples were genotyped with the Porcine SNP60 BeadChip v2. The quality control of raw data (191 samples and 61,565 SNP markers) were performed with SNP & Variation Suite v8.x (Golden Helix, Inc., Bozeman, MT, USA 2015) (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) using the following parameters: 90% Call Rate for samples (11 animals excluded), 95% Call Rate and 0.05% Minor Allele Frequency (MAF) for the markers (21,605 excluded). Additional Linkage Disequilibrium (LD) pruning was performed using a window size=50, window increment=5 and  $r^2$  threshold=0.05, which eliminated a further 11,646 SNPs. The final data had 28,860 SNP markers with a SNP density of 1/87,026 kb.

### **Environmental variables**

The environmental variables from the Brazilian territory were obtained at the World ClimProject (<http://www.worldclim.org/>), GTOPO30 (<https://lta.cr.usgs.gov/GTOPO30>) and Harvest Choice (<http://harvestchoice.org/>) and 30 sec geographical information system (GIS) layer using Qgis v2.6 (QGIS Development Team 2009). Monthly maximum, average and minimum temperature, annual average and median for maximum, average and minimum temperature, seasonal averages and medians for maximum, average and minimum temperature, monthly solar radiation, annual average and median solar radiation, seasonal averages and medians solar radiation, monthly precipitation, annual average and median precipitation

seasonal averages and medians precipitation, 19 bioclimatic variables (BIO1-19), elevation, PETannual (annual potential evapor-transpiration) and Aridity (ratio of precipitation to PET) were obtained for each sample from this layer (Table S2).

### **Relationship between Samples**

The individual and populational levels of expected heterozygosity ( $H_e$ ) and observed heterozygosity ( $H_o$ ) were computed by the SNP & Variation Suite v8.x (Golden Helix, Inc., Bozeman, MT, USA 2015) and used to identify a source of genetic variance by means of analysis of molecular variance (AMOVA) between and within the geopolitical groups, as well as between native and commercial breeds (Tables S3 and S4), performed in Arlequin V 3.5.2.2 (Excoffier *et al.* 2005). The discontinuity of genetic composition was evaluated using a Mantel test between the geographical distance and pairwise inbreeding matrices calculated in SAM v4.0 (Rangel *et al.* 2010). An additional Mantel test was performed between Edward's genetic distance and Euclidian geographic distances among samples using ADEGENT R package (Jombart 2008).

### **Signatures of Selection and Outlier Detection**

Loci with high or low allelic differentiation in relation to the expected neutrality were used as an indication of selection (Hoffmann & Willi 2008) and were tested by two different methodologies of outliers identification.

BayeScan software V 2.1 (Foll & Gaggiotti 2008) used a Bayesian approach via Markov Chain Monte Carlo (MCMC), assuming a prior Dirichlet distribution of alleles within populations and using a hierarchical Bayesian model. The program calculates posterior odds, from the posterior probability of



the models, with and without selection on a locus, by using the proportion of loci with a strong increase in  $F_{ST}$  relative to other loci among the MCMC outputs of its simulations (Beaumont & Balding 2004). The software was set up with 5,000 burn in interactions, followed by 10,000 interactions with thinning interval of 10. The second approach was performed using the software Samβada (Joost *et al.* 2007; Stucki *et al.* 2014) that used logistic regression models to determine the probability of allele presence/absence in the specific environment. The models were considered significant when the G Score and Wald Score were significant at  $\alpha=0.01$  threshold with a Bonferroni correction. The G Score can be defined as the ratio between maximum log likelihood of model with the presence of the independent variable and the maximum log likelihood of model without independent variable, or as the independent variable affects log likelihood model. The Wald Score tests if adjustment goodness is affected when independent variable is removed from the model. The agreement between the two methods was evaluated through the Kappa index calculated by FREQ procedure (Proc FREQ) of SAS v9.3 (SAS Institute Inc. 2011). Gene annotations within candidate regions were obtained using the data provided by Ensembl ([www.ensembl.org](http://www.ensembl.org)) and NCBI (<http://www.ncbi.nlm.nih.gov>).

To measure the degree of spatial association for marker signaled as  $F_{ST}$  outliers by both methods, the Global spatial autocorrelation (Moran's I) was calculated. Moran's I describes the autocorrelation between the values of a variable in a certain location with the values of this same variable in a neighboring location (Druck *et al.* 2004), with null hypothesis being that there is no spatial clustering.

## Results

Molecular variance analysis among states grouped into regions (Table S3) showed 93.35% of the genetic variance was contained within states and only 0.87% among regions. The genetic variance among a group of animals from commercial breeds and a group of locally adapted genetic groups (Table S4) showed individual variance (81.85%) was larger than variance between groups (3.66%) or from individuals within groups (8.27%).

BayeScan software identified 71 SNP markers as  $F_{ST}$  outliers while Samβada software identified 60 genotypes (from 58 SNP markers) in 371 univariate logistic models, using 112 environmental variables. No multivariate model was significant with  $\alpha=0.05$ . The markers MARC0021990; ASGA0033717; MARC0007678 were responsible for 42% of all models generated by Samβada (Fig.1). Five markers were found using both methods (ALGA0032795; ALGA0054315; ASGA0026250; ASGA0029202; BGIS0004952) with Kappa 0.073 (95% CI: 0.011 to 0.134). The markers identified as signatures of selection were near or in genes responsible for intracellular transport, immune response, cell respiration and related with the circulatory system, probably as a physiological response to cellular stress (Table 2). The marker ASGA0029202 was associated with precipitation and thermal amplitude and is near ( $\pm 0.2$ Mb) the CDH2 gene that has an effect on the formation of blood vessels, while the marker MARC0021990 (responsible for 20% of the models) was close to ( $\pm 0.27$ Mb) the gene CYP7B1 which is evolved, among other functions, with cofactor HEME, suggesting indirect

evidence of importance of circulatory system on genetic adaptation to fluctuation on temperature, Bio18 and solar radiation in the Brazilian territory (Table2).

Global spatial correlation for environmental variables was high, with 5 (five) neighbors windows, (average Moran's  $I=0.89$ , from 0.55 to 0.97), and reaching close to zero with 15 (fifteen) neighbors windows. The highest value for "Moran's  $I$ " was associated with solar radiation in the summer months. The selection signal markers have had a high global spatial correlation between 5 to 10 neighbors and presents a rapid decrease to zero with 35 neighbors (Fig 2.). With 5 neighbors the maximum local  $I$  was 0.7072 from marker CASI0001257 and the smallest was -0.04346 from marker ASGA0002592 (Fig.2).

For these five markers, considered as selection markers in BayeScan and Samβada, the influence of homozygotes in their neighbors was larger than heterozygotes, but only up to 30 neighbors. A regionalization of these markers was observed around a nucleus of climatic variation (Fig. 4), with a loss of influence when geographic distance between samples was increased, or when distancing from the climatic influence center was decreased.

The Mantel test (Fig.S2) between individual pairwise genetic distances and individual geographic distance had correlation coefficient of  $r=0.20$  ( $P(r)>1e^{-05}$ , with 99999 replications). The Mantel test between inbreeding coefficients and geographic distance showed a correlation of  $r=0.02$  ( $P(r)>1e^{-03}$  with 9999 replications), while the Mantel between pairwise  $F_{ST}$  matrix (genetic similarity) and geographic distance was  $r=0.07$ .

## Discussion

As the evolution and adaptation of pigs is subject to environmental influences as has been observed in humans (Storz 2010), humans and cattle (Beja-Pereira *et al.* 2003), fish (Nielsen *et al.* 2009) and other species (Manthey & Moyle 2015). A high genetic variability among local adapted breeds was found for Vietnamese (Pham *et al.* 2014), Indian (De *et al.* 2013) and Colombian pigs. This genetic diversity could be connected with the environment through years of selection leaving marks on the swine genome. There are many different methodologies for detection of genetic markers or genomic regions under influence of natural selection and one of these approaches is the identification of populational theoretical  $F_{ST}$  outliers. The use of georeferenced environmental data associated with  $F_{ST}$  outlier helps in the understanding of the evolutionary process and the influence of the environment on the process.

For this work we used two methods to detect Outliers in  $F_{ST}$ . According to Pérez-Figueroa *et al.*(2010), BayeScan's algorithm, under neutral hypothesis admits less than 1% of false discoveries, when we assume the Dirichlet distribution and that population has a neutral structure. Those presuppositions on distribution and structure may become biased due to the existence of more than one sample within the population, or when individuals share a common ancestor in the recent past (Lotterhos & Whitlock 2014). Feng *et al.* (2015) argues that some BayeScan configurations can affect the proportion and the direction of the markers in selection. This kind of bias does not occur with Samβada, because it translates samples in alleles frequencies associated with ambient data and use these outliers to calculate logistic regression which explains allele presence in a specific environment (Stucki *et*

*al.* 2014). As the Samβada algorithm is based at individual and local levels, taking into consideration the p-value after Bonferroni correction to determinate the significance of the models, the probability of mistakenly considering significant an association between marker and environmental variables decreases (Stucki 2014; Stucki *et al.* 2014).

The rates of spacial autocorrelation (Fig.2) showed that 5 to 10 closest neighbors tend to have high spatial autocorrelation among each other. This behavior was possibly motivated by the habit of breeders interchanging sires and dams, trying to maintain inbreeding at low levels (Favero & Figueiredo 2009; Gama *et al.* 2013). The probability of genetic similarity at a distance higher than 10 neighbors decreases and this might be related today with limited dispersion due to sanitary legislation within the country for swine species as well the market organization.

The pattern of spatial distribution of the genotypes, identified as selection markers (fig. 4), associated with environmental conditions such as temperature, solar radiation and BIO18- precipitation of the warmest quarter (Table 1), during some periods in the year, shows adaptive selection linked to seasonality. The genotypic frequency of these signatures of selection divides the territory into two regions (Table 3), one in the north where we have predominantly the occurrence of one of the genotypes and the other to the south where the alternative genotype occurs. Accordingy to Nimer (1979), these two regions are identified by different climates: the north shows “equatorial”, “tropical” and “northeast occidental tropical” climates; the south shows “temperate” and “central Brazil tropical” climates.

Although the markers MARC0021990; ASGA0033717; MARC0007678 were responsible for a high number of significant models identified in Samβada, we did not find any significant multivariate model (Fig. 1). When one marker is linked with some environmental variables, this suggests many evolutionary steps within the environment, throughout the year, influence the presence of markers. Despite only univariate models being found, there were associations between these alleles and the variation of temperature throughout the year, but not among the seasons as discussed by Johnson *et al.* (2015b). The environmental temperature is closely linked to welfare (Lee & Phillips 1948) and animal productivity (Collier & Gebremedhin 2015), affecting pigs in all stages of life (Wildt *et al.* 1975; Ross *et al.* 2015), including intra-uterine development, with consequences in the postnatal development of animals (Johnson *et al.* 2013, 2015a). The significant models found by Samβada for mean diurnal range (BIO2) associated with the marker ALGA0012967 in an intronic region of the LGR4 gene, which directly influences the testicular development and spermatogenesis, was in accordance with Petrocelli *et al.* (2015) who reported seasonal variation of seminal quality parameters affecting the reproductive performance of females. Once the survival and adaptation of the species in the environment are limited by reproductive success from individuals, and knowing when environmental conditions such as temperature and humidity are outside thermal comfort limits, we can see physiological alterations leading to reproductive failure in females (Nteebe *et al.* 2015) and males (Wettemann & Bazer 1985; Flowers 2015).

Ai *et al.*(2013), working with Chinese pigs in Tibet, and Burgos-Paz *et al.* (2012) with American pigs, found selection signatures correlated with the extremes of environmental conditions (high-altitude adaptation), linked to altitude and circulatory system, respectively. Different from these authors, we found selection signatures for variation in temperature, radiation solar and BIO18 (Fig. 4). The identification of selection signatures helps us to understand the relationship between climate and adaptive genetic variation, informing the conservation of both putatively neutral and adaptive components of genetic diversity (Bradbury *et al.* 2013) across a dynamic and heterogeneous unpredictable landscape. Selection signatures from autochthone breeds may be a tool to improve livestock production through changes in the frequencies of these alleles in commercial herds, improving the adaptation in different environments. This is important in a world marked by environmental change that acts by altering the composition of the community and shifting range boundaries, phenology, genetic diversity and genetic structure of organisms (Manel *et al.* 2012), probably imposing strong selection pressures on traits important for fitness (Gienapp *et al.* 2008).

### **Conclusion**

The distribution of SNP alleles over the Brazilian territory demonstrate a clear north-south orientation, dividing the country into two distinct regions. The allele frequency of markers from Brazilian locally adapted swine breeds was seem to be under influence of environmental conditions showing footprints of divergent selection in at least eight SNP markers, associated with

temperature, solar radiation and BIO18 linked with intracellular activity and circulatory system and were considered important for species adaptation.



Table 1: Samβada output to Environmental association to markers detected as signatures of selection in both methods.

Marker	Env_1	Loglikelihood	Gscore	WaldScore	Beta_0	Beta_1
ALGA0032795	TMinoutMedinan	-89.93	50.10	38.44	3.85	-0.03
ALGA0032795	TMINMai	-89.93	50.10	38.44	3.85	-0.03
ALGA0054315	TMAXAbr	-88.34	50.37	38.83	11.20	-0.04
ASGA0026250	TMinoutMedinan	-88.46	51.60	39.19	3.92	-0.03
ASGA0026250	TMINMai	-88.46	51.60	39.19	3.92	-0.03
ASGA0026250	TMinoutMed	-88.62	51.28	38.71	4.17	-0.03
ASGA0026250	TMINAbr	-89.15	50.22	38.57	5.26	-0.03
ASGA0029202	Bio18	-99.32	49.84	39.36	-2.41	0.01
BGIS0004952	Bio18	-94.83	60.58	45.77	-2.87	0.01
BGIS0004952	RadSolPrimMed	-95.64	58.96	41.26	-36.92	2.30
BGIS0004952	RadSolPrimMediana	-95.75	58.75	42.09	-30.26	1.87
BGIS0004952	RadSolNov	-95.87	58.51	41.34	-30.74	1.90
BGIS0004952	RadSolJAn	-96.23	57.80	41.04	-25.05	1.54
BGIS0004952	RadSolDez	-96.57	57.11	40.59	-20.93	1.29

Table 2: Biological function of markers identified as selection signatures in Brazilian locally adapted swine breeds

Marker	Chr	Most severe consequence	Gene	Markers is between (in reverse strand)			
				Gene (SNP distance)	Function	Gene(SNP distance)	Function
ALGA0032795	5	Intergenic		ENSSSCG00000024523 (±0.003 Mb)	transmembrane transporter activity	ENSSSCG00000000783 (±0.3 Mb)	glucose transmembrane transporter activity
		Upstream gene	ENSSSCG00000015405	cell surface receptor signaling pathway /immune response/			
ALGA0054315	9	Intergenic		ENSSSCG00000015405 (± 0.8Mb)	cell surface receptor signaling pathwaySource: InterPro	ENSSSCG00000015406 (± 0.08 Mb)	immune response
ASGA0026250	5	Intergenic		ENSSSCG00000024523 (± 0.05 Mb)	glucose transmembrane transporter activity	ENSSSCG00000000783 (± 0.1Mb)	glucose transmembrane transporter activit/ sugar:proton symporter activity.
ASGA0029202	6	Intergenic		ENSSSCG000000003720 (±0.6 Mb)	transporter activity	ENSSSCG000000003722 (±0.2 Mb)	calcium ion binding (blood vessel morphogenesis....)
BGIS0004952	8	Intergenic	COMMD8	ENSSSCG00000019505 (±0.1 Mb)	RNA genes	ENSSSCG00000019140 (±0.0075 Mb)	RNA genes
MARC0021990	4	Intergenic		ENSSSCG00000022092 / ENSSSCG00000006222 (±0.2 Mb)	Iron (heme axial ligand)/RNA polymerase II transcription factor activity, sequence-specific DNA binding		
ASGA0033717	7	Intron	ENSSSCG00000001771	Rho guanyl-nucleotide exchange factor activity			
MARC0007678	3	Intergenic		ENSSSCG00000008605 (±0.4 Mb)	Uncharacterized protein	ENSSSCG00000008606 (±0.0001 Mb)	metal ion binding

Source: [www.ensembl.org](http://www.ensembl.org) and <http://www.ncbi.nlm.nih.gov/>

Table 3: Regional frequencies for genotypes from markers detected as signature of selection.

Marker		ALGA0032795			ALGA0054315			ASGA0026250			ASGA0029202			BGIS0004952		
Marker	Genotype	GG	GA	AA	TT	CT	CC	TT	CT	CC	AA	AC	CC	TT	CT	CC
Region	North	2%	14%	35%	2%	17%	26%	35%	13%	2%	2%	20%	50%	4%	35%	21%
	Northeast	20%	39%	42%	26%	49%	30%	42%	40%	19%	27%	43%	42%	20%	33%	76%
	Midwest	28%	32%	20%	24%	19%	34%	20%	31%	29%	39%	20%	3%	41%	17%	3%
	Southeast	8%	2%	0%	3%	0%	5%	0%	1%	8%	5%	2%	0%	5%	2%	0%
	South	42%	14%	4%	45%	15%	4%	4%	13%	42%	28%	14%	6%	30%	13%	0%
Genotype frequency Total		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table S1: Number of swine breeds and individuals per breed sampled per Brazilian State

		BREED <sup>1</sup>																	Total
		SBA	SCB	SCR	SCT	SDC	SLD	SLW	SMA	SME	SMEc	SMO	SMT	SNI	SPE	SPI	SRB	SUR	Total
S T A T E <sup>2</sup>	PE	10	0	4	18	0	3	0	0	5	0	8	0	9	0	8	0	1	66
	GO	0	4	3	0	0	3	1	0	0	0	1	5	3	0	8	2	0	30
	PA	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	29
	SC	0	1	0	0	4	7	2	0	0	0	9	0	0	0	0	0	0	23
	MS	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	18
	RS	0	0	3	0	0	0	0	0	1	0	4	0	5	0	1	0	0	14
	BA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3
	MG	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	3
	MT	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	3
	PB	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
Total		10	5	10	18	4	13	3	29	6	2	22	26	17	2	21	2	1	191
%		5.24	2.62	5.24	9.42	2.09	6.81	1.57	15.18	3.14	1.05	11.52	13.61	8.9	1.05	10.99	1.05	0.52	

1- SBA= Baé; SCB= Casco de Burro; SCR= Caruncho; SMEc= Crioulo; SCT= Canastra; SCTA= Canastrão; SDC= Duroc; SLD= Landrace; SLW= Large Withe; SMA= Marajó; SME= Mestiço; SMO= Moura; SMT= Monteiro; SNI= Nilo; SPI= Piau; SPN= Pietran; SRP= Rabo de Peixe

2- PA=Pará; BA=Bahia; PE=Pernambuco; PB=Paraíba; GO=Goiás; MS=Mato Grosso do Sul; MT=Mato Grosso; MG= Minas Gerais; RS=Rio Grande do Sul; SC=Santa Catarina.

Table S2: Environmental variables used in the present study (Geographic Data Base).

Data Base	Environmental Variable	Data obtained	Data Calculated
Harvest Choise	Aridity	Ratio of precipitation to PET	-
	PETannual	Annual potential evaporo-transpiration	-
World Clim	Temperature	Monthly maximum temperature	Means and medians from annual and seasonal periods
		Monthly average temperature	Means and medians from annual and seasonal periods
		Monthly minimum temperature	Means and medians from annual and seasonal periods
	Solar Radiation	Monthly solar radiation	Means and medians from annual and seasonal periods
	Precipitation	Monthly precipitation	Means and medians from annual and seasonal periods
Gtopo30	Bioclimatic	Bio1-19	-
	Elevation	Elevation	-

Data from last 50 years, resolution 30 arc.

PETannual=annual potential evaporo-transpiration; Aridity= ratio of precipitation to PET

Source: <http://www.worldclim.org>; <https://lta.cr.usgs.gov/GTOPO30>; <http://harvestchoise.org>.

Table S3: AMOVA values among Brazilian regions.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	4	90740.541	46.30876Va	0.87
Among populations within groups	5	57448.151	309.73284Vb	5.79
Within populations	352	1759100.502	4997.44461Vc	93.35
Total	361	1907289.193	5353.48620	

Groups were formed with any animal sampled in a state within regions. Group 1: South Region (Santa Catarina and Rio Grande do Sul); Group 2: Midwest Region (Goiais, Mato Grosso and Mato Grosso do Sul); Group 3: Southeast Region (Minas Gerais); Group 4: Northeast Region (Bahia, Paraíba and Pernambuco); Group 5 : North Region (Pará)

Vc; Vb: P(rand. value > obs. value) = 0.00000, P-value = 0.00000+-0.00000

Va: P(rand. value > obs. value) = 0.24772, P-value = 0.24772+-0.00443

Table S4: AMOVA values among commercial and locally adapted genetic groups

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	26664.93	200.6518Va	3.66
Among populations within groups	16	191515.1	340.3295Vb	6.22
Among individuals with in populations	163	877994.2	452.5846Vc	8.27
Within individuals	181	811115	4481.298Vd	81.85
Total	361	1907289.193	5474.864	

Group 1=Brazilian Locally adapted genetic groups (Bae, Canastra, Canastrao, Caruncho, Casco de Burro, Crioulo, Mammelado, Marajoara, Mestico, Monteiro, Moura, Nilo, Piau, Rabo de Peixe)

Group 2= Globally commercial breeds (Large White, Duroc, Pietran, Landrace)

Vd; Vc; Vb: P(rand. value < obs. value) = 0.00000; P-value = 0.00000+-0.00000.

Va: P(rand. value > obs. value) = 0.00574; P-value = 0.00574+-0.00071



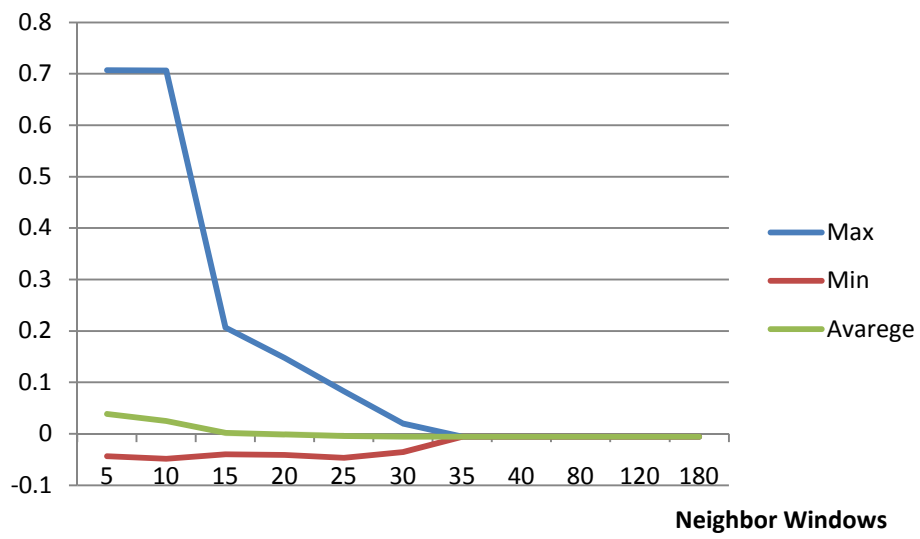


Figure 2: Correlogram of Maximum (Max), Minimum (Min) and Average values of Moran's I from all markers.



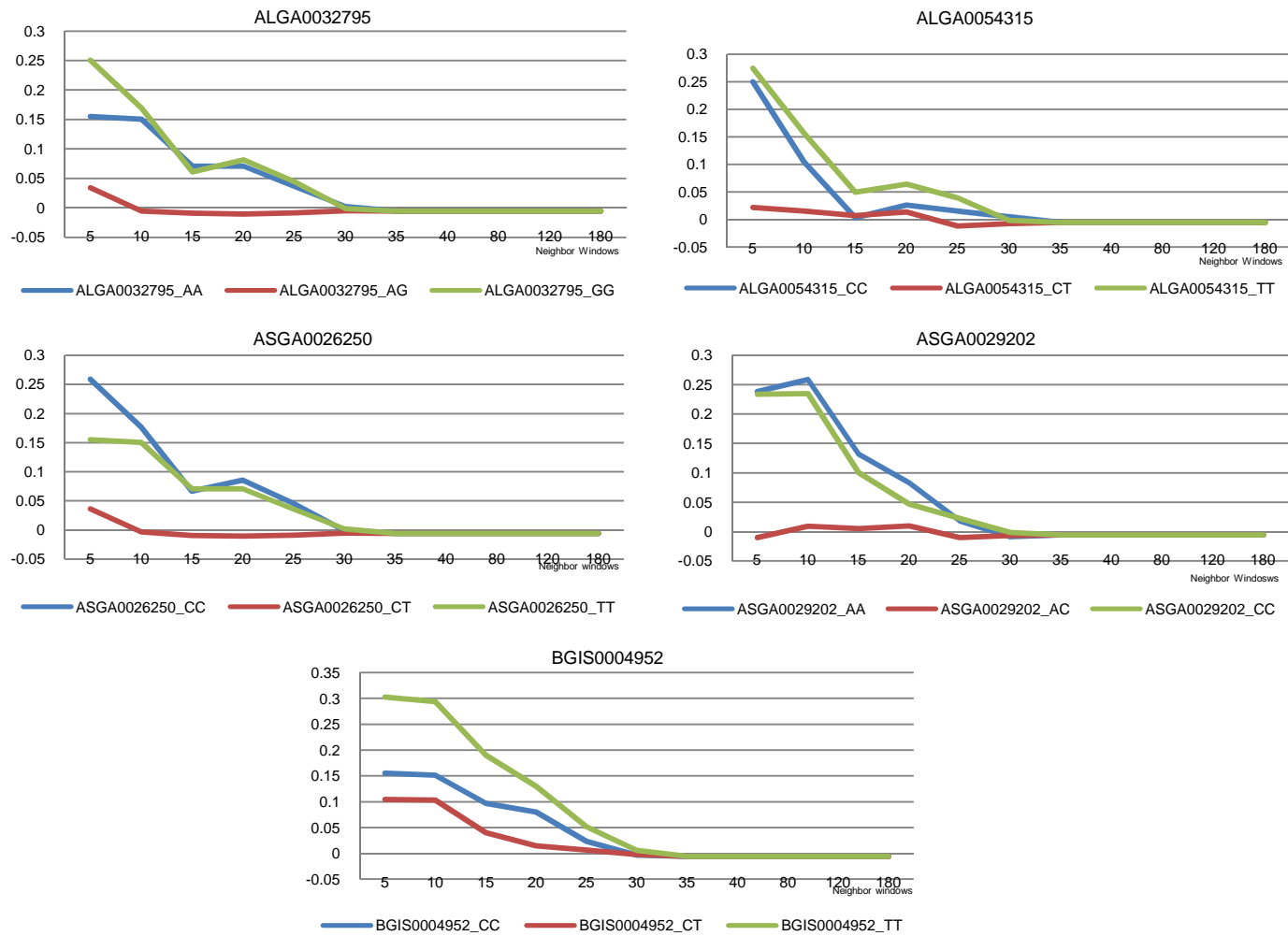


Figure 3: Correlogram from Genotypes of the markers identified as selection signatures in Brazilian locally adapted swine breeds by BayeScan and Samβada.

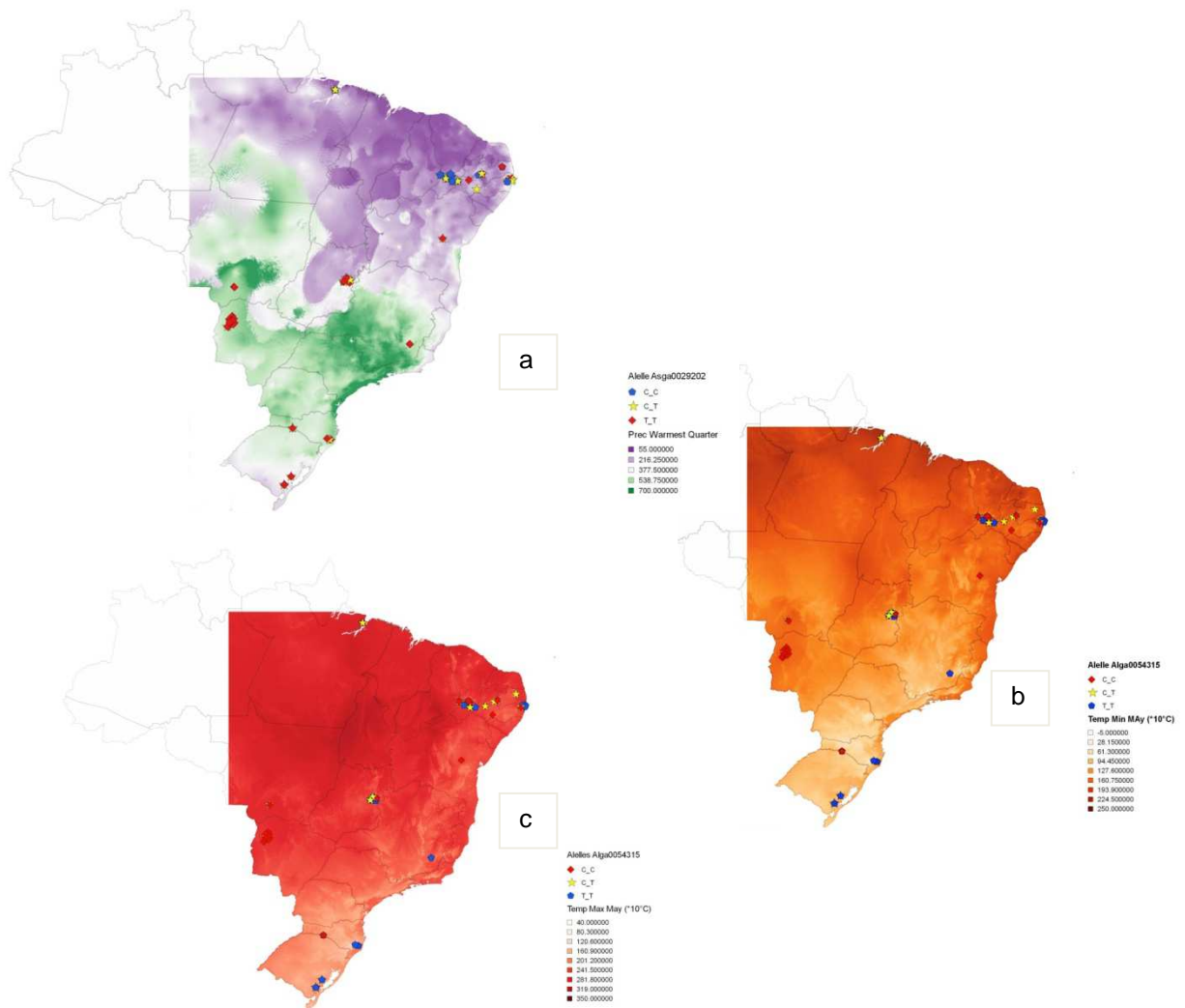


Figure 4: Genotype distribution maps from markers identified as selection signatures by Samβada/BayeScan. (a) Marker Asga0029202 in Bio18 (precipitation of Warmest Quarter). (b) Marker Alga0054315 in Minimum Temperature in May and (c) Marker Alga0054315 in Maximum Temperature in May.

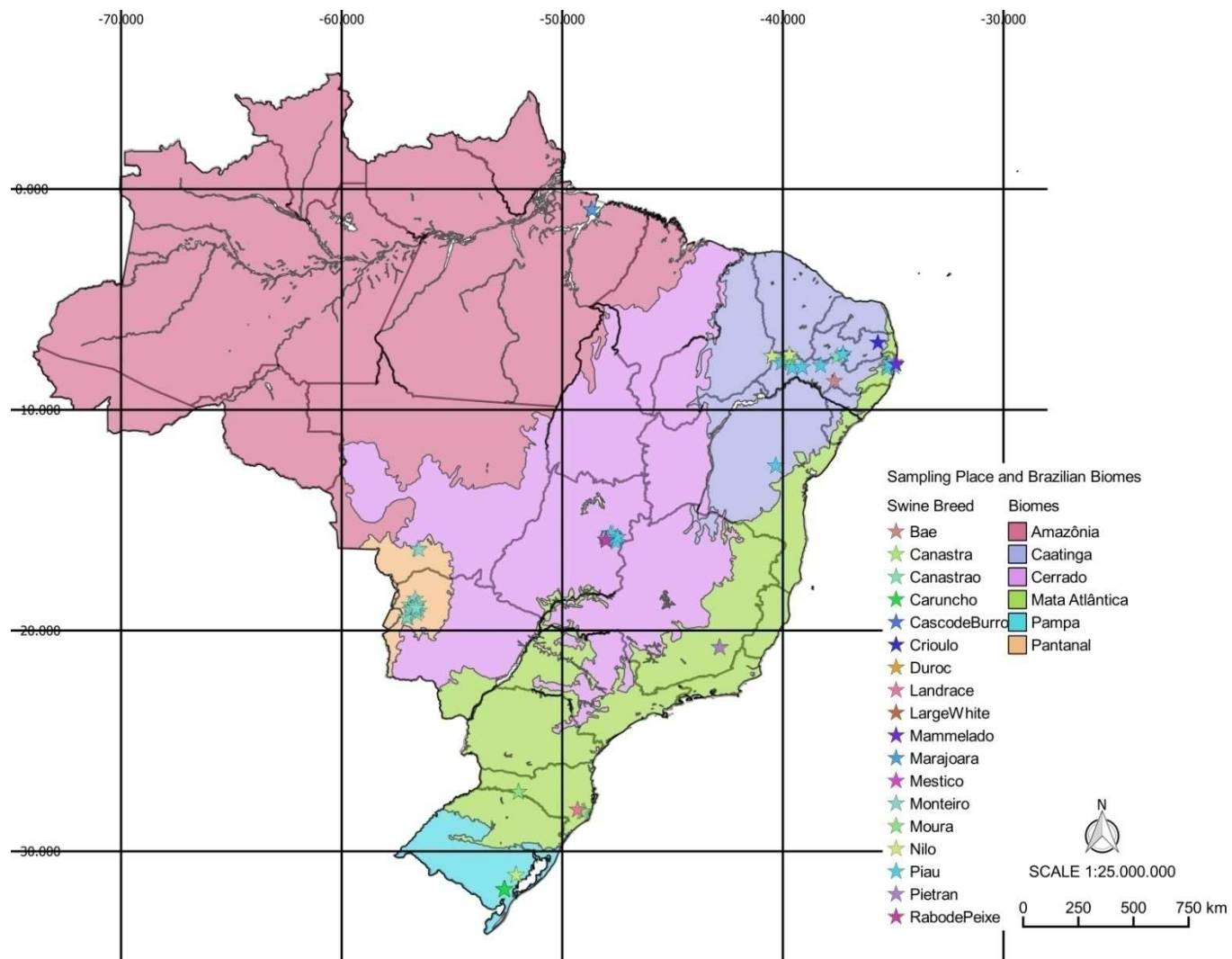


Figure S1: Sampling location for Brazilian locally adapted swine breeds with biomes and states on the Brazilian territory.

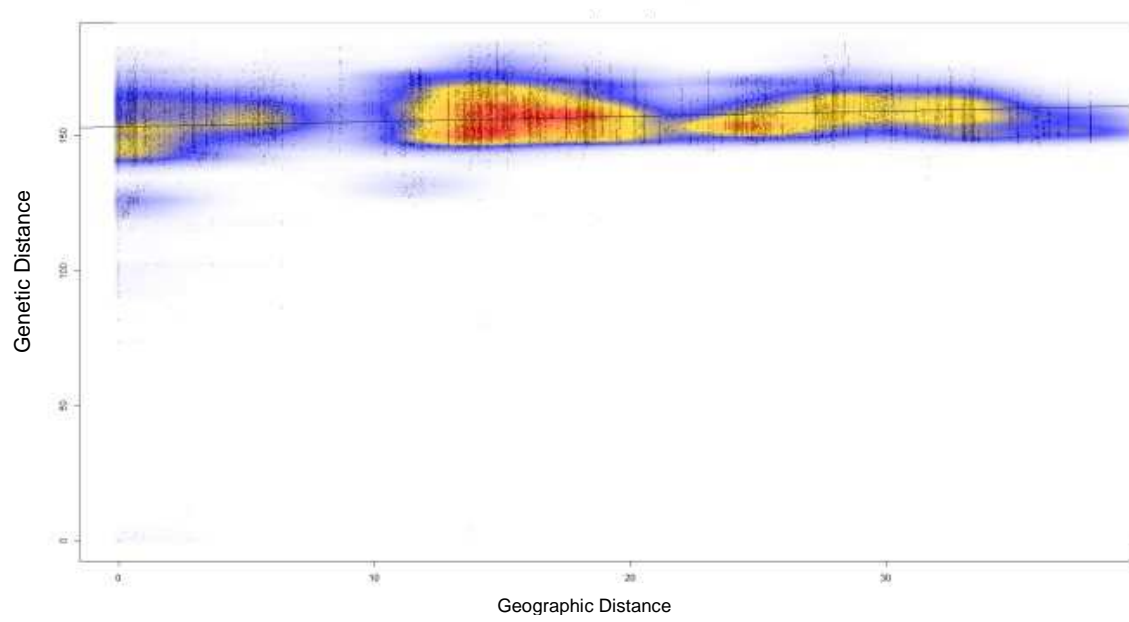


Figure S2: Density of points (samples) around Mantel correlation line.

## Acknowledgements

We wish to thank Centro Nacional de Recursos Geneticos e Biotecnologia for the provision of DNA samples from Brazilian Pig Germoplasm Bank as well as Elizabete C. Silva and Gleison R. De Biazio for their support in the preparation and submission of material for genotyping and Kevin Leempoel for their support in the analyses. Thanks also to CAPES, CNPq and INCT-Pecuária for support of this research.

## Bibliography

- Ai H, Huang L, Ren J (2013) Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers (CA Kozak, Ed.). *PLoS ONE*, **8**, e56001.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Beja-Pereira A, Luikart G, England PR *et al.* (2003) Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genetics*, **35**, 311–313.
- Bradbury D, Smithson A, Krauss SL (2013) Signatures of diversifying selection at EST-SSR loci and association with climate in natural Eucalyptus populations. *Molecular Ecology*, **22**, 5112–5129.
- Burgos-Paz W, Souza C, Megens H *et al.* (2012) Porcine colonization of the Americas: a 60k SNP story.
- Collier RJ, Gebremedhin KG (2015) Thermal Biology of Domestic Animals. *Annual Review of Animal Biosciences*, **3**, 513–532.
- De AK, Kundu A, Kundu MS *et al.* (2013) Genetic characterization of Andaman Desi pig, an indigenous pig germplasm of Andaman and Nicobar group of islands, India by microsatellite markers. *Veterinary World*, **6**, 750–753.
- Druck S, Carvalho MS, Câmara G, Monteiro AMV (2004) *Análise espacial de dados geográficos*. Planaltina, DF.

- Excoffier L, Laval G, Stefan S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **2005:1**, 47–50.
- Favero JA, Figueiredo EAP de (2009) Evolução do melhoramento genético de suínos no Brasil. *Ceres*, **56**.
- Feng X-J, Jiang G-F, Fan Z (2015) Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. *Scientific Reports*, **5**, 13758.
- Flowers W (2015) Factors Affecting the Efficient Production of Boar Sperm. *Reproduction in Domestic Animals*, **50**, 25–30.
- Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, **180**, 977–993.
- Gama LT, Martínez AM, Carolino I *et al.* (2013) Genetic structure, relationships and admixture with wild relatives in native pig breeds from Iberia and its islands. *Genet Sel Evol*, **45**, e18.
- Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167–178.
- Giovambattista G, Ripoli MV, Peral-Garcia P, Bouzat JL (2001) Indigenous domestic breeds as reservoirs of genetic diversity: the Argentinean Creole cattle. *Animal Genetics*, **32**, 240–247.
- Golden Helix, Inc., Bozeman, MT, USA (2015) *SNP & Variation Suite*. Golden Helix, Inc, Bozeman, MT.
- Hall SJG, Ruane J (1993) Livestock Breeds and Their Conservation: A Global Overview. *Conservation Biology*, **7**, 815–825.
- Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.
- Johnson JS, Boddicker RL, Sanz-Fernandez MV *et al.* (2013) Effects of mammalian in utero heat stress on adolescent body temperature. *International Journal of Hyperthermia*, **29**, 696–702.

- Johnson JS, Fernandez MVS, Gutierrez NA *et al.* (2015a) Effects of in utero heat stress on postnatal body composition in pigs: I. Growing phase. *Journal of Animal Science*, **93**, 71–81.
- Johnson JS, Sanz Fernandez MV, Patience JF *et al.* (2015b) Effects of mammalian in utero heat stress on adolescent body temperature. *Journal of Animal Science*, **93**, 82–92.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.
- Lee DHK, Phillips RW (1948) ASSESSMENT OF THE ADAPTABILITY OF LIVESTOCK TO CLIMATIC STRESS. *Journal of Animal Science*, **7**, 391–425.
- Lewontin RC, Krakauer J (1973) DISTRIBUTION OF GENE FREQUENCY AS A TEST OF THE THEORY OF THE SELECTIVE NEUTRALITY OF POLYMORPHISMS. *Genetics*, **74**, 175–195.
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.
- Luijkart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet*, **4**, 981–994.
- Manel S, Gugerli F, Thuiller W *et al.* (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–3738.
- Manthey JD, Moyle RG (2015) Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*, **24**, 3628–3638.
- Mariante a. d. S, Castro STR, Albuquerque M do SM, Paiva SR, Germano JL (2003) PIG BIODIVERSITY IN BRAZIL. *Arch. Zootec*, **52**, 245–248.
- Mirkena T, Duguma G, Haile A *et al.* (2010) Genetics of adaptation in domestic farm animals: A review. *Livestock Science*, **132**, 1–12.

- Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.
- Nimer E (1979) Um modelo metodológico da classificação de climas. *Revista Brasileira de Geografia*, **41**, 59–89.
- Nteeba J, Sanz-Fernandez MV, Rhoads RP *et al.* (2015) Heat Stress Alters Ovarian Insulin-Mediated Phosphatidylinositol-3 Kinase and Steroidogenic Signaling in Gilt Ovaries. *Biology of Reproduction*, **92**, 148.
- Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers: Comparing methods to detect selective loci. *Journal of Evolutionary Biology*, **23**, 2267–2276.
- Petrocelli H, Batista C, Gosálvez J (2015) Seasonal variation in sperm characteristics of boars in southern Uruguay. *Revista Brasileira de Zootecnia*, **44**, 1–7.
- Pham LD, Do DN, Nam LQ *et al.* (2014) Molecular genetic diversity and genetic structure of Vietnamese indigenous pig populations. *Journal of Animal Breeding and Genetics*, **131**, 379–386.
- QGIS Development Team (2009) *QGIS Geographic Information System*. Open Source Geospatial Foundation.
- Rangel TF, Diniz-Filho JAF, Bini LM (2010) SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography*, **33**, 46–50.
- Ross JW, Hale BJ, Gabler NK *et al.* (2015) Physiological consequences of heat stress in pigs. *Animal Production Science*, **55**, 1381–1390.
- SAS Institute Inc. (2011) *SAS/STAT® 9.3 User's Guide*. SAS Inst., Cary, NC.
- Shabtay A (2015) Adaptive traits of indigenous cattle breeds: The Mediterranean Baladi as a case study. *Meat Science*, **109**, 27–39.
- Storz JF (2005) INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Storz JF (2010) Genes for High Altitudes. *Science (New York, N. Y.)*, **329**, 40–41.



- Stucki S (2014) Développement d'outils de géo-calcul haute performance pour l'identification de régions du génome potentiellement soumises à la sélection naturelle: analyse spatiale de la diversité de panels de polymorphismes nucléotidiques à haute densité (800k) chez *Bos taurus* et *B. indicus* en Ouganda. EPFL, Suisse.
- Stucki S, Orozco-terWengel P, Bruford MW *et al.* (2014) High performance computation of landscape genomic models integrating local indices of spatial association. *arXiv preprint arXiv:1405.7658*.
- Vitalis R, Gautier M, Dawson KJ, Beaumont MA (2014) Detecting and Measuring Selection from Gene Frequency Data. *Genetics*, **196**, 799–817.
- Wettemann RP, Bazer FW (1985) Influence of environmental temperature on prolificacy of pigs. *Journal of Reproduction and Fertility. Supplement*, **33**, 199–208.
- Wildt DE, Riegler GD, Dukelow WR (1975) Physiological temperature response and embryonic mortality in stressed swine. *The American Journal of Physiology*, **229**, 1471–1475.

#### **CAPITULO IV**

“...Ouço o mugindo, o barulho da cordeona  
E a velha porca rabona, retouçando no salão,...”  
(*Knelmo Alves, Francisco Alves*)

#### 4.1 CONSIDERAÇÕES FINAIS

Desde que o homem domesticou os primeiros animais, há milhares de anos atrás, a diversidade genética daquelas populações tem sido dividida em diferentes raças, de forma a atender as necessidades locais das populações humanas. De forma geral, as raças locais são fruto de um longo período de seleção ambiental, sendo criadas pelas comunidades locais, praticamente sem manejo especializado. Com a globalização da produção mundial, muitas destas raças localmente adaptadas, têm sido substituídas pelas raças mundiais, mais produtivas, porém, mais exigentes quanto ao manejo, alimentação e condições ambientais. Esse é o caso das raças suínas brasileiras que deixaram de ser a preferência dos consumidores e, por consequência, dos produtores, anos atrás. Este processo de substituição teve como consequências, não somente a extinção de várias raças locais ao redor do mundo, mas a perda de uma parcela da diversidade genética total desta importante espécie doméstica e, que dificilmente será recuperada. Esta preocupação com a erosão dos recursos genéticos mundiais é antiga, sendo demonstrada já em 1993 com a ratificação da “Convention on Biological Diversity,” onde um consenso internacional sobre a conservação da diversidade das espécies zootécnicas foi estabelecido.

Por outro lado, raças locais/regionais são utilizadas mundialmente na elaboração de produtos com certificação de origem protegida (PDO, PGI e TSG). Ainda que famosos, muitos destes produtos são oriundos de rebanhos pequenos e ameaçados. Fraudes e incertezas, quanto à pureza racial dos rebanhos, são uma constante que pode ser resolvida com o uso da informação genética dos rebanhos e regiões.

O uso de informações genéticas georeferenciadas e, por conseguinte, da genética de paisagem, permite identificar marcas de seleção: associações entre o genoma e o meio ambiente, deixadas durante o processo de adaptação. Essas informações podem ser usadas tanto em programas de melhoramento das raças comerciais, pela inclusão destes marcadores na matriz de seleção genômica, como em programas de conservação de recursos genéticos, principalmente daquelas raças sob risco de extinção pela

identificação dos alelos favoráveis as condições ambientais de cada unidade de manejo ou conservação.

As análises realizadas, utilizando a genotipagem de 60k de SNPs de animais das raças localmente adaptadas de suínos brasileiros, mostraram que, apesar do grande número de raças, diversidade genética e de haver estruturação nos diversos níveis populacionais estudados, existe certa similaridade genética dentre as raças suínas brasileiras. Podemos observar também, que a maior fonte de variabilidade genética foi o indivíduo dentro das raças, sendo possível a diferenciação genética dos grupos. Porém, em alguns casos, essa diferenciação genética entre os grupos pode ser muito sutil, pois as frequências alélicas dos marcadores foram muito similares e o grau de separação/diferenciação entre as populações, pequeno. Níveis elevados de endogamia foram encontrados em algumas populações, e se tratando de populações isoladas ou grupos pequenos de animais, tem como consequência depressão endogâmica e, risco de extinção. Este conhecimento sobre as frequências alélicas deve ser levado em conta na formação de rebanhos de conservação, evitando endogamia e respeitando as diferenças regionais das frequências alélicas.

A presença de associações genéticas com precipitação, radiação solar e temperatura ambiental, além da divisão do território em duas metades pelas frequências alélicas, confirma a influência do meio ambiente no processo de adaptação dos suínos ao território brasileiro. Considerando os resultados obtidos, o uso das assinaturas de seleção (presentes no genoma das raças suínas nacionais) pode ser incluído nos programas de melhoramento suíno tradicionais. A região de destino das matrizes produzidas nos núcleos de genética pode ser orientada pelas assinaturas de seleção identificadas em associadas ao meio ambiente, da mesma forma, a frequência destes alelos deve ser levada em conta na formação de unidade de manejo e conservação das raças nacionais.

Estudos adicionais, com uma amostragem mais abrangente, com a presença de dados fenotípicos de produção e caracterização racial se fazem necessários para uma melhor distinção genética dos grupos raciais, assim

como, melhorar a acurácia e precisão na identificação das assinaturas de seleção.

## 4.2 REFERÊNCIAS BIBLIOGRÁFICAS

ASSOCIAÇÃO BRASILEIRA DE CRIADORES DE SUÍNOS. Relatório do serviço de registro genealógico de suínos 2014. Estrela: ABCS, 2014. p. 46.

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*, San Francisco, v. 8, n. 2, p. e56001, 2013.

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

BARKER, H. R.; BARKER, B. M. *Multivariate analysis of variance (manova): a practical guide to its use in scientific decision-making*. [S.l.]: University of Alabama Press, 1984.

BEAUMONT, M. A.; BALDING, D. J. Identifying adaptive genetic divergence among populations from genome scans. *Molecular ecology*, Oxford, v. 13, n. 4, p. 969–980, 2004.

BEIGUELMAN, B. *Genética de populações humanas*. Ribeirão preto: sbg, 2008. p. 483–488.

BEJA-PEREIRA, A. et al. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature genetics*, New York, v. 35, n. 4, p. 311–313, 2003.

BLACKBURN, H. Genetic selection and conservation of genetic diversity\*: genetic selection and conservation of genetic diversity. *Reproduction in domestic animals*, Berlin, v. 47, suppl. 4, p. 249–254, 2012.

BOITARD, S. et al. Genetic variability, structure and assignment of spanish and french pig populations based on a large sampling: variability of spanish and french pig populations. *Animal genetics*, Oxford, v. 41, n. 6, p. 608–618, 2010.

BRADBURY, D.; SMITHSON, A.; KRAUSS, S. L. Signatures of diversifying selection at est-ssr loci and association with climate in natural eucalyptus populations. *Molecular ecology*, Oxford, v. 22, n. 20, p. 5112–5129, 2013.

BURGOS-PAZ, W. et al. Porcine colonization of the americas: a 60k snp story. *Heredity (Edinb)*, London, v. 110, n. 4, p. 321-330, 2012. Disponível em: <<http://dspace.unl.edu.ec/jspui/handle/123456789/291>>. Acesso em: 17 mar. 2016.

CAÑÓN, J. et al. Genetic diversity measures of local european beef cattle breeds for conservation purposes. *Genetics selection evolution*, v. 33, n. 3, p. 311–332, 2001.

- CAVALCANTI, S. M. C.; GESE, E. M. Kill rates and predation patterns of jaguars ( *panthera onca* ) in the southern pantanal, brazil. *Journal of mammalogy*, Lawrence, v. 91, n. 3, p. 722–736, 2010.
- CAVALCANTI, S. S. *Produção de suínos*. 2. ed. Campinas: Instituto Campineiro de Ensino Agrícola, 1985.
- CERQUEIRA, R. et al. *Fragmentação de ecossistemas: causas, efeitos sobre a biodiversidade e recomendações de políticas públicas*. Brasília: MMA/SBF, 2003, p. 510.
- CHARLESWORTH, B. *Evolution in age-structured populations*. 2. ed. Cambridge: Cambridge University Press, 1994.
- COLLIER, R. J.; GEBREMEDHIN, K. G. Thermal biology of domestic animals. *Annual review of animal biosciences*, Palo Alto, v. 3, n. 1, p. 513–532, 2015.
- CROSBY, A. W. *The columbian exchange: biological and cultural consequences of 1492*. [S.l.]: Greenwood Publishing Group, 2003.
- DE, A. K. et al. Genetic characterization of andaman desi pig, an indigenous pig germplasm of andaman and nicobar group of islands, india by microsatellite markers. *Veterinary world*, v. 6, n. 10, p. 750–753, 2013.
- DE GRUIJTER, J. J.; TER BRAAK, C. J. F. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical geology*, New York, v. 22, n. 4, p. 407–415, 1990.
- DESBIEZ, A. L. J. et al. *Ecologia de populações de porco monteiro no pantanal do brasil*. Corumbá, MS: Embrapa, 2009. p. 40. (Documentos, 106).
- DINIZ-FILHO, J. A. F. et al. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and molecular biology*, Ribeirão Preto, v. 32, n. 2, p. 203–211, 2009.
- DOBZHANSKY, T. *Evolution in the tropics*, by theodosius dobzhansky. *American scientist*, New Haven, v. 38, p. 209-221, 1950.
- DRUCK, S. et al. *Análise espacial de dados geográficos*. Planaltina, DF: Embrapa, 2004.
- DRUML, T. et al. Genetic diversity, population structure and subdivision of local balkan pig breeds in austria, croatia, serbia and bosnia-herzegovina and its practical value in conservation programs. *Genetics selection evolution*, Paris, v. 44, n. 5, p. 1–9, 2012.
- EMIGH, T. H. A comparison of tests for hardy-weinberg equilibrium. *Biometrics*, Alexandria, v. 36, n. 4, p. 627-642, 1980.

- EXCOFFIER, L.; LAVAL, G.; STEFAN, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, Auckland, v. 1, p. 47–50, 2005.
- FAHRENKRUG, S. C. et al. Single nucleotide polymorphism (snp) discovery in porcine expressed genes. *Animal genetics*, Oxford, v. 33, n. 3, p. 186–195, 2002.
- FAVERO, J. A.; FIGUEIREDO, E. A. P. De. Evolução do melhoramento genético de suínos no brasil. *Ceres*, Viçosa, v. 56, n. 4, p. 420-427, 2009. Disponível em: <<http://www.ceres.ufv.br/ojs/index.php/ceres/article/view/3447>>. Acesso em: 17 mar. 2016.
- FENG, X.-J.; JIANG, G.-F.; FAN, Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *ceracris kiangsu*. *Scientific reports*, London, v. 5, p. 13758, 2015.
- FLOWERS, W. Factors affecting the efficient production of boar sperm. *Reproduction in domestic animals*, Berlin, v. 50, suppl. 2, p. 25–30, 2015.
- FOLL, M.; GAGGIOTTI, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, Baltimore, v. 180, n. 2, p. 977–993, 2008.
- FORD, S. P. Embryonic and fetal development in different genotypes in pigs. *Journal of reproduction and fertility. supplement*, Cambridge, v. 52, p. 165–176, 1997.
- FREIRE-MAIA, N. *Genética de populações humanas*. [S.l.]: Hucitec, 1974.
- GAMA, L. T. et al. Genetic structure, relationships and admixture with wild relatives in native pig breeds from iberia and its islands. *Genetics selection evolution*, v. 45, n. 1, 2013.
- GIENAPP, P. et al. Climate change and evolution: disentangling environmental and genetic responses. *Molecular ecology*, Oxford, v. 17, n. 1, p. 167–178, 2008.
- GIOVAMBATTISTA, G. et al. Indigenous domestic breeds as reservoirs of genetic diversity: the argentinean creole cattle. *Animal genetics*, Oxford, v. 32, n. 5, p. 240–247, 2001.
- GOLDBERG, C. S.; WAITS, L. P. Using habitat models to determine conservation priorities for pond-breeding amphibians in a privately-owned landscape of northern idaho, usa. *Biological conservation*, Essex, v. 142, n. 5, p. 1096–1104, 2009.
- GOLDEN HELIX. *Snps & variation suite*. Bozeman, MT: Golden Helix, Inc, 2015.



- GREENBERG, J. A.; DOBROWSKI, S. Z.; USTIN, S. L. Shadow allometry: estimating tree structural parameters using hyperspatial image analysis. *Remote sensing of environment*, New York, v. 97, n. 1, p. 15–25, 2005.
- GROENEN, M. A. M. et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, Basingstoke, v. 491, n. 7424, p. 393–398, 2012.
- HALL, L. A.; BEISSINGER, S. R. A practical toolbox for design and analysis of landscape genetics studies. *Landscape ecology*, Dordrecht, v. 29, n. 9, p. 1487–1504, 2014.
- HALL, S. J. G.; RUANE, J. Livestock breeds and their conservation: a global overview. *Conservation biology*, Boston, v. 7, n. 4, p. 815–825, 1993.
- HAMRICK, J. L. Plant population genetics and evolution. *American journal of botany*, Lancaster, v. 69, n. 10, p. 1685–1693, 1982.
- HANOTTE, O.; DESSIE, T.; KEMP, S. Time to tap africa's livestock genomes. *Science*, Whashington, v. 328, n. 5986, p. 1640–1641, 2010.
- HOFFMANN, A. A.; WILLI, Y. Detecting genetic responses to environmental change. *Nature reviews genetics*, London, v. 9, n. 6, p. 421–432, 2008.
- HOHN, M. E. An introduction to applied geostatistics. *Computers & geosciences*, New York, v. 17, n. 3, p. 471–473, 1991.
- IBGE. Instituto brasileiro de geografia e estatística., [S.l.], [s.d.]. Disponível em: <<http://www.ibge.gov.br/home/>>. Acesso em: 17 jun. 2016.
- JOHNSON, J. S. et al. Effects of mammalian in utero heat stress on adolescent body temperature. *International journal of hyperthermia*, London, v. 29, n. 7, p. 696–702, 2013.
- JOHNSON, J. S. et al. Effects of in utero heat stress on postnatal body composition in pigs: i. growing phase. *Journal of animal science*, Champaign, v. 93, n. 1, p. 71–81, 2015.
- JOMBART, T. Adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics*, Oxford, v. 24, n. 11, p. 1403–1405, 2008.
- JOOST, S. et al. A spatial analysis method (sam) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular ecology*, Oxford, v. 16, n. 18, p. 3955–3969, 2007.
- JOOST, S. et al. Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Animal genetics*, Oxford, v. 41, suppl. 1, p. 47–63, 2010.

KAWECKI, T. J.; EBERT, D. Conceptual issues in local adaptation. *Ecology letters*, Oxford, v. 7, n. 12, p. 1225–1241, 2004.

KIM, T. H. et al. Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. *Journal of animal science*, Champaign, v. 83, n. 10, p. 2255–2263, 2005.

KIRK, H.; FREELAND, J. R. Applications and implications of neutral versus non-neutral markers in molecular ecology. *International journal of molecular sciences*, Basel, v. 12, n. 6, p. 3966–3988, 2011.

LEE, D. H. K.; PHILLIPS, R. W. Assessment of the adaptability of livestock to climatic stress. *Journal of animal science*, Champaign, v. 7, n. 4, p. 391–425, 1948.

LEEMPOEL, K. et al. Very high-resolution digital elevation models: are multi-scale derived variables ecologically relevant? *Methods in ecology and evolution*, v. 6, n. 12, p. 1373–1383, 2015.

LEGENDRE, P.; LEGENDRE, L. *Numerical ecology*. 2nd. ed. Amsterdam: Elsevier Science B.V., 1998.

LEWONTIN, R. C.; KRAKAUER, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, v. 74, n. 1, p. 175–195, 1973.

LOTTERHOS, K. E.; WHITLOCK, M. C. Evaluation of demographic history and neutral parameterization on the performance of *fst* outlier tests. *Molecular ecology*, Oxford, v. 23, n. 9, p. 2178–2192, 2014.

LUIKART, G. et al. The power and promise of population genomics: from genotyping to genome typing. *Nature reviews genetics*, London, v. 4, n. 12, p. 981–994, 2003.

LYNCH, M.; CONERY, J.; BURGER, R. Mutation accumulation and the extinction of small populations. *American naturalist*, Chicago, v. 145, n. 4, p. 489–518, 1995.

MANEL, S. et al. Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution*, Amsterdam, v. 18, n. 4, p. 189–197, 2003.

MANEL, S. et al. A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular ecology*, Oxford, v. 16, n. 10, p. 2031–2043, 2007.

MANEL, S. et al. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field: adaptive genetic variation and landscape genetics. *Molecular ecology*, Oxford, v. 19, n. 17, p. 3760–3772, 2010.

- MANEL, S. et al. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular ecology*, Oxford, v. 21, n. 15, p. 3729–3738, 2012.
- MANEL, S.; HOLDEREGGER, R. Ten years of landscape genetics. *Trends in ecology & evolution*, Barking, v. 28, n. 10, p. 614–621, 2013.
- MANTHEY, J. D.; MOYLE, R. G. Isolation by environment in white-breasted nuthatches (*sitta carolinensis*) of the madrean archipelago sky islands: a landscape genomics approach. *Molecular ecology*, Oxford, v. 24, n. 14, p. 3628–3638, 2015.
- MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. *Nature reviews genetics*, London, v. 11, n. 7, p. 499–511, 2010.
- MARIANTE, A. D. S. et al. Pig biodiversity in brazil. *Arch. zootec*, v. 52, p. 245–248, 2003.
- MARIANTE, A. Da S.; CAVALCANTE, N. Animais do descobrimento: raças domésticas da história do brasil. 2. ed. Brasília: Embrapa Sede / Embrapa Recursos Genéticos e Biotecnologia, 2006.
- MARTÍNEZ, A. M. et al. The southwestern fringe of europe as an important reservoir of caprine biodiversity. *Genetics selection evolution*, Paris, v. 47, p. 86, 2015.
- MATHERON, G. Principles of geostatistics. *Economic geology*, Lancaster, v. 58, n. 8, p. 1246–1266, 1963.
- MAURO, R. Faunistic studies in the embrapa pantanal. *Arch. zootec*, v. 51, p. 175–185, 2002.
- MAURO, R.; PEREIRA, M.; PAES, R. Perspectivas da criação de porcos monteiro no Brasil. In: SIMPÓSIO SOBRE ANIMAIS SILVESTRES E EXÓTICOS, 41., 2004, Campo Grande - MS. Anais eletrônicos... Campo Grande - MS: [s.n.], 2004. p. 511–516. Disponível em: <<http://www.sbz.org.br/reuniaoanual/anais/?idiom=pt>>. Acesso em: 17 mar 2016
- MCMANUS, C. et al. Phenotypic characterization of naturalized swine breeds in brazil, uruguay and colombia. *Brazilian archives of biology and technology*, Curitiba, v. 53, n. 3, p. 583–591, 2010.
- MCMANUS, C. et al. Geographical distribution of sheep breeds in brazil and their relationship with climatic and environmental factors as risk classification for conservation. *Journal of science and technology*, v. 1, n. 3, p. 15, 2013.
- MEDUGORAC, I. et al. Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and evolution*, v. 1, n. 3, p. 408–420, 2011.

MIRKENA, T. et al. Genetics of adaptation in domestic farm animals: a review. *Livestock science*, Amsterdam, v. 132, n. 1–3, p. 1–12, 2010.

NEI, M. Bottlenecks, genetic polymorphism and speciation. *Genetics*, Baltimore, v. 170, n. 1, p. 1–4, 2005.

NEI, M.; MARUYAMA, T.; CHAKRABORTY, R. The bottleneck effect and genetic variability in populations. *Evolution*, Lawrence, v. 29, n. 1, p. 1-10, 1975.

NIEDERLE, P. A.; GELAIN, J. Geographical indications in brazilian food markets: quality conventions, institutionalization, and path dependence. *Journal of rural social sciences*, v. 28, n. 1, p. 26-53, 2013.

NIELSEN, E. E. et al. Genomic signatures of local directional selection in a high gene flow marine organism; the atlantic cod (*gadus morhua*). *Bmc evolutionary biology*, London, v. 9, n. 1, p. 276, 2009.

NIMER, E. Um modelo metodológico da classificação de climas. *Revista brasileira de geografia*, Rio de Janeiro, v. 41, n. 4, p. 59–89, 1979.

NORTH, A. et al. Local adaptation in a changing world: the roles of gene-flow, mutation, and sexual reproduction. *Evolution*, Lawrence, v. 65, n. 1, p. 79–89, 2011.

NTEEBBA, J. et al. Heat stress alters ovarian insulin-mediated phosphatidylinositol-3 kinase and steroidogenic signaling in gilt ovaries. *Biology of reproduction*, New York, v. 92, n. 6, p. 148, 2015.

ODUM, E. P. *Fundamentals of ecology*. Philadelphia and London: W. B. Saunders Co, 1959.

PAIVA, S. R. et al. Viabilidade de loci microssatélites para exclusão de paternidade em raças de ovinos. In: *SIMPÓSIO NACIONAL DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL*, 5., 2004, Pirassununga, SP. *Anais eletrônicos...* Pirassununga, SP: [s.n.], 2004. Disponível em: <<http://sbmaonline.org.br/anais/v/trabalhos/pdfs/bm018.pdf>>. Acesso em: 17 jun. 2016.

PÉREZ-FIGUEROA, A. et al. Comparing three different methods to detect selective loci using dominant markers: comparing methods to detect selective loci. *Journal of evolutionary biology*, Basel, v. 23, n. 10, p. 2267–2276, 2010.

PETROCELLI, H.; BATISTA, C.; GOSÁLVEZ, J. Seasonal variation in sperm characteristics of boars in southern uruguay. *Revista brasileira de zootecnia*, Viçosa, v. 44, n. 1, p. 1–7, 2015.

PHAM, L. D. et al. Molecular genetic diversity and genetic structure of vietnamese indigenous pig populations. *Journal of animal breeding and genetics*, Berlin, v. 131, n. 5, p. 379–386, 2014.

- PURCELL, S. Plink: whole genome data analysis toolset. [S.l.], 2014. Disponível em: <<http://pngu.mgh.harvard.edu/~purcell/plink/>>. Acesso em: 4 ago. 2016.
- PYRCZ, M. J.; DEUTSCH, C. V. Declustering and debiasing. Newsletter, v. 19, 2003. Disponível em: <<http://www.gaa.org.au/pdf/DeclusterDebias-CCG.pdf>>. Acesso em: 2 fev. 2016.
- QGIS DEVELOPMENT TEAM. Qgis geographic information system. [S.l.]: Open Source Geospatial Foundation, 2009.
- RAMOS, A. M. et al. Design of a high density snp genotyping assay in the pig using snps identified and characterized by next generation sequencing technology. Plos one, San Francisco, v. 4, n. 8, p. e6524, 2009.
- RANGEL, T. F.; DINIZ-FILHO, J. A. F.; BINI, L. M. Sam: a comprehensive application for spatial analysis in macroecology. Ecography, Copenhagen, v. 33, n. 1, p. 46–50, 2010.
- RAUFASTE, N.; ROUSSET, F. Are partial mantel tests adequate? Evolution; international journal of organic evolution, Lawrence, v. 55, n. 8, p. 1703–1705, 2001.
- ROHLFS, R. V.; WEIR, B. S. Distributions of hardy–weinberg equilibrium test statistics. Genetics, v. 180, n. 3, p. 1609–1616, 2008.
- ROSS, J. W. et al. Physiological consequences of heat stress in pigs. Animal production science, v. 55, n. 12, p. 1381–1390, 2015.
- SAS INSTITUTE INC. Sas/stat® 9.3 user's guide. Cary, NC.: SAS Inst., 2011.
- SCHRÖDER, W. Gis, geostatistics, metadata banking, and tree-based models for data analysis and mapping in environmental monitoring and epidemiology. International journal of medical microbiology: ijmm, Jena, v. 296, Suppl 40, p. 23–36, 2006.
- SHABTAY, A. Adaptive traits of indigenous cattle breeds: the mediterranean baladi as a case study. Meat science, Barking, v. 109, p. 27–39, 2015.
- SILVA, M. et al. Utilização de marcadores genéticos em suínos. i. características reprodutivas e de resistência a doenças. Archivos latinoamericanos de producción animal, Mayaguez, v. 11, n. 1, p. 1–10, 2003.
- SILVA-FILHA, O. L. et al. Caracterização da criação de suínos locais em sistema de utilização tradicional no estado da paraíba, brasil. Arch zootec, Cordoba, v. 54, n. 206-207, p. 523–528, 2005.
- SILVEIRA, P. R. S. et al. Identificação de paternidade para avaliação da contribuição da primeira e segunda doses inseminantes na composição da leitegada suína. R. bras. zootec, Viçosa, v. 37, n. 4, p. 609–615, 2008.

- SOLLERO, B. P. et al. Genetic diversity of brazilian pig breeds evidenced by microsatellite markers. *Livestock science, Amsterdam*, v. 123, n. 1, p. 8–15, 2009.
- STEIN, A.; ETTEMA, C. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, ecosystems & environment*, v. 94, n. 1, p. 31–47, 2003.
- STERN, C. The hardy-weinberg law. *Science*, v. 97, n. 2510, p. 137–138, 1943.
- STORFER, A. et al. Putting the “landscape” in landscape genetics. *Heredity, Cary*, v. 98, n. 3, p. 128–142, 2007.
- STORZ, J. F. Invited review: using genome scans of dna polymorphism to infer adaptive population divergence. *Molecular ecology, Oxford*, v. 14, n. 3, p. 671–688, 2005.
- STORZ, J. F. Genes for high altitudes. *Science, New York*, v. 329, n. 5987, p. 40–41, 2010.
- STUCKI, S. et al. High performance computation of landscape genomic models integrating local indices of spatial association. Arxiv preprint arxiv:1405.7658, 2014. Disponível em: <<http://arxiv.org/abs/1405.7658>>. Acesso em: 8 dez. 2015.
- STUCKI, S. Développement d’outils de géo-calcul haute performance pour l’identification de régions du génome potentiellement soumises à la sélection naturelle: analyse spatiale de la diversité de panels de polymorphismes nucléotidiques à haute densité (800k) chez bos taurus et b. indicus en ouganda. Suisse: EPFL, 2014. Disponível em: <[http://infoscience.epfl.ch/record/196921/files/EPFL\\_TH6014.pdf](http://infoscience.epfl.ch/record/196921/files/EPFL_TH6014.pdf)>. Acesso em: 18 mar 2015
- TOCHER, K. D. Extension of the neyman-pearson theory of tests to discontinuous variates. *Biometrika, Oxford*, v. 37, n. 1/2, p. 130–144, 1950.
- VECHT, J. Van Der. Studies on indo-australian and east-asiatic eumenidae (hymenoptera, vespoidea). Leiden: E.J. Brill, 1963.
- VICENTE, A. A. et al. Genetic diversity in native and commercial breeds of pigs in portugal assessed by microsatellites. *Journal of animal science, Champaign*, v. 86, n. 10, p. 2496–2507, 2008.
- VITALIS, R. et al. Detecting and measuring selection from gene frequency data. *Genetics*, v. 196, n. 3, p. 799–817, 2014.
- WETTEMANN, R. P.; BAZER, F. W. Influence of environmental temperature on prolificacy of pigs. *Journal of reproduction and fertility. Supplement, Cambridge*, v. 33, p. 199–208, 1985.

WIGGINTON, J. E.; CUTLER, D. J.; ABECASIS, G. R. A note on exact tests of hardy-weinberg equilibrium. *American journal of human genetics*, Chicago, v. 76, n. 5, p. 887–893, 2005.

WILDT, D. E.; RIEGLE, G. D.; DUKELOW, W. R. Physiological temperature response and embryonic mortality in stressed swine. *The american journal of physiology*, Bethesda, v. 229, n. 6, p. 1471–1475, 1975.

WRIGHT, S. Evolution in mendelian populations. *Bulletin of mathematical biology*, Elmsford, v. 52, n. 1, p. 241–295, 1931.

WRIGHT, S. Evolution and the genetics of populations. Chicago: University of Chicago Press, 1978. v. 4.

WULDER, M. A. et al. High spatial resolution remotely sensed data for ecosystem characterization. *Bioscience*, Uberlândia, v. 54, n. 6, p. 511–521, 2004.

YAMAMOTO, J. K.; LANDIM, P. M. B. *Geoestatística: conceitos e aplicações*. [S.I.]: Oficina de Textos, 2013.

YARO, M. et al. Molecular identification of livestock breeds: a tool for modern conservation biology. *Biological reviews*, London, s/v., s/n., s/p., 2016.

ZAPATA, C.; ALVAREZ, G. On fisher's exact test for detecting gametic disequilibrium between dna polymorphisms. *Annals of human genetics*, Oxford, v. 61, n. 1, p. 69–75, 1997.

## **APENDICE**



## Anexo 1

### **Porquê Preservar a Raças Suínas Brasileiras**

Cesconeto, R.J.<sup>1</sup>

Dentre as espécies domesticas, os suínos estão entre as primeiras a serem domesticadas<sup>1</sup>, há cerca de 9000 mil anos atrás. É a única espécie de grandes mamíferos domesticados que ainda convive com o ancestral selvagem<sup>2</sup>. Assim, podemos dizer que o conjunto de todas as raças<sup>3</sup> de suínos modernos representa uma fragmentação da variabilidade genética dos javalis europeus e asiáticos existentes na época da domesticação, ou seja, toda aquela variabilidade genética inicial está subdividida nas raças suínas existentes nos dias de hoje. Seria o mesmo que comparar a variabilidade desta espécie com um grande quebra cabeças, onde, cada uma das raças seria um das peças da figura final.

#### **Origem da variabilidade genética**

Esta variabilidade genética, tão importante para a preservação das raças nos dias atuais, não é simples de ser criada. Ela começa com a ocorrência de eventos conhecidos como mutação<sup>4</sup> ainda na formação dos gametas<sup>5</sup>. Porém, estes eventos são raros, demandam grandes períodos de tempo para acontecer e os mecanismos de reparo do DNA<sup>6</sup> são altamente eficazes, fazendo com que uma parte muito pequena destas alterações seja encontrada no material genético dos gametas. Para que uma mutação seja transmitida aos descendentes, gameta portador da mutação deve participar da formação de um novo indivíduo e, no mínimo não ser prejudicial à vida. Entretanto, a grande maioria das mutações não tem efeito sobre a habilidade de sobrevivência do indivíduo e, somente uma parcela muito pequena tem algum efeito positivo (Teoria Neutralista<sup>7</sup>). Uma vez que a mutação tenha ocorrido, os mecanismos de reparo falhados e, o gameta tenha participado da formação de um novo indivíduo, ela ainda deve ser disseminada na população. Essa disseminação só irá acontecer se o indivíduo portador da mutação tiver sucesso reprodutivo, transmitindo essa mutação aos seus descendentes e,

assim sucessivamente. Mas ainda há risco de desaparecimento deste novo alelo pela ação da deriva genética<sup>8</sup>.

Com todos estes fatores contra o surgimento de variabilidade, somado ao tempo necessário para que uma mutação ocorra e seja fixada, podemos entender porque a perda de uma raça suína constitui a perda de uma parcela importante da diversidade total da espécie.

1- Doutorando em Melhoramento Genético Animal, Faculdade de Agronomia, UFRGS  
**Origem das raças suínas Brasileiras**

Os suínos, assim como, a grande maioria dos mamíferos domésticos, não é originária do continente americano, sendo introduzida nas expedições de colonização há aproximadamente cinco séculos atrás. Estes animais tiveram papel importante para o sucesso das expedições de colonização, uma vez que forneciam proteína de alto valor biológico, gordura para conservação e processamento dos alimentos e, ainda atuavam como recicladores de alimentos. Com o passar dos anos e a interiorização da população no nosso país, a criação de suínos foi sendo levada dos povoados litorâneos para o interior, o que permitiu o estabelecimento de raças adaptadas as diferentes condições ambientais do território brasileiro, bem como, às necessidades econômicas da época, onde a produção de carne e, principalmente, de banha eram prioridades para as famílias.

Essa foi a principal função das raças suínas nacionais até meados da década de 70, quando a introdução dos óleos vegetais para uso doméstico, aliado à publicidade contra as gorduras animais na alimentação humana gerou a necessidade de mudar o perfil de produção da suinocultura do país. Com a introdução das raças européias e americanas a indústria suinícola deixava de ser produtora de banha e passava a produzir carne, relegando as raças nacionais às criações de subsistência, miscigenação e praticamente a extinção.

### **Uso das raças nacionais Brasileiras pela indústria**

As raças comerciais são simplesmente o fruto de um intenso trabalho de melhoramento genético, focado em características de produtividade, em animais menos produtivos que os atuais, de raças locais de outros países. Assim as raças nacionais também podem passar por este processo de melhoramento genético, gerando animais altamente produtivos e

prolíficos, com uma vantagem das raças nacionais brasileiras locais sobre as raças importadas: 500 anos de adaptação ao meio ambiente brasileiro. Esta melhor adaptação pode ser explorada nos programas de melhoramento, produzindo linhagens com maior manejo mais fácil, além de uma possível diminuição dos custos fixos da atividade (instalações e equipamentos).

Para o uso efetivo de linhagens comerciais de suínos Brasileiros em programas de melhoramento genético, criação de produtos com certificação de origem ou, simplesmente a elaboração de estratégias de preservação, precisamos:

- 1- Identificação dos padrões raciais, produtivos e genéticos;
- 2- Identificação dos rebanhos;
- 3- Seleção dos animais mais produtivos;
- 4- Identificação de marcadores moleculares ligados as características ambientais e direcionamento de seleção para as diferentes regiões;
- 5- Identificação e introgreção de genes favoráveis nas linhagens comerciais.

### **Identificação dos padrões**

A identificação dos padrões raciais se torna indispensável, uma vez que a maioria das raças é encontrada ao longo de todo o território nacional e, muitas vezes uma mesma raça é conhecida por diferentes nomes nas diferentes regiões, podendo levar ao registro equivocado dos animais. As raças mais populares têm padrão racial registrado ou conhecido desde a década 50. Algumas tiveram até os padrões produtivos descritos, tanto no passado como em trabalhos recentes, como por exemplo, a raça Moura, Piau e Sorocaba. Porém, a maioria não apresenta registros de padrões raciais ou produtivos, além daqueles encontrados nos livros de suinocultura da década de 50. O aumento do vigor híbrido dos animais nativos cruzados com animais ditos brancos (e.g.: landrace) na primeira geração, pode ser perdido nas gerações seguintes, pela perda de adaptação ao ambiente local. Todo progresso conseguido em anos de seleção natural pode ser perdido em algumas poucas gerações pela falta de adaptação dos animais cruzados, por exemplo: a

melhoria na conversão alimentar e precocidade dos animais cruzados é contrabalanceada pela menor resistência às doenças e parasitas, o que aumentaria a mortalidade dos leitões. Assim, a identificação dos padrões genéticos das raças pode ajudar a formação de rebanhos puros, sem a introgressão<sup>7</sup> de “genes comerciais” e, até mesmo, de javalis. Normalmente a identificação de mestiços javali-suíno (Javaporco) se dá pela análise do cariótipo, porém, diferentes estudos apontam que, apesar de 100% dos animais filhos do cruzamento de javalis com suínos terem um cariótipo 17, a F2<sup>9</sup> desses animais apresenta uma proporção 50% com cariótipo 16. Ou seja, em algumas poucas gerações os animais cruzados com javali apresentam um cariótipo tipicamente suíno, sendo a diferenciação possível apenas pela presença de marcadores moleculares exclusivos, dos animais selvagens.

#### **Identificação dos rebanhos**

A existência de alguns rebanhos de conservação, mantidos por instituições de pesquisa como a Embrapa, por exemplo, se mostra de grande importância na preservação dos rebanhos, pois, além de manter a pureza racial destes animais, permite a difusão destes animais entre os criadores. Por outro lado, as características da suinocultura industrial e o uso das raças nativas somente em criações do tipo Hobbie/subsistência, somado a ausência de registros destas criações junto às associações de criadores e a falta de identificação desses rebanhos pelos órgãos de vigilância sanitária e institutos de sanidade animal, torna difícil sua localização, assim como, o acompanhamento destas populações. Identificar rebanhos e promover o acompanhamento dos mesmos é uma das mais importantes etapas de um programa bem sucedido de conservação de recursos genéticos, pois preservar as raças, facilita a difusão de tecnologias de base, muitas vezes ignoradas pelos produtores devido a falta de interesse do sistema produtivo neste tipo de produção.

#### **Seleção dos animais, Identificação de marcadores moleculares e introgressão de genes favoráveis**

A partir dos rebanhos de conservação existentes pode-se iniciar o estudo das características produtivas das raças e, iniciar programas de

melhoramento genético baseados em raças locais. Entender e quantificar as características de rusticidade e resistência à doenças das raças nativas, para que estes animais possam ser utilizados, tanto no sistema produtivo tradicional, em cruzamentos industriais ou ainda, na criação de híbridos. Outra frente de trabalho seriam os estudos de associação genômica, onde se buscaria identificar regiões do genoma ligadas às variações no meio ambiente, resistência às doenças, adaptação ao frio/calor, etc. Estas regiões do genoma podem ser exploradas de forma racional sendo introduzido em híbridos comerciais, incrementando a produtividade dos rebanhos comerciais brasileiros. Da mesma forma, raças como o Piau, que são sabidamente hiper prolíficas, poderiam ser utilizadas na produção de linhagens mãe, a exemplo do que foi feito com a raça chinesa Meishan, porém, com a vantagem de serem mais adaptadas às condições nacionais.

Dentro do Melhoramento Genético Animal, a equação básica ( $F=G+A$ ), nos ensina que podemos manipular o ambiente para dar as melhores condições aos animais, de forma que eles possam expressar todo o potencial de seus genótipos. Porém, quando os animais são selecionados dentro de um ambiente e, a produção será em ambiente diferente, podemos esperar dificuldades de adaptação e perdas consideráveis, tanto de tempo, como de recursos, para tornar o meio ambiente mais favorável para eles. Ao identificarmos marcadores altamente relacionados com as variações ambientais de uma região, estes marcadores podem ser utilizados na seleção de animais geneticamente adaptados, voltados para cada região, diminuindo assim, os prejuízos causados pela baixa adaptação regional. Os rebanhos de genética podem ser centralizados em uma única região do país, seja por motivos financeiros ou sanitários, e, a produção das matrizes e cachasos direcionada através da incorporação destes alelos na matriz de seleção, especificamente para cada uma das regiões produtoras. Da mesma forma, a construção de núcleos de genética com condições mais próximas daquelas encontradas nos locais para onde os animais serão direcionados, respeitando as necessidades locais.

A partir da identificação, tanto dos rebanhos de raças naturalizadas, como dos padrões genéticos de cada uma das raças, as associações de criadores e outros organismos podem criar as certificações de origem protegida (POD, PIG, etc) dos produtos nacionais, a exemplo dos Presuntos Parma ou Pata Negra, e de tantos outros produtos, de várias regiões e raças, ao redor do mundo. Estes produtos apresentam valores agregados superiores aos produtos industrializados, comumente encontrados no comércio, o que pode, além de aumentar a renda das pequenas propriedades rurais, melhorar o status de conservação das raças localmente adaptadas.

### **Concluindo**

A redução da variabilidade genética em raças de suínos adaptados brasileiros é uma questão de preocupação não só para pesquisadores, mas também para empresas comerciais, uma vez que, os 500 anos de adaptação às condições ambientais no Brasil, podem ajudar a aumentar a produtividade dos rebanhos comerciais. Os programas de conservação genéticos das raças suínas brasileiras devem ser capazes de reconhecer a variabilidade individual dos animais dentro e entre as raças, preservando a diversidade genética e alélica globais. Ao preservar a diversidade neutra e adaptativa das raças suínas localmente adaptadas, conservamos o potencial destas raças para reagir aos desafios futuros, que são importantes para o desenvolvimento de sistemas de produção animal mais eficientes e sustentáveis.

### **Glossário**

**1- Domesticação:** é o processo pelo qual o homem consegue que certas espécies se adaptem a sua convivência para desempenhar alguma função útil que facilite a vida dos humanos. Causa modificação no comportamento e na aparência das espécies.

**2- Ancestral selvagem:** espécie em seu estado primitivo, antes do processo de domesticação. Indivíduos cujos descendentes passaram pelo processo de domesticação.

**3- Raças:** classificação de diferentes populações de uma mesma espécie biológica, de acordo com suas características genéticas ou fenotípicas.

**4- Mutação:** alteração súbita no material genético existente. As mutações são divididas em diferentes tipos: **gênicas, extranucleares e cromossômicas**. Na mutação gênica as modificações ocorrem nas bases nitrogenadas do DNA. As mutações extranucleares ocorrem no DNA de organelas do citoplasma (mitocôndrias e cloroplastos). Nas mutações cromossômicas as alterações acontecem tanto na estrutura (deleção, duplicação, inversão e translocação) quanto no número de cromossomos (aneuploidia e euploidia). Quanto à sua origem, as mutações podem ser espontâneas ou induzidas (Bespalkok Filho *et al.*)

**5- Gametas:** são células que se unem na fecundação e apresentam a metade do número cromossômico da espécie. Eles são produzidos nas gônadas masculinas (testículos) e femininas (ovários).

**6- Reparo de DNA:** consiste em mecanismos essenciais que protegem a integridade do genoma. É a capacidade da maquinaria celular corrigir os erros nas fitas de DNA.

**7- Teoria Neutralista:** A teoria neutra da evolução assume que a maioria da variação encontrada em seqüências de DNA e proteínas tanto dentro como entre diferentes espécies é neutra com relação à seleção e considera que boa parte dos diferentes alelos para um mesmo loco possui o mesmo valor adaptativo

**8- Deriva genética:** é a mudança ao acaso das frequências alélicas. Ela pode ser definida também como oscilação genética aleatória. Ao longo do tempo, essa oscilação pode eliminar alelos (frequência = 0) ou fixá-los (frequência = 1) nas populações.

**9- F2:** em genética os dois indivíduos que dão origem a uma família são chamados de geração parental. O produto do cruzamento da geração parental é conhecido como F1, e os descendentes do cruzamento dos F1 são conhecidos como F2.

### Referências Bibliográficas

Allendorf, F.W.. (1986). **Genetic drift and the loss of alleles versus heterozygosity**. Zoo biology. vol:5, 2 pg:181 -190

Arnaud J-F.. (2003) **Metapopulation genetic structure and migration pathways in the land snail *Helizaspersa*: influence of landscape heterogeneity**. *Landscape Ecol* 18: 333–346.

Beiguelman, B.. (2008) **Genética de populações Humanas**, Ribeirao Preto, SBG. pp 235.

Blackburn, H.. (2012) **Genetic Selection and Conservation of Genetic Diversity**. *Reproduction in Domestic Animals*, 2012, V 47: 249–254

Bespalhok Filho JC, Guerra EP, Oliveira R DOMESTICAÇÃO DAS PLANTAS CULTIVADAS. In: *Melhoramento de Plantas*, pp. 11–17

Cañón, J.. (2001) **Genetic diversity measures of local European beef cattle breeds for conservation purposes**. *Genet. Sel. Evol.* 33 (2001). INRA, EDP Sciences.

Cerqueira, R.; Brant, A.; Nascimento, M.T. & Pardini, R.. (2003). **Fragmentação: alguns conceitos**. Pp. 24-40. In: **Fragmentação de Ecossistemas: Causas, efeitos sobre a biodiversidade e recomendações de políticas públicas**. Biodiversidade 6. Ministério do Meio Ambiente, Brasília/DF. 510p.

Diniz-Filho J. A.F., Telles, M.P.C., Bonatto, S.L., Eizirik, E, Freitas, T.R.O., Marco Jr, P., Santos, F.R., Sole-Cava, A., Soares, T.N.. (2008) **Mapping the evolutionary twilight zone: molecular markers, populations and geography**. *J. Biogeogr.* 35, 753–763.

Egito, A.A, Mariante, A.S., Albuquerque, M.S.M.. (2002) **The Brazilian genetic recourses conservation program**. *Archivos de Zootecnia*.v 51, 39-52.

Ellegren, H. & Sheldon, B.C.. (2008). **Genetic basis of fitness differences in natural populations**. *Nature* 452, 169-175.

Foll, M., and O.E. Gaggiotti.. (2006). **Identifying the environmental factors that determine the genetic structure of Populations**. *Genetics* 174: 875-891.

Joost, S., Colli, L., Baret, P.V., Garcia, J.F., Boettcher, P.J., Tixier-Boichard, M., Ajmone-Marsan, P. & the GLOBALDIV Consortium. (2010). **Integrating geo-referenced multiscale and multidisciplinary data for the**



**management of biodiversity in livestock genetic resources**, Animal Genetics, Volume 41, Issue s1.

Kawecki ,T.J., Ebert, D.. (2004) Conceptual issues in local adaptation. *Ecology Letters*, 7, 1225–1241.

Lynch, M., J. Conery, and Burger, R.. (1995). **Mutation accumulation and the extinction of small populations**. *American Naturalist* 146:489-518

Martínez AM, Gama LT, Delgado JV et al. (2015) The Southwestern fringe of Europe as an important reservoir of caprine biodiversity. *Genetics Selection Evolution*, 47, 86.

McManus, C. M. ; Paludo, G.R.; Louvandini, H.; Gugel, R.; Sasaki, L.C.B.; Paiva, S.R.. (2009). **Heat Tolerance in Brazilian Sheep: Physiological and Blood Parameters**. *Tropical Animal Health and Production*, v. 9, p. 1.

Medugorac I, Veit-Kensch CE, Ramljak J et al. (2011). Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and Evolution*, 1, 408–420.

Nei, M.. (2005). **Bottlenecks, Genetic Polymorphism and Speciation**. *Genetics*, 170(1), 1–4.

Scherf, B.. (2008). **Description of production environments to support management of farm animal breeds in marginal areas**. FAO/ARC/ILRI Workshop on Conservation, management and use of marginal farm animal genetic resources, held at the 10th World Conference for Animal Production, Capetown, South Africa.

Sollero, B. P.. (2006) **Diversidade genética das raças naturalizadas de suínos no Brasil por meio de marcadores microssatélites**, Dissertação, Universidade de Brasília.

*Wikipédia, a enciclopédia livre* (2016)

Wright, S.. (1931). **Evolution in Mendelian populations**. *Genetics* 10:97-159.

Wright, S.. (1950). **Genetical structure of populations.**Nature 166 (4215): 247–9. Bibcode:1950Natur.166..247W. doi:10.1038/166247a0. PMID 15439261.

Wright, S.. (1978). **Variability between and among natural populations.**Evolution and the genetics of populations. University of Chicago Press, Chicago. Volume 4.

Yaro, M., Munyard, K.A., Stear, M.J., Groth, D.M.. (2016) Molecular identification of livestock breeds: a tool for modern conservation biology. Biological Reviews, n/a-n/a.

*Wikipédia, a enciclopédia livre* (2016)

## 6 VITA

Nascido aos quatro dias do mês de março do ano de 1977, na capital do estado do Paraná, Curitiba, é o segundo dos três filhos que José Vicente Cesconeto e Maria Helena F. Cesconeto tiveram.

Iniciou os estudos no Colégio Toneza Cascaes na cidade de Orleans, SC, cidade em que seus pais decidiram criar e educar seus filhos. Ao concluir o ensino básico iniciou o 2º grau em Urussanga-SC, mais precisamente no Colégio Rainha do Mundo, onde formou-se auxiliarem Análises Químicas no ano de 1994.

Prestou vestibular para Medicina Veterinária ainda naquele ano para duas conceituadas universidades: Universidade do Estado de Santa Catarina - UDESC e Universidade Federal de Pelotas - UFPel, tendo sido aprovado em ambas, optou pela Universidade Federal de Pelotas. Iniciou os estudos em no curso de Medicina Veterinária no ano de 1995, e final do quinto ano, após o estágio obrigatório na Empresa Perdigão S.A unidade de Marau-RS, graduou-se Médico Veterinário. Durante o curso foi bastante eclético, estagiou nos Hospitais da Universidade Federal do Paraná (campus Curitiba) e no hospital São Bernardo (também em Curitiba) nas áreas de clínica e cirurgia de pequenos animais; estagiou ainda na Parceria Agrícola Cerito (Cerito-RS) nas áreas de clínica e manejo de suínos – UPL; na Companhia Integrada de Desenvolvimento Rural de Santa Catarina – CIDASC, nas áreas de inspeção de carnes e derivados. Durante praticamente todo o curso foi colaborador no departamento de Microbiologia e Parasitologia da Universidade Federal de Pelotas, trabalhando com controle biológico de *Musca Domestica*.

Após sua formatura passou um ano prestando assistência técnica a produtores de gado leiteiro e suinocultores na região de Braço do Norte – SC, período que pode amadurecer a idéia de seguir a vida acadêmica. Decidido, iniciou o Mestrado em Zootecnia na Universidade Federal de Pelotas, Pelotas, RS, na área de produção animal sob orientação dos doutores Paulo de Sousa da Silveira (CNPSA-EMBRAPA) e Fernando Rutz (UFPel), obtendo o grau de Mestre em Ciências em abril de 2001.

Voltou ao campo na região de Braço do Norte prestando assistência ao distribuidor Shering Plough-Coopers da região por 6 anos nas áreas Manejo, Nutrição e Sanidade Suína. Foi então convidado pelo distribuidor Prezzoto Nutrição animal para prestar o mesmo serviço junto aos parceiros, até o ano de 2011. Neste ano foi convidado a ministrar aulas de Melhoramento Animal para o curso de Medicina Veterinária da Universidade do Extremo Sul de Santa Catarina- UNISUL, quando então decidiu voltar para a academia.

Iniciou o curso de Doutorado em Zootecnia na Universidade Federal do Rio Grande do Sul – UFGRS, sob a orientação do DSc José Braccini Neto, onde desenvolveu pesquisa envolvendo genética de paisagem em suínos naturalizados brasileiros. Entre outubro de 2014 e outubro de 2015 realizou estágio de doutoramento no exterior (Doutorado Sanduíche), no Laboratory of Geographic Information Systems – LASIG da École Polytechnique Fédérale de Lausanne – EPFL em Lausanne – Suíça, sob supervisão do professor DSc Stephane Joost, onde realizou análises em genética de paisagens necessárias para seu trabalho de tese.

Retornando ao Brasil finalizou os trabalhos do programa de doutorado para submeter-se à avaliação de defesa da Tese no Programa de Pós-Graduação em Zootecnia, área de concentração Produção Animal da Universidade Federal do Rio Grande do Sul em Porto Alegre-RS